PALGRAVE ADVANCED TEXTS IN ECONOMETRICS

Series Editor: Michael Clements

palgrave pivot

SINGULAR SPECTRUM ANALYSIS

Using R

Hossein Hassani
Rahim Mahmoudvand

Palgrave Advanced Texts in Econometrics

Series Editor
Michael Clements
ICMA Centre, Henley Business School
University of Reading
Wheatley, UK

Palgrave Advanced Texts in Econometrics is a series that provides coverage of econometric techniques, applications and perspectives at an advanced research level. It will include research monographs that bring current research to a wide audience; perspectives on econometric themes that develop a long term view of key methodological advances; textbook style presentations of advanced teaching and research topics. An over-riding theme of this series is clear presentation and accessibility through excellence in exposition, so that it will appeal not only to econometricians, but also to professional economists and, particularly, to Ph.D students and MSc students undertaking dissertations. The texts will include developments in theoretical and applied econometrics across a wide range of topics and areas including time series analysis, panel data methods, spatial econometrics and financial econometrics.

More information about this series at
http://www.palgrave.com/gp/series/14722

Hossein Hassani · Rahim Mahmoudvand

Singular Spectrum Analysis

Using R

palgrave
macmillan

Hossein Hassani
Research Institute of Energy Management and Planning
University of Tehran
Tehran, Iran

Rahim Mahmoudvand
Department of Statistics
Bu-Ali Sina University
Hamedan, Iran

Palgrave Advanced Texts in Econometrics
ISBN 978-1-137-40950-8 ISBN 978-1-137-40951-5 (eBook)
https://doi.org/10.1057/978-1-137-40951-5

Library of Congress Control Number: 2018941884

Cover illustration: Pattern adapted from an Indian cotton print produced in the 19th century

Printed on acid-free paper

This Palgrave Pivot imprint is published by the registered company Macmillan Publishers Ltd. part of Springer Nature
The registered company address is: The Campus, 4 Crinan Street, London, N1 9XW, United Kingdom

Preface

Time series analysis is crucial in the modern world as time series data emerge naturally in the field of statistics. As a result, the application of time series analysis covers diverse areas, including those relating to ecological and environmental data, medicine and more importantly economic and financial time series analysis. In the past, time series analysis was restricted by the necessity to meet certain assumptions, for example, normality. In addition, the presence of outlier events, such as the 2008 recession, which causes structural changes in time series data, has further implications by making the time series non-stationary. Whilst methods have been developed using condemning time series models, such as variations of autoregressive moving average models, ARIMA models, such methods are largely parametric. In contrast, Singular Spectrum Analysis (SSA) is a non-parametric technique and requires no prior statistical assumptions such as stationarity or linearity of the series and works with both linear and non linear data. In addition, SSA has outperformed methods such as ARIMA, ARAR and Holt-Winters in terms of forecast accuracy in a number of applications. The SSA method consists of two complementary stages, known as decomposition and reconstruction, and both stages include two separate steps. At the first stage the time series is decomposed and at the second stage the original series is reconstructed and this series, which is noise free, is then used to forecast new data points. The practical benefits of SSA have resulted in its wide using over the last decade. As a result, the successful applications of SSA can now be identified across varying disciplines such as physics, meteorology, oceanology, astronomy, medicine, climate data, image processing, physical sciences, economics and finance. Practically there are few programs, such as SAS and Caterpillar, which allow performing the SSA technique, but these require payments which are sometimes not economical for an individual researcher. R is an open-source software package that was developed by Robert Gentleman and Ross Ihaka at the University of Auckland in 1999. Since then, it has experienced a huge growth in popularity within a short span of time. R is a programme which allows the user to create their own objects, functions and packages. The R system is command driven and it documents the analysis steps making it easy to reproduce or update the analysis and figure errors. R can be installed on any platform and is license free. A major advantage with R is that it allows integrating and interacting with other paid platforms such as SAS, Stata, SPSS and Minitab. Although there are some books in the market relating to SSA, this book is unique as it not only details the theoretical aspects underlying SSA, but also provides a comprehensive guide enabling the user to apply the theory in practice using the R software. This book provides the user with step-by-step coding and guidance for the practical application of the SSA technique to analyse their time series databases using R. We provided some basic R commands in Appendix, so the readers who are not familiar with this language please learn the very basics in the Appendix at first.

The help of Prof. Kerry Patterson and Prof. Michael Clements in editing the text is gratefully acknowledged. Discussions with Kerry and Michael helped to clarify various questions treated on the following pages. We thank both for their encouragement.

As this book endeavours to provide a concise introduction to SSA, as well as to its application procedures to time series analysis, it is mainly aimed at masters and Ph.D.'s students with a reasonably strong stats/maths background who wants to learn SSA, and is already acquainted with R. It is also appropriate for practitioners wishing to revive their knowledge of times series analysis or to quickly learn about the main mechanisms of SSA. On the time series side, it is not necessary to be an expert on what is popularly called Box-Jenkins modelling. In fact this could be a disadvantage since SSA modelling start from a somewhat different point and in doing so challenges some of the underlying assumptions of the Box-Jenkins approach.

Tehran, Iran — Hossein Hassani
Hamedan, Iran — Rahim Mahmoudvand
June 2018

Contents

List of Figures

List of Tables

CHAPTER 1

Univariate Singular Spectrum Analysis

Abstract A concise description of univariate Singular Spectrum Analysis (SSA) is presented in this chapter. A step-by-step guide for performing filtering, forecasting as well as forecasting interval using univariate SSA and associated R codes is also provided. After reading this chapter, the reader will be able to select two basic, but very important, choices of SSA: window length and number of singular values. The similarity and dissimilarity between SSA and principal component analysis (PCA) is also briefly deliberated.

Keywords Univariate SSA · Window length · Singular values
Reconstruction · Forecasting

1.1 Introduction

There are several different methods for analysing time series all of which have sensible applications in one or more areas. Many of these methods are largely parametric, for example, requiring linearity or nonlinearity of a particular form. An alternative approach uses non-parametric techniques that are neutral with respect to problematic areas of specification, such as linearity, stationarity and normality. As a result, such techniques can provide a reliable and often better means of analysing time series data. Singular Spectrum Analysis (SSA) is a relatively new non-parametric method that has proved its capability in many different time series applications ranging from economics to physics. For the history of SSA, see Broomhead et al. (1987), and Broomhead and King (1986a, b). SSA has subsequently

H. Hassani and R. Mahmoudvand, *Singular Spectrum Analysis*,
Palgrave Advanced Texts in Econometrics,
https://doi.org/10.1057/978-1-137-40951-5_1

been developed in several ways including multivariate SSA (Hassani and Mahmoudvand 2013), SSA based on minimum variance (Hassani 2010) and SSA based on perturbation (Hassani et al. 2011b) (for more information, see Sanei and Hassani (2016)).

The increased application of SSA is further influenced by the following. Firstly, the emergence of Big Data may increase noise in time series, which in turn results in a distortion of the signal, thereby hindering the overall forecasting process. Secondly, volatile economic conditions ensure that time series (in most cases) are no longer stationary in mean and variance, especially following recessions which have left behind structural breaks. This in turn results in a violation of the parametric assumptions of stationarity and prompts data transformations when adopting classical time series methods. Such data transformations result in a loss of information and by relying on a technique such as SSA, which is not bound by any assumptions, users can overcome the restrictions imposed by parametric models in relation to the structure of the data. It is also noteworthy that recently it has been shown that SSA can provide accurate forecasts before, during and after recessions. Thirdly, SSA can be extremely useful as it enables the user to decompose a time series and extract components such as the trend, seasonal components and cyclical components (Sanei and Hassani 2016), which can then be used for enhancing the understanding of the underlying dynamics of a given time series. Fourthly, SSA is also known for its ability to deal with short time series where classical methods fail due to a lack of observations (Hassani and Thomakos 2010).

A common problem in economics is that most of the times series we study contain many components such as trend, harmonic and cyclical comments, and irregularities. Trend extraction or filtering are difficult even if we assume there is a time series with additive components. In general, as in SSA too, the trend of a time series is considered as a smooth additive component that contains information about the general tendency of the series. The most frequently used approaches for trend extraction are, for instance, simple linear regression model, moving average filtering, Tramo-Seats, X-11, X-12, and the most common one, the Hodrick-Prescott (HP) filter. To apply each method, one needs to consider model's specification or parameters. Generally, one can classify trend extraction approaches into tow main categorizes; the Model-Based approach, and non-parametric approaches including SSA. The Model-Based approach assumes the specification of a stochastic time series model for the trend, which is usually either an ARIMA model or a state space model. On the other hand, the non-parametric filtering methods (i.e. the Henderson, and Hodrick-Prescott filters) do not require specification of a model; they are quite easy to apply and are used in all applied areas of time series analysis. However, there are a few disadvantages of using HP filter; (i) "the HP filter produces series with spurious dynamic relations that have no basis in the underlying data-generating process; (ii) a one-sided version of the filter reduces but does not eliminate spurious predictability and moreover produces series that do not have the properties sought by most potential users of the HP filter" Hamilton (2017).

Two main important applications of SSA are filtering and smoothing, and forecasting, which will be discussed in the following sections.

1.2 Filtering and Smoothing

The SSA technique decomposes the original time series into the sum of a small number of interpretable components, such as a slowly varying trend, oscillatory components and noise. The basic SSA method consists of two complementary stages: decomposition and reconstruction, of which each stage includes two separate steps. At the first stage the series is decomposed and, in the second stage, the filtered series is reconstructed; the reconstructed series is then used for forecasting new data points. A short description of the SSA technique is given below (for more details, see Hassani et al. 2012).

Stage I. Decomposition

We consider a stochastic process Y generating a sequence comprising N random variables: $Y_N \equiv \{Y_t\} \equiv \{Y_t\}_{t=1}^{N}$. The sequence is ordered in time. In practice, we deal with realizations, or outcomes, from this process which we index by $t = 1, \ldots, N$, and distinguish them from the underlying random variables by using lower case y, that is $Y_N = (y_1, \ldots, y_N)$.

1st Step: Embedding. Embedding can be considered as a mapping which transfers a one-dimensional time series $Y_N = (y_1, \ldots, y_N)$ into a multi-dimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1})^T \in \mathbf{R}^L$, where L is the *window length* (see Sect. 1.4.1), and $2 \leq L \leq N/2$ and $K \equiv N - L + 1$. The single input at this stage is the SSA choice of L. The result of this step is the trajectory matrix: