Algorithm 808: ARFIT—A Matlab Package for the Estimation of Parameters and Eigenmodes of Multivariate Autoregressive Models

TAPIO SCHNEIDER
New York University
and
ARNOLD NEUMAIER
Universität Wien

ARFIT is a collection of Matlab modules for modeling and analyzing multivariate time series with autoregressive (AR) models. ARFIT contains modules for fitting AR models to given time series data, for analyzing eigenmodes of a fitted model, and for simulating AR processes. ARFIT estimates the parameters of AR models from given time series data with a stepwise least squares algorithm that is computationally efficient, in particular when the data are high-dimensional. ARFIT modules construct approximate confidence intervals for the estimated parameters and compute statistics with which the adequacy of a fitted model can be assessed. Dynamical characteristics of the modeled time series can be examined by means of a decomposition of a fitted AR model into eigenmodes and associated oscillation periods, damping times, and excitations. The ARFIT module that performs the eigendecomposition of a fitted model also constructs approximate confidence intervals for the eigenmodes and their oscillation periods and damping times.

Categories and Subject Descriptors: G.3 [**Mathematics of Computing**]: Probability and Statistics—*Markov processes*; *Multivariate statistics*; *Statistical software*; *Stochastic processes*; *Time series analysis*; G.4 [**Mathematics of Computing**]: Mathematical Software—*Documentation*; I.6.4 [**Simulation and Modeling**]: Model Validation and Analysis; J.2 [**Computer Applications**]: Physical Sciences and Engineering—*Earth and atmospheric sciences*; *Mathematics and statistics*

General Terms: Algorithms, Documentation

Additional Key Words and Phrases: Confidence intervals, eigenmodes, least squares, model identification, Matlab, order selection, parameter estimation, principal oscillation pattern

Authors' addresses: T. Schneider, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012; email: tapio@cims.nyu.edu; A. Neumaier, Institute for Mathematics, Universität Wien, Strudlhofgasse 4, A-1090, Wien, Austria; email: neum@cma.univie.ac.at.

1. OVERVIEW

ARFIT is a collection of Matlab modules for modeling and analyzing multivariate time series with autoregressive (AR) models. The stochastic model that underlies the ARFIT modules is the m-variate autoregressive model of order p (AR(p) model),

$$v_\nu = w + \sum_{l=1}^{p} A_l v_{\nu-l} + \varepsilon_\nu, \quad \varepsilon_\nu = \text{noise}(C), \tag{1}$$

a model for a stationary time series of m-dimensional state vectors v_ν that have been observed at equally spaced instants ν. The matrices $A_1, \ldots, A_p \in \mathcal{R}^{m\times m}$ are the coefficient matrices of the AR model, and the m-dimensional vectors $\varepsilon_\nu = \text{noise}(C)$ are uncorrelated random vectors with mean zero and covariance matrix $C \in \mathcal{R}^{m\times m}$. The m-dimensional vector w is a vector of intercept terms, which allows for a nonzero mean of the time series,

$$\langle v_\nu \rangle = (I - A_1 - \cdots - A_p)^{-1} w, \tag{2}$$

where $\langle\cdot\rangle$ denotes an expected value. (For an introduction to modeling multivariate time series with such AR models, see Lütkepohl [1993].) ARFIT contains modules (i) for estimating, from a sample of time series data v_ν, the order p of an AR(p) model, the intercept vector w, the coefficient matrices $A_1, \ldots, A_p$, and the noise covariance matrix C; (ii) for assessing the adequacy of a fitted AR model; (iii) for analyzing the eigendecomposition of a fitted AR(p) model; and (iv) for simulating AR processes.

1.1 Estimating the Parameters of an AR Model

For the selection of the order p of an AR(p) model and for the estimation of the parameters w, $A_1, \ldots, A_p$, and C, the stepwise least squares algorithm described by Neumaier and Schneider [2001] is implemented in ARFIT. Given a time series of $N + p$ state vectors v_ν ($\nu = 1 - p, \ldots, N$) and a lower bound $p_{\min}$ and an upper bound $p_{\max}$ on the model order, the ARFIT module **arfit** evaluates criteria for the selection of the model order for a sequence of AR models of successive orders $p_{\min}, \ldots, p_{\max}$ and computes the parameters w, $A_1, \ldots, A_{p_{\text{opt}}}$, and C for a model of the optimum order p_{opt}.

The optimum order p_{opt} of an AR model is generally chosen as the optimizer of an order selection criterion [Lütkepohl 1993, Chapter 4]. The order selection criteria implemented in ARFIT are Akaike's [1971] Final Prediction Error (FPE) criterion and Schwarz's [1978] Bayesian Criterion (SBC). Lütkepohl [1985] compared these and other order selection criteria in a simulation study and found that Schwarz's Bayesian Criterion chose the correct model order most often and led, on the average, to the smallest mean-squared prediction error of the fitted AR models. Schwarz's Bayesian Criterion is therefore the default order selection criterion of ARFIT.

In the stepwise least squares algorithm, the order selection criteria are evaluated for models of order $p_{\min}, \ldots, p_{\max}$ by stepwise downdating a regularized QR factorization of a data matrix for a model of the maximum order $p_{\max}$. From the QR factorization for a model of order $p_{\max}$, approximate least squares estimates of the parameters $w, A_1, \ldots, A_{p_{\text{opt}}}$, and C are computed for the model of the order p_{opt} that optimizes the order selection criterion. The stepwise least squares estimation is computationally efficient in particular when the time series data v_ν are high-dimensional. Neumaier and Schneider [2001] discuss properties of the stepwise least squares algorithm and compare this algorithm with other methods for the estimation of AR parameters.

Basing finite-sample inferences on the asymptotic distribution of the least squares estimator makes it possible to construct approximate confidence intervals for the intercept vector w and for the coefficient matrices $A_1, \ldots, A_{p_{\text{opt}}}$ (e.g., Lütkepohl [1993]). The ARFIT module **arconf** constructs approximate 95% confidence intervals for the intercept vector w and for the coefficient matrices $A_1, \ldots, A_{p_{\text{opt}}}$ (see Neumaier and Schneider [2001] for details).

1.2 Assessing the Adequacy of a Fitted Model

Before the structure of a fitted AR model is analyzed or a fitted AR model is used for predictions, it is necessary to assess whether the fitted model provides an adequate representation of the given time series (cf. Tiao and Box [1981]). Various tests of the adequacy of a fitted model are described by Brockwell and Davis [1991] and by Wei [1994]. Tests of model adequacy are usually tests of whether the statistics of the residuals

$$\hat{\varepsilon}_\nu = v_\nu - \hat{w} - \sum_{l=1}^{p} \hat{A}_l v_{\nu-l}, \qquad \nu = 1, \ldots, N, \tag{3}$$

are consistent with the assumptions intrinsic to the AR model (1). (The hat-accent $\hat{A}$ designates an estimate of the quantity A.) A principal assumption intrinsic to AR models is that the noise vectors ε_ν be uncorrelated. Uncorrelatedness of the noise vectors is, for example, invoked in the derivation of the least squares estimator. With the ARFIT module **acf**, the autocorrelation function of the residuals (3) can be examined graphically (cf. Brockwell and Davis [1991]). With the ARFIT module **arres**, the hypothesis that the residuals are uncorrelated can be tested (cf. Li and McLeod [1981]).

The uncorrelatedness of the residuals is tested using estimates $\hat{R}(l)$ of the lag l correlation matrices that consist of the elements

$$\hat{R}_{ij}(l) = \frac{\hat{c}_{ij}(l)}{\sqrt{\hat{c}_{ii}(0)\hat{c}_{jj}(0)}}, \qquad l = 1, \ldots, k,$$

where the matrices

$$\hat{c}(l) = \sum_{\nu=l+1}^{N} (\hat{\varepsilon}_{\nu-l} - \hat{\mu})(\hat{\varepsilon}_{\nu} - \hat{\mu})^T$$

contain the lagged residual cross-products and the vector

$$\hat{\mu} = \frac{1}{N}\sum_{\nu=1}^{N}\hat{\varepsilon}_{\nu}$$

is the mean of the residuals. Li and McLeod [1981] show that under the null hypothesis of model adequacy, for Gaussian noise, and for sufficiently large k the quantity

$$Q_k = N\sum_{l=1}^{k} x_{\hat{R}(l)}^T\left(\hat{R}(0)^{-1} \otimes \hat{R}(0)^{-1}\right)x_{\hat{R}(l)} + \frac{m^2k(k+1)}{2N} \tag{4}$$

is asymptotically χ^2-distributed with

$$f = m^2(k - p)$$

degrees of freedom. Here, the vector x_A consists of the components of the matrix A, arranged as a vector by stacking adjacent columns; the superscript ^{T} denotes transposition; and $P \otimes Q$ is the Kronecker product of P and Q (as returned by the Matlab function **kron**). The maximum lag k up to which residual correlation matrices are computed should be chosen such that, for models of order $l > k$, the estimated AR parameter matrices $\hat{A}_l$ (often called the lag-l partial autocorrelation matrices [Tiao and Box 1981]) are consistent with zero. For models of low order, the choice $k = 20$ should, in practice, suffice. In doubt, it is better that the lag k be chosen too large than too small.

From the asymptotic distribution of the Li-McLeod statistic Q_k, it follows that the hypothesis that the residuals are uncorrelated is rejected with approximate significance level β if the statistic Q_k exceeds the $(1-\beta)$-quantile $\chi^2_{1-\beta}(f)$ of a χ^2-distribution with f degrees of freedom,

$$Q_k > \chi^2_{1-\beta}(f). \tag{5}$$

For a significance level β, the critical value $\chi^2_{1-\beta}(f)$ of the statistic Q_k is a solution of

$$\beta = 1 - \Phi\left(\frac{f}{2}, \frac{\chi^2_{1-\beta}(f)}{2}\right)$$

where

$$\Phi(\alpha, x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t}\, dt$$

is the incomplete gamma function. Equivalently, the hypothesis that the residuals are uncorrelated is rejected if the estimated significance level

$$\hat{\beta} = 1 - \Phi\left(\frac{f}{2}, \frac{Q_k}{2}\right) \tag{6}$$

satisfies $\hat{\beta} < \beta$, where β indicates the probability that the test rejects a true hypothesis. As a typical value one may choose $\beta = 0.05$.

This extension of the univariate portmanteau test (cf. Box and Jenkins [1970] and Box and Pierce [1970]) to the multivariate case is implemented in the ARFIT module **arres**. We will refer to this test of the uncorrelatedness of the residuals as the modified Li-McLeod portmanteau (LMP) test.

1.3 Analyzing the Eigendecomposition of a Fitted Model

Tiao and Box [1981] discuss how structural analyses of a fitted AR model can yield insight into dynamical characteristics of the system being modeled. The eigendecomposition of a fitted AR model, described by Neumaier and Schneider [2001], is a structural analysis of an AR model that allows one to examine characteristics of oscillatory dynamics of a system. The ARFIT module **armode** performs the eigendecomposition of a fitted AR model. It computes the estimated eigenmodes and their oscillation periods, damping times, and excitations, as well as approximate confidence intervals for the eigenmodes, periods, and damping times. Eigenmodes computed with **armode** might, for example, be analyzed graphically (cf. Neumaier and Schneider [2001] and von Storch and Zwiers [1999, Chapter 15]).

1.4 Simulating AR Processes

A realization of an AR process can be simulated by substituting Gaussian pseudorandom vectors with covariance matrix C for the noise vectors ε_ν in the AR model (1). Gaussian pseudorandom vectors with covariance matrix C can be obtained by multiplying Gaussian white noise vectors by the Cholesky factor R of the covariance matrix $C = R^T R$. In this way, the ARFIT module **arsim** generates Gaussian pseudorandom vectors and simulates realizations of AR processes (cf. Lütkepohl [1993]).

2. DESCRIPTION OF MODULES

The Matlab implementations of the above methods are extensively annotated and include online documentation with information on the usage of the different ARFIT modules. Neumaier and Schneider [2001] give detailed descriptions of the implemented algorithms and the results of numerical tests. Presented here is a summary of what functions the different ARFIT modules fulfill.

Matlab modules come in the form of what are called M-files, files with the generic name *module*.**m**, where *module* is the module name. ARFIT consists of the following modules:

arfit Given a minimum model order $p_{\min}$ and a maximum model order $p_{\max}$, **arfit** uses the stepwise least squares algorithm of Neumaier and Schneider [2001] both to evaluate the order selection criteria FPE and SBC for AR(p) models of order $p_{\min} \le p \le p_{\max}$ and to compute estimates of the parameters $A_1, \ldots, A_{p_{\text{opt}}}$, w, and C of the AR model of the optimum order p_{opt}. The optimum order p_{opt} is chosen as the optimizer either of Schwarz's Bayesian Criterion or of Akaike's Final Prediction Error.

arres Given the time series of state vectors v_ν and estimates of the parameters $A_1, \ldots, A_p$, and w of an AR(p) model, **arres** computes the time series of residuals (3) and the significance level $\hat{\beta}$ of the LMP statistic (6).

acf The module **acf** plots the sample autocorrelation function of a univariate time series. In assessing the adequacy of a fitted model, **acf** may be used to test whether the time series of residuals show significant autocorrelations. (The module **acf** requires the module **xcorr** from the Matlab *Signal Processing Toolbox*, which is not included in the standard distribution of Matlab.)

arconf The module **arconf** computes approximate 95% confidence intervals for the intercept vector w and for the AR coefficient matrices $A_1, \ldots, A_p$. The confidence coefficient for which **arconf** computes confidence intervals is an adjustable program parameter.

armode Given estimates of the coefficient matrices $A_1, \ldots, A_p$ of an AR(p) model, **armode** computes the eigenmodes of the fitted model and the associated oscillation periods, damping rates, and excitations. The module **armode** also computes approximate 95% confidence intervals for the eigenmodes, periods, and damping times. The confidence coefficient for which **armode** computes confidence intervals is an adjustable program parameter.

arsim Monte-Carlo simulation of AR processes.

ardem Demonstration of the modules contained in the ARFIT package.

These are the modules that a user will typically access. Some of these modules, however, require lower-level modules, which are also part of ARFIT:

arqr Regularized QR factorization for an AR model. The module **arqr** is required by the module **arfit**.

arord Evaluates order selection criteria for a sequence of AR models by successively downdating a QR factorization. The module **arord** is required by the module **arfit**.

adjph Multiplies a complex vector by a phase factor such that its real part and its imaginary part are orthogonal and the norm of the real part is greater than or equal to the norm of the imaginary part. The module **adjph** is required by the module **armode** to normalize the eigenmodes of the AR model.

tquant Calculates quantiles of Student's t distribution. The modules **arconf** and **armode** require the module **tquant** in the construction of approximate confidence intervals.

If one places the M-files in a directory that Matlab can access and invokes the online help function of Matlab,

help *module*,

detailed information on the usage of the M-file *module* will be displayed. The script **ardem** demonstrates the main features of the M-files listed above. It illustrates with a simulated time series how ARFIT can be used in modeling and analyzing multivariate time series.

REFERENCES

AKAIKE, H. 1971. Autoregressive model fitting for control. *Ann. Inst. Stat. Math. 23*, 163–180.

BOX, G. E. P. AND JENKINS, G. M. 1970. *Time Series Analysis: Forecasting and Control*. Holden-Day, Inc., San Francisco, CA.

BOX, G. E. P. AND PIERCE, D. A. 1970. Distribution of the residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc. 65*, 1509–1526.

BROCKWELL, P. J. AND DAVIS, A. 1991. *Time Series: Theory and Methods*. 2nd ed. Springer-Verlag, New York, NY.

LI, W. K. AND MCLEOD, A. I. 1981. Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Statist. Soc. B43*, 231–239.

LÜTKEPOHL, H. 1985. Comparison of criteria for estimating the order of a vector autoregressive process. *J. Time Ser. Anal. 6*, 35–52. Correction, Vol 8 (1987), page 373.

LÜTKEPOHL, H. 1993. *Introduction to Multiple Time Series Analysis*. 2nd. Springer-Verlag, Berlin, Germany.

NEUMAIER, A. AND SCHNEIDER, T. 2001. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw. 27*, 1 (Mar.), 27–57.

SCHWARZ, G. 1978. Estimating the dimension of a model. *Ann. Stat. 6*, 461–464.
TIAO, G. C. AND BOX, G. E. P. 1981. Modeling multiple time series with applications. *J. Amer. Statist. Assoc. 76*, 802–816.
VON STORCH, H. AND ZWIERS, F. W. 1999. *Statistical Analysis in Climate Research*. Cambridge University Press, New York, NY.
WEI, W. W. S. 1994. *Time Series Analysis*. Addison-Wesley Publishing Co., Inc., Redwood City, CA.

Received: September 1997; revised: January 2000; accepted: October 2000

Guest Editor: Geoff Miller