Classification of stationary time series by functional data analysis

Jonas Andersson
Uppsala University, Department of Information Science
Box 513
SE-751 20 Uppsala, Sweden
jonas.andersson@dis.uu.se

This paper presents a method for classification of stationary time series. The method is based on functional principal component analysis and exploits features of estimators of continuous spectral densities. It is evaluated by classifying simulated time series from the same process but with different parameter values as well as time series from structurally different processes. An illustration with data of Swedish stock return volatility is given.

Introduction

The need to classify observed time series occur in many scientific problems. One example is in cardiology, where ECG signals from different patients, are classified in order to see whether a particular type of patients, e.g. patients with a history of some disease, has a different pattern than a control group. Another example is the classification of signals, coming either from earthquakes or from nuclear explosions, in order to monitor a nuclear test ban treaty. Ways to do this, by means of discrimination- and cluster analysis has been presented in e.g. Shumway (1982) and Kakizawa, Shumway and Taniguchi (1998). A third example is in empirical finance when the time dynamics of asset return volatility is investigated for different type of companies.

In this paper an exploratory approach to classify stationary time series on the basis of their spectral densities is introduced. The mehtod is based on functional principal component analysis (e.g., Ramsay & Silverman, 1997). The idea is to use smoothed versions of the periodograms and consider them as observations on continuous curves. Thus, an underlying assumption about a continuous spectral density is used without imposing further assumptions. The method is investigated both for situations when the data generating process differs only by their parameter values and when they are structurally different.

The method

Consider n covariance stationary time series processes $\{x_{k,t}\}_{t=-\infty}^{\infty}$ with autocovariance functions $\{\gamma_{k,j}\}_{j=0}^{\infty}$ for $k = 1, ..., n$ and spectral densities $h_k(\omega)$. I also assume that the autocovariance function is square summable implying a continuous spectral density. Suppose further that a sample of size T from each of these processes is available.

The first step in a functional data analysis is to represent the observed data as the smooth functions they are assumed to be sampled from. The term functional data refer to data in this form, i.e. a function, determined by the observed data, which can be calculated for all arguments in a predefined interval. The function used in this paper for the classification is the periodogram of the time series, smoothed by a common lag window λ

$$(1) \qquad S_k(\omega) = \frac{1}{2\pi} \sum_{s=-(T-1)}^{T-1} \lambda(s)\widehat{R}_k(s)cos(s\omega)$$

If λ is such that the number of used lags in the right hand side of (1) tends to infinity with T, but slower, (1) can be shown to be a consistent estimator of $h_k(\omega)$, see e.g. Priestley (1981). The representation (1) is the functional representation of the data matrix. From it $S_k(\omega)$ can be calculated for any ω. The classification is made by functional principal component analysis. The procedure used is the one by Ramsay and Silverman (1997) who suggest to do this by first representing the data by a system of basis functions

$$(2) \qquad S_k(\omega) = c_k' \phi(\omega)$$

where c_k and $\phi(\omega)$ are $Jx1$ vectors. The same is done with the principal component functions which will be calulated.

$$(3) \qquad \xi(\omega) = \phi(\omega)' b$$

By considering (1) obvious candidates for c_k and $\phi(\omega)$ are obtained, namely

$$(4) \qquad c_{k,s} = \lambda(s) \widehat{R}_k(s)$$

and

$$(5) \qquad \phi_s(\omega) = cos(s\omega)$$

where $s = 1, ..., S$. The approach is to perform principal component analysis on a rotated version of a matrix, in which the rows consist of the transposed coefficient vectors, c_k, $k = 1, ..., n$. By (3) the obtained principal component vectors, b_j, are transformed into principal component functions (PCF), $\xi_j(\omega)$, $j = 1, ..., J$. Scores can also be calculated for each time series and PCF. The scores of the first or the first two PCF's can be used to classify the time series.

Simulation study and application

The performance of the method when different time series are generated by the same data generating process (DGP) but with different parameter values is investigated as well as the situation where the DGP's are structurally different. The method is finally illustrated by applying it to a data set of volatilities of Swedish stocks.

REFERENCES

Kakizawa, Y., Shumway, R.H., and Taniguchi, M. (1998) Discrimination and clustering for multivariate time series, Journal of the American Statistical Association 93, 328-340.

Priestley, M.B. (1981) Spectral Analysis and Time Series. Academic press. San Diego.

Ramsay, J.O. and Silverman, B.W. (1997) Functional Data Analysis. Springer. New York.

Shumway R.H. (1982) Discriminant analysis for time series. In Classification, Pattern Recognition and Reduction of Dimensionality, Handbook of Statistics Vol 2, pp 1-46. P.R. Krishnaiah and Kanal, L.N. eds. Amsterdam, North Holland.

ACKNOWLEDGEMENTS

The author would like to thank Johan Lyhagen for valuable comments and Carolina Meier-Hirmer for her help translating the résume into French. This research was partly financed by The Sasakawa Young Leadership Fund.

RÉSUMÉ

Cet article présente une méthode de classification de séries temporelles stationnaires. La méthode est basée sur l'analyse en composantes principales fonctionnelle et elle utilise les propriets des estimateurs de densités spéctrales continues. Elle est evaluée par la classifications de séries temporelles simulées, issues du mme processus avec differents paramtres et des séries temporelles issues de processus structuellement differents. Des données sur la volatitilé du rendement d'actions suédoises servent d'exemple.