Iraqi Journal of Statistical Science (14) 2008
p.p. [32-47]

Time Series Forecasting with UCM Model ; A Comparative Study using the Tigris River Data

Thafer Ramathan Muttar* **Elham Abdul-kareem Hussain****

ABSTRACT

In this paper,we build two basic models to forecast a flow water of the Tigris river which enters to mosul city . The first model is Unobserved Components Model which is abbreviated as UCM ,the second is Autoregressive and Moving Average model which is mentioned as ARMA,we built 10 primary models from ARMA to data of the time series of flow Tigris river after we transfer the data to a standrized form to remove seasonal effects . The best model which represented the data among ARMA models which are mentionted above is ARMA(2,2) by depending on the correction of Akaike information criterion which is symbolized by AIC_c.while ARMA(1,2) model is the best model for forecasting because it has a minimum mean absolute error which is symbolized by MAE.We obtained that the forecasting of flow water by UCM model is better than the results of ARMA(1,2) model by depending on the criterion MAE .

تكهن السلاسل الزمنية بنموذج UCM ؛ دراسة مقارنة باستخدام بيانات نهر دجلة

الملخص

يتناول هذا البحث بناء نموذجين رياضيين أساسيين للتكهن بتدفق مياه نهر دجلة الداخلة الى مدينة الموصل ،الأول هو نموذج المكونات غير المشاهدة Unobserved Components Model والذي يرمز له بـ UCM والثاني هو نموذج الانحدار الذاتي والمتوسطات المتحركة Autoregressive and Moving

* Ass. Prof. -Dep, of Statistics and Informatics- College of Computer Sciences and mathematics

** Ass. Lecturer-Insitiute of Technical / Mosul-Department of Systems of computers

Received: 3/1 /2008 ____________________Accepted: 18/ 6 / 2008

Average والذي يُرمز له بـ ARMA ، اذ تم بناء 10 نماذج أولية من نماذج ARMA لبيانات السلسلة الزمنية لتدفق نهر دجلة بعد تحويل هذه البيانات الى الصيغة القياسية للتخلص من التاثيرات الموسمية،وكان افضل نموذج يمثل البيانات من هذه النماذج هو نموذج ARMA(2,2) اعتمادا" على معيار اكاكي المصحح بينما كان نموذج ARMA(1,2) افضل نموذج تكهني لامتلاكه اقل قيمة لمتوسط الخطأ المطلق Mean Absolute Error والذي يُرمز له بالرمز MAE و تم التوصل الى ان نتائج التكهن بنموذج UCM افضل من نتائج التكهن بنموذج ARMA اعتمادا" على المعيار الاحصائي MAE.

1 - Introduction

The development and use of stochastic model of hydrological phenomena plays an important role in water resources engineering including their use to forecast river flow. The choice of the right model for a given hydrological series is an important aspect of the modeling process. (Mujumdar and Kumar, 1990) . The effectiveness of any decision depends upon the nature of sequence of events preceding the decision . The ability to predict the uncontrollable aspects of these events prior to making the decision should permit an improved choice over that which can otherwise be made . The purpose of forecasting is to reduce the risk in decision making . Information regarding stream flow , at any given point of interest, is necessary in the analysis and design of several water resources projects such as dam construction , reservoir operation , flood control and wastewater disposal.The most widely used stochastic models for river flow forcasting belong to the class of ARMA models proposed by Box and Jenkins.(Kumar,et al.,(2004))

Box and Jenkins give a method to estimate the orders of the ARand MA terms of a model based on autocorrelation and partial autocorrelation.Procedures for estimsting these orders from the given data based on testing residuals.(Mujumdar and Kumar, (1990)).

The UCM procedure analyzes and forecasts equally spaced univariate time series data using the unobserved components models (UCM). (UCMs are also called structural models in the time series literature.) A UCM decomposes the response series into components such as trend, seasonals, cycles, and the regression effects due to predictor series. The components in the model are supposed to capture the salient features of the series that are useful in explaining and predicting its behavior. In the time series literature, the components in a UCM are sometimes called the stylized facts about the series under consideration. It is fair to say that UCMs capture the versatility of the ARIMA models while possessing the interpretability of the smoothing models.(Selukar (2007)).

This paper is organised as follows.In section 1 a comman introduction about the modelling time series and forecasting, while in section 2, we handled with UCM model ;section 3, speciefied to ARIMA model ;Section 4 introduced the select of the model,whilst section 5 tackled an evaluation of forecast performance by a specific statistical criterion ; Finally section 6 focused on application of forecasting of flow water of the Tigris river which enters to Mosul city , also this section includes , a compartion between forecasting UCM and ARMA models depending on the statistical criterion .

2 - UCM Model

One thing that makes UCM useful is its similarity to regression. A useful conceptual framework for UCM is that of a regression model :

$$Y= B_o +B_1X_1 + B_2X_2 + e \qquad (1)$$

where the betas are allowed to be time varying , Y:dependent variable and X_1,X_2 are independent variables.A major difference between data properly modeled with regression and data typically modeled by time series techniques is the presence of autocorrelation, or serial correlation. ln "time series data" observations close together tend to behave similarly. If observation number n is above a fitted regression line , it is likely that observations n-1 and n+1 will also be above the regression line. This pattern of correlation between observations (and errors) breaks down as observations get farther apart in time. These characteristics suggest that a model for the data should place more "weight" or "importance" on "recent" observations and not give all observations in the data set equal importance .

The model for UCM is:

$Y_t = \mu_t + \gamma_t + \psi_t + r_t + \Sigma \varphi_i Y_{t-1}$

Y_t = trend + Seasonal + Cycle + Autoregressive term + A regressive terms involving lagged dep. Variables

$+ \Sigma \beta_j X_{jt} + \varepsilon_t$ (2)

+ A regressive term on indep. vars. + error term

The model components μ_t, , y_t , ψ_t and r_t are assumed to be independent of each other and model underlying “drivers” of the time series.

We can display components of two sides of equation(2) as (Lvery,1995) :

Y_t : Dependent Variable .

μ_t : Trend is the natural tendency of a series in the absence of seasonality,cycles
or the effect of any independent variable . It is modeled in two ways :

One method is a random walk :

$\mu_t = \mu_{t-1} + \acute{\eta}$ (3)

(where ή is an i.i.d error term).

The second method is a locally linear trend with a slope that varies,
only,with time.

$\mu_t = \mu_{t-1} + \beta_{t-1} + \acute{\eta}$ (4)

(where ή is ~ $N(0,\sigma^2{}_{\acute{\eta}}$, i.i.d error term).

γ_t : Season.The main characteristic of seasonally is that it's period (the time it takes to get through one full cycle) is known, it is modeled as a trigonomeric form. It is the sum of different cycles.

ψ_t : Cycles are like seasonals, but with an unknown period, UCM has implemented cycles as having fixed periods but time varying amplitude and phase.

r_t : Autoregressive term,UCM considers an autoregressive term as a cycle where frequency is either 0 or Π.

The expression for UCM autoregression is:

$$r_t = \rho\, r_{t-1} + \upsilon_t \qquad (5)$$

(where υ_t is ~i.i.d N(0, $\sigma^2{}_\upsilon$ IID error term).

$\Sigma\varphi_i Y_{t-1}$: A regressive terms involving lagged dep. Variable

$\Sigma\ \beta_j\ X_{jt}$: A regressive term on indep. vars.

ε_t : Irregular term or error term , ε_t is ~ N(0,$\sigma^2{}_\varepsilon$, i.i.d error term) .

but the basic structural time series model has three components . It is decomposed by a trend (μ_t),cycle (ψ_t) and seasonal (γ_t), addition to irregular component, cycle (ψ_t) is omitted as we mentioned above , then the components of UCM are :

$$Y_t = \mu_t + \gamma_t + \varepsilon_t \qquad (6)$$

with their details . (Marcelo,(1995))

The parameter vector in a UCM consists of the variances of the disturbance terms of the unobserved components, the damping coefficients and frequencies in the cycles, the damping coefficient in the autoregression, and the regression coefficients in the regression terms. These parameters are estimated by maximizing the likelihood. It is possible to restrict the values of the model parameters to user specified values.(Harvy,1989)

3 - ARIMA models

The ARIMA methodology was proposed by Box and Jenkins (1976) and it is now a quite popular tool in economic forecasting. The basic idea is that a stationary time series can be modeled as having both an autoregressive (AR) and moving average (MA) components. Non-stationary integrated series can also be handeled in the ARIMA framework , but it has to be reduced to stationary beforehand by difference the data . The multiplicative ARIMA representation can be written as :

$\Phi_P(L)\varphi_p(L)(1 - L^s)^D (1 - L)^d) y_t = \delta + \Theta_Q(L)\theta_q(L)\varepsilon_t$ (7)

Where $\Phi_p(L)$, $\varphi_p(L)$, $\Theta_Q(L)$ and $\theta_q(L)$ are polynomials in the lag operator (L) ,d and D are the number of consecutive and seasonal differences needed to make the series stationary , p,P ,q and Q are the degree of the autoregressive and moving average polynomials and ε_t is a normally distributed random error with mean and constant variance . The model is multipticative in the sense that the observed data result form the successive filtering of a random noise (ε_t) through the non- stationary ($\varphi_p(L)$) and the seasonal ($\Phi_P(L)$) . (Marcelo (1995)) . Many books divided the modeling into three parts. In the first stage,the order of differencing and the degrees of the AR and MA polynomials are determined using both the estimated autocorrelation and partial autocorrelation functions. In the second stage, the paramerers (φ, Φ, θ and Θ) are estimated and several tests are performed to assure that the residuals are white noise. Likelihood methods are generally used for estimation purposes, but if the model has only an AR part , least squares estimation will be appropriate . Finally, in the third stage, the best and most parsirnonious model is used to obtain the forecasts. The likelihood function for the non-seasonal stationary model can be written as :

$L(\varphi,\theta,\sigma^2/ y_t) = (2\pi)^{-T/2}(\sigma^2)^{-T/2} \exp[(-1/2\sigma^2)\Sigma\varepsilon_t^2]$ (8)

(Marcelo (1995), Box and Jenkins (1976))

4 - Select of the Model

We don't depend on Akaike Information Criterion (AIC) in the comparison because it has some flaws . Firstly , the AIC has no optimal property, i.e. it does not minimize the averagr value of any criterion function . Secondly, the AIC rule is not consistent, i.e. the probability that the decision rule will choose a wrong model does not go to zero even when the number of observations tends to infinity (Mujumdar and Kumar, (1990)).

Therefore,the selection of a best model depends on criterion AIC_c which is used to compare among models , It has been suggested by Brockwell and Davis (1993) the bias-corrected version of the AIC,and is obtained by :

$$AIC_c = AIC + \frac{2(p+1)(p+2)}{n-p-2} \qquad (9)$$

where :

AIC : Akaike Information Criterion.

p : the number of parameters to the model .

n : the observations which are used to build the model .

The additional to the AIC makes little difference for small values of p but, for larger values of p,penalizes extra parameters much more severely than the AIC and more effective on model selection (Chatfield ,(1998)) .

5 - Evaluation of Forecast Performance

To evaluate the forecast perfomance of model , it is common to reserve a small portion of the data at the end of a time series solely for forecast comparison and use a number of criteria which are available for the computation of forecast performance,including mean absolute error (MAE) as defined below : (Liu,2006)

$$MAE = \left(\frac{1}{m} \sum_{t=1}^{m} | y_t - \hat{y}_t | \right) \qquad (10)$$

where y_t is the actual observation at time t , $\hat{y}_t$ is the forecast value of y_t based on a particular model or method ,and m is the total number of observation in the post-sample period .

6 - Application :

In this section we present an empirical results for the two models proposed in the early part of the paper .The series we are going to use perform the comparative forecasting exercises is flow water of Tigris river that enters Mosul city in Iraq and we can see this series in figure (1) :

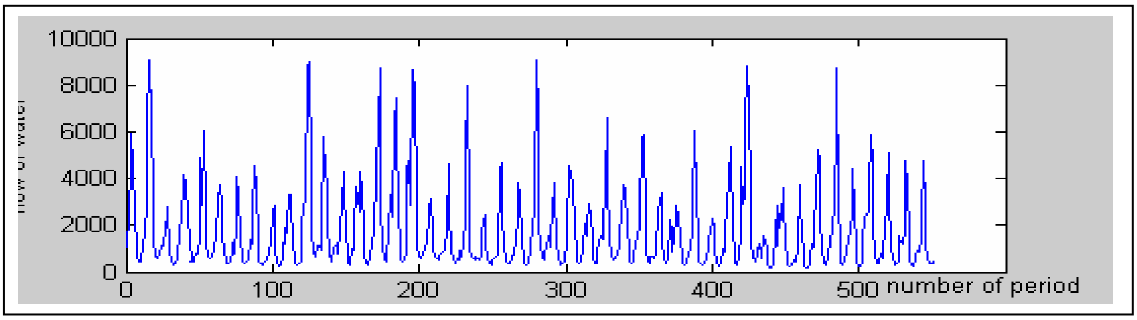


Figure (1) : The series of Tigris river from 1950 – 1995

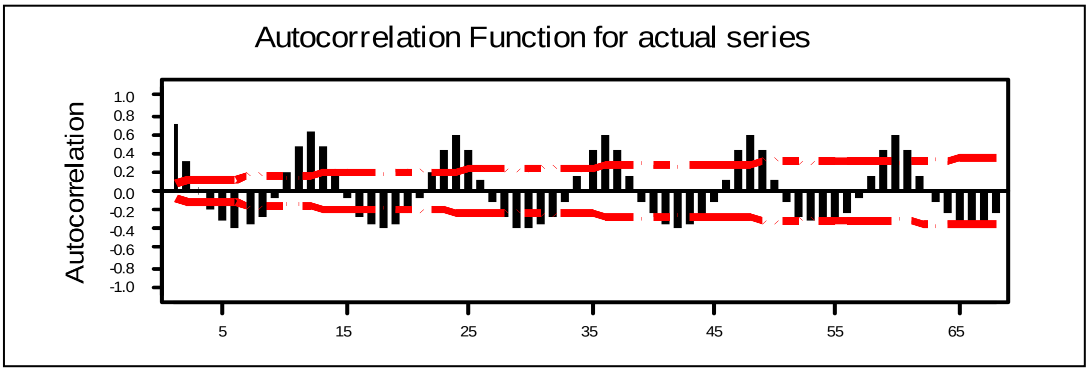


Figure (2)
Autocorrelation Function of the series of Tigris river from 1950 – 1995

Figure (1) displays the time plot of the series,and from figure (2) we can see clearly that seasonality effecte does not decay,then we transfered the data to standardize form to ensure the removal of periodicitied inherent in the process (Mujumdar and Kumar,(1990)) . This series (consists of monthly data from January 1950 to December 1995) is used after standardization (subtracting monthly mean and dividing it by standard deviation of the corresponding month) as shown in figure (3):

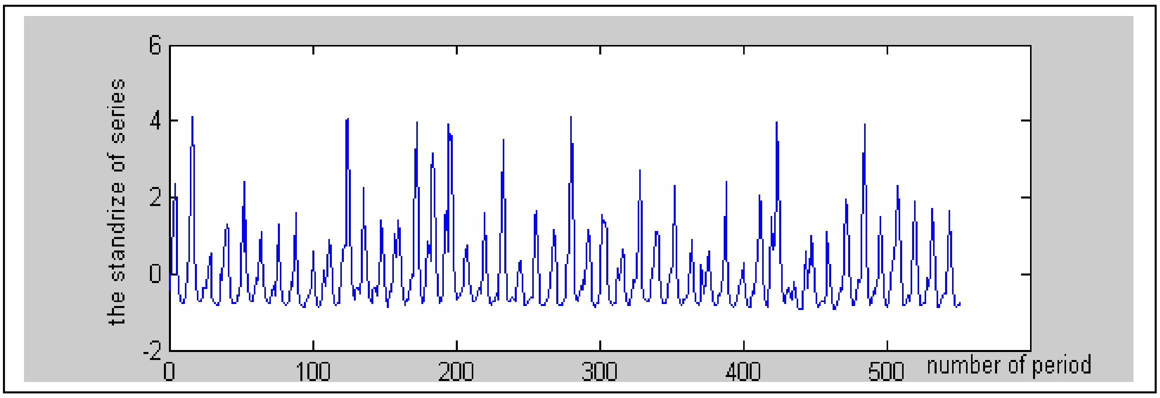


Figure (3)
The standardize series of Tigris river from 1950 – 1995

Now,for all the two models,the sample estimation period was from January 1950 to December 1994, leaving the last year 1995 consecutive observation to be used for out of sample forecasting comparisons . We have generated a set of twelve one step ahead forecasts .

Ten ARMA models was estimated by making many comparison among them using SAS program,these models are : AR(1) , AR(2) , AR(3) , AR(4) , MA(1) , MA(2) , ARMA(1,1) , ARMA(1,2) , ARMA(2,1), and ARMA(2,2) (Mujumdar and Kumar,(1990)), then we removed the models which their residuals (MA(1),MA(2),)) are significant and its parameter

(ARMA(2,1)) are not .The choice of the best model depends on AIC_c . The final results are shown in the table below :

Table (1)

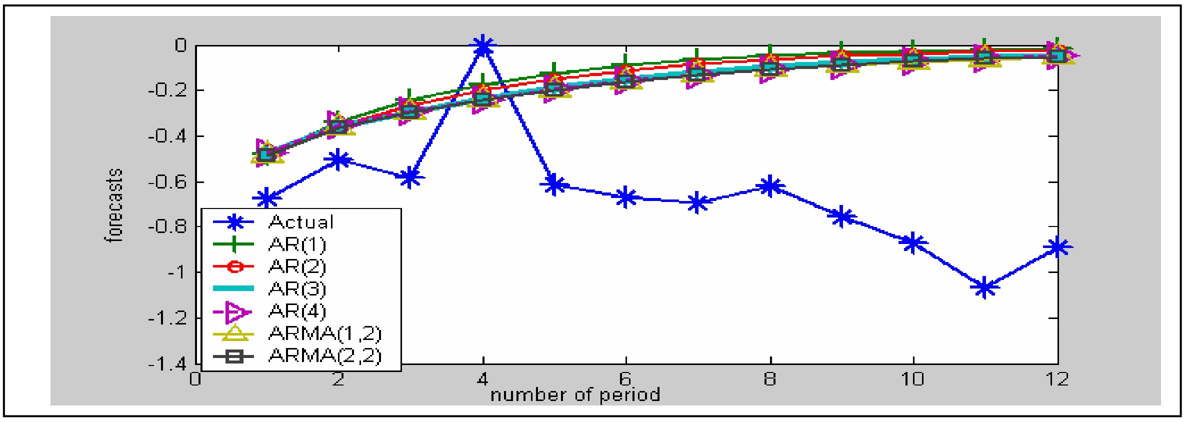
The final ARMA models

Model	AIC_C
AR(1)	-386.091
AR(2)	-481.778
AR(3)	-481.771
AR(4)	-481.247
ARMA(1,1)	-458.031
ARMA(1,2)	-469.424
ARMA(2,2)	-497.216

From the above tabel, we can choose the model which has a minimum value of AIC_c,that is ARMA(2,2) :

$$Y_t = - 0.00045 + 1.4Y_{t-1} - 0.63Y_{t-2} + 0.45a_{t-1} + 0.15a_{t-2} + \varepsilon_t \quad (11)$$

Until the objective of time series analysis is selected a model or the representation of data and for one step ahead forecasting,the best models for these needs are often not the same(Mujumdar and Kumar,1990),hence,the aim of this research is to make a comparison between the forecasting of the models ARMA and UCM , then firstly , we make a comparison among the forecasting of ARMA models which is mentioned in table (1) and the actual data to obtain the best model for using it in the comparison with UCM model,we can see this comparison in the figure (4) :



Figure(4)

Acomparison of one step ahead forecasts among ARMA models and the acatual data

From figure (4) we can see that the forecasts of all the models of ARMA are bad if they are compared with the actual data,hence,all of the forecasts are almost equal ,then we choose the model ARMA(1,2)which is used in the comparison with UCM model because it has a least of MAE as it is shown in table (2) :

Table (2)

The final ARMA forecast

Model	MAE
AR(1)	0.5023
AR(2)	0.5022
AR(3)	0.5028
AR(4)	0.5020
ARMA(1,1)	0.5023
ARMA(1,2)	0.5019
ARMA(2,2)	0.5022

The forecasts of ARMA(1,2) are shown in tabel(3) and it is used for a comparison with the actual data:

Tabel (3) : ARMA(1,2) Model Forecasts

Period	Actual	**One stepAhead/ARMA(1,2)**	Percental Erroe
541	884	1341.57	-35%
542	1439	1788.79	-20%
543	2397	2017.45	18%
544	4730	2093.01	125%
545	2938	2048.32	43%
546	1231	1935.30	-37%
547	549	1819.91	-70%
548	348	1740.19	-81%
549	285	1706.51	-84%
550	348	1711.70	-80%
551	381	1739.05	-79%
552	522	1771.19	-71%

The forecasts based on the ARMA(1,2) model are bad and there is a large forecast error in each period .

For the estimation of the UCM model we started by trying to estimate a general basic structural time series model . If we are looking to the figure (1) , we will see that the slop coefficient should be fixed or even be a set equal to zero.The initial results show that $\sigma^2_\varepsilon = 0$ confirming , therefor, that a fixed slop would be appropriate.The estimated slop in the fixed slop model is close to zero(3.4866E-11)) and the seasonal pattern is also close to zero because,the data orginal is transfered to the standardize a form as a building ARMA model.The estimation of the fixed slop basic structural time was performed using a time domain exact maximum likelihood method and the results shown by the SAS program below :

```
proc ucm data=series_g;
id date interval=month;
model logair;
irregular;
level;
slope;
season length=12 type=trig print=smooth;
estimate;
forecast lead=12 print=decomp;
run;
```

Tabel(4):Estimates of UCM model

Cpmponent	Parameter	Estimate	Standard Error
Irregular	Error Variance	0.16455	0.02940
level	Error Variance	0.24304	0.04037
slop	Error Variance	3.486E-11	1.8803E-8
seasonal	Error Variance	0.0000418	0.0000251

and we can show the forecasts depending on UCM model in table (5) :

Tabel(5) : UCM model Forecasts

Period	Actual	One stepAhead/UCM	Percental Error
541	884	686.84	28 %
542	1439	2670.9	- 46 %
543	2397	797.25	200 %
544	4730	1430.8	230 %
545	2938	866.85	238 %
546	1231	718.81	71 %
547	549	616.7	- 10 %
548	348	518.13	- 32 %
549	285	270.45	5 %
550	348	214.63	62 %
551	381	286.88	32 %
552	522	162.16	221 %

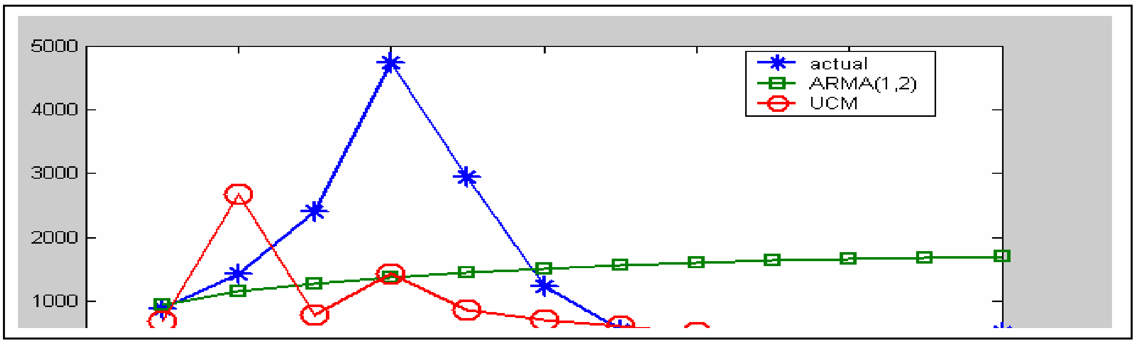
The results seem to be good if they are compared with the results of ARMA model .The percental error for the one step ahead shows that the period 543-545 and the period 552 are larger than other periods in UCM ,while in ARMA model all the periods are smaller or larger than the actuals except the period 543 which is considered acceptable.Also, The criterion MAE in ARMA model is larger than UCM model as shown in the tabel below :

Tabel(6) :MAE and MSE for the two models ARMA(1,2), and UCM

Model	MAE
ARMA(1,2)	0.501
UCM	0.159

We can show the difference between the observasion which are forecasted by the two models ARMA(1,2) and UCM by depending on tabel (3) and tabel (5). It is a obvious that the forecasting by UCM model is better than ARMA model and we can show this in figure(5) :

Figure (5)
Comparison of one step ahead forecasts using ARMA and UCM models with Actual data



7- Conclusion :

1- The best ARMA model is ARMA(2,2) to represent the data, after the comparision among 10 Primary models based on AIC_c criterion.

2- The best model for forecasting among the ARMA models is ARMA(1,2).

3- In UCM model , the slop closed to zero,and the seasonal closed to zero after the data is transfered to standardized form , then the components consist a level and irregular only .

4- The UCM model presented a better forecasting than ARMA(1,2) model .

References :

1-Box,G.E.P.and Jenkins, G.M.(1976): "Time Series Analysis, Forecasting and Control " , 2nd ed., Holden-Day, San Francisco,U.S.A.

2- Chatfield,Chris & Faraway,Julian (1998) : " Time series forecasting with neural networks : a comparative study using the airline data " , University of Bath , UK , University of Michigan ,Ann Arbor , USA.

3- Harvey,A.C. (1989): "Forecasting Structural Time Series Models and the Kalman Filter ", Cambridge University Press,Cambridge.(Chapter 28 The UCM Procedure (Experimental))

4- Kumar,D.N. ,Raju,K.S. and Sathish,T. ,(2004) : " River Flow Forecasting Using Recurrent Neural Netmork " , Water Resources Management , Journal , 18 : 143 – 161.

5- Lavery , Russell (1995), " Proc UCM (Unobserved Components Model" , Contractor for ASG,Inc.9 statin Ave. Apt 1, Ardmore , PA 19003 , www.asg-inc.com

6- Lin-Me Lou ,(2006) " Time Series Analaysis and Forecasting " , University of Illinois at Chicago , Scientific Computing Associates U.S.A .

7- Marcelo, S. P. (1995) :" Neural Networks Versus Time Series Methods : A Forecasting Exercise" .Universidade Federal do Rio Grande do Sul. Brazil .

8- Mujumdar,P.P. & Kumar,D.N. (1990) :"Stochastic Model Streamflow: some case studies" , Hydrological Sciences, Journal,35,4,8/1990 , p.p 395-410 .

9- Selukar, R. (2007): "Modeling Trend, Cycles, and Seasonality in Time Series Data Using PROC UCM " ,unpublished Ph,D. Thesis. Research statistician at SAS Institute ,U.S.A .