# TIMSAC for R package

The Institute of Statistical Mathematics

November 1, 2006

## 1 Introduction

The TIMSAC (TIMe Series Analysis and Control) is a general program package for analysis, prediction and contorol of time series and has beeen developed at the Institute of Statistical Mthematics. The original TIMSAC or TIMSAC-72 was published in Akaike and Nakagawa (1972). After that, TIMSAC-74, TIMSAC-78 and TIMSAC-84 were published as the TIMSAC series in Computer Science Monograph [1] . Many programs in the TIMSAC series were developed to provide procedures for analysing practical data, e.g., optimal control of an industrial process, analysis of economic fluctuations and so on. In this package several information criteria are used for model selection. In TIMSAC-72, FPE (Final Prediction Error) is used. After TIMSAC-74, AIC (Akaike Information Criterion) is used for model selection. TIMSAC-78 contains several programs based on Bayesian modeling where ABIC (Akaike Bayesian Information Criterion) is also used for model selection.

The programs of the TIMSAC series are written in FORTRAN. Recently a DLL (Dynamic Link Library) on Windows and a shared library on Linux has been developed for providing procedures of part of programs of the TIMSAC series. Programs written in FORTRAN, C or Java can use these libraries.

R is a free programming language or an environment that includes many statistical techniques. R has facilities for data manipulation on arrays and matrices, graphic and foreign language interfaces.

We provide *timsac* R package for using TIMSAC libraries from R. All functions in *timsac* R package use .C function of R to communicate between timsac.dll or libtimsac.so and R. And if necessary some functions display statistical graphs using R graphical procedures.

[1]H. Akaike, E. Arahata, T. Ozaki: TIMSAC-74, A Time series analysis and control program package (1) & (2), *Computer Science Monographs*, No.5 & 6, The Institute of Statistical Mathematics, Tokyo, 1975-1976

H. Akaike, G. Kitagawa, E. Arahata, F. Tada: TIMSAC-78, *Computer Science Monographs*, No.11, The Institute of Statistical Mathematics, Tokyo, 1979

H. Akaike, T. Ozaki, M. Ishiguro, Y. Ogata, G. Kitagawa, Y.-H. Tamura, E. Arahata, K. Katsura, Y. Tamura: TIMSAC-84 Part 1 & Part 2,*Computer Science Monographs*, No.22 & 23, The Institute of Statistical Mathematics, Tokyo, 1985

This research is partly supported by "Function and Induction Research Project" held by the Transdisciplinary Research Integration Center at the Research Organization of Information and Sciences, Japan.

# 2 Functions in this package

This section briefly describes functions, thier models and information criteria for model selection in chronological order. For further information and examples about the *timsac* function, see the help documentation.

## 2-1 R functios in TIMSAC-72

**autcor()** : Autocovariance and autocorrelation computation by direct method
**mulcor()** : Multiple covariance and correlation computation by direct method
**fftcor()** : Auto and/or cross correlation via FFT
**auspec()** : Power spectrum estimation by Blackman-Tukey type procedure
**mulspe()** : Cross spectrum estimation by Blackman-Tukey type procedure
**sglfre()** : Frequency response function computation (single input)
**mulfre()** : Multiple frequency response function computation (multiple inputs)
**fpeaut()** : FPE computation for uni-variate AR model
**fpec()** : FPE computation for control system model or multivariate AR model
**mulnos()** : Relative power contribution computation
**raspec()** : Rational spectrum computation (uni-variate)
**mulrsp()** : Rational spectrum computation (multi-variate)
**optdes()** : Optimal controller design
**optsim()** : Optimal controller simulation
**wnoise()** : White noise simulation

### Uni-variate AR (autoregressive) model

For the uni-variate stationary time series $y(t)$, an AR model is given by

$$y(t) = a(1)y(t-1) + \cdots + a(p)y(t-p) + u(t)$$

where $u(t)$ is a Gaussian white noise with mean 0 and variance $\sigma^2$. For fitting AR model, we estimate coefficients $a(1), ..., a(p)$ and innovation variance $\sigma^2$.

### Multi-variate AR (autoregressive) model

In the same way as the uni-variate case, an AR model of $k$-dimensional stationary time series is given by

$$y(t) = \sum_{m=1}^{p} A(m)y(t-m) + u(t)$$

where $A(m)$ is the $k \times k$ matrix and $u(t) = (\epsilon_1(t), \epsilon_2(t), \cdots, \epsilon_k(t))'$ is a k-dimensional Gaussian white noise with zero mean vector and variance covariance matrix $[\sigma_{ij}]$.

### FPE (Final Prediction Error)

The information criterion FPE is given as the square of the expected prediction error. FPE of a uni-variate AR model of order $p$ is calculated by

$$FPE = \frac{N+p+1}{N-p-1}\hat{\sigma}^2$$

where $N$ is the data length and $\hat{\sigma}^2$ is the estimate of innovation variance. The order which gives minimum FPE is used for model selection. (fpeaut)

Similarly FPE of the multi-variate AR model is defined. (fpec)

## 2-2 R functions in TIMSAC-74

**armafit()** : ARMA model fitting (uni-variate)
**autoarmafit()** : Automatic ARMA model fitting (uni-variate)
**canarm()** : Canonical correlation analysis (uni-variate)
**covgen()** : Covariance generation from gain function
**canoca()** : Canonical correlation analysis (multi-variate)
**markov()** : Automatic ARMA model fitting (multi-variate)
**prdctr()** : Prediction by ARMA model
**simcon()** : Optimal controller design and simulation
**nonst()** : Locally stationary AR model fitting (uni-variate)
**thirmo()** : Third order moment computation
**bispec()** : Bi-spectrum computation

### Uni-variate ARMA (autoregressive moving average) model

An ARMA model representation of time series $y(t)$ is given by

$$y(t) - \sum_{l=1}^{p} a(l)y(t-l) = u(t) - \sum_{m=1}^{q} b(m)u(t-m)$$

where $u(t)$ is a Gaussian white noise with mean 0 and variance $\sigma^2$. (armafit, autoarmafit)

**Multi-variate ARMA (autoregressive moving average) model**

An ARMA model representation of the $i$-th time series $y_i(t)$ is given by

$$y_i(t) - \sum_{l=1}^{p} \sum_{j=1}^{k} A_{ij}(l) y_j(t-l) = u_i(t) - \sum_{m=1}^{q} B_i(m) u_i(t-m) \qquad (i = 1, 2, ..., k)$$

where $(u_1(t), ..., u_k(t))$ is a $k$-dimensional Gaussian white noise. (markov)

**AIC (Akaike Information Criterion)**

The information criterion AIC is given by

$$AIC = -2l(\hat{\theta}) + 2k$$

$$= -2(maximum \ \ log\text{-}likelihood) + 2(the \ \ number \ \ of \ \ parameters).$$

In the case that the Yule-Walker equation is used for the uni-variate AR model, AIC is given by

$$AIC = N \log(2\pi\hat{\sigma}_p^2) + N + 2(p+1)$$

where $N$ is the data length and $p$ is the order of the AR model. Since the data length $N$ is much larger than the AR order $p$, FPE can be approximated by

$$\texttt{log}FPE(p) = \texttt{log}(\hat{\sigma}_p^2) + \frac{2p}{N}.$$

We can get the following relation:

$$N\texttt{log}FPE(p) = AIC - N\texttt{log}(2\pi) - N - 2.$$

## 2-3 R functions in TIMSAC-78

**unimar()** : AR model fitting (uni-variate, minimum AIC method)
**unibar()** : AR model fitting (uni-variate, Bayesian method)
**bsubst()** : Subset regression analysis by a model linear in parameters
(uni-variate, Bayesian method)
**mulmar()** : AR model fitting (multi-variate, minimum AIC method)
**mulbar()** : AR model fitting (multi-variate, Bayesian method)
**perars()** : Periodic autoregression (uni-variate, minimum AIC method)
**mlocar()** : Locally stationary AR model fitting (uni-variate, minimum AIC method)
**blocar()** : Locally stationary AR model fitting (uni-variate, Bayesian method)
**mlomar()** : Locally stationary AR model fitting (multi-variate, minimum AIC method)
**blomar()** : Locally stationary AR model fitting (multi-variate, Bayesian method)
**exsar()** : AR model fitting (uni-variate, exact maximum likelihood method)
**xsarma()** : ARMA model fitting (uni-variate, exact maximum likelihood method)

Least squares methods via Householder Transformations are used here. For the univariate AR model

$$y(t) = a(1)y(t-1) + \cdots + a(p)y(t-p) + u(t)$$

the matrix $Z$ and vectors $y$ and $a$ are defined as follows.

$$Z = \begin{bmatrix} y(p) & y(p-1) & \ldots & y(1) \\ y(p+1) & y(p) & \ldots & y(2) \\ \vdots & \vdots & & \vdots \\ y(N-1) & y(N-2) & \ldots & y(N-p) \end{bmatrix}, y = \begin{bmatrix} y(p+1) \\ y(p+2) \\ \ldots \\ y(N) \end{bmatrix}, a = \begin{bmatrix} a(1) \\ \vdots \\ a(p) \end{bmatrix} \tag{1}$$

The least square estimate of vector $a$ is obtained by minimizing $\left\| Za - y \right\|^2$.

We defined the $(n-p) \times (p+1)$ matrix X by

$$X = [Z|y].$$

By the adoption of the Householder transformation $U$, the following upper triangle matrix $S$ is obtained.

$$UX = S = \begin{bmatrix} s_{11} & \cdots & s_{1p} & s_{1,p+1} \\ & \ddots & \vdots & \vdots \\ & & s_{pp} & s_{p,p+1} \\ & & & s_{p+1,p+1} \end{bmatrix} \tag{2}$$

Since the matrix $U$ is orthogonal, the Euclidean norm is not changed.

$$\left\| Za - y \right\|^2 = \left\| UZa - Uy \right\|^2 = \left\| \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ & \ddots & \vdots \\ 0 & & s_{pp} \end{bmatrix} \begin{bmatrix} a(1) \\ \vdots \\ a(p) \end{bmatrix} - \begin{bmatrix} s_{1,p+1} \\ \vdots \\ s_{p,p+1} \end{bmatrix} \right\|^2 + s^2_{p+1,p+1} \tag{3}$$

The estimates of the AR coefficients $a_p(1), \cdots, a_p(p)$ are given by the solution of the following equation.

$$\begin{bmatrix} s_{11} & \cdots & s_{1p} \\ & \ddots & \vdots \\ 0 & & s_{pp} \end{bmatrix} \begin{bmatrix} a_p(1) \\ \vdots \\ a_p(p) \end{bmatrix} = \begin{bmatrix} s_{1,p+1} \\ \vdots \\ s_{p,p+1} \end{bmatrix} \tag{4}$$

The estimate of the innovation variance $\hat{\sigma}^2_p$ is given by

$$\hat{\sigma}^2_p = \frac{s^2_{p+1,p+1}}{N-p}.$$

That is, the estimate of the innovation variance of the AR model of order k $(k \leq p)$ is given by

$$\hat{\sigma}^2_k = \frac{1}{N-p} \sum_{i=k+1}^{p+1} s^2_{i,p+1}$$

and maximum log-likelihood is given by

$$-\frac{N-p}{2}\mathtt{log}(2\pi\hat{\sigma}^2_k) - \frac{N-p}{2}.$$

AIC for the uni-variate AR model of order k $(k \leq p)$ is given by

$$AIC(k) = (N-p)\mathtt{log}(2\pi\hat{\sigma}^2_k) + N - p + 2(k+1).$$

Above AIC is not equal to that given for AR model fitting with the aid of the Yule-Walker equation. Because in the least squares computation an AR model of order $p$ is considerd, but in the Yule-Walker eauation all data is estimated. (unimar)

For the multi-variate AR model fitting AIC is defined in the same way. (mulmar)

**ABIC (Akaike Bayesian Information Criterion)**

The Bayesian AR model fitting is based on the idea that AIC is the estimate of the expected log-likelihood. For an AR model of order $k$ the likelihood of the model is given by

$$f(y|k) = \mathtt{exp}(-\frac{1}{2}AIC(k)).$$

The posterior probability of the $k$-th AR model is given by

$$\pi(k|y) = \frac{f(y|k)\pi(k)}{\int f(y|k)\pi(k)dk}$$

where $\pi(k)$ is the prior proability. The Bayesian estimate of the AR model is given as the average of the models of various orders weighted with the posterior probabilities. The information criterion ABIC is defined by

$$ABIC = (N-p)\mathtt{log}\hat{\sigma}_B^2 + 2(\sum_{i=1}^{p} d(i)^2 + 1)$$

where $\hat{\sigma}_B^2$ is the estimate of innovation variance and $d(i)$ is given by

$$d(i) = \sum_{k=i}^{p} \pi(k|y)$$

and $\pi(k|y)$ is the posterior probability as mentioned above. For the Bayesian model the goodness of fit is evaluated by ABIC. (unibar, mulbar)

**Locally stationary AR model fitting**

The uni-variate locally stationary AR model for the $l$-th span is given by

$$y(t) = a^l(1)y(t-1) + \cdots + a^l(p)y(t-p) + \epsilon^l(t), \quad for \;\; s_{l-1} < s \leq s_l.$$

The basic idea of locally stationary AR model fitting is as follows. We divide $N$ data into some spans (sets of data) and consider two competing models. The first model is two independent AR models fitted to two sets of data respectively. The second model is an AR model fitted to the set of pooled data. For model selection following two AIC are compared.

$$AIC_1 = (n_1 - p)\mathtt{log}(2\pi\hat{\sigma}_{p_0}^2) + n_2\mathtt{log}(2\pi\hat{\sigma}_{p_1}^2) + n_1 + n_2 - p + 2(p_0 + p_1 + 2)$$

$$AIC_2 = (n_1 + n_2 - p)\mathtt{log}(2\pi\hat{\sigma}_{p_2}^2) + n_1 + n_2 - p + 2(p_2 + 1)$$

Here $n_1$ and $n_2$ are the length of data of each span and $p_0$, $p_1$ and $p_2$ are the orders of AR models fitted to the sets of data $\{y(1), \cdots, y(n_1)\}$, $\{y(n_1+1), \cdots, y(n_1+n_2)\}$ and $\{y(1), \cdots, y(n_1+n_2)\}$ respectively. (mlocar)

For the multi-variate AR model fitting AIC is defined in the same way. (mlomar)

## 2-4 R function in TIMSAC-84

**decomp()** : Decomposition of a nonstationary time series

*decomp* is a procedure for the decomposition of a nonstationary time series into several possible components. that is, a time series $y(t)$ is represented by the sum of components

$$y(t) = T(t) + AR(t) + S(t) + TD(t) + W(t)$$

where $T(s)$ : a trend component
$AR(t)$ : an autoregressive (AR) component
$S(t)$ : a seasonal component
$TD(t)$ : a trading day effect component
$W(t)$ : a white noise