STSA 2.1

Statistical Time Series Analysis

An O-Matrix Toolbox

Toolbox developed by Dimitrios D. Thomakos

1. Summary

The STSA toolbox is a collection of O-Matrix functions that can be used in analyzing time-dependent observations (time series). The functions are placed in eight different directories and can be used to perform a large variety of different types of analysis, including descriptive and graphical analysis, model identification, fitting and forecasting, residual diagnostic checking, spectral analysis, filtering and smoothing, optimization etc. The functions of the toolbox combine breadth of applications, easy of use and the speed and computing power of O-Matrix to provide you with an integrated working environment for analyzing your time series data. The functions can easily be incorporated into other functions by the user thus limiting the time need for writing programs or prototyping new routines. The STSA toolbox has functions that cover most of standard time series analysis plus a number of related functions not directly found in the main O-Matrix distribution or related products. A directory with examples using real-world and simulated data is provided and can be used as a starting point for immediately using the power of STSA.

2. What's new in version 2.1?

A number of enhancements and additional functions have been incorporated in the current version of STSA. When compared to version 2.0, the current version now has:

- There are 2 new subdirectories (POD and NONPAR) that expand the capabilities of STSA.
- All subdirectories have been updated and enlarged with new functions – STSA can now handle a greater array of time series problems.
- STSA now contains additional functions that can be used and in non-time series contexts (random numbers, statistical tools, generic optimization).
- Many functions now contain formatted screen output and automated graphics that greatly enhances the speed and quality of any analysis.

Some specific enhancements include:

- Many functions for performing singular spectrum analysis (SSA) of a time series in the POD directory, including decomposition, reconstruction and forecasting.
- Many functions for handling nonlinear time series using nonparametric models in the NONPAR directory, including local polynomials, cubic splines, functional coefficient models, partially linear models and various cross validation methods for automated bandwidth selection.
- Enhanced statistical tools (logistic regression for handling binary time series, enhanced QQ plot function).
- More examples that illustrate and expand the functions in the new directories.
- Extensive use of real and simulated datasets.

3. Where can STSA be of help to you?

If you use O-Matrix and any kind of dependent observations you will find that STSA is a valuable addition to your existing function toolboxes. The range of applications that can be tackled directly or indirectly with STSA includes:

- Financial and economic forecasting.
- Sales forecasting and inventory control.
- Modeling and forecasting of physical time series (hydrology, earth sciences, astronomy, oceanography and marine biology etc.)
- Simulation of time series models and their examination of their properties.
- Filtering and smoothing of any type of time series.
- Statistical estimation of parameters of time series models, either linear or nonlinear, with several types of optimization methods.

4. Function Directories

There are ten directories that hold the function collection of STSA. Each directory groups functions that “belong together”, in the sense that they are usually used together in modeling a sample realization. Of course they can be combined in any way the user wants and the examples directory shows how this can be done easily and effectively. Please note that the range of problems that can be addressed with STSA is much larger than the range of problems covered in the examples directory – if you don't see what you need contact us for additional support. The function directories are as follows:

- **ARMA:** functions for identification, estimation, testing and forecasting of time series based on the popular class of **A**uto**R**egressive **M**oving **A**verage univariate models. It also has functions for the analysis of bivariate models using transfer functions and multivariate models using **V**ector **A**uto**R**egressive models and transfer functions models.
- **BAYES:** functions for the identification and forecasting of structural time series based on an underlying Bayesian methodology.
- **FILTER:** functions for filtering and forecasting of univariate time series.
- **NONLIN:** functions for estimation and forecasting of univariate time series models based on nonlinear and nonparametric models.
- **NONPAR:** functions for nonparametric, nonlinear time series analysis.
- **OPTIMIZE:** functions for nonlinear optimization not available in the main O-Matrix distribution.
- **POD:** functions for singular spectrum analysis of a time series.
- **RNG:** function for generating random numbers from various statistical distributions.
- **SPECTRAL:** functions for spectral analysis of univariate and bivariate time series.
- **STATS:** various statistical functions that aid in the analysis of time series data. This subdirectory on its own greatly extends the statistical capabilities of the main distribution of O-Matrix.

5. Methods and Models

Below you will find a comprehensive list of the methods and models available with STSA.

a. ***Descriptive, graphical and statistical methods and models***

- Fast, formatted plot of a time series with rapid setting of default values.
- Descriptive statistics with formatted screen output.
- Statistical tests for assessing Gaussianity (normality).
- Box-Cox estimation of the optimal exponent to near Gaussianity.
- Box plots.
- Least squares and least absolute deviations regression.

- Regression with formatted screen output.
- Trend regression.
- Rolling least squares regression.
- Principal components analysis.
- Factor analysis using principal components.
- Generic nonlinear optimization (without constraints) using Newton-Raphson, BFGS and BHHH algorithms.
- Empirical probability distribution and cumulative distribution functions.
- Goodness-of-fit test for the equality of distributions between two time series with bootstrapped critical values.
- Maximum entropy bootstrap of a time series.
- Generate random numbers from various statistical distributions.

b. *Analytical methods and models– time domain*

i. Univariate

- Simulate a sample time series from a Gaussian ARMA model.
- Compute sequential lags of any order for a time series.
- Compute and plot the theoretical and sample autocovariance/autocorrelation and partial autocorrelations function of a time series (two methods).
- Compute the roots of the characteristic polynomials for an ARMA model.
- Compute the coefficients of the pure AR or the pure MA representation of a generic ARMA model.
- Identify the optimal lag order for an AR model using order selection criteria (corrected AIC and BIC).
- ARMA filter with known coefficients for a time series.
- Estimation of an AR model using Yule-Walker, nonlinear least squares or least absolute deviations.
- Estimation of the parameters of an ARMA model using nonlinear least squares with formatted screen output.
- Diagnostic checking of the residuals from an estimated ARMA model with formatted screen output.

- Forecasting using an estimated ARMA model.
- Durbin-Levinson-Whittle algorithm for computing innovations.
- Generic, finite impulse response filtering of a time series.
- Savitzky-Golay filtering of a time series.
- Moving average and exponential smoothing of a time series.
- Smoothing and forecasting using the Holt-Winters method (with or without seasonal/cyclical component).
- Generic trend+cycle+noise modeling and forecasting of a time series.
- Bayesian modeling and forecasting of a time series.
- Estimation of an ARMA-GARCH model for a time series.
- Test for linearity of a time series.
- Nonparametric modeling and forecasting of a nonlinear time series.
- Estimation and forecasting using TAR and FCAR models for a nonlinear time series.
- Simulate a sample time series following fractional Gaussian noise.
- Simulate a sample time series following a generic ARFIMA model.
- Fractionally difference a time series with known fractional order.
- Estimation and forecasting using an ARFIMA model for a time series.
- Formal comparison of forecasts between two competing models using the Diebold-Mariano test.

ii. Multivariate

- Compute the cross-correlation function between two time series.
- Compute the impulse response coefficients of the linear filter between two time series.
- Test for Granger-type causality between two time series.

- Identify the optimal lag order for a VAR model using order selection criteria (corrected AIC and BIC).
- Estimate and forecast using a VAR model.
- Compute the coefficient matrices of the pure MA representation of a VAR model.
- Estimate and forecast using a transfer function model.
- Estimate and forecast any model that can be put in a time-invariant state space form using Kalman filtering.

c. ***Analytical methods – spectral domain***

   i. **Univariate**

   - Discrete and fast Fourier transforms of a time series.
   - Compute the sample spectrum (periodogram) of a time series.
   - Compute the power spectrum of a time series using the sample autocovariances, an autoregressive approximation or the smoothed periodogram.
   - Plot the periodogram and the power spectrum of a time series.
   - Estimate the fractional order of a time series using the GPH regression or the Whittle approximation methods.

   ii. **Multivariate**

   - Compute the cross-spectrum of a time series.
   - Compute the co-spectrum, quad-spectrum, amplitude and phase between two time series.
   - Compute the squared coherence between two time series.

6. Examples and References

There are 33 example files that illustrate the use of most STSA functions located in the examples directory of the STSA distribution. Browsing through the examples is the fastest way of getting started using STSA for your own needs. The examples use real-world data, replicating previously published solutions, as well as simulated data.

A number of standard references in the field, which were used in constructing the functions in the toolbox, are given below.

- [1] "Time Series Analysis and its Applications" (2000) by R. H. Shumway and D. S. Stoffer, New York, Springer-Verlag.
- [2] "Introduction to Multiple Time Series Analysis" (1993), 2nd. edition, by H. Lutkepohl, New York, Springer-Verlag.
- [3] "Time Series Analysis: Forecasting and Control" (1994), 3rd edition, by G. E. P. Box, G. M. Jenkins and G. C. Reinsel, New Jersey, Prentice-Hall.
- [4] "Introduction to Statistical Time Series" (1996), 2nd edition, by W. A. Fuller, New York, John Wiley.
- [5] "Time Series: Theory and Methods" (1991), 2nd. edition, by P. J. Brockwell and R. A. Davis, New York, Springer-Verlag.
- [6] "Bayesian Forecasting and Dynamic Models" (1997) by M. West and J. Harrison, New York, Springer-Verlag.
- [7] "Nonlinear Time Series: A Dynamical System Approach", (1990), by H. Tong, Oxford University Press.
- [8] "Functional Coefficient Time Series Models for Nonlinear Time Series" (2000) by Z. Cai, J. Fan and Q. Yao, Journal of the American Statistical Association, 95, pp. 941-956.
- [9] "Nonparametric Econometrics" (1999) by A. Pagan and A. Ullah, Cambridge University Press.
- [10] "Statistics for Long-Memory Processes" (1994) by Jan Beran, CRC Press.
- [11] "Nonlinear Time Series: Parametric and Nonparametric Methods" (2003) by J. Fan and Q. Yao, New York, Springer-Verlag.