MULTIVARIATE TIME SERIES ANALYSIS

With R and Financial Applications
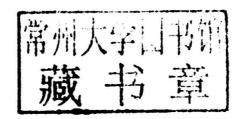
RUEY S. TSAY

WILEY

Multivariate Time Series Analysis

With R and Financial Applications

RUEY S. TSAY

Booth School of Business
University of Chicago
Chicago, IL

WILEY

Multivariate Time Series Analysis

To my teacher and friend
George

Preface

This book is based on my experience in teaching and research on multivariate time series analysis over the past 30 years. It summarizes the basic concepts and ideas of analyzing multivariate dependent data, provides econometric and statistical models useful for describing the dynamic dependence between variables, discusses the identifiability problem when the models become too flexible, introduces ways to search for simplifying structure hidden in high-dimensional time series, addresses the applicabilities and limitations of multivariate time series methods, and, equally important, develops a software package for readers to apply the methods and models discussed in the book.

Multivariate time series analysis provides useful tools and methods for processing information embedded in multiple measurements that have temporal and cross-sectional dependence. The goal of the analysis is to provide a better understanding of the dynamic relationship between variables and to improve the accuracy in forecasting. The models built can also be used in policy simulation or in making inference. The book focuses mainly on linear models as they are easier to comprehend and widely applicable. I tried to draw a balance between theory and applications and kept the notation as consistent as possible. I also tried to make the book self-contained. However, given the complexity of the subject, the level of coverage on selected topics may vary throughout the book. This reflects in part my own preference and understanding of the topics and in part my desire to keep the book at a reasonable length.

The field of high-dimensional data analysis is still under rapid developments, especially for dependent data. Omission of some important topics or methods is not avoidable for a book like this one. For instance, nonlinear models are not discussed, nor the categorical time series. Readers are advised to consult recent articles or journals for further development.

The book starts with some general concepts of multivariate time series in Chapter 1, including assessing and quantifying temporal and cross-sectional dependence. As the dimension increases, the difficulty in presenting multivariate data quickly becomes evident. I tried to keep the presentation in a compact form if possible. In some cases, scalar summary statistics are given. Chapter 2 focuses on vector

autoregressive (VAR) models as they are, arguably, the most widely used multivariate time series models. My goal is to make the chapter as comprehensive as possible for readers who are interested in VAR models. Both Bayesian and classical analyses of VAR models are included. Chapter 3 studies stationary vector autoregressive moving-average (VARMA) models. It begins with properties and estimation of vector moving-average (VMA) models. The issue of identifiability of VARMA models is investigated and properties of the models are given. Chapter 4 investigates the structural specification of a multivariate time series. Two methods are introduced to seek the simplifying structure hidden in a vector time series. These methods enable users to discover the *skeleton* of a linear multivariate time series. Chapter 5 deals with unit-root nonstationarity and cointegration. It includes the basic theory for understanding unit-root time series and some applications. In Chapter 6, I discuss factor models and some selected topics of multivariate time series. Both the classical and approximate factor models are studied. My goal is to cover all factor models currently available in the literature and to provide the relationship between them. Chapter 7 focuses on multivariate volatility modeling. It covers volatility models that are relatively easy to use and produce positive-definite volatility matrices. The chapter also discusses ways to detect conditional heteroscedasticity in a vector time series and methods for checking a fitted multivariate volatility model. Throughout the book, real examples are used to demonstrate the analysis. Every chapter contains some exercises that analyze empirical vector time series.

Software is an integral part of multivariate time series analysis. Without software packages, multivariate time series becomes a pure theoretical exercise. I have tried my best to write R programs that enable readers to apply all methods and models discussed in the book. These programs are included in the MTS package available in R. Readers can duplicate all the analyses shown in the book with the package and some existing R packages. Not a professional programmer, I am certain that many of the codes in MTS are not as efficient as they can be and are likely to have bugs. I would appreciate any suggestions and/or corrections to both the package and the book.

RUEY S. TSAY

Chicago, Illinois
September 2014

Acknowledgements

Contents

7   Multivariate Volatility Models   **399**

CHAPTER 1

Multivariate Linear Time Series

1.1 INTRODUCTION

Multivariate time series analysis considers simultaneously multiple time series. It is a branch of multivariate statistical analysis but deals specifically with dependent data. It is, in general, much more complicated than the univariate time series analysis, especially when the number of series considered is large. We study this more complicated statistical analysis in this book because in real life decisions often involve multiple inter-related factors or variables. Understanding the relationships between those factors and providing accurate predictions of those variables are valuable in decision making. The objectives of multivariate time series analysis thus include

1. To study the dynamic relationships between variables
2. To improve the accuracy of prediction

Let $z_t = (z_{1t}, \cdots, z_{kt})'$ be a k-dimensional time series observed at equally spaced time points. For example, let z_{1t} be the quarterly U.S. real gross domestic product (GDP) and z_{2t} the quarterly U.S. civilian unemployment rate. By studying z_{1t} and z_{2t} jointly, we can assess the temporal and contemporaneous dependence between GDP and unemployment rate. In this particular case, $k = 2$ and the two variables are known to be instantaneously negatively correlated. Figure 1.1 shows the time plots of quarterly U.S. real GDP (in logarithm of billions of chained 2005 dollars) and unemployment rate, obtained via monthly data with averaging, from 1948 to 2011. Both series are seasonally adjusted. Figure 1.2 shows the time plots of the real GDP growth rate and the changes in unemployment rate from the second quarter of 1948 to the fourth quarter of 2011. Figure 1.3 shows the scatter plot of the two time series given in Figure 1.2. From these figures, we can see that the GDP