

SPRINGER BRIEFS IN STATISTICS

Masanobu Taniguchi

Tomoyuki Amano

Hiroaki Ogata

Hiroyuki Taniai

Statistical Inference for Financial Engineering



Springer

SpringerBriefs in Statistics

For further volumes:
<http://www.springer.com/series/8921>

Masanobu Taniguchi · Tomoyuki Amano
Hiroaki Ogata · Hiroyuki Taniai

Statistical Inference for Financial Engineering

 Springer

Masanobu Taniguchi
Department of Applied Mathematics
Waseda University
Tokyo
Japan

Hiroaki Ogata
School of Business Administration,
Faculty of Urban Liberal Arts
Tokyo Metropolitan University
Tokyo
Japan

Tomoyuki Amano
Faculty of Economics
Wakayama University
Wakayama
Japan

Hiroyuki Taniai
School of International Liberal Studies
Waseda University
Tokyo
Japan

ISSN 2191-544X ISSN 2191-5458 (electronic)
ISBN 978-3-319-03496-6 ISBN 978-3-319-03497-3 (eBook)
DOI 10.1007/978-3-319-03497-3
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014934588

Mathematics Subject Classification (2010): 62P05, 91G70

© The Author(s) 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To our families

Preface

This book provides the foundation of statistical inference for financial engineering, which is a huge integration of economics, probability theory, statistics, time series analysis, operation research, etc. Black and Scholes introduced the modern option pricing theory assuming that the price process of an underlying asset follows a geometric Brownian motion. However, empirical studies for the price processes show that they do not follow the geometric Brownian motion, and that they often behave like as non-Gaussian dependent processes. Motivated by this, we investigate what stochastic models can describe the actual financial time series data sufficiently, and how to estimate the proposed models optimally. Concretely, we introduce various stochastic processes, e.g., non-Gaussian linear processes, nonlinear processes, long-memory processes, locally stationary processes, etc. For them we will develop the theory of optimal statistical inference based on local asymptotic normality (LAN), which is due to Le Cam. This book also describes a variety of statistical approaches, e.g., discriminant analysis, empirical likelihood method, control variate method, quantile regression, realized volatility, etc. The philosophy is that financial engineering should be constructed on optimal statistical approaches for plausible stochastic processes.

[Chapter 1](#) discusses time series modeling for financial data, and how to estimate such models optimally. The models include general non-Gaussian vector linear processes, nonlinear time series models, e.g., ARCH, CHARN, etc., and locally stationary processes, etc. Their inference is developed by LAN-based optimal theory. We give an approach based on integral functionals of nonparametric spectral estimators, which are applicable to a lot of problems in financial statistics. Option pricing will be discussed for a class of discretized diffusion processes. Also we address the problem of classification for locally stationary processes, and apply the results to actual financial data by use of dendrogram.

[Chapter 2](#) deals with empirical likelihood approaches for financial data. Since the observations are often supposed to be dependent, we will introduce an estimating function which corresponds to the differential functional of Whittle likelihood, i.e., empirical likelihood in frequency domain. After explaining the asymptotic properties of the frequency domain empirical likelihood estimator, we also introduce the extensions of the empirical likelihood method such as Cressie-Read power-divergence statistic and generalized empirical likelihood. As an application, we consider the generalized empirical likelihood estimation

method for the multivariate stable distributions. Several real data analyses are also included.

Chapter 3 gives applications of two statistical methods to the financial time series data. The first one is the control variate method. The control variate method is a method to reduce the variance of estimators by use of some information of another process, which is correlated with the process concerned. This method has been discussed in the case when the data are i.i.d. Since financial time series data are often dependent, we extend this method to dependent data and give the application to financial econometrics. The second one is the instrumental variable method. We introduce a stochastic regression model where the explanatory and disturbance processes are correlated. For this we apply the instrumental variable method to estimate the regression coefficients consistently. We also address the problem of CAPM to our analysis.

Chapter 4 introduces two techniques, which can be utilized in study of financial risks. The first one is the method called Quantile Regression (QR), which can be used to analyze the conditional quantile of financial assets. There, by means of rank-based semiparametrics, we provide the statistically efficient version of QR inference under the autoregressive conditional heteroskedasticity (ARCH). The second technique, the Realized Volatility (RV), estimates the conditional variance, or “volatility” of financial assets. Revealing the fact that its inference can be greatly affected by the existence of additional noise called market microstructure, we introduce and study the asymptotics of some appropriate estimator under the microstructure with ARCH property.

This book is suitable as a professional reference book on finance, statistics, and statistical financial engineering, or a text book for students who specialize these topics.

A part of this book was done in a collaboration between Research Institute for Science and Engineering, Waseda University and Government Pension Investment Fund (GPIF) of Japan. We thank all the members, especially, Prof. Takeru Suzuki and Dr. Takashi Yamashita (GPIF) for their cooperation. Our research was partially supported by the following Japanese Grant-in-Aids: A2324401 (Taniguchi, M.), B22700291 (Ogata, H.), B22700296 (Amano, T.).

Finally, we thank the editors of *SpringerBriefs in Statistics* for their kindness.

January 2013

Masanobu Taniguchi
Tomoyuki Amano
Hiroaki Ogata
Hiroyuki Taniai

Contents

1	Features of Financial Data	1
1.1	Introduction	1
1.2	Time Series Modeling for Financial Data	2
1.3	Optimal Inference for Various Return Processes	19
1.4	Introduction to Time Series Financial Engineering	27
	References	37
2	Empirical Likelihood Approaches for Financial Returns	41
2.1	Introduction	41
2.2	Empirical Likelihood Method for i.i.d. Data	42
2.3	Estimation with Frequency Domain Empirical Likelihood for Stationary Processes	43
2.4	Extensions of Empirical Likelihood	46
2.5	GEL for Multivariate Stable Distributions	51
2.6	Appendix	55
	2.6.1 Proof of Theorem 2.2	57
	2.6.2 Proof of Theorem 2.3	60
	References	63
3	Various Methods for Financial Engineering	65
3.1	Introduction	65
3.2	Control Variate Methods for Financial Data	66
3.3	Statistical Estimation for Stochastic Regression Models with Long Memory Dependence	74
	References	83
4	Some Techniques for ARCH Financial Time Series	85
4.1	Introduction	85
4.2	Quantile Regression and Its Semiparametric Efficiency for ARCH Series	86
	4.2.1 Model, Estimators, and Some Asymptotics	86
	4.2.2 Semiparametrically Efficient Inference	89
	4.2.3 Numerical Studies	95

- 4.3 Asymptotics of Realized Volatility with Non-Gaussian
 - ARCH(∞) Microstructure Noise 102
 - 4.3.1 Model, Estimators, and Main Results 103
 - 4.3.2 Numerical Studies 109
- 4.4 Appendix 112
 - 4.4.1 Proof of Theorem 4.1 112
- References 115
- Index** 117

Chapter 1

Features of Financial Data

Abstract This chapter discusses actual features of financial time series data, and how to model them statistically. Because the mechanism of financial market is obviously complicated, modeling for financial time series is difficult. For this, first, we look at some empirical characteristics of financial data. Then, we review and examine various time series models (e.g., ARCH, general linear process, non-stationary process, etc.), which show plausibility. Their estimation theory is provided in a unified fashion. Optimality of the estimation and testing, etc., is described based on the local asymptotic normality (LAN) due to Le Cam. The theory and models are very general and modern.

Keywords Nonlinear time series models · Vector linear processes · LAN approach · Option pricing · Classification for time series · Locally stationary processes

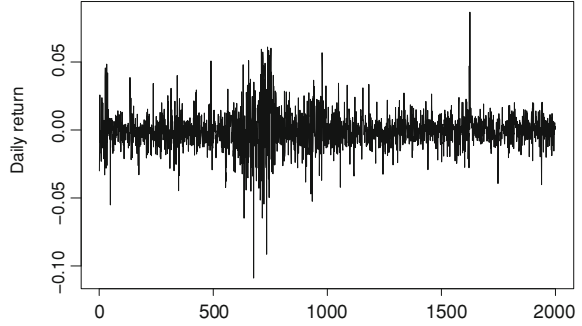
1.1 Introduction

In this chapter we address the problems of modeling for financial time series data and the optimal inference for various models. Then an introduction to financial engineering based on time series analysis is provided.

Section 1.2 studies statistical features of financial returns. They naturally lead to nonlinear time series models, e.g., ARCH, GARCH, EGARCH, CHARN, etc. If we assume their stationarity, Wold decomposition theorem shows that they can be expressed as linear processes. For general non-Gaussian vector-valued linear processes, we develop the asymptotic theory for Whittle estimators and integral functional of nonparametric spectral density estimators. Then the problems of h-step ahead prediction and testing a strength of causality are discussed.

Lucien LeCam established one of the most important foundations of the general statistical asymptotic theory (e.g., see LeCam (1986)). He introduced the concept of local asymptotic normality (LAN) for the likelihood ratio of general statistical model

Fig. 1.1 Daily log return of IBM $\{X_t\}$



in i.i.d. case. Beautiful thing is that once LAN is proved, the asymptotic optimality of estimators and tests is described in terms of the LAN property.

Section 1.3 introduces an important class of nonlinear time series models, called ARCH(∞) with stochastic mean (ARCH(∞)-SM) models, which includes the usual ARCH and GARCH as special cases. For this class, we establish the LAN, and describe the framework of the optimal inference based on LAN, proving the optimality of MLE. This stream of discussion is extended to a very general class of conditional heteroscedastic autoregressive nonlinear (CHARN) models and nonstationary processes, called locally stationary processes.

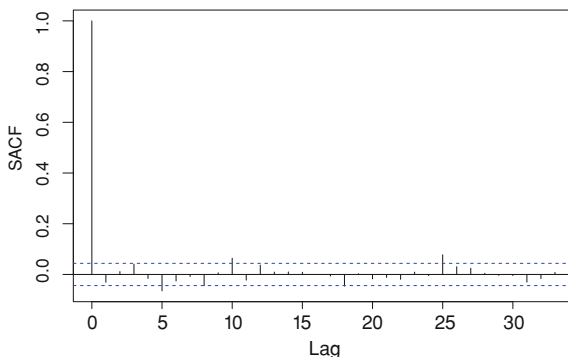
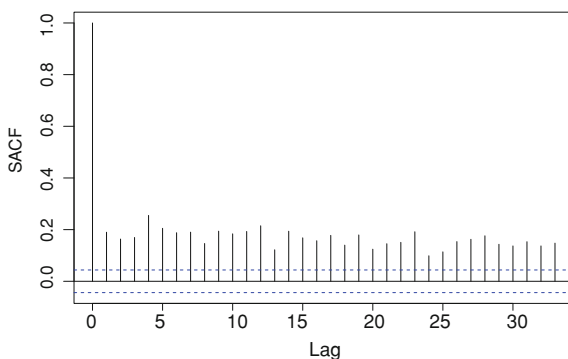
Section 1.4 develops the arguments of financial engineering based on time series analysis. Fundamental concepts, i.e., arbitrage-free, self-financing portfolio, completeness, etc., are explained. We discuss the pricing a European call option for discretized diffusion returns. Classification method of locally stationary processes is introduced. Using a distance between their nonparametric spectral estimators, we executed the hierarchical clustering for daily log-stock returns of 13 companies. The dendrogram classifies the type of industry clearly, which implies that the method will be useful for the problem of credit rating.

Because this book is mainly based on the research papers of the authors, it is recommendable to refer Gouriéroux and Jasiak (2001) for financial econometrics, Van der Vaart (1998) for the LAN approach and Brockwell and Davis (1991) for general time series analysis.

1.2 Time Series Modeling for Financial Data

We begin by looking at actual stock price data. Figure 1.1 plots the daily log return of IBM from October 16, 2003 to September 26, 2011. Write the observed stretch as X_1, X_2, \dots, X_n . As a fundamental analysis we often examine the behavior of the sample autocorrelation function (SACF):

$$\hat{\rho}_{X_t}(l) \equiv \frac{\sum_{t=1}^{n-l} (X_{t+l} - \bar{X}_n)(X_t - \bar{X}_n)}{\sum_{t=1}^n (X_t - \bar{X}_n)^2} \quad (1.1)$$

Fig. 1.2 SACF of $\{X_t\}$ **Fig. 1.3** SACF of $\{X_t^2\}$ 

where $\bar{X}_n \equiv n^{-1} \sum_{t=1}^n X_t$. The SACF $\hat{\rho}_{X_t}(l)$ for $\{X_t\}$ is shown in Fig. 1.2. The SACF $\hat{\rho}_{X_t^2}(l)$ for square transformed data X_t^2 is shown in Fig. 1.3. From (1.1), $\hat{\rho}_{X_t}(l)$ shows a strength of interrelation between X_{t+l} and X_t , hence, if X_t 's are mutually independent or uncorrelated, $\hat{\rho}_{X_t}(l)$, ($l \neq 0$), will be near zero. Figures 1.2 and 1.3, respectively, suggest that X_t 's are almost uncorrelated, and that X_t^2 's are not uncorrelated, which leads to the following conclusion:

X_t 's are not mutually independent and the distribution of $\{X_t\}$ is non-Gaussian.

Observing these symptoms of financial data, Engle (1982) proposed an autoregressive conditional heteroscedastic model (ARCH(q)), which is defined as

$$\begin{cases} E(X_t | \mathcal{F}_{t-1}) = 0 & a.e., \\ Var(X_t | \mathcal{F}_{t-1}) = a_0 + \sum_{j=1}^q a_j X_{t-j}^2 & a.e., \end{cases} \quad (1.2)$$

where \mathcal{F}_{t-1} is σ -algebra generated by $\{X_{t-1}, X_{t-2}, \dots\}$ and $a_0 > 0$, $a_j \geq 0$, $j = 1, \dots, q$. A concrete representation of it is given by

$$\begin{cases} X_t = \epsilon_t \sqrt{h_t}, \\ h_t = a_0 + \sum_{j=1}^q a_j X_{t-j}^2, \end{cases} \quad (1.3)$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. $(0, \sigma^2)$ random variables. Evidently, X_t 's are uncorrelated, and it is seen that they are dependent and non-Gaussian generally. Bollerslev (1986) generalized ARCH(q) to

$$\begin{cases} E(X_t | \mathcal{F}_{t-1}) = 0 \quad a.e., \\ Var(X_t | \mathcal{F}_{t-1}) \equiv a_0 + \sum_{j=1}^q a_j X_{t-j}^2 + \sum_{j=1}^p b_j h_{t-j}^2 \quad a.e., \end{cases} \quad (1.4)$$

where $a_0 > 0$, $a_j \geq 0$, $j = 1, \dots, q$, $b_j \geq 0$, $j = 1, \dots, p$, which is called a generalized autoregressive conditional heteroscedastic model (GARCH(p, q)). Further, generalizing ARCH and GARCH, Giraitis et al. (2000) introduced the ARCH(∞) defined by

$$\begin{cases} X_t = \epsilon_t \sqrt{h_t} \\ h_t = a_0 + \sum_{j=1}^{\infty} a_j X_{t-j}^2 \end{cases} \quad (1.5)$$

where $a_0 > 0$, $a_j \geq 0$, $j = 1, \dots$, $\{\epsilon_t\}$ is a sequence of i.i.d. random variables, and ϵ_t is \mathcal{F}_t -measurable and independent of \mathcal{F}_{t-1} .

Empirically, it is known that stock returns are negatively correlated with changes in returns volatility, i.e., volatility tends to rise in response to “bad news” and to fall in response to “good news.” ARCH and GARCH models cannot describe this aspect. To allow the asymmetric effect, Nelson (1991) proposed the exponential GARCH model(EGARCH(p, q)):

$$\begin{cases} X_t = \epsilon_t \cdot \sigma_t \\ \log \sigma_t^2 = a_0 + \sum_{j=1}^p a_j \frac{|X_{t-j}| + \gamma_j X_{t-j}}{\sigma_{t-j}} + \sum_{j=1}^q b_j \log \sigma_{t-j}^2 \end{cases} \quad (1.6)$$

where a_j, b_j, γ_j are unknown parameters, and are allowed to be negative unlike ARCH and GARCH parameters. If $a_j \gamma_j < 0$, $j = 1, \dots, p$, then we can understand the asymmetry of EGARCH.

Recently, the following stochastic volatility model(SV(m)) has been introduced, and is defined by

$$\begin{cases} X_t = \sigma_t \epsilon_t \\ \log \sigma_t^2 - \alpha_1 \log \sigma_{t-1}^2 - \dots - \alpha_m \log \sigma_{t-m}^2 = \alpha_0 + v_t \end{cases} \quad (1.7)$$

where $\{\epsilon_t\} \sim \text{i.i.d.}(0, 1)$, $\{v_t\} \sim \text{i.i.d.}(0, \sigma_v^2)$, and $\{\epsilon_t\}$ and $\{v_t\}$ are mutually independent.

Incorporating ARCH with SV, Härdle et al. (1998) introduced the following conditional heteroscedastic autoregressive nonlinear model (denoted by CHARN):

$$\mathbf{X}_t = \mathbf{F}_\theta(\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}) + \mathbf{H}_\theta(\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-q})\boldsymbol{\epsilon}_t \quad (1.8)$$

where $\mathbf{F}_\theta : \mathbb{R}^{mp} \rightarrow \mathbb{R}^m$ and $\mathbf{H}_\theta : \mathbb{R}^{mq} \rightarrow \mathbb{R}^m \times \mathbb{R}^m$, are measurable functions depending on unknown parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^r$, $\{\mathbf{X}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$ are sequences of m -dimensional random vectors, $\{\boldsymbol{\epsilon}_t\} \sim \text{i.i.d.}(\mathbf{0}, \mathbf{V})$ and $\boldsymbol{\epsilon}_t$ is independent of $\{\mathbf{X}_s, s < t\}$. Although CHARN has been used for financial data, Kato et al. (2006) applied this model to analysis of brain and muscular waves. Hence, CHARN models are very general.

Besides the above, a lot of nonlinear time series models have been proposed. Among them, ARCH(∞) and CHARN models are typical and general. If we discuss estimation theory for nonlinear time series, their stationarity is one of the most fundamental assumptions. Lu and Jiang (2001) gave a sufficient condition for strict stationarity of CHARN model (1.8). Giraitis et al. (2000) gave sufficient conditions for strict stationarity and second-order stationarity of ARCH(∞) model (1.5).

For second-order stationary processes, we have the following convenient result (for proof, e.g., Hannan (1970), p.137).

Proposition 1.1 (Wold decomposition Theorem) *Any zero-mean second-order stationary process $\{X_t\}$ can be represented in the form of linear process:*

$$X_t = \sum_{j=0}^{\infty} a_j u_{t-j} + \eta_t, \quad (1.9)$$

where $a_0 = 1$, $\sum_{j=0}^{\infty} a_j^2 < \infty$, and

- (i) $E\{u_s u_t\} = \delta(t - s)\sigma^2$
- (ii) $E\{u_s \eta_t\} = 0$, for any $s, t \in \mathbb{Z}$
- (iii) η_t is purely deterministic, i.e., $\eta_t = E\{\eta_t | X_{t-1}, X_{t-2}, \dots\}$.

In view of Proposition 1.1, the class of linear processes is a sufficiently rich one, which contains classes of various stationary nonlinear processes. Thus, in what follows, we provide some fundamentals of linear processes.

Let $\{\mathbf{X}(t) = (X_1(t), \dots, X_m(t))'; t \in \mathbb{Z}\}$ be generated by

$$\mathbf{X}(t) = \sum_{j=0}^{\infty} \mathbf{A}(j)\mathbf{U}(t - j), \quad t \in \mathbb{Z} \quad (1.10)$$

where $\{\mathbf{U}(t) = (u_1(t), \dots, u_m(t))'\}$ is a sequence of m -vector random variables satisfying

$$E\{\mathbf{U}(t)\} = \mathbf{0}, \quad E\{\mathbf{U}(t)\mathbf{U}(s)'\} = \mathbf{0} \quad (t \neq s)$$

and

$$\text{Var}\{U(t)\} = K = \{K_{ab} : a, b = 1, \dots, m\} \text{ for all } t \in \mathbb{Z}.$$

Here, $A(j) = \{A_{ab}(j); a, b = 1, \dots, m\}$, $j \in \mathbb{Z}$, are $m \times m$ matrices with $A(0) = I_m$. We set

Assumption 1.1 $\sum_{j=0}^{\infty} \text{tr}\{A(j)KA(j)'\} < \infty$.

Then, $\{X(t)\}$ becomes a second-order stationary process with spectral density matrix

$$f(\lambda) = \{f_{ab}(\lambda); a, b = 1, \dots, m\} = (2\pi)^{-1} A(\lambda)KA(\lambda)^*,$$

where $A(\lambda) = \sum_{j=0}^{\infty} A(j)e^{ij\lambda}$.

In the inference for (1.10), we often deal with the integral functional of periodogram matrix $I_n(\lambda)$:

$$F_n \equiv \int_{-\pi}^{\pi} \text{tr}\{\phi(\lambda)I_n(\lambda)\}d\lambda, \quad (1.11)$$

where $\phi(\lambda)$ is an $m \times m$ matrix-valued continuous function on $[-\pi, \pi]$, and

$$I_n(\lambda) = (2\pi n)^{-1} \left\{ \sum_{t=1}^n X(t)e^{it\lambda} \right\} \left\{ \sum_{t=1}^n X(t)e^{it\lambda} \right\}^*.$$

To describe the asymptotics of $\sqrt{n}(F_n - F)$, where $F = \int_{-\pi}^{\pi} \text{tr}\{\phi(\lambda)f(\lambda)\}d\lambda$, we mention two approaches. The first approach is due to Hosoya and Taniguchi (1982). Consider the linear process (1.10). Denote by \mathcal{F}_t the σ -algebra generated by $\{U(s); s \leq t\}$.

Assumption (HT)

- (i) $\{U(t)\}$ is fourth-order stationary, and the fourth-order cumulants $c_{a_1 \dots a_4}^U(t_1, t_2, t_3) \equiv \text{cum}\{u_{a_1}(0), u_{a_2}(t_1), u_{a_3}(t_2), u_{a_4}(t_3)\}$, $(a_1, \dots, a_4 = 1, \dots, m)$, satisfy

$$\sum_{t_1, t_2, t_3 = -\infty}^{\infty} \left| c_{a_1 \dots a_4}^U(t_1, t_2, t_3) \right| < \infty.$$

- (ii) For each a_1, a_2 and s ,

$$\text{Var}[E\{u_{a_1}(t)u_{a_2}(t+s)|\mathcal{F}_{t-\tau}\} - \delta(s)K_{a_1 a_2}] = O(\tau^{-2-\epsilon}) \text{ for } \epsilon > 0,$$

uniformly in t .

(iii) For each a_1, a_2, a_3 and a_4 ,

$$E|E\{u_{a_1}(t_1)u_{a_2}(t_2)u_{a_3}(t_3)u_{a_4}(t_4)|\mathcal{F}_{t_1-\tau}\} - E\{u_{a_1}(t_1)u_{a_2}(t_2)u_{a_3}(t_3)u_{a_4}(t_4)\}| \\ = O(\tau^{-1-\eta}),$$

uniformly in t_1 , where $t_1 \leq t_2 \leq t_3 \leq t_4$ and $\eta > 0$.

(iv) For any $\rho > 0$ and for any integer $L \geq 0$, there exists $B_\rho > 0$ such that

$$E[T(n, s)^2 I\{T(n, s) > B_\rho\}] < \rho,$$

uniformly in n and s , where

$$T(n, s) = \left[\sum_{a_1, a_2=1}^n \sum_{r=0}^L \left\{ \sum_{t=1}^n \frac{u_{a_1}(t+s)u_{a_2}(t+s+r) - \delta(r)K_{a_1 a_2}}{\sqrt{n}} \right\}^2 \right]^{\frac{1}{2}},$$

and $I\{\cdot\}$ is the indicator function of $\{\cdot\}$.

(v) $f(\lambda) \in Lip(\alpha)$, the Lipschitz class of degree α , $\alpha > 1/2$.

In the literature, a higher order martingale difference condition for $\{U(t)\}$ is imposed, i.e.,

$$E\{u_{a_1}(t)|\mathcal{F}_{t-1}\} = 0 \quad a.s., \quad E\{u_{a_1}(t)u_{a_2}(t)|\mathcal{F}_{t-1}\} = K_{a_1 a_2} \quad a.s., \\ E\{u_{a_1}(t)u_{a_2}(t)u_{a_3}(t)|\mathcal{F}_{t-1}\} = \gamma_{a_1 a_2 a_3} \quad a.s., \quad \text{etc.} \quad (1.12)$$

(e.g., Dunsmuir and Hannan (1976)). The assumption (HT) (ii) and (iii) mean a kind of ‘‘asymptotically’’ higher order martingale difference condition, which is more natural than (1.12).

To elucidate the asymptotics of $\sqrt{n}(F_n - F)$, Hosoya and Taniguchi (1982) showed:

$\sqrt{n}(F_n - F)$ is approximated by a finite linear combination of

$$\sqrt{n} \left\{ \frac{1}{n} \sum_{t=1}^n X_{a_1}(t)X_{a_2}(t+s) - E\{X_{a_1}(t)X_{a_2}(t+s)\} \right\} \quad (1.13)$$

where $s = 0, 1, \dots, n-1, a_1, a_2 = 1, \dots, m$.

Recalling that $\{X(t)\}$ is generated by (1.10), we can see that (1.13) is approximated by a finite linear combination of

$$\frac{1}{\sqrt{n}} \left\{ \sum_{t=1}^n u_{a_1}(t) u_{a_2}(t+l) - \delta(l) K_{a_1 a_2} \right\}, \quad l \leq s. \quad (1.14)$$

Applying Brown's central limit theorem (Brown (1971)) to (1.14), Hosoya and Taniguchi (1982) elucidated the asymptotics of $\sqrt{n}(F_n - F)$.

The second approach is due to Brillinger (2001). Assume that $\{X(t)\}$ is strictly stationary and all the moments exist.

Let

$$c_{a_1 \dots a_k}^X(t_1, \dots, t_{k-1}) = \text{cum}\{X_{a_1}(0), X_{a_2}(t_1), \dots, X_{a_k}(t_{k-1})\},$$

for $a_1, \dots, a_k = 1, \dots, m; k = 2, 3, \dots$

Assumption (B) For each $j = 1, 2, \dots, k-1$ and any k -tuple a_1, a_2, \dots, a_k we have

$$\sum_{t_1, \dots, t_{k-1} = -\infty}^{\infty} (1 + |t_j|) |c_{a_1 \dots a_k}^X(t_1, \dots, t_{k-1})| < \infty, \quad k = 2, 3, \dots \quad (1.15)$$

Under Assumption (B), Brillinger (2001) evaluated the J th order cumulant $c_n(J)$ of $\sqrt{n}(F_n - F)$, and showed that $c_n(J) \rightarrow 0$ as $n \rightarrow \infty$ for all $J \geq 3$, which implies the asymptotic normality of $\sqrt{n}(F_n - F)$.

Summarizing the two approaches, we have

Proposition 1.2 (Hosoya and Taniguchi (1982), Brillinger (2001)) *Suppose that one of the following conditions (HT) and (B) holds:*

(HT) $\{X(t) : t \in \mathbb{Z}\}$ is generated by (1.10) and satisfies Assumption (HT).

(B) $\{X(t) : t \in \mathbb{Z}\}$ is strictly stationary, and satisfies Assumption (B).

Let $\phi_j(\lambda)$, $j = 1, \dots, q$, be $m \times m$ matrix-valued continuous functions on $[-\pi, \pi]$ such that $\phi_j(\lambda) = \phi_j(\lambda)^*$ and $\phi_j(-\lambda) = \phi_j(\lambda)'$. Then

(i) for each $j = 1, \dots, q$,

$$\int_{-\pi}^{\pi} \text{tr}\{\phi_j(\lambda) \mathbf{I}_n(\lambda)\} d\lambda \xrightarrow{p} \int_{-\pi}^{\pi} \text{tr}\{\phi_j(\lambda) \mathbf{f}(\lambda)\} d\lambda \quad (1.16)$$

as $n \rightarrow \infty$,

(ii) the quantities

$$\sqrt{n} \int_{-\pi}^{\pi} \text{tr}[\phi_j(\lambda) \{\mathbf{I}_n(\lambda) - \mathbf{f}(\lambda)\}] d\lambda, \quad j = 1, \dots, q$$

have, asymptotically, a normal distribution with zero-mean vector and covariance matrix \mathbf{V} whose (j, l) th element is

$$4\pi \int_{-\pi}^{\pi} \text{tr}\{\mathbf{f}(\lambda)\boldsymbol{\phi}_j(\lambda)\mathbf{f}(\lambda)\boldsymbol{\phi}_l(\lambda)\}d\lambda \\ + 2\pi \sum_{r,t,u,v=1}^m \int \int_{-\pi}^{\pi} \phi_{rt}^{(j)}(\lambda_1)\phi_{uv}^{(l)}(\lambda_2)Q_{rtuv}^{\mathbf{X}}(-\lambda_1, \lambda_2, -\lambda_2) d\lambda_1 d\lambda_2.$$

where $\phi_{rt}^{(j)}(\lambda)$ is the $(r;t)$ th element of $\boldsymbol{\phi}_j(\lambda)$, and

$$Q_{rtuv}^{\mathbf{X}}(\lambda_1, \lambda_2, \lambda_3) = (2\pi)^{-3} \sum_{t_1, t_2, t_3=-\infty}^{\infty} \exp\{-i(\lambda_1 t_1 + \lambda_2 t_2 + \lambda_3 t_3)\}c_{rtuv}^{\mathbf{X}}(t_1, t_2, t_3).$$

Let us think about the case when $m = 1$ (scalar process), and $\{\mathbf{X}(t)\}$ is a Gaussian process. In the case we write $\mathbf{X}(t)$, $\mathbf{I}_n(\lambda)$, $\mathbf{f}(\lambda)$ by $X(t)$, $I_n(\lambda)$ and $f(\lambda)$, respectively. Let $\mathbf{X} = (X(1), X(2), \dots, X(n))'$ be a partial realization of $\{X(t) : t \in \mathbb{Z}\}$. Denote the probability density function of \mathbf{X} by $p(\mathbf{X})$. Under appropriate regularity conditions, Liggett (1971) showed

$$\frac{1}{n} \log\{p(\mathbf{X})\} = -\log 2\pi - \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\log f(\lambda) + \frac{I_n(\lambda)}{f(\lambda)} \right] d\lambda + o_p(1). \quad (1.17)$$

Hence, if the spectral density is parameterized by q -dimensional unknown parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^q$, an approximated maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ may be defined by

$$\hat{\boldsymbol{\theta}} \equiv \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \int_{-\pi}^{\pi} \{\log f_{\boldsymbol{\theta}}(\lambda) + f_{\boldsymbol{\theta}}(\lambda)^{-1} I_n(\lambda)\} d\lambda. \quad (1.18)$$

Returning to the original setting (1.10), we may understand that the multivariate version of the right-hand side integral part in (1.18) is

$$D(\mathbf{f}_{\boldsymbol{\theta}}, \mathbf{I}_n) \equiv \int_{-\pi}^{\pi} [\log \det \mathbf{f}_{\boldsymbol{\theta}}(\lambda) + \text{tr}\{\mathbf{f}_{\boldsymbol{\theta}}(\lambda)^{-1} \mathbf{I}_n(\lambda)\}] d\lambda. \quad (1.19)$$

Let

$$\hat{\boldsymbol{\theta}}_{QGM L} \equiv \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} D(\mathbf{f}_{\boldsymbol{\theta}}, \mathbf{I}_n), \quad (1.20)$$

which is called a quasi-Gaussian maximum likelihood estimator. Although in the derivation of (1.17), we used Gaussianity of the process, the assumption of Gaussianity may be dropped when we use $\hat{\boldsymbol{\theta}}_{QGM L}$.

Let $\mathcal{F} = \{f_\theta(\lambda) : \theta \in \Theta \subset \mathbb{R}^q\}$ be a family of parametric spectral density models, which may not contain the true model $f(\lambda)$. We fit $f_\theta(\lambda) \in \mathcal{F}$ to $f(\lambda)$ by the criterion $D(f_\theta, f)$ defined by (1.19) replacing \mathbf{I}_n by f . This setting is very natural for practical situations, and has expectedly wide applications. Let

$$\underline{\theta} = \arg \min_{\theta \in \Theta} D(f_\theta, f), \quad (1.21)$$

which is called a quasi-true value of the parameter. If $f_\theta(\lambda)$ is sufficiently smooth with respect to θ , and if $\hat{\theta}_{QGML}$ satisfies

$$\frac{\partial}{\partial \theta} D(f_\theta, \mathbf{I}_n) \Big|_{\theta = \hat{\theta}_{QGML}} = \mathbf{0}, \quad (1.22)$$

then Taylor's expansion of (1.22) at $\underline{\theta}$ leads to

$$\mathbf{0} = \frac{\partial}{\partial \theta} D(f_{\underline{\theta}}, \mathbf{I}_n) + \frac{\partial^2}{\partial \theta \partial \theta'} D(f_{\underline{\theta}}, \mathbf{I}_n) (\hat{\theta}_{QGML} - \underline{\theta}) + \cdots \quad (1.23)$$

If (1.21) implies

$$\frac{\partial}{\partial \theta} D(f_\theta, f) \Big|_{\theta = \underline{\theta}} = \mathbf{0}, \quad (1.24)$$

then, it is seen from (1.23) that

$$\sqrt{n}(\hat{\theta}_{QGML} - \underline{\theta}) \approx - \left[\frac{\partial^2}{\partial \theta \partial \theta'} D(f_{\underline{\theta}}, \mathbf{I}_n) \right]^{-1} \sqrt{n} \left[\frac{\partial}{\partial \theta} D(f_{\underline{\theta}}, \mathbf{I}_n) - \frac{\partial}{\partial \theta} D(f_{\underline{\theta}}, f) \right]. \quad (1.25)$$

Because the derivatives of $D(f_\theta, \mathbf{I}_n)$ are integral functionals of \mathbf{I}_n , hence, application of Proposition 1.2 to (1.25) yields,

Proposition 1.3 (Hosoya and Taniguchi (1982)) *Suppose that $\underline{\theta}$ exists uniquely and lies in $\text{Int}\Theta$, and that*

$$\mathbf{M}_f \equiv \int_{-\pi}^{\pi} \left[\frac{\partial^2}{\partial \theta \partial \theta'} \text{tr} \left\{ f_\theta(\lambda)^{-1} f(\lambda) \right\} + \frac{\partial^2}{\partial \theta \partial \theta'} \log \det f_\theta(\lambda) \right]_{\theta = \underline{\theta}} d\lambda$$

is a nonsingular matrix. Then, under Assumption(HT), it holds that, as $n \rightarrow \infty$,

- (i) $\hat{\theta}_{QGML} \xrightarrow{P} \underline{\theta}$
- (ii) the distribution of $\sqrt{n}(\hat{\theta}_{QGML} - \underline{\theta})$ tends to the normal distribution with zero mean vector and covariance matrix $\mathbf{M}_f^{-1} \tilde{\mathbf{V}} \mathbf{M}_f^{-1}$, where $\tilde{\mathbf{V}} = \{\tilde{V}_{jl}\}$ is a $q \times q$ matrix such that

$$\begin{aligned}
\tilde{V}_{jl} &= 4\pi \int_{-\pi}^{\pi} \text{tr} \left[\mathbf{f}(\lambda) \left\{ \frac{\partial}{\partial \theta_j} \mathbf{f}(\lambda)^{-1} \right\} \mathbf{f}(\lambda) \left\{ \frac{\partial}{\partial \theta_l} \mathbf{f}(\lambda)^{-1} \right\} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}} d\lambda \\
&\quad + 2\pi \sum_{r,t,u,v=1}^m \iint_{-\pi}^{\pi} \left\{ \frac{\partial}{\partial \theta_j} f_{\boldsymbol{\theta}}^{(r,t)}(\lambda_1) \frac{\partial}{\partial \theta_l} f_{\boldsymbol{\theta}}^{(u,v)}(\lambda_2) \right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}} \\
&\quad \times Q_{rtuv}^{\mathbf{X}}(-\lambda_1, \lambda_2, -\lambda_2) d\lambda_1 d\lambda_2
\end{aligned} \tag{1.26}$$

and $f_{\boldsymbol{\theta}}^{(r,t)}(\lambda)$ is the (r, t) th element of $\mathbf{f}_{\boldsymbol{\theta}}(\lambda)^{-1}$.

If $\{\mathbf{X}(t)\}$ is Gaussian, then the non-Gaussian quantities satisfy $Q_{rtuv}^{\mathbf{X}}(\omega_1, \omega_2, \omega_3) = 0$, hence, the second term in (1.26) vanishes. If $\mathbf{f}(\lambda) \equiv \mathbf{f}_{\boldsymbol{\theta}}$, i.e., there is no misspecification, then differential calculus for $(\partial/\partial \boldsymbol{\theta}) \mathbf{f}_{\boldsymbol{\theta}}^{-1}$ and $(\partial/\partial \boldsymbol{\theta}) \log \det \mathbf{f}_{\boldsymbol{\theta}}$ yield,

Corollary 1.3 *Suppose that $\{\mathbf{X}(t)\}$ is Gaussian, and the spectral density matrix $\mathbf{f}(\lambda) \in \mathcal{F}$, i.e., $\mathbf{f}(\lambda) = \mathbf{f}_{\boldsymbol{\theta}}(\lambda)$. Then, under Assumption (HT),*

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{QGM L} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{F}(\boldsymbol{\theta})^{-1}), \tag{1.27}$$

where

$$\mathbf{F}(\boldsymbol{\theta}) = \left[\frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ \mathbf{f}_{\boldsymbol{\theta}}(\lambda)^{-1} \left(\frac{\partial}{\partial \theta_j} \mathbf{f}_{\boldsymbol{\theta}}(\lambda) \right) \mathbf{f}_{\boldsymbol{\theta}}(\lambda)^{-1} \left(\frac{\partial}{\partial \theta_l} \mathbf{f}_{\boldsymbol{\theta}}(\lambda) \right) \right\} d\lambda; j, l = 1, \dots, q \right],$$

which is called the (Gaussian) Fisher information matrix in time series.

In view of this corollary, tentatively, we say that an estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is Gaussian asymptotically efficient of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{F}(\boldsymbol{\theta})^{-1})$. The rigorous asymptotic efficiency will be discussed in the next section based on the concept of LAN.

We saw in Proposition 1.3 that the asymptotics of $\hat{\boldsymbol{\theta}}_{QGM L}$ depend on the integral of the fourth-order cumulant spectra $Q_{rtuv}^{\mathbf{X}}(\cdot, \cdot, \cdot)$ which show a degree of non-Gaussianity. We will show that there is a case when the integrals vanish even if the process $\{\mathbf{X}(t)\}$ is non-Gaussian.

Assumption (NGR)

(i) The innovation process $\{\mathbf{U}(t)\}$ in (1.10) satisfies

$$\text{cum}\{u_a(t_1), u_b(t_2), u_c(t_3), u_d(t_4)\} = \begin{cases} \kappa_{abcd} & \text{if } t_1 = t_2 = t_3 = t_4, \\ 0 & \text{otherwise.} \end{cases} \tag{1.28}$$

(ii) $\mathbf{f}(\lambda) = \mathbf{f}_{\boldsymbol{\theta}}(\lambda)$.

- (iii) $\mathbf{f}_\theta(\lambda)$ has the form $(2\pi)^{-1} \mathbf{A}_\theta(\lambda) \mathbf{K} \mathbf{A}_\theta(\lambda)^*$, where $\mathbf{A}_\theta(\lambda) = \sum_{j=0}^{\infty} \mathbf{A}_\theta(j) e^{-itj}$ with $\mathbf{A}_\theta(0) = \mathbf{I}_m$, and the innovation covariance matrix \mathbf{K} is independent of θ , then we say that θ is “innovation-free.”

Evaluating the second integral of (1.26), we can see that all of them vanish, hence,

Corollary 1.4 *Under Assumption (NGR), the asymptotic distribution of $\sqrt{N}(\hat{\theta}_{QGMLE} - \theta)$ is independent of non-Gaussianity of the process (we say that the estimator is non-Gaussian robust (NGR)).*

Our approach due to $D(\mathbf{f}_\theta, \mathbf{f})$ is very wide. For example, we are interested in prediction of $\mathbf{X}(t+h)$ based on a linear combination of $\mathbf{X}(t)$, $\mathbf{X}(t-1)$, \dots , $\mathbf{X}(t-r)$, i.e., $\mathbf{B}(h)\mathbf{X}(t) + \mathbf{B}(h+1)\mathbf{X}(t-1) + \dots + \mathbf{B}(h+r)\mathbf{X}(t-r)$, where $\mathbf{B}(h)$, \dots , $\mathbf{B}(h+r)$ are $m \times m$ matrices satisfying that $\mathbf{I}_m - \mathbf{B}(h)z^h - \dots - \mathbf{B}(h+r)z^{h+r}$ is holomorphic on $\mathcal{D} = \{z \in \mathbb{C}; |z| \leq 1\}$.

Let $\mathbf{B}_\theta(\lambda) = \mathbf{I}_m - \mathbf{B}(h)e^{ih\lambda} - \dots - \mathbf{B}(h+r)e^{i(h+r)\lambda}$, and $\mathbf{f}_\theta(\lambda) \equiv \mathbf{B}_\theta(\lambda)^{-1} \{\mathbf{B}_\theta(\lambda)^{-1}\}^*$, where $\theta = \text{vec}\{\mathbf{B}(h), \dots, \mathbf{B}(h+r)\}$. Then the h -step ahead prediction error of $\{\mathbf{X}(t+h)\}$ is

$$\begin{aligned}
 & \text{tr} E[\{\mathbf{X}(t+h) - \mathbf{B}(h)\mathbf{X}(t) - \dots - \mathbf{B}(h+r)\mathbf{X}(t-r)\} \\
 & \quad \times \{\mathbf{X}(t+h) - \mathbf{B}(h)\mathbf{X}(t) - \dots - \mathbf{B}(h+r)\mathbf{X}(t-r)\}'] \\
 &= \text{tr} \int_{-\pi}^{\pi} \mathbf{B}_\theta(\lambda) \mathbf{f}(\lambda) \mathbf{B}_\theta(\lambda)^* d\lambda \\
 &= \int_{-\pi}^{\pi} \text{tr}\{\mathbf{B}_\theta(\lambda)^* \mathbf{B}_\theta(\lambda) \mathbf{f}(\lambda)\} d\lambda \\
 &= \int_{-\pi}^{\pi} \text{tr}\{\mathbf{f}_\theta^{-1}(\lambda) \mathbf{f}(\lambda)\} d\lambda \tag{1.29}
 \end{aligned}$$

From Theorem 3''' of Hannan (1970), p.162 it follows that

$$\int_{-\pi}^{\pi} \log \det \mathbf{f}_\theta(\lambda) d\lambda = 0 \tag{1.30}$$

which, together with (1.29), implies that the h -step ahead prediction error is equal to $D(\mathbf{f}_\theta, \mathbf{f})$. Let $\underline{\theta}$ be the quasi-true value of θ , i.e.,

$$\underline{\theta} = \text{vec}\{\underline{\mathbf{B}}(h), \dots, \underline{\mathbf{B}}(h+r)\} \equiv \arg \min_{\theta \in \Theta} D(\mathbf{f}_\theta, \mathbf{f}).$$

Then, the corresponding predictor

$$\underline{\mathbf{B}}(h)\mathbf{X}(t) + \cdots + \underline{\mathbf{B}}(h+r)\mathbf{X}(t-r) \quad (1.31)$$

becomes the best h -step ahead linear predictor on $\{\mathbf{X}(t), \dots, \mathbf{X}(t-r)\}$. Observing the data $\{\mathbf{X}(t), \mathbf{X}(t-1), \dots, \mathbf{X}(t-n+1)\}$, we can estimate $\underline{\boldsymbol{\theta}}$ by

$$\hat{\boldsymbol{\theta}}_{QGML} \equiv \text{vec}\{\hat{\underline{\mathbf{B}}}_{QGML}(h), \dots, \hat{\underline{\mathbf{B}}}_{QGML}(h+r)\} \quad (1.32)$$

whose asymptotics are given in Proposition 1.3 with $\dim \boldsymbol{\theta} = m(r+1)$, i.e.,

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{QGML} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{M}_f^{-1} \tilde{\mathbf{V}} \mathbf{M}_f^{-1}). \quad (1.33)$$

Let us consider the problem of portfolio. Suppose that we have m -asset returns described by $X_1(t), \dots, X_m(t)$. Let $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)'$ be a portfolio weight vector satisfying $\alpha_1 + \cdots + \alpha_m = 1$. Assuming that the present time is t , we want to estimate the future portfolio return $\boldsymbol{\alpha}'\mathbf{X}(t+h)$. In view of above, this is estimated by

$$\boldsymbol{\alpha}' \left\{ \hat{\underline{\mathbf{B}}}_{QGML}(h)\mathbf{X}(t) + \cdots + \hat{\underline{\mathbf{B}}}_{QGML}(h+r)\mathbf{X}(t-r) \right\}. \quad (1.34)$$

Hence our setting by $D(\mathbf{f}_\theta, \mathbf{f})$ is unexpectedly general and wide.

We next discuss the problem of nonparametric spectral estimation. We estimate $\mathbf{f}(\lambda)$ by weighted averages of the periodogram matrix $\mathbf{I}_n(\lambda)$ with a spectral window $W_n(\lambda)$ as weight, i.e.,

$$\hat{\mathbf{f}}_n(\lambda) = \int_{-\pi}^{\pi} W_n(\lambda - \mu) \mathbf{I}_n(\mu) d\mu \quad (1.35)$$

where $W_n(\cdot)$ satisfies

Assumption 1.2

(i) $W_n(\lambda)$ can be expanded as

$$W_n(\lambda) = \frac{1}{2\pi} \sum_{l=-M}^M w\left(\frac{l}{M}\right) \exp(-il\lambda).$$

(ii) $w(x)$ is continuous, even function with $w(0) = 1$, and satisfies

$$\left\{ \begin{array}{l} |w(x)| \leq 1, \\ \int_{-\infty}^{\infty} w(x)^2 dx < \infty, \\ \lim_{x \rightarrow 0} \frac{1 - w(x)}{|x|^2} = \kappa_2 < \infty. \end{array} \right.$$

(iii) $M = M(n)$ satisfies

$$n^{1/4}/M + M/n^{1/2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Concrete examples of $W_n(\cdot)$ satisfying Assumption 1.2 are found in e.g., Hannan (1970), and it is known that

$$E \left[\|\hat{f}_n(\lambda) - f(\lambda)\|^2 \right] = \left(\frac{M}{n} \right) + O(M^{-4}) \quad (1.36)$$

uniformly in λ , hence, $\hat{f}_n(\lambda) - f(\lambda) = O_p\{(M/n)^{1/2}\}$, i.e., $\hat{f}(\lambda)$ is a $\sqrt{n/M}$ -consistent estimator of $f(\lambda)$ (e.g., Hannan (1970)). As we saw in Proposition 1.3, parametric estimators usually have \sqrt{n} -consistency, so the result above means a disadvantage of nonparametric spectral estimators. But, in what follows, we will show that this disadvantage disappears if we consider appropriate integral functionals of $\hat{f}_n(\lambda)$.

Let D be an open subset of \mathbb{C}^{m^2} .

Assumption 1.3 A mapping $K : D \rightarrow \mathbb{R}$ is holomorphic.

The integral functional of $f(\lambda)$

$$\int_{-\pi}^{\pi} K\{f(\lambda)\}d\lambda \quad (1.37)$$

can represent a wide variety of important time series indices.

The example in financial econometrics will be given later. If we are interested in estimation of (1.37), it is very natural to use $\int_{-\pi}^{\pi} K\{\hat{f}_n(\lambda)\}d\lambda$ as an estimator of (1.37). Denote by $\mathbf{K}^{(1)}\{f(\lambda)\} = [K_{ab}^{(1)}\{f(\lambda)\}; a, b = 1, \dots, m]$ the first-order derivative of $K\{f(\lambda)\}$ at $f(\lambda)$ (see Magnus and Neudecker (1988)). Showing the relation

$$\begin{aligned} & \sqrt{n} \left[\int_{-\pi}^{\pi} K\{\hat{f}_n(\lambda)\}d\lambda - \int_{-\pi}^{\pi} K\{f(\lambda)\}d\lambda \right] \\ &= \sqrt{n} \int_{-\pi}^{\pi} \left[\text{tr} \{ \mathbf{I}_n(\lambda) - f(\lambda) \} \mathbf{K}^{(1)}\{f(\lambda)\} \right] d\lambda + o_p(1), \quad (1.38) \end{aligned}$$

Taniguchi et al. (1996) gave the following.

Proposition 1.4 *Suppose that Assumption (B) or (HT) holds. Then, under Assumptions 1.2 and 1.3, as $n \rightarrow \infty$,*

$$\sqrt{n} \left[\int_{-\pi}^{\pi} K\{\hat{f}_n(\lambda)\}d\lambda - \int_{-\pi}^{\pi} K\{f(\lambda)\}d\lambda \right] \xrightarrow{d} N\left(0, v_1(f) + v_2(Q^X)\right), \quad (1.39)$$

where

$$v_1(f) = 4\pi \int_{-\pi}^{\pi} \text{tr} \left[f(\lambda) \mathbf{K}^{(1)}\{f(\lambda)\} \right]^2 d\lambda, \quad (1.40)$$

and

$$v_2(Q^X) = 2\pi \sum_{a,b,c,d=1}^m \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} K_{ab}^{(1)}\{f(\lambda)\} K_{cd}^{(1)}\{f(\lambda)\} Q_{abcd}^X(-\lambda_1, \lambda_2, -\lambda_2) d\lambda_1 d\lambda_2. \quad (1.41)$$

It may be noted that \sqrt{n} -consistency holds in Proposition 1.4 despite $\hat{f}_n(\lambda)$ being a $\sqrt{n/M}$ -consistent estimator of $f(\lambda)$. This is due to the fact that integration of $\hat{f}_n(\lambda)$ recovers \sqrt{n} -consistency.

Consider the problem of testing

$$\begin{aligned} \text{H} : \int_{-\pi}^{\pi} K\{f(\lambda)\}d\lambda &= c \\ &\text{against} \\ \text{A} : \int_{-\pi}^{\pi} K\{f(\lambda)\}d\lambda &\neq c, \end{aligned} \quad (1.42)$$

where c is a given constant. Under H, it follows from Proposition 1.4 that

$$S_n \equiv \frac{\sqrt{n} \left[\int_{-\pi}^{\pi} K\{\hat{f}_n(\lambda)\}d\lambda - c \right]}{\sqrt{v_1(f) + v_2(Q^X)}} \xrightarrow{d} N(0, 1), \quad (1.43)$$

as $n \rightarrow \infty$. However, we have to estimate the denominator if S_n is feasible for testing (1.42). Regarding $v_1(f)$, $v_1(\hat{f}_n)$ becomes a consistent estimator of it. Taniguchi (1982) gave a consistent estimator for quantities of the form of $v_2(Q^X)$, hence, we write this estimator as $\widehat{v_2(Q^X)}$. Consequently, it is seen that, as $n \rightarrow \infty$,

$$T_n \equiv \frac{\sqrt{n}[\int_{-\pi}^{\pi} K\{\hat{f}_n(\lambda)\}d\lambda - c]}{\sqrt{v_1(\hat{f}_n) + v_2(\widehat{Q^X})}} \xrightarrow{d} N(0, 1), \quad (1.44)$$

under H. We can use T_n as a test statistic for (1.42). It may be noted that T_n is essentially nonparametric, and has \sqrt{n} -consistency.

Because our integral functional $\int_{-\pi}^{\pi} K\{f(\lambda)\}d\lambda$ can represent so many important indices in time series, the test setting in (1.42) is very wide. In what follows, a few examples are provided. Suppose the process $\{X(t) : t \in \mathbb{Z}\}$ is of the form

$$X(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$$

with $x(t)$, q vector-valued, and $y(t)$, r vector-valued; $q + r = m$ and has the spectral density matrix

$$f(\lambda) = \begin{bmatrix} f_{xx}(\lambda) & f_{xy}(\lambda) \\ f_{yx}(\lambda) & f_{yy}(\lambda) \end{bmatrix}. \quad (1.45)$$

We denote by $H\{\cdot\}$ the linear closed manifold generated by $\{\cdot\}$, and denote by $\mathcal{P}[x(t)|H\{\cdot\}]$ the linear projection of $x(t)$ on $H\{\cdot\}$.

Consider the residual process

$$u_1(t) = x(t) - \mathcal{P}[x(t)|H\{x(t-1), x(t-2), \dots\}],$$

$$v_1(t) = y(t) - \mathcal{P}[y(t)|H\{y(t-1), y(t-2), \dots\}],$$

$$u_2(t) = x(t) - \mathcal{P}[x(t)|H\{x(t-1), x(t-2), \dots; y(t-1), y(t-2), \dots\}],$$

$$v_2(t) = y(t) - \mathcal{P}[y(t)|H\{x(t-1), x(t-2), \dots; y(t-1), y(t-2), \dots\}],$$

and

$$u_3(t) = x(t) - \mathcal{P}[x(t)|H\{x(t-1), x(t-2), \dots; y(t), y(t-1), \dots\}].$$

The measure of linear feedback from $Y = \{y(t)\}$ to $X = \{x(t)\}$ is defined by

$$F_{Y \rightarrow X} = \log[\det\{Var(u_1(t))\} / \det\{Var(u_2(t))\}] \quad (1.46)$$

Symmetrically, we can define

$$F_{X \rightarrow Y} = \log[\det\{Var(v_1(t))\} / \det\{Var(v_2(t))\}] \quad (1.47)$$

The measure of instantaneous linear feedback

$$F_{X,Y} = \log[\det\{Var(u_2(t))\} / \det\{Var(u_3(t))\}]$$

has motivation similar to that of the above two measures. The following

$$F_{X,Y} = \log[\det\{Var(u_1(t))\} \det\{Var(v_1(t))\} / \det\{Var(U(t))\}]$$

is called the measure of linear dependence. Then it is shown that

$$F_{X,Y} = F_{Y \rightarrow X} + F_{X \rightarrow Y} + F_{X \cdot Y} \quad (1.48)$$

and

$$F_{X \cdot Y} = \int_{-\pi}^{\pi} K\{\mathbf{f}(\lambda)\} d\lambda, \quad (1.49)$$

where

$$K\{\mathbf{f}(\lambda)\} = -\frac{1}{2\pi} \log[\det\{\mathbf{I}_q - \mathbf{f}_{xy}(\lambda)\mathbf{f}_{yy}(\lambda)^{-1}\mathbf{f}_{yx}(\lambda)\mathbf{f}_{xx}(\lambda)^{-1}\}],$$

(see Geweke (1982).) Because $F_{Y \rightarrow X}$, $F_{X \rightarrow Y}$ and $F_{X \cdot Y}$ are important econometric measures which represent “strength of causality,” the problem

$$H : F_{X,Y} = c, \quad \text{against} \quad A : F_{X,Y} \neq c, \quad (1.50)$$

is exactly an example of our testing (1.42). Therefore, we can test (1.50) by use of T_n given in (1.44).

Next we discuss another interrelation analysis. Let $\mathbf{X}(t) = \{\mathbf{x}(t)', \mathbf{y}(t)', \mathbf{z}(t)'\}'$ be the m -dimensional linear process given in (1.10), where $\mathbf{x}(t)$, $\mathbf{y}(t)$ and $\mathbf{z}(t)$ are q , r and s ($q + r + s = m$) component processes, respectively. Correspondingly we write the spectral density matrix and the spectral representation, respectively, as

$$\mathbf{f}(\lambda) = \begin{bmatrix} \mathbf{f}_{xx}(\lambda) & \mathbf{f}_{xy}(\lambda) & \mathbf{f}_{xz}(\lambda) \\ \mathbf{f}_{yx}(\lambda) & \mathbf{f}_{yy}(\lambda) & \mathbf{f}_{yz}(\lambda) \\ \mathbf{f}_{zx}(\lambda) & \mathbf{f}_{zy}(\lambda) & \mathbf{f}_{zz}(\lambda) \end{bmatrix},$$

and

$$\mathbf{X}(t) = \int_{-\pi}^{\pi} e^{-it\lambda} d\boldsymbol{\xi}(\lambda),$$

where $d\boldsymbol{\xi}(\lambda) = (d\xi_x(\lambda)', d\xi_y(\lambda)', d\xi_z(\lambda)')'$. Hannan (1970) considered a test for association for $\mathbf{x} = \{\mathbf{x}(t)\}$ with $\mathbf{y} = \{\mathbf{y}(t)\}$ (at frequency λ) after allowing for any affects of $\mathbf{z} = \{\mathbf{z}(t)\}$. The hypothesis is given by

$$H_\lambda : \mathbf{f}_{xy}(\lambda) - \mathbf{f}_{xz}(\lambda)\mathbf{f}_{zz}(\lambda)^{-1}\mathbf{f}_{zy}(\lambda) = \mathbf{0}, \quad (1.51)$$

which means that $d\xi_x(\lambda) - \mathbf{f}_{xz}(\lambda)\mathbf{f}_{zz}(\lambda)^{-1}d\xi_z(\lambda)$ is incoherent with $d\xi_y(\lambda) - \mathbf{f}_{yz}(\lambda)\mathbf{f}_{zz}(\lambda)^{-1}d\xi_z(\lambda)$, and all of the apparent association between \mathbf{x} and \mathbf{y} is truly due only to their common association with \mathbf{z} . For a given λ , Hannan (1970) developed the testing theory for H_λ based on the asymptotic normality of the finite

Fourier transformations of $\{X(t)\}$ in a neighborhood of λ . In the context of (1.42), we can introduce a test for association for \mathbf{x} and \mathbf{y} at “all the frequency $\lambda \in [-\pi, \pi]$ ” after allowing for any effects \mathbf{z} . The hypothesis is written as

$$H : \mathbf{f}_{xy}(\lambda) - \mathbf{f}_{xz}(\lambda) \mathbf{f}_{zz}(\lambda)^{-1} \mathbf{f}_{zy}(\lambda) = \mathbf{0} \quad \text{for all } \lambda \in [-\pi, \pi],$$

equivalently,

$$H : \int_{-\pi}^{\pi} K\{\mathbf{f}(\lambda)\} d\lambda = 0, \quad (1.52)$$

where

$$K\{\mathbf{f}(\lambda)\} = \text{tr}[\{\mathbf{f}_{xy}(\lambda) - \mathbf{f}_{xz}(\lambda) \mathbf{f}_{zz}(\lambda)^{-1} \mathbf{f}_{zy}(\lambda)\} \{\mathbf{f}_{yx}(\lambda) - \mathbf{f}_{yz}(\lambda) \mathbf{f}_{zz}(\lambda)^{-1} \mathbf{f}_{zx}(\lambda)\}].$$

Then, we can test (1.52) by use of T_n , hence, we can grasp Hannan’s problem above as ours.

Now, let us think of estimation based on integral functional of $\hat{\mathbf{f}}_n$. Kakizawa (1996) introduced the following generalized disparity measure between spectral density matrices $\mathbf{f}(\lambda)$ and $\mathbf{g}(\lambda)$;

$$D_H(\mathbf{f} : \mathbf{g}) = \frac{1}{4\pi} \int_{-\pi}^{\pi} H\{\mathbf{f}(\lambda) \mathbf{g}(\lambda)^{-1}\} d\lambda, \quad (1.53)$$

where $H(\mathbf{Z})$ is a holomorphic function of $m \times m$ matrix \mathbf{Z} , and has a unique minimum zero at $\mathbf{Z} = \mathbf{I}_m$, such as

$$H_1(\mathbf{Z}) = -\log \det(\mathbf{Z}) + \text{tr}(\mathbf{Z}) - m,$$

$$H_2(\mathbf{Z}) = \frac{1}{2} \text{tr}(\mathbf{Z} - \mathbf{I}_m)^2,$$

$$H_3(\mathbf{Z}) = \frac{1}{\alpha(1-\alpha)} [\log \det\{(1-\alpha)\mathbf{I}_m + \alpha\mathbf{Z}\} - \alpha \log \det \mathbf{Z}], \quad \alpha \in (0, 1).$$

In fact, $H(\mathbf{Z})$ is defined by

$$H(\mathbf{Z}) = \sum_{l \neq 0} \sum_{j=1}^3 p_{j,l} \frac{H_j(\mathbf{Z}^l)}{l^2}, \quad (1.54)$$

where $p_{j,l} \geq 0$ for $j = 1, 2, 3; l = \pm 1, \pm 2, \dots$ and $\sum_l \sum_{j=1}^3 p_{j,l} = 1$. Thus $D_H(\mathbf{f} : \mathbf{g})$ includes many other criteria as special cases. Let

$$\hat{\boldsymbol{\theta}}_H = \arg \min_{\boldsymbol{\theta} \in \Theta} D_H(\mathbf{f}_{\boldsymbol{\theta}}, \hat{\mathbf{f}}_n) \quad (1.55)$$

The following proposition is due to Kakizawa (1996) and Taniguchi and Kakizawa (2000).

Proposition 1.5 *Suppose all the assumptions in Proposition 1.3 hold, and the true spectral density matrix is of the form f_{θ} . Then, under Gaussianity of the process,*

$$\sqrt{n}(\hat{\theta}_H - \theta) \xrightarrow{d} N(\mathbf{0}, \mathbf{F}(\theta)^{-1}), \quad (1.56)$$

as $n \rightarrow \infty$, i.e., $\hat{\theta}_H$ has the same asymptotics as $\hat{\theta}_{QGML}$ does (see Corollary 1.3).

Proposition 1.5 implies that we can construct infinitely many Gaussian asymptotically efficient estimators. In Sect. 1.3 we will show that $\hat{\theta}_H$ has a robustness which $\hat{\theta}_{QGML}$ does not share.

1.3 Optimal Inference for Various Return Processes

As we saw in the previous subsection, there are a lot of time series models (e.g., ARMA, linear process, ARCH, GARCH, EGARCH, SV, CHARN, etc.). Here, we develop the asymptotic optimal estimation theory based on local asymptotic normality (LAN) for typical time series models.

Lucien Le Cam established one of the most important foundations of the general statistical asymptotic theory (e.g., see LeCam (1986)). He introduced the concept of LAN for the likelihood ratio of general statistical model. Once LAN is proved, the asymptotic optimality of estimators and tests is described in terms of the LAN property.

We briefly review the LAN results for stochastic processes. Swensen (1985) showed that the likelihood ratio of an autoregressive process of finite order with a trend is LAN. Regarding ARMA processes, Hallin et al. (1985) and Kreiss (1987) showed the LAN property, and applied the results to test and estimation theory. Further, Kreiss (1990) developed the LAN theory for a class of autoregressive processes with infinite order. Garel and Hallin (1995) proved the LAN for multivariate general linear models with ARMA residual.

As an important class of nonlinear time series models, Giraitis et al. (2000) introduced a class of ARCH(∞) models, which includes the ARCH and GARCH models as special cases, and gave sufficient conditions for the existence of a stationary solution and its explicit representation.

Here we deal with the more general ARCH(∞) model with stochastic mean (ARCH(∞)-SM model). In what follows we provide the LAN results and asymptotic optimal estimation for ARCH(∞)-SM model in line with Lee and Taniguchi (2005).

Suppose that (Ω, \mathcal{F}, P) is a probability space, and $\{\mathcal{F}_t : t \in \mathbb{Z}\}$ is a sequence of sub- σ -algebras of \mathcal{F} satisfying $\mathcal{F}_t \subset \mathcal{F}_{t+1}$, $t \in \mathbb{Z}$. Consider the ARCH(∞)-SM model $\{Y_t : t \in \mathbb{Z}\}$ defined by

$$\begin{cases} Y_t - \boldsymbol{\beta}'\mathbf{z}_t = \sigma_t u_t, & t \in \mathbb{Z}, \\ \sigma_t^2 = a + \sum_{j=1}^{\infty} b_j (Y_{t-j} - \boldsymbol{\beta}'\mathbf{z}_{t-j})^2. \end{cases} \quad (1.57)$$

where $a > 0$, $b_j \geq 0$, $j = 1, 2, \dots$, $\{u_t : t \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables with density $g(\cdot)$, and u_t is \mathcal{F}_t -measurable and independent of \mathcal{F}_{t-1} . Here $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is an unknown vector and the $\mathbf{z}_t = (z_{t1}, \dots, z_{tp})'$'s are observable $p \times 1$ random vectors which are \mathcal{F}_{t-1} -measurable. If $\boldsymbol{\beta} = \mathbf{0}$, this model reduces to the ARCH(∞) model proposed by Giraitis et al. (2000). We can see that the class of ARCH(∞)-SM models is larger than that of GARCH(r, s)-SM models defined by

$$\begin{cases} Y_t - \boldsymbol{\beta}'\mathbf{z}_t = \sigma_t u_t, \\ \sigma_t^2 = a + \sum_{j=1}^r a_j \sigma_{t-j}^2 + \sum_{j=1}^s b_j (Y_{t-j} - \boldsymbol{\beta}'\mathbf{z}_{t-j})^2. \end{cases} \quad (1.58)$$

where the associated polynomials of the second equation of (1.58) satisfy the invertible condition. If we take $\mathbf{z}_t = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})'$, then (1.58) becomes the AR(p)-GARCH(r, s) models, which implies that the class of ARCH(∞)-SM models is sufficiently extensive. To develop the asymptotic theory for (1.58) we impose,

Assumption 1.4

- (i) $E\{u_t\} = 0$, $\text{Var}\{u_t\} = 1$ and $E\{u_t^4\} < \infty$
- (ii) a and b_j 's are function of an unknown parameter $\boldsymbol{\eta} = (\eta_1, \dots, \eta_q)'$, i.e., $a = a(\boldsymbol{\eta})$ and $b_j = b_j(\boldsymbol{\eta})$ for $j \geq 1$ and $\boldsymbol{\eta} \in \mathcal{H}$, where \mathcal{H} is an open subset of \mathbb{R}^q . The functions $a(\boldsymbol{\eta})$ and $b_j(\boldsymbol{\eta})$ are twice continuously differentiable with respect to $\boldsymbol{\eta}$.
- (iii) There exist $\tilde{a} > 0$ and $\tilde{b}_j \geq 0$ satisfying $\sum_{j=1}^{\infty} \tilde{b}_j < 1$, such that $a(\boldsymbol{\eta}) \geq \tilde{a}$ and $b_j(\boldsymbol{\eta}) \leq \tilde{b}_j$ for all $j \geq 1$ and $\boldsymbol{\eta} \in \mathcal{H}$, which entails

$$\sum_{j=1}^{\infty} b_j(\boldsymbol{\eta}) < 1 \quad \text{for all } \boldsymbol{\eta} \in \mathcal{H}. \quad (1.59)$$

- (iv) There exist $\tilde{a}^{(1)} > 0$, $\tilde{a}^{(2)} > 0$ and $\tilde{b}_j^{(1)} \geq 0$, $\tilde{b}_j^{(2)} \geq 0$ satisfying $\sum_{j=1}^{\infty} \tilde{b}_j^{(i)} < \infty$, $i = 1, 2$, such that $\|\partial a(\boldsymbol{\eta})/\partial \boldsymbol{\eta}\| \leq \tilde{a}^{(1)}$, $\|(\partial^2/\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}')a(\boldsymbol{\eta})\| \leq \tilde{a}^{(2)}$ for all $\boldsymbol{\eta} \in \mathcal{H}$, and $\|(\partial/\partial \boldsymbol{\eta})b_j(\boldsymbol{\eta})\| \leq \tilde{b}_j^{(1)}$, $\|(\partial^2/\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}')b_j(\boldsymbol{\eta})\| \leq \tilde{b}_j^{(2)}$ for all $j \geq 1$ and $\boldsymbol{\eta} \in \mathcal{H}$, where $\|\mathbf{a}\|$ denotes the Euclidean norm of a vector or matrix \mathbf{a} , i.e., $\sqrt{\text{tr}(\mathbf{a}'\mathbf{a})}$.

The condition (1.59) guarantees the existence of strictly stationary solution for $\{Y_t - \boldsymbol{\beta}'\mathbf{z}_t\}$ in (1.58) (c.f. Giraitis et al. (2000)).

Assumption 1.5 $\{E(u_t^4)\}^{1/2} \sum_{j=1}^{\infty} b_j < 1$ and $E\|\mathbf{z}_t\|^4 < \infty$.

The condition $\{E(u_t^4)\}^{1/2} \sum_{j=1}^{\infty} b_j < 1$ implies $E(Y_t^4) < \infty$ (see Giraitis et al. (2000)). In a special case of GARCH models in (1.58), a necessary and sufficient

condition for the existence of the fourth moment of $Y_t - \beta' z_t$ was established by Ling and Li (1997), Chen and An (1998) and Ling and McAleer (2003).

Assumption 1.6 The innovation density $g(\cdot)$ is symmetric, twice continuously differentiable, and satisfies

- (i) $0 < I(g) \equiv \int_{-\infty}^{\infty} \{g'(u)/g(u)\}^2 g(u) du < \infty$, and $\int_{-\infty}^{\infty} \{g'(u)/g(u)\}^4 g(u) du < \infty$.
- (ii) $\lim_{|u| \rightarrow \infty} u g(u) = 0$, $\lim_{|u| \rightarrow \infty} u^2 g'(u) = 0$.

Assumption 1.7 The matrix $n^{-1} \sum_{t=1}^n z_t z_t' / \sigma_t^2$ converges to a finite limit $\mathbf{M}(0)$ in L^2 -sense, where $\mathbf{M}(0)$ is positive definite.

We write $\theta = (\beta', \eta')' \in \Theta$ and $\dim \Theta = r = p + q$, where Θ is an open subset of \mathbb{R}^r . Then the $\sigma_t = \sigma_t(\theta)$'s are measurable functions of θ and Y_{t-j} , $j \geq 1$. Let $P_\theta^{(n)}$ be the distribution of $(u_s, s \leq 0, Y_1, \dots, Y_n)$. For two hypothetical values $\theta, \theta' \in \Theta$ the log-likelihood ratio is written as

$$\Lambda_n(\theta, \theta') \equiv \log \frac{dP_{\theta'}^{(n)}}{dP_\theta^{(n)}} = 2 \sum_{t=1}^n \log \Phi_t^{(n)}(\theta, \theta'), \quad (1.60)$$

where $\Phi_t^{(n)}(\theta, \theta') = [\{g\{\phi_t(\theta')\}\sigma_t(\theta)\} / \{g\{\phi_t(\theta)\}\sigma_t(\theta')\}]^{1/2}$ with $\phi_t(\theta) = (Y_t - \beta' z_t) / \sigma_t(\theta)$. We denote by $H(g; \theta)$ the hypothesis under which the underlying parameter is $\theta \in \Theta$ and the density of u_t is $g = g(\cdot)$. Define

$$\theta_n = (\beta'_n, \eta'_n)' \equiv \theta + \frac{1}{\sqrt{n}} \xi, \quad \xi = (\kappa', h')' \in S \subset \mathbb{R}^r,$$

where $\kappa = (\kappa_1, \dots, \kappa_p)'$, $h = (h_1, \dots, h_q)'$, and S is an open subset of \mathbb{R}^r . In what follows we denote by $\mathbb{R}^{\mathbb{Z}}$ the product space $\dots \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \dots$, whose component spaces correspond to the coordinate spaces of $(\dots, u_{-1}, u_0, Y_1, Y_2, \dots)$, and write its Borel σ -algebra by $\mathcal{B}^{\mathbb{Z}}$.

The following proposition is due to Lee and Taniguchi (2005).

Proposition 1.6 (LAN for ARCH(∞)-SM models) *Suppose that Assumptions 1.4–1.7 hold. Then the sequence of experiments $\mathcal{E}_n = \{\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \{P_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^r\}\}$, $n \in \mathbb{N}$, is locally asymptotically normal and equicontinuous on compact subset C of S . That is,*

- (i) *For all $\theta \in \Theta$, the log-likelihood ratio $\Lambda_n(\theta, \theta_n)$ admits the following stochastic expansion under $H(g; \theta)$:*

$$\Lambda_n(\theta, \theta_n) = (\kappa', h') \frac{1}{\sqrt{n}} \sum_{t=1}^n (\Delta'_{1,t}, \Delta'_{2,t})' - \frac{1}{2} \xi' F \xi + o_p(1), \quad (1.61)$$

where $\mathbf{\Delta}_{1,t} = -(\mathbf{z}_t/\sigma_t)(g'(\phi_t)/g(\phi_t)) - (2\sigma_t^2)^{-1}(\partial\sigma_t^2/\partial\boldsymbol{\beta})\{1 + \phi_t(g'(\phi_t)/g(\phi_t))\}$, $\mathbf{\Delta}_{2,t} = -(2\sigma_t^2)^{-1}\{1 + \phi_t(g'(\phi_t)/g(\phi_t))\}(\partial/\partial\boldsymbol{\eta})\sigma_t^2$, and

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{pmatrix}.$$

Here

$$\mathbf{F}_{11} = I(g) \cdot \mathbf{M}(0) + \{J(g) - 1\} E[(4\sigma_t^4)^{-1}(\partial\sigma_t^2/\partial\boldsymbol{\beta})(\partial\sigma_t^2/\partial\boldsymbol{\beta}')],$$

$$\mathbf{F}_{12} = \{J(g) - 1\} E[(4\sigma_t^4)^{-1}(\partial\sigma_t^2/\partial\boldsymbol{\beta})(\partial\sigma_t^2/\partial\boldsymbol{\eta}')],$$

$$\mathbf{F}_{22} = \{J(g) - 1\} E[(4\sigma_t^4)^{-1}(\partial\sigma_t^2/\partial\boldsymbol{\eta})(\partial\sigma_t^2/\partial\boldsymbol{\eta}')],$$

where $J(g) = E\{u_t^2(g'(u_t)/g(u_t))^2\}$.

(ii) Under $H(g; \boldsymbol{\theta})$, $\mathbf{\Delta}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{F})$, where $\mathbf{\Delta}_n = n^{-1/2} \sum_{t=1}^2 (\mathbf{\Delta}'_{1,t}, \mathbf{\Delta}'_{2,t})'$.

(iii) For all $n \in \mathbb{N}$ and $\boldsymbol{\xi} \in S$, the mapping $\boldsymbol{\xi} \rightarrow P_{\boldsymbol{\theta}_n^{(n)}}$ is continuous with respect to the variational distance $\|P - Q\| = \sup\{|P(A) - Q(A)|; A \in \mathcal{B}^{\mathbb{Z}}\}$.

The term $\mathbf{\Delta}_n$, called the central sequence, is measurable with respect to u_s , $s \leq 0$, Y_j , \mathbf{z}_j , $j = 1, \dots, n$, but it is not so with respect to the observable sequences, Y_j , \mathbf{z}_j , $j = 1, \dots, n$. Therefore, we will construct a $(Y_1, \dots, Y_n, \mathbf{z}_1, \dots, \mathbf{z}_n)$ -measurable version $\tilde{\mathbf{\Delta}}_n$ of $\mathbf{\Delta}_n$. For this, we introduce the truncated versions of σ_t^2 and ϕ_t by $\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\boldsymbol{\beta}, \boldsymbol{\eta}) \equiv a + \sum_{j=1}^{t-1} b_j(Y_{t-j} - \boldsymbol{\beta}'\mathbf{z}_{t-1})^2$ and $\tilde{\phi}_t \equiv (Y_t - \boldsymbol{\beta}'\mathbf{z}_t)/\tilde{\sigma}_t$, respectively.

Assumption 1.8 (i) For some $r \in [0, 1)$, $b_j(\boldsymbol{\eta}) = O(r^j)$ for all $j \in \mathbb{N}$ and $\boldsymbol{\eta} \in \mathcal{H}$.
(ii) For some $v_1, \dots, v_q \in [0, 1)$, $\partial b_j(\boldsymbol{\eta})/\eta_k$, $k = 1, \dots, q$, are of order $O(v_k^j)$, for $j \in \mathbb{N}$ and $\boldsymbol{\eta} \in \mathcal{H}$.

The following proposition shows that the unobservable values $\{u_s; s \leq 0\}$ have no influence on the LAN form below.

Proposition 1.7 (Lee and Taniguchi (2005)) *Under Assumptions 1.4–1.8, the log-likelihood ratio $\Lambda_n(\boldsymbol{\theta}, \boldsymbol{\theta}_n)$ admits, under $H(g; \boldsymbol{\theta})$, as $n \rightarrow \infty$, the stochastic expansion*

$$\Lambda_n(\boldsymbol{\theta}, \boldsymbol{\theta}_n) = (\boldsymbol{\kappa}', \mathbf{h}') \tilde{\mathbf{\Delta}}_n - \frac{1}{2} \boldsymbol{\xi}' \mathbf{F} \boldsymbol{\xi} + o_p(1), \quad (1.62)$$

where $\tilde{\mathbf{\Delta}}_n = n^{-1/2} \sum_{t=1}^n (\tilde{\mathbf{\Delta}}'_{1,t}, \tilde{\mathbf{\Delta}}'_{2,t})'$,

$$\tilde{\mathbf{\Delta}}_{1,t} = -(\mathbf{z}_t \tilde{\sigma}_t)(g'(\tilde{\phi}_t)/g(\tilde{\phi}_t)) - (2\tilde{\sigma}_t^2)^{-1}(\partial\tilde{\sigma}_t^2/\partial\boldsymbol{\beta})\{1 + \tilde{\phi}_t(g'(\tilde{\phi}_t)/g(\tilde{\phi}_t))\},$$

$$\tilde{\mathbf{\Delta}}_{2,t} = -(2\tilde{\sigma}_t^2)^{-1}\{1 + \tilde{\phi}_t(g'(\tilde{\phi}_t)/g(\tilde{\phi}_t))\}(\partial/\partial\boldsymbol{\eta})\tilde{\sigma}_t^2.$$

Here, $\tilde{\mathbf{\Delta}}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{F})$ under $H(g; \boldsymbol{\theta})$.

Next we discuss the estimation of θ . In what follows distribution law of a random vector Y_n under $P_\theta^{(n)}$ is denoted by $\mathcal{L}(Y_n|P_\theta^{(n)})$, and the weak convergence to Z is denoted by $\mathcal{L}(Y_n|P_\theta^{(n)}) \xrightarrow{d} Z$. Define the class \mathcal{A} of sequences of estimators $\{S_n\}$ of θ as

$$\mathcal{A} = [\{S_n\} : \mathcal{L}\{\sqrt{n}(S_n - \theta_n)|P_{\theta_n}^{(n)}\} \xrightarrow{d} Z_\theta, \text{ a probability distribution }],$$

where Z_θ depends on $\{S_n\}$ generally. Let \mathcal{L} be the class of all loss functions $l : \mathbb{R}^r \rightarrow [0, \infty)$ of the form $l(\mathbf{x}) = \tau(|\mathbf{x}|)$ which satisfies $\tau(0) = 0$ and $\tau(a) \leq \tau(b)$ if $a \leq b$. Typical examples are $l(\mathbf{x}) = I(|\mathbf{x}| > a)$ and $l(\mathbf{x}) = |\mathbf{x}|^p$, $p \geq 1$, where $I(\cdot)$ is the indicator function of (\cdot) .

Assume that the LAN property (1.61) holds. Then, a sequence $\{\hat{\theta}_n\}$ of estimators of θ is said to be a sequence of asymptotically centering estimators if

$$\sqrt{n}(\hat{\theta}_n - \theta) - \mathbf{F}^{-1} \Delta_n = o_p(1) \quad \text{in } P_\theta^{(n)}. \quad (1.63)$$

The following proposition can be proved by following the arguments in Section 83 of Strasser (1985), Jeganathan (1995), and p. 69 of Taniguchi and Kakizawa (2000).

Proposition 1.8 *Assume that the LAN property (1.61) for the ARCH(∞)-SM model (1.57) holds, and that $\{S_n\} \in \mathcal{A}$. Let Δ be a random vector, distributed as $N(\mathbf{0}, \mathbf{F})$. Then the following statements hold:*

(i) *For any $l \in \mathcal{L}$ with $E\{l(\Delta)\} < \infty$,*

$$\liminf_{n \rightarrow \infty} E[l\{\sqrt{n}(S_n - \theta)\}|P_\theta^{(n)}] \geq E\{l(\mathbf{F}^{-1} \Delta)\}. \quad (1.64)$$

(ii) *If*

$$\limsup_{n \rightarrow \infty} E[l\{\sqrt{n}(S_n - \theta)\}|P_\theta^{(n)}] \leq E\{l(\mathbf{F}^{-1} \Delta)\}, \quad (1.65)$$

for a nonconstant $l \in \mathcal{L}$ with $E\{l(\Delta)\} < \infty$, then S_n is a sequence of asymptotically centering estimators.

From Proposition 1.8, it follows that $\{\hat{\theta}_n\} \in \mathcal{A}$ is asymptotically efficient if it is asymptotically centering. Let us construct an asymptotically efficient estimator. For any sequence of estimators $\tilde{\theta}_n$, the discretized estimator $\bar{\theta}_n$ of $\tilde{\theta}_n$ is defined by the nearest vertex of $\{\theta; \theta = n^{-1/2}(i_1, \dots, i_r)', i_j \text{ integers}\}$. First, we assume that the innovation density $g(\cdot)$ is known. Denote the Fisher information matrix \mathbf{F} and the central sequence $\tilde{\Delta}_n$ by $\mathbf{F}(\theta, g)$ and $\tilde{\Delta}_n(\theta, g)$, respectively. Let

$$\hat{\theta}_n = \bar{\theta}_n + n^{-1/2} \mathbf{F}(\bar{\theta}_n, g)^{-1} \tilde{\Delta}_n(\bar{\theta}_n, g)$$

where $\bar{\theta}_n$ is a discrete and \sqrt{n} -consistent estimator of θ (for technical justification for use of discrete estimators, see p. 120 of Kreiss (1987)). In (1.57), we can use the

least squares estimator(LSE) $\hat{\beta}_{LS}$ for β . Then η is estimated by conditional LSE (see Tjøstheim (1986)),

$$\hat{\eta}(\hat{\beta}_{LS}) = \arg \min_{\eta} \sum_{t=2}^n \left[(Y_t - \hat{\beta}'_{LS} z_t)^2 - a(\eta) - \sum_{j=1}^{t-1} b_j(\eta)(Y_{t-j} - \hat{\beta}'_{LS} z_{t-j})^2 \right]^2.$$

Then it is seen that the estimator $(\hat{\beta}'_{LS}, \hat{\eta}'(\hat{\beta}_{LS}))'$ becomes a candidate of $\tilde{\theta}_n$. Similar to LeCam (1986) and Linton (1993), it can be shown that $\hat{\theta}_n$ is asymptotically efficient. If $g(\cdot)$ is unknown, substituting an appropriate nonparametric density estimator $\hat{g}_n(\cdot)$ for $g(\cdot)$, we can propose

$$\hat{\theta}_n = \bar{\theta}_n + n^{-1/2} \mathbf{F}(\bar{\theta}, \hat{g}_n)^{-1} \tilde{\Delta}_n(\bar{\theta}_n, \hat{g}_n).$$

Next, we discuss the problem of testing. Let $\mathcal{M}(\mathbf{B})$ be the linear space spanned by the columns of a matrix \mathbf{B} . Consider the problem of testing the null hypothesis H , under which $\sqrt{n}(\theta - \theta_0) \in \mathcal{M}(\mathbf{B})$ for some given $r \times (r-l)$ matrix \mathbf{B} of full rank and a given vector $\theta_0 \in \mathbb{R}^r$. Then, similar to Sect. 8.2 of Strasser (1985) and p.78 of Taniguchi and Kakizawa (2000), it is seen that the test

$$T_n = \left\| [\mathbf{I}_r - \mathbf{F}^{1/2} \mathbf{B}(\mathbf{B}' \mathbf{F} \mathbf{B})^{-1} \mathbf{B}' \mathbf{F}^{1/2}] \mathbf{F}^{-1/2} \tilde{\Delta}_n \right\|_{\theta = \bar{\theta}_n}^2 \quad (1.66)$$

is asymptotically χ_l^2 -distributed under H , and is locally asymptotic optimal. Here \mathbf{I}_r is the $r \times r$ identity matrix.

As a very general nonlinear model, we introduced the following m -dimensional CHARN model in (1.8):

$$\mathbf{X}_t = \mathbf{F}_\theta(\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}) + \mathbf{H}_\theta(\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-q}) \cdot \boldsymbol{\epsilon}_t \quad (1.67)$$

where $\dim \theta = r$. In what follows, without loss of generality we assume $p = q$, and set down

Assumption 1.9

(i)

$$E_\theta \left\| \mathbf{F}_\theta(\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}) \right\|^2 < \infty, \quad E_\theta \left\| \mathbf{H}_\theta(\mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}) \right\|^2 < \infty, \\ \text{for all } \theta \in \Theta.$$

(ii) There exists $c > 0$ such that

$$c \leq \left\| \mathbf{H}_{\theta'}^{-1/2}(\mathbf{x}) \mathbf{H}_\theta(\mathbf{x}) \mathbf{H}_{\theta'}^{-1/2}(\mathbf{x}) \right\| < \infty,$$

for all $\theta, \theta' \in \Theta$ and for all $\mathbf{x} \in \mathbb{R}^{mp}$.

- (iii) \mathbf{F}_θ and \mathbf{H}_θ are continuously differentiable with respect to θ , and their derivatives $\partial_j \mathbf{F}_\theta$ and $\partial_j \mathbf{H}_\theta$ ($\partial_j = \partial/\partial\theta_j$, $j = 1, \dots, r$) satisfy the condition that there exist square-integrable functions A_j and B_j such that

$$\|\partial_j \mathbf{F}_\theta\| \leq A_j \quad \text{and} \quad \|\partial_j \mathbf{H}_\theta\| \leq B_j, \quad j = 1, \dots, r, \quad \text{for all } \theta \in \Theta.$$

- (iv) the innovation density $p(\cdot)$ of ϵ_t satisfies

$$\lim_{\|\mathbf{u}\| \rightarrow \infty} \|\mathbf{u}\| p(\mathbf{u}) = 0, \quad \int \mathbf{u} \mathbf{u}' p(\mathbf{u}) d\mathbf{u} = \mathbf{I}_m.$$

- (v) the continuous derivative $\mathbf{D}p$ of $p(\cdot)$ exists on \mathbb{R}^m , and

$$\int \|\mathbf{p}^{-1} \mathbf{D}p\|^4 p(\mathbf{u}) d\mathbf{u} < \infty, \quad \int \|\mathbf{u}\|^2 \|\mathbf{p}^{-1} \mathbf{D}p\|^2 p(\mathbf{u}) d\mathbf{u} < \infty.$$

Suppose that an observed stretch $\mathbf{X}^{(n)} = (X_1, \dots, X_n)$ from (1.67) is available. Denote by $\mathcal{P}_{n,\theta}$ the probability distribution of $\mathbf{X}^{(n)}$. For two hypothetical value $\theta, \theta' \in \Theta$, the log-likelihood ratio is

$$\begin{aligned} \Lambda_n(\theta, \theta') &\equiv \log \frac{d\mathcal{P}_{n,\theta'}}{d\mathcal{P}_{n,\theta}} \\ &= \sum_{t=p}^n \log \frac{p\left\{\mathbf{H}_{\theta'}^{-1}(X_t - \mathbf{F}_{\theta'})\right\} \det \mathbf{H}_{\theta'}}{p\left\{\mathbf{H}_\theta^{-1}(X_t - \mathbf{F}_\theta)\right\} \det \mathbf{H}_{\theta'}}. \end{aligned} \quad (1.68)$$

Let $H(p; \theta)$ be the hypothesis under which the concerned model is (1.67) with unknown parameter $\theta \in \Theta$ and the innovation density $p(\cdot)$. We introduce the sequence of contiguous alternatives by

$$\theta_n = \theta + \frac{1}{\sqrt{n}} \mathbf{h}, \quad \mathbf{h} \in \mathcal{S} \subset \mathbb{R}^r, \quad (1.69)$$

where $\mathbf{h} = (h_1, \dots, h_r)'$ and \mathcal{S} is an open set of \mathbb{R}^r .

Then, deriving the asymptotics of $\Lambda_n(\theta, \theta_n)$, Kato et al. (2006) showed

Proposition 1.9 *Under Assumption 1.9, the family $\{\mathcal{P}_{n,\theta} : \theta \in \Theta \subset \mathbb{R}^r\}$, $n \in \mathbb{N}$, is locally asymptotically normal.*

To estimate θ , we may use the maximum likelihood estimator given by

$$\hat{\theta}_{ML} \equiv \arg \max_{\theta} \Lambda(\underline{\theta}, \theta), \quad (1.70)$$

whose asymptotic optimality is shown in the sense of Proposition 1.8.

We often observe that financial time series data are nonstationary. When we deal with nonstationary processes, one of the difficult problems is how to set up an adequate asymptotic theory. For this Dahlhaus (1996a,b, 1997) introduced an important class of nonstationary processes and developed the statistical inference. The following definition is due to Dahlhaus (1996a,b, 1997).

Definition 1.1 A sequence of stochastic processes $X_{t,n}$ ($t = 1, \dots, n : n \geq 1$) is called locally stationary with transfer function A° if there exists a representation

$$X_{t,n} = \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,n}^\circ(\lambda) d\xi(\lambda) \quad (1.71)$$

where

- (i) $\xi(\lambda)$ is a stochastic process on $[-\pi, \pi]$ with $\overline{\xi(\lambda)} = \xi(-\lambda)$ and

$$\text{cum} \{d\xi(\lambda_1), \dots, d\xi(\lambda_k)\} = \eta \left(\sum_{j=1}^k \lambda_j \right) h_k(\lambda_1, \dots, \lambda_{k-1}) d\lambda_1 \cdots d\lambda_{k-1},$$

where $h_1 = 0$, $h_2(\lambda) = 1/(2\pi)$, $h_k(\lambda_1, \dots, \lambda_{k-1}) = h_k/(2\pi)^{k-1}$ for all k , and $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period 2π extension of the Dirac delta function.

- (ii) There exists a constant K and a 2π -periodic function $A : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}$ with $A(u, -\lambda) = \overline{A(u, \lambda)}$ and

$$\sup_{t,\lambda} \left| A_{t,n}^\circ(\lambda) - A \left(\frac{t}{n}, \lambda \right) \right| \leq K n^{-1} \quad (1.72)$$

for all n . $A(u, \lambda)$ is assumed to be continuous in u .

In what follows, we assume that $A(u, \lambda)$ depends on an unknown parameter $\theta = (\theta_1, \dots, \theta_r)' \in \Theta \subset \mathbb{R}^r$, i.e., $A(u, \lambda) = A_\theta(u, \lambda)$. The function $f_\theta(u, \lambda) \equiv |A_\theta(u, \lambda)|^2$ is called the time varying spectral density. Let $X_{1,n}, \dots, X_{n,n}$ be realizations of a locally stationary process with $f_\theta(u, \lambda)$. Writing

$$\epsilon_t = \int_{-\pi}^{\pi} \exp(it\lambda) d\xi(\lambda) \quad (1.73)$$

suppose that ϵ_t 's are i.i.d.(0,1) with probability density $p(\cdot)$. Further, we assume $\{X_{t,n}\}$ has the $AR(\infty)$ representation

$$a_{\theta,t,n}^0(0)\epsilon_t = \sum_{k=0}^{\infty} b_{\theta,t,n}^0(k) X_{t-k,n}. \quad (1.74)$$

Under appropriate regularity conditions, Hirukawa and Taniguchi (2006) showed,

Proposition 1.10 (i) *The log-likelihood ratio $\Lambda_n(\boldsymbol{\theta}, \boldsymbol{\theta}_n)$ between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_n$ given in (1.69) has the LAN properties.*

(ii) *For the class of regular estimators, Proposition 1.8 holds.*

Since the log-likelihood ratio $\Lambda_n(\cdot, \cdot)$ contains $\{\epsilon_s : s \leq 0\}$ which are unobservable, Hirukawa and Taniguchi (2006) introduced a quasi-maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{QML}$ of $\boldsymbol{\theta}$ which is defined by maximizing the feasible one:

$$L_n(\boldsymbol{\theta}) = \prod_{t=1}^n \frac{1}{a_{\boldsymbol{\theta},t,n}^0(0)} P \left\{ \frac{\sum_{k=0}^{t-1} b_{\boldsymbol{\theta},t,n}^0(k) X_{t-k,n}}{a_{\boldsymbol{\theta},t,n}^0(0)} \right\}, \quad (1.75)$$

and showed that $\hat{\boldsymbol{\theta}}_{QML}$ is asymptotically efficient.

For a class of vector linear processes (1.10), assuming that the coefficient matrices satisfy long-range dependence and that $\{\mathbf{U}(t)\}$'s are i.i.d. with probability density $p(\cdot)$, Taniguchi and Kakizawa (2000) showed the results stated in Proposition 1.10, and proved that a feasible quasi-maximum likelihood estimator is asymptotically efficient.

1.4 Introduction to Time Series Financial Engineering

Financial engineering is the construction of various financial positions to manage financial risks. This section uses the terminology “time series financial engineering” because we will develop the arguments based on statistical inference of stochastic processes, i.e., time series analysis.

Assets are defined as contracts that give the right to receive or obligation to provide monetary cash flows (e.g., stock, bank account, bond). Stocks are called risky assets, and bank accounts and bonds are called risk-free assets.

Now the mathematical description is given. Let (Ω, \mathcal{F}, P) be a probability space, and let $\{\mathcal{F}_t\}$ be a family of sub σ -fields of \mathcal{F} satisfying $\mathcal{F}_s \subset \mathcal{F}_r$ ($s \leq r$). Suppose that $\mathbf{S}_t = (S_t^0, S_t^1, \dots, S_t^m)'$ is the price of $(m+1)$ assets at time t ($t = 0, 1, \dots, T$). Usually S_t^0 is taken to be a risk-free asset, and $(S_t^1, \dots, S_t^m)'$ is taken to be a collection of m risky assets. In what follows we assume that $\{\mathbf{S}_t\}$ is a stochastic process on (Ω, \mathcal{F}, P) , and that each \mathbf{S}_t is \mathcal{F}_t -measurable.

If one invests the assets S_t^i with fraction weights $w_{i,t}$ ($i = 0, 1, \dots, m$) satisfying $\sum_{i=0}^m w_{i,t} = 1$, the fraction vector $\mathbf{w}_t = (w_{0,t}, w_{1,t}, \dots, w_{m,t})'$ is called the portfolio. Here \mathbf{w}_t is assumed to be \mathcal{F}_{t-1} -measurable. Then the total investment at time t is

$$v_t(\mathbf{w}_t) = \sum_{i=0}^m w_{i,t} S_t^i \quad (1.76)$$

which is called the value process. If a portfolio \mathbf{w}_t satisfies

$$\sum_{i=0}^m w_{i,t-1} S_t^i = \sum_{i=0}^m w_{i,t} S_t^i \quad (1.77)$$

then it is said to be self-financing, which means that after the initial investment no further capital is either invested or withdrawn. A collection of assets $\mathbf{S}_t = (S_t^0, S_t^1, \dots, S_t^m)'$ is said to admit an arbitrage opportunity if there exists a self-financing portfolio \mathbf{w}_t such that

$$\begin{aligned} v_0(\mathbf{w}_0) &= 0, & v_T(\mathbf{w}_T) &\geq 0, & (\text{P} - \text{a.s.}) \\ P\{v_T(\mathbf{w}_T) > 0\} &> 0. \end{aligned} \quad (1.78)$$

If there is no self-financing portfolio for which (1.78) holds, then the collection of assets \mathbf{S}_t is said to be arbitrage-free. “Arbitrage-free” means the impossibility of achieving a sure, strictly positive gain with a zero initial endowment.

For a self-financing portfolio \mathbf{w}_t , we have $v_t(\mathbf{w}_t) = v_t(\mathbf{w}_{t-1})$. In what follows we understand $v_t(\mathbf{w}_s) = \sum_{i=0}^m w_{i,s} S_t^i$. Hence, for any $l < T$, it holds that

$$\begin{aligned} v_T(\mathbf{w}_T) &= v_l(\mathbf{w}_l) + \sum_{t=l+1}^T [v_t(\mathbf{w}_{t-1}) - v_{t-1}(\mathbf{w}_{t-1})] \\ &= v_l(\mathbf{w}_l) + \sum_{t=l+1}^T \sum_{i=0}^m w_{i,(t-1)} (S_t^i - S_{t-1}^i). \end{aligned} \quad (1.79)$$

Let Q be the probability distribution of $\mathcal{S}_T \equiv \{\mathbf{S}_t : t = 0, 1, \dots, T\}$. Suppose there exists another probability distribution Q^* of \mathcal{S}_T which satisfies

- (i) Q^* is equivalent to Q (i.e., $Q(A) = 0 \Leftrightarrow Q^*(A) = 0$ for any $A \in \mathcal{F}$),
- (ii) $\{\mathbf{S}_t\}$ is a martingale with respect to Q^* , i.e.,

$$E^*\{\mathbf{S}_t | \mathcal{F}_{t-1}\} = \mathbf{S}_{t-1}, \quad \text{a.e.}, \quad (1.80)$$

where $E^*\{\cdot\}$ is the expectation with respect to Q^* .

From (1.79) and (1.80) it is not difficult to see that, for any $l < T$,

$$E^*\{v_T(\mathbf{w}_T) | \mathcal{F}_l\} = v_l(\mathbf{w}_l) \quad Q^* - \text{a.e.}, \quad (1.81)$$

which implies that $v_l(\mathbf{w}_l)$ is a martingale with respect to Q^* . Next we show that $\{\mathbf{S}_t\}$ is arbitrage-free. For this we assume that $\{\mathbf{S}_t\}$ is not arbitrage-free, i.e., (1.78) holds. Then,

$$\begin{aligned} Q(v_0(\mathbf{w}_0) = 0) &= 1, \\ Q(v_T(\mathbf{w}_T) \geq 0) &= 1, \\ Q(v_T(\mathbf{w}_T) > 0) &> 0. \end{aligned} \quad (1.82)$$

Since Q^* is equivalent to Q , (1.82) implies

$$\begin{aligned} Q^*(v_0(\mathbf{w}_0) = 0) &= 1, \\ Q^*(v_T(\mathbf{w}_T) \geq 0) &= 1, \\ Q^*(v_T(\mathbf{w}_T) > 0) &> 0. \end{aligned} \tag{1.83}$$

If we set $l = 0$ in (1.81), then

$$E^*\{v_T(\mathbf{w}_T)|\mathcal{F}_0\} = v_0(\mathbf{w}_0) = 0 \quad Q^* - \text{a.e.}, \tag{1.84}$$

which contradicts (1.83). Hence, $\{S_t\}$ must be arbitrage-free.

Proposition 1.11 (e.g., Kariya and Liu (2003)) *If a collection of assets $S_t = (S_t^0, S_t^1, \dots, S_t^m)'$ ($t = 0, 1, \dots, T$) is a martingale with respect to a probability distribution Q^* which is equivalent to the probability distribution Q of $\{S_0, S_1, \dots, S_T\}$, then the collection of assets S_t is arbitrage-free.*

The probability distribution Q^* is called the equivalent martingale measure. In Proposition 1.11, the reverse statement “if the collection of assets is arbitrage-free, there exists an equivalent martingale measure” holds.

- Definition 1.2** (i) A contingent claim is a nonnegative random variable X representing a payoff at some future time T . We can regard it as a contract that an investor makes at time $t < T$ (e.g., option).
(ii) For a contingent claim X , if there exists a self-financing portfolio \mathbf{w}_T such that

$$X = v_T(\mathbf{w}_T) \tag{1.85}$$

then \mathbf{w}_T is called a replicating portfolio of X .

- (iii) For any contingent claim, if there exists the replicating portfolio, then the market of financial assets is said to be complete.

In the case of arbitrage-free, the statement “the market is complete if and only if there exists a unique equivalent martingale measure” holds.

A derivative is a financial instrument whose value is derived from the value of some underlying instrument such as stock price, interest rate, or foreign exchange rate. Options are one example of many derivatives on the market. A call option gives one the right to buy the underlying asset by a certain date, called the maturity, for a certain price, called the strike price, while a put option gives one the right to sell the underlying asset by a maturity for a strike price. American options can be exercised at any time up to the maturity, but European options can be exercised only at their maturity. European options are easier to price than American options since one does not need to consider the possibility of early exercise. Most of the options traded on exchanges are American.

In what follows we explain these options concretely. Let S_t be the price of an underlying asset at time t . A European call with maturity T and strike price K on this asset is theoretically equivalent to a contingent claim

$$C_T = \max(S_T - K, 0). \quad (1.86)$$

If $S_T > K$ at maturity T , we will buy the asset for the strike price K and sell it immediately in the market, whence the gain is $S_T - K$. On the other hand, if $S_T \leq K$, we will not exercise the right, whence the gain is 0. But, actually, we have to pay the initial cost of purchasing the options, which is called the premium. Therefore the essential gain is $C_T - (\text{premium})$. A European put with maturity T and strike price K is equivalent to a contingent claim

$$P_T \equiv \max(K - S_T, 0). \quad (1.87)$$

European options can only be exercised at maturity date T , but American options can be exercised at any time before or at the maturity date T . An American call option is equivalent to a contingent claim

$$C_{n^*} \equiv \max(S_{n^*} - K, 0), \quad (1.88)$$

where n^* is not determined in advance but depends on the path of the underlying process, i.e., n^* is a random variable such that the event $\{n^* = n\} \in \mathcal{F}_n = \sigma(S_1, S_2, \dots, S_n)$, $n \leq T$. Similarly, we can define an American put option. Although there are various options, finally we just mention an Asian call option whose payoff function is given by

$$\max \left[T^{-1} \sum_{t=1}^T S_t - K, 0 \right]. \quad (1.89)$$

Next we consider the problem of pricing options. We denote the present time and the maturity date of options by t and T , respectively. For a contingent claim X , assume that there exists a replicating portfolio w_t of X constructed on an asset process S_t , i.e.,

$$X = v_T(w_T). \quad (1.90)$$

Suppose that S_t is arbitrage-free. Then, converting values of future payments into their present values by risk-free interest rate r we obtain

$$E^* \{ e^{-r(T-t)} X | \mathcal{F}_t \} = v_t(w_t), \quad (1.91)$$

where $E^* \{ \cdot \}$ is the expectation with respect to an equivalent martingale measure Q^* . Therefore, if X is an option, and if there exists a replicating portfolio of it, then (1.91) implies that the initial capital of the portfolio is

$$E^* \{ e^{-r(T-t)} X | \mathcal{F}_t \}. \quad (1.92)$$

Hence, the reasonable price of the option X should be (1.92). Concrete valuation of (1.92) has been done for the geometric Brownian motion:

$$S_t = S_0 \exp \left[\mu t + \sigma \int_0^t dW_u \right], \quad (t \in [0, T]), \quad (1.93)$$

where $\{W_t\}$ is the Wiener process. Dividing $[0, T]$ into N subintervals with length h , Kariya and Liu (2003) introduced a discretized version of (1.93):

$$S_n = S_0 \exp \left\{ \mu n h + \sigma \sum_{k=1}^n u_k \sqrt{h} \right\}, \quad (\{u_k\} \sim \text{i.i.d. } N(0, 1)), \quad (1.94)$$

where $n = 0, 1, \dots, N$ and $Nh = T$. Then they evaluated the price of a European call option (1.86) at time $t = nh$;

$$C = \exp\{-r(T-t)\} E^* \{\max(S_N - K, 0) | \mathcal{F}_n\} \quad (1.95)$$

as follows. From (1.94), it is seen that

$$S_n = S_{n-1} \exp \left\{ \mu h + \sigma \sqrt{h} u_n \right\}, \quad (1.96)$$

hence,

$$\frac{S_n}{\exp(rnh)} = \frac{S_{n-1}}{\exp\{r(n-1)h\}} \exp \left\{ (\mu - r)h + \sigma \sqrt{h} u_n \right\}. \quad (1.97)$$

If the discounted process $S_n / \exp(rnh)$ becomes a martingale with respect to an equivalent martingale measure Q^* , it should hold that

$$E^* \left[\exp\{(\mu - r)h + \sigma \sqrt{h} u_n\} | \mathcal{F}_{n-1} \right] = 1 \quad \text{a.e.} \quad (1.98)$$

For this, letting the distribution of u_n under Q^* be $N(m, 1)$, we can see that the left-hand side of (1.98) is

$$\exp\{(\mu - r)h\} \times \exp \left\{ m\sigma\sqrt{h} + \frac{\sigma^2}{2}h \right\}, \quad (1.99)$$

which implies that if we take

$$m = -\frac{1}{\sigma\sqrt{h}} \left\{ (\mu - r)h + \frac{\sigma^2}{2}h \right\}, \quad (1.100)$$

then (1.98) holds. Let $u_n^* = u_n - m$, then $u_n^* \sim N(0, 1)$ under Q^* . From (1.96) and (1.100) it follows that

$$S_n = S_{n-1} \exp \left\{ \left(rh - \frac{\sigma^2 h}{2} \right) + \sigma \sqrt{h} u_n^* \right\}, \quad (1.101)$$

under Q^* . Recursively, we obtain

$$\begin{aligned} S_N &= S_n \exp \left\{ \left(r - \frac{\sigma^2}{2} \right) (N - n)h + \sigma \sqrt{h} \sum_{j=n+1}^N u_j^* \right\} \\ &= S_n \exp\{A + BZ\}, \end{aligned} \quad (1.102)$$

whence

$$\begin{aligned} A &= \left(r - \frac{\sigma^2}{2} \right) (N - n)h, \\ B &= \sigma \sqrt{(N - n)h}, \\ Z &= \frac{1}{\sqrt{N - n}} \sum_{j=n+1}^N u_j^* \sim N(0, 1) \quad \text{under } Q^*. \end{aligned}$$

Then we can evaluate the call option (1.95) by (1.102), leading to the Black-Scholes formula (Black and Scholes (1973)):

$$C = S_n \Phi(d_t) - \exp\{-(T - t)r\}K \Phi(d_t - \sigma \sqrt{T - t}), \quad (1.103)$$

where $d_t = \{\log \frac{S_n}{K} + (r + \frac{\sigma^2}{2})(T - t)\} / (\sigma \sqrt{T - t})$, $T = Nh$, $t = nh$ and $\Phi(\cdot)$ is the distribution function of $N(0, 1)$.

Let us return to (1.94) with $n = N$ and $h = T/N$, i.e.,

$$S_N = S_0 \exp \left\{ \mu T + \sigma \sqrt{\frac{T}{N}} \sum_{k=1}^N u_k \right\}. \quad (1.104)$$

As we saw in Sect. 1.2, it is natural to assume that financial returns are non-Gaussian and dependent. In view of this, Tamaki and Taniguchi (2007) derived the third-order Edgeworth expansion of the distribution of $\sqrt{\frac{T}{N}} \sum_{k=1}^N u_k$ assuming that T is fixed and $N \nearrow \infty$, and that $\{u_k\}$ is a non-Gaussian fourth-order stationary process. Based on the Edgeworth expansion they evaluated (1.95) as in the form of

$$A_1 + N^{-1/2} A_2 + N^{-1} A_3 + o(N^{-1}), \quad (1.105)$$

without martingale adjustment. Next, imposing the martingale constraint on $S_n/\exp(rnh)$, Tamaki and Taniguchi (2007) derived an Edgeworth approximation for the arbitrage-free price (1.95) in the form:

$$C = M_1 + N^{-1/2}M_2 + N^{-1}M_3 + o(N^{-1}), \tag{1.106}$$

where M_1 equals the Black-Scholes formula. They discussed the influence of non-Gaussianity and dependence of $\{u_k\}$ on the higher order terms M_2 and M_3 .

In the field of financial engineering the problem of credit rating is an important one. Usually, the credit rating has been done by use of the discriminant analysis in i.i.d. settings, although financial data are often supposed to be non-Gaussian dependent. For dependent observations the discriminant analysis has been developed by Shumway and Unger (1974), Zhang and Taniguchi (1994), and Kakizawa et al. (1998). For locally stationary processes, Sakiyama and Taniguchi (2004) addressed the problem of classification. Because financial data are often nonstationary, in what follows, we state discrimination and classification for dependent observations in line with Sakiyama and Taniguchi (2004).

Recall that we already introduced a scalar-valued locally stationary process in (1.71). Here we generalize it to the case of vector-valued.

Definition 1.3 A sequence of vector-valued stochastic processes $X_{t,n} = (X_{t,n}^{(1)}, \dots, X_{t,n}^{(d)})'$ ($t = 1, \dots, n$) is called locally stationary with transfer function matrix $A_{t,n}(\lambda) = \{A_{t,n}(\lambda)_{ab} : a, b = 1, \dots, d\}$ and mean $\mathbf{0}$ if there exists a representation

$$X_{t,n} = \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,n}(\lambda) d\xi(\lambda), \tag{1.107}$$

where the following holds:

(i) $\xi(\lambda) = (\xi_1(\lambda), \dots, \xi_d(\lambda))'$ is a complex-valued vector process on $[-\pi, \pi]$ with $\xi_a(\lambda) = \xi_a(-\lambda)$, $E\{\xi_{aj}(\lambda)\} = 0$ and

$$\text{cum}\{d\xi_{a_1}(\lambda_1), \dots, d\xi_{a_k}(\lambda_k)\} = \eta \left(\sum_{j=1}^k \lambda_j \right) g_{a_1, \dots, a_k}(\lambda_1, \dots, \lambda_{k-1}) d\lambda_1 \cdots d\lambda_k \tag{1.108}$$

for all $a_1, \dots, a_k \in \{1, \dots, d\}$ and $\eta(\lambda) = \sum_{l=-\infty}^{\infty} \delta(\lambda + 2\pi l)$, ($\delta(\cdot)$ is the delta function).

(ii) There exists a constant K and a 2π -periodic matrix-valued function $A(u, \lambda) = \{A(u, \lambda)_{ab} : a, b = 1, \dots, d\} : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}^{d \times d}$ with $\overline{A(u, \lambda)} = A(u, -\lambda)$ and

$$\sup_{t, \lambda} \left| A_{t,n}(\lambda)_{a,b} - A \left(\frac{t}{n}, \lambda \right)_{a,b} \right| \leq Kn^{-1} \tag{1.109}$$

for all $a, b \in \{1, \dots, d\}$ and $n \in \mathbb{N}$, where $A(u, \lambda)$ is assumed to be continuous in u .

We call $f(u, \lambda) \equiv A(u, \lambda)A(u, \lambda)^*$ the time-varying spectral density matrix of $\{\mathbf{X}_{t,n}\}$. Letting

$$d_N^{(a)}(u, \lambda) = \sum_{s=0}^{N-1} X_{[un]-N/2+s+1,n}^{(a)} \exp(-i\lambda s) \quad (1.110)$$

we introduce the periodogram matrix $I_N(u, \lambda) = \{I_N(u, \lambda)_{a,b} : a, b = 1, \dots, d\}$ over a segment of length N with midpoint $[un]$, where

$$I_N(u, \lambda)_{a,b} = \frac{1}{2\pi N} d_N^{(a)}(u, \lambda) d_N^{(b)}(u, -\lambda) \quad (1.111)$$

The shift from segment to segment is denoted by N . $I_N(u_j, \lambda)$ is calculated over segments with midpoints $u_j n = t_j = N(j-1/2)$ ($j = 1, \dots, M$), where $n = NM$.

For $\phi : [0, 1] \times [-\pi, \pi] \rightarrow \mathbb{C}^{d \times d}$, define

$$J_n(\phi) \equiv \frac{1}{M} \sum_{j=1}^M \int_{-\pi}^{\pi} \text{tr}\{\phi(u_j, \lambda) I_N(u_j, \lambda)\} d\lambda \quad (1.112)$$

and

$$J(\phi) \equiv \int_0^1 \int_{-\pi}^{\pi} \text{tr}\{\phi(u, \lambda) f(u, \lambda)\} dud\lambda \quad (1.113)$$

Under appropriate regularity conditions, Dahlhaus (1997) showed that

$$(i) \quad J_n(\phi) \xrightarrow{P} J(\phi), \quad (n \rightarrow \infty), \quad (1.114)$$

$$(ii) \quad \sqrt{n}\{J_n(\phi) - J(\phi)\} \xrightarrow{\mathcal{L}} N(0, V), \quad (n \rightarrow \infty), \quad (1.115)$$

where V is written in terms of ϕ , $f(u, \lambda)$ and $g_{a_1, \dots, a_4}(\cdot)$.

Next, we discuss the problem of classifying a vector-valued locally stationary process $\{\mathbf{X}_{t,n}\}$ into one of two categories described by two hypotheses:

$$\Pi_1 : f(u, \lambda), \quad \Pi_2 : g(u, \lambda), \quad (1.116)$$

where $f(u, \lambda)$ and $g(u, \lambda)$ are $d \times d$ time-varying spectral density matrices. For this we use

$$D(\mathbf{f} : \mathbf{g}) \equiv \frac{1}{4\pi M} \sum_{j=1}^M \int_{-\pi}^{\pi} \left[\log \left\{ \frac{|\mathbf{g}(u_j, \lambda)|}{|\mathbf{f}(u_j, \lambda)|} \right\} + \text{tr}[\mathbf{I}_N(u_j, \lambda)\{\mathbf{g}^{-1}(u_j, \lambda) - \mathbf{f}^{-1}(u_j, \lambda)\}] \right] \quad (1.117)$$

as a classification statistic. That is, if $D(\mathbf{f} : \mathbf{g}) > 0$ we choose category Π_1 . Otherwise we choose category Π_2 .

If we use $D(\mathbf{f} : \mathbf{g})$ as a classification criterion, the misclassification probabilities are $P(2|1) = P\{D(\mathbf{f} : \mathbf{g}) \leq 0 | \Pi_1\}$, $P(1|2) = P\{D(\mathbf{f} : \mathbf{g}) > 0 | \Pi_2\}$. Then, our classification statistic is asymptotically consistent, i.e.,

Proposition 1.12 (Sakiyama and Taniguchi (2004))

$$\lim_{n \rightarrow \infty} P(2|1) = 0, \quad \lim_{n \rightarrow \infty} P(1|2) = 0 \quad (1.118)$$

Proposition 1.12 implies that the classification criterion $D(\mathbf{f} : \mathbf{g})$ has the fundamental goodness. To evaluate more delicate goodness of $D(\mathbf{f} : \mathbf{g})$ we assume that $\mathbf{g}(u, \lambda)$ is contiguous to $\mathbf{f}(u, \lambda)$. Now we set the spectral densities as

$$\Pi_1 : \mathbf{f}(u, \lambda) = \mathbf{f}(u, \lambda), \quad \Pi_2 : \mathbf{g}(u, \lambda) = \mathbf{f}(u, \lambda) + n^{-1/2} \mathbf{h}(u, \lambda), \quad (1.119)$$

where $\mathbf{h}(u, \lambda)$ is a $d \times d$ matrix-valued function, and assumed to be nonnegative definite on $[0, 1] \times [-\pi, \pi]$.

Proposition 1.13 (Sakiyama and Taniguchi (2004)) *Suppose that $\mathbf{f}(u, \lambda)$ and $\mathbf{h}(u, \lambda)$ are continuous on $[0, 1] \times [-\pi, \pi]$, and that N and n fulfill the relations $n^{1/4} \ll N \ll n^{1/2} / \log n$. Then, under (1.119),*

$$\lim_{n \rightarrow \infty} P(2|1) = \lim_{n \rightarrow \infty} P(1|2) = \Phi \left[\frac{-F/2}{\{F + H\}^{1/2}} \right], \quad (1.120)$$

where $\Phi(\cdot)$ is the distribution function of $N(0, 1)$ and

$$\begin{aligned} F &= \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \text{tr}\{\mathbf{h}(u, \lambda) \mathbf{f}^{-1}(u, \lambda)\}^2 du d\lambda, \\ H &= \frac{1}{8\pi} \int_0^1 \sum_{b_1, b_2, b_3, b_4=1}^d \\ &\quad \times \left[\int_{-\pi}^{\pi} \mathbf{A}(u, \lambda)^* \{\mathbf{f}^{-1}(u, \lambda) \mathbf{h}(u, \lambda) \mathbf{f}^{-1}(u, \lambda)\} \mathbf{A}(u, \lambda) d\lambda \right]_{b_2 b_1} \\ &\quad \times \left[\int_{-\pi}^{\pi} \mathbf{A}(u, \mu)^* \{\mathbf{f}^{-1}(u, \mu) \mathbf{h}(u, \mu) \mathbf{f}^{-1}(u, \mu)\} \mathbf{A}(u, \mu) d\mu \right]_{b_3 b_4} \\ &\quad \times g_{b_1 b_2 b_3 b_4}(\lambda, -\lambda, -\mu) du \end{aligned} \quad (1.121)$$

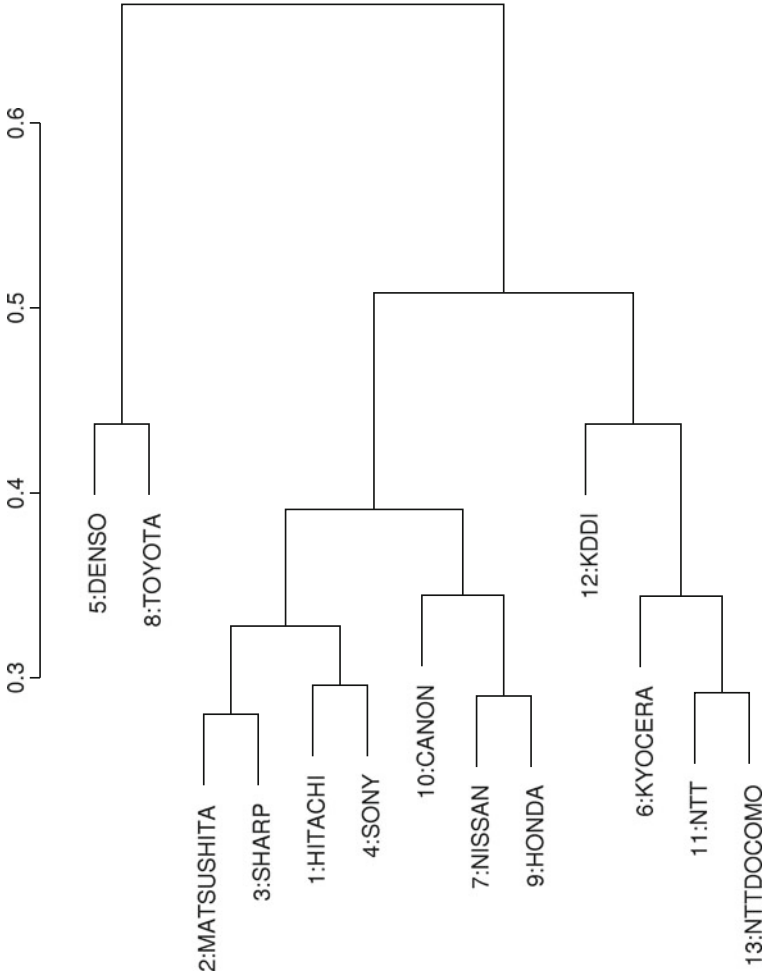


Fig. 1.4 Dendrogram for 13 financial data. Taken from Hirukawa (2006). Published with the kind permission of ©World Scientific Publishing Co. 2006. All Rights Reserved

In Proposition 1.13, the misclassification probabilities depend on H which shows an influence of non-Gaussianity of the process. Sakiyama and Taniguchi (2004) discussed non-Gaussian robustness of $D(\mathbf{f} : \mathbf{g})$ checking whether $H = 0$ or $H \neq 0$.

Suppose $\{X_{t,n}\}$ belongs to categories $\Pi_j : f(u, \lambda) = f_j(u, \lambda)$. For actual stock data, Hirukawa (2006) introduced the following classification statistic:

$$D_H(\hat{f}_j : \hat{f}_k) \equiv \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} H \left\{ \frac{\hat{f}_j(u, \lambda)}{\hat{f}_k(u, \lambda)} \right\} d\lambda du, \quad (1.122)$$

where $\hat{f}_j(u, \lambda)$ is a consistent nonparametric estimator of $f_j(u, \lambda)$, and $H(\cdot)$ is a smooth function. Hirukawa (2006) executed the hierarchical clustering based on $D_H(\hat{f}_j : \hat{f}_k)$ for daily log-returns of 13 companies: 1. HITACHI, 2. MATSUSHITA, 3. SHARP, 4. SONY, 5. DENSO, 6. KYOCERA, 7. NISSAN, 8. TOYOTA, 9. HONDA, 10. CANON, 11. NTT, 12. KDDI, 13. NTTDOCOMO.

Figure 1.4 shows the dendrogram for $H(z) = z - \log z - 1$. The figure classifies the type of industry clearly, which implies that the method will be useful for the problem of credit rating.

References

- Black, F., Scholes, M.: The pricing of options and corporate liabilities. *J. Polit. Econom.* **81**, 637–654 (1973)
- Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**(3), 307–327 (1986)
- Brillinger, D.R.: *Time Series: Data Analysis and Theory*, expanded edn. Holden-Day, San Francisco (2001)
- Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*, 2nd edn. Springer, New York (1991)
- Brown, B.M.: Martingale central limit theorems. *Ann. Math. Stat.* **42**, 59–66 (1971)
- Chen, M., An, H.Z.: A note on the stationarity and the existence of moments of the GARCH model. *Stat. Sinica* **8**, 505–510 (1998)
- Dahlhaus, R.: Maximum likelihood estimation and model selection for nonstationary processes. *J. Nonparametric Stat.* **6**, 171–191 (1996a)
- Dahlhaus, R.: On the Kullback-Leibler information divergence of locally stationary processes. *Stoch. Process. Appl.* **62**(1), 139–168 (1996b)
- Dahlhaus, R.: Fitting time series models to nonstationary processes. *Ann. Stat.* **25**, 1–37 (1997)
- Dunsmuir, W., Hannan, E.J.: Vector linear time series model. *Adv. Appl. Probab.* **8**, 339–364 (1976)
- Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007 (1982)
- Garel, B., Hallin, M.: Local asymptotic normality of multivariate ARMA processes with a linear trend. *Ann. Inst. Stat. Math.* **47**(3), 551–579 (1995)
- Geweke, J.: Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.* **77**, 304–313 (1982)
- Giraitis, L., Surgailis, D.: A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Relat. Fields* **86**(1), 87–104 (1990)
- Giraitis, L., Kokoszka, P., Leipus, R.: Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory* **16**, 3–22 (2000)
- Gouriéroux, C., Jasiak, J.J.: *Financial Econometrics: Problems, Models, and Methods*. Princeton University Press, Princeton (2001)
- Hallin, M., Ingenbleek, J.F., Puri, M.L.: Linear serial rank tests for randomness against ARMA alternatives. *Ann. Stat.* **13**, 1156–1181 (1985)
- Hannan, E.J.: *Multiple Time Series*. Wiley, New York (1970)
- Härdle, W., Tsybakov, A., Yang, L.: Nonparametric vector autoregression. *J. Stat. Plann. Infer.* **68**(2), 221–245 (1998)
- Hirukawa, J.: Cluster analysis for non-Gaussian locally stationary processes. *Int. J. Theor. Appl. Finance* **9–1**, 113–132 (2006)

- Hirukawa, J., Taniguchi, M.: LAN theorem for non-Gaussian locally stationary processes and its applications. *J. Stat. Plann. Infer.* **136**(3), 640–688 (2006)
- Hosoya, Y., Taniguchi, M.: A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Stat.* **10**, 132–153 (1982). Correction: **21**, 1115–1117 (1993).
- Jeganathan, P.: Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* **11**(05), 818–887 (1995)
- Kakizawa, Y.: Discriminant analysis for non-Gaussian vector stationary processes. *J. Nonparametric Stat.* **7**(2), 187–203 (1996)
- Kakizawa, Y., Shumway, T.H., Taniguchi, M.: Discrimination and clustering for multivariate time series. *J. Am. Stat. Assoc.* **93**, 328–340 (1998)
- Kariya, T., Liu, R.Y.: *Asset Pricing: Discrete Time Approach*. Kluwer Academic Publishers, Boston (2003)
- Kato, H., Taniguchi, M., Honda, M.: Statistical analysis for multiplicatively modulated nonlinear autoregressive model and its applications to electrophysiological signal analysis in humans. *IEEE Trans. Signal Process.* **54**(9), 3414–3425 (2006)
- Kreiss, J.P.: On adaptive estimation in stationary ARMA processes. *Ann. Stat.* **15**, 112–133 (1987)
- Kreiss, J.P.: Local asymptotic normality for autoregression with infinite order. *J. Stat. Plann. Infer.* **26**(2), 185–219 (1990)
- Le Cam, L.M.: *Asymptotic Methods in Statistical Decision Theory*. Springer, New York (1986)
- Lee, S., Taniguchi, M.: Asymptotic theory for ARCH-SM models: LAN and residual empirical processes. *Stat. Sinica* **15**, 215–234 (2005)
- Liggett Jr, W.: On the asymptotic optimality of spectral analysis for testing hypotheses about time series. *Ann. Math. Stat.* **42**, 1348–1358 (1971)
- Ling, S., Li, W.: On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Am. Stat. Assoc.* **92**, 1184–1194 (1997)
- Ling, S., McAleer, M.: On adaptive estimation in nonstationary ARMA models with GARCH errors. *Ann. Stat.* **31**, 642–674 (2003)
- Linton, O.: Adaptive estimation in ARCH models. *Econometric Theory* **9**(4), 539–569 (1993)
- Lu, Z., Jiang, Z.: L_1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term. *Stat. Probab. Lett.* **51**(2), 121–130 (2001)
- Magnus, J.R., Neudecker, H.: *Matrix differential calculus with applications in statistics and econometrics*. Wiley, New York (1988)
- Nelson, D.B.: Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**, 347–370 (1991)
- Sakiyama, K., Taniguchi, M.: Discriminant analysis for locally stationary processes. *J. Multivariate Anal.* **90**, 282–300 (2004)
- Shumway, R.H., Unger, A.N.: Linear discriminant functions for stationary time series. *J. Am. Stat. Assoc.* **69**, 948–956 (1974)
- Strasser, H.: *Mathematical Theory of Statistics*. Walter de Gruyter, Berlin (1985)
- Swensen, A.R.: The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* **16**(1), 54–70 (1985)
- Tamaki, K., Taniguchi, M.: Higher order asymptotic option valuation for non-Gaussian dependent return. *J. Statist. Plan. Inf. Special Issue Honor of Professor Madan L. Puri.* **137**, 1043–1058 (2007).
- Taniguchi, M.: On estimation of the integrals of the fourth order cumulant spectral density. *Biometrika* **69**(1), 117–122 (1982)
- Taniguchi, M., Hirukawa, J., Tamaki, K.: Optional statistical inference in financial engineering. In: *Financial Mathematics Series*. Chapman & Hall, New York (2008)
- Taniguchi, M., Kakizawa, Y.: *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York (2000)
- Taniguchi, M., Puri, M.L., Kondo, M.: Nonparametric approach for non-Gaussian vector stationary processes. *J. Multivariate Anal.* **56**(2), 259–283 (1996)

- Tjøstheim, D.: Estimation in nonlinear time series models. *Stoch. Process. Appl.* **21**(2), 251–273 (1986)
- Van der Vaart, A.W.: *Asymptotic Statistics*. Cambridge University Press, Cambridge (1998)
- Zhang, G., Taniguchi, M.: Discriminant analysis for stationary vector time series. *J. Time Ser. Anal.* **15**, 117–126 (1994)

Chapter 2

Empirical Likelihood Approaches for Financial Returns

Abstract We deal with an empirical likelihood and apply it to several financial problems. Empirical likelihood is one of the nonparametric methods of statistical inference. It allows us to use likelihood methods although we do not assume that the data comes from a known family. Consequently, the empirical likelihood has both effectiveness and flexibility of the likelihood method, and reliability of the nonparametric methods. The construction of this chapter is as follows. We briefly look at the history of the empirical likelihood in Sect. 2.1 and review its method for i.i.d. data in Sect. 2.2. The frequency domain approach of empirical likelihood for multivariate non-Gaussian linear processes is discussed in Sect. 2.3. Section 2.4 gives extensions of the empirical likelihood such as Cressie-Read power-divergence statistic and generalized empirical likelihood. Section 2.5 considers application of the generalized empirical likelihood to an inference problem for multivariate stable distributions. Technical proofs of the theorems are given in Sect. 2.6.

Keywords Empirical likelihood · Non-Gaussian linear process · Frequency domain · Portfolio · Generalized empirical likelihood · Stable distribution

2.1 Introduction

Empirical likelihood is originally proposed by Owen (1988, 1990) and its overview is found in Owen (2001). Because of its advantages, many researchers investigated the empirical likelihood. For example, Qin and Lawless (1994) linked the estimating functions and equations to empirical likelihood. Newey and Smith (2004) considered generalized empirical likelihood, which is a richer class than empirical likelihood, and gave its higher order properties. For the application of empirical likelihood to the dependent data, Kitamura (1997) proposed blockwise empirical likelihood for the weakly dependent processes. Monti (1997) gave the frequency domain approach to the empirical likelihood methods for linear processes and Ogata and Taniguchi (2010)

extended it to the multivariate one. Ogata and Taniguchi (2009) investigated the asymptotic properties of Cressie-Read divergence statistics, which encompasses the empirical likelihood, for multivariate linear processes. A frequency domain empirical likelihood for the processes including long-range dependence is considered by Nordman and Lahiri (2006).

2.2 Empirical Likelihood Method for i.i.d. Data

Given $X_1, \dots, X_n \in \mathbb{R}^m$, assumed independent with common distribution function F_0 , the nonparametric likelihood of distribution function F is defined by

$$L(F) = \prod_{i=1}^n F(\{X_i\}), \quad (2.1)$$

where $F(\{X_i\})$ is the probability of getting the value X_i in a sample from F . Apparently, only the distributions which have the positive point mass probability on each observation constitute positive nonparametric likelihoods. Therefore, we restrict F to the one having the probability $p_i = F(\{X_i\}) > 0$ on each observation X_i . By a simple calculation, we find the maximizer of the nonparametric likelihood (2.1) turns to be the empirical distribution function F_n , placing probability $1/n$ on each observation. Therefore, similar to the parametric case, nonparametric likelihood ratio of F to the maximizer F_n is defined by:

$$R(F) = \frac{L(F)}{L(F_n)} = \frac{\prod_{i=1}^n p_i}{\prod_{i=1}^n 1/n} = \prod_{i=1}^n (np_i).$$

Suppose that we are interested in a parameter $\theta \in \Theta \subset \mathbb{R}^p$ and that the true value θ_0 satisfies the following estimating equation

$$E[\mathbf{m}(X; \theta_0)] = \mathbf{0} \quad (2.2)$$

where $\mathbf{m}(X; \theta) \in \mathbb{R}^q$ is a vector-valued function, called estimating function. As examples of estimating functions, we take $\mathbf{m}(X; \theta) = X - \theta$ to indicate a vector mean by Eq. (2.2). For $\Pr(X \in A)$, we take $\mathbf{m}(X; \theta) = 1_{X \in A} - \theta$. For a continuously distributed scalar X and $\theta \in \mathbb{R}$, the function $m(X; \theta) = 1_{X < \theta} - \alpha$ defines θ as the α quantile of X . Now we define the empirical likelihood ratio function for θ by

$$\mathcal{R}(\theta) = \max_p \left\{ \prod_{i=1}^n (np_i) \left| \sum_{i=1}^n p_i \mathbf{m}(X_i; \theta) = \mathbf{0}, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right. \right\}, \quad (2.3)$$

where $\mathbf{p} = (p_1, \dots, p_n)$. This is the maximum of the nonparametric likelihood ratio with the restriction that the mean of the estimating function is zero under the distribution F . $\mathcal{R}(\boldsymbol{\theta})$ is calculated by the Lagrange multiplier method as follows. Write

$$G = \sum_{i=1}^n \log(np_i) - n\boldsymbol{\xi}' \sum_{i=1}^n p_i \mathbf{m}(X_i; \boldsymbol{\theta}) + \gamma \left(\sum_{i=1}^n p_i - 1 \right),$$

where $\boldsymbol{\xi} \in \mathbb{R}^q$ and $\gamma \in \mathbb{R}$ are Lagrange multipliers. Setting $\partial G / \partial p_i = 0$ gives

$$\frac{\partial G}{\partial p_i} = \frac{1}{p_i} - n\boldsymbol{\xi}' \mathbf{m}(X_i; \boldsymbol{\theta}) + \gamma = 0.$$

Therefore, the equation $\sum_{i=1}^n p_i (\partial G / \partial p_i) = 0$ gives $\gamma = -n$. Then, we may write

$$p_i = \frac{1}{n} \frac{1}{1 + \boldsymbol{\xi}' \mathbf{m}(X_i; \boldsymbol{\theta})}$$

where the vector $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta})$ satisfies q equations given by

$$\frac{1}{n} \sum_{i=1}^n \frac{\mathbf{m}(X_i; \boldsymbol{\theta})}{1 + \boldsymbol{\xi}' \mathbf{m}(X_i; \boldsymbol{\theta})} = \mathbf{0}. \quad (2.4)$$

Finally, we may write

$$\mathcal{R}(\boldsymbol{\theta}) = \prod_{i=1}^n (np_i) = \prod_{i=1}^n \frac{1}{1 + \boldsymbol{\xi}' \mathbf{m}(X_i; \boldsymbol{\theta})}. \quad (2.5)$$

The following theorem is due to Owen (2001).

Theorem 2.1 (Owen (2001, Theorem 3.4)) *Assume that $\text{Var}(\mathbf{m}(X; \boldsymbol{\theta}_0))$ is finite and has rank r . Then $-2 \log \mathcal{R}(\boldsymbol{\theta}_0) \rightarrow \chi_{(r)}^2$ in distribution as $n \rightarrow \infty$.*

2.3 Estimation with Frequency Domain Empirical Likelihood for Stationary Processes

Here we are concerned with the m -dimensional linear process $\{X(t)\}$ in (1.10) and consider the problem of estimating parameter $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. Suppose that the information of $\boldsymbol{\theta}$ exists through a system of general estimating equations in frequency domain as follows. Let $\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta})$, ($j = 1, \dots, q$) be $m \times m$ matrix-valued continuous functions on $[-\pi, \pi]$ satisfying $\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta}) = \boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta})^*$ and $\boldsymbol{\phi}_j(-\lambda; \boldsymbol{\theta}) = \boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta})'$. We assume that each $\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta})$ satisfies the spectral moment condition

$$\int_{-\pi}^{\pi} \text{tr}\{\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta}_0) \mathbf{f}(\lambda)\} d\lambda = 0 \quad (j = 1, \dots, q) \quad (2.6)$$

where $\boldsymbol{\theta}_0 = (\theta_{10}, \dots, \theta_{p0})'$ is the true value of parameter and $\mathbf{f}(\lambda)$ is the true spectral density matrix of $\mathbf{X}(t)$. By taking an appropriate function for $\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta})$, the $\boldsymbol{\theta}_0$ specified by Eq. (2.6) can express various important indices for time series.

As an estimating function, we consider the following quantity

$$\mathbf{m}(\lambda_t; \boldsymbol{\theta}) = \left(\text{tr}\{\boldsymbol{\phi}_1(\lambda_t; \boldsymbol{\theta}) \mathbf{I}_n(\lambda_t)\}, \dots, \text{tr}\{\boldsymbol{\phi}_q(\lambda_t; \boldsymbol{\theta}) \mathbf{I}_n(\lambda_t)\} \right)' \quad (2.7)$$

where $\lambda_t = (2\pi t)/n$, ($t = 1, \dots, n$) and $\mathbf{I}_n(\lambda)$ is the periodogram matrix. Following (2.3), we define the empirical likelihood ratio function for frequency domain by

$$\tilde{\mathcal{R}}(\boldsymbol{\theta}) = \max_{\mathbf{p}} \left\{ \prod_{t=1}^n (np_t) \left| \sum_{t=1}^n p_t \mathbf{m}(\lambda_t; \boldsymbol{\theta}) = \mathbf{0}, p_t \geq 0, \sum_{t=1}^n p_t = 1 \right. \right\}. \quad (2.8)$$

Now we impose the following assumption.

Assumption 2.1

- (i) $\{\mathbf{X}(t)\}$ satisfies (B) in Proposition 1.2.
- (ii) For the sequence $\{C_k\}$ defined by

$$C_k = \sup_{a_1 \dots a_k} \sum_{t_1, \dots, t_{k-1} = -\infty}^{\infty} |c_{a_1 \dots a_k}^X(t_1, \dots, t_{k-1})|,$$

it holds

$$\sum_{k=1}^{\infty} \frac{C_k z^k}{k!} < \infty$$

for z in a neighborhood of zero. Here, $c_{a_1 \dots a_k}^X(t_1, \dots, t_{k-1})$ is the k -th-order cumulant of \mathbf{X} defined in Chap. 1.

Applying Proposition 1.2, we obtain the following theorem.

Theorem 2.2 (Ogata and Taniguchi (2010)) *Let Assumption 2.1 hold. Then*

$$-2 \log \tilde{\mathcal{R}}(\boldsymbol{\theta}_0) \rightarrow (\boldsymbol{\Sigma} N)' (\boldsymbol{\Sigma} N),$$

where N has a q -dimensional standard normal distribution and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{\Sigma}_1^{1/2}$. Here $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are constant $q \times q$ matrices whose (i, j) -th components are

$$\begin{aligned} & \frac{1}{\pi} \int_{-\pi}^{\pi} \text{tr}\{\mathbf{f}(\lambda)\boldsymbol{\phi}_i(\lambda; \boldsymbol{\theta}_0)\mathbf{f}(\lambda)\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta}_0)\}d\lambda \\ & + \frac{1}{2\pi} \sum_{r,t,u,v=1}^m \int \int_{-\pi}^{\pi} \phi_{rt}^{(i)}(\lambda_1; \boldsymbol{\theta}_0)\phi_{uv}^{(j)}(\lambda_2; \boldsymbol{\theta}_0)\mathcal{Q}_{rtuv}^X(-\lambda_1, \lambda_2, -\lambda_2)d\lambda_1d\lambda_2 \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{2\pi} \left[\int_{-\pi}^{\pi} \text{tr}\{\mathbf{f}(\lambda)\boldsymbol{\phi}_i(\lambda; \boldsymbol{\theta}_0)\mathbf{f}(\lambda)\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta}_0)\}d\lambda \right. \\ & \left. + \int_{-\pi}^{\pi} \text{tr}\{\mathbf{f}(\lambda)\boldsymbol{\phi}_i(\lambda; \boldsymbol{\theta}_0)\}\text{tr}\{\mathbf{f}(\lambda)\boldsymbol{\phi}_j(\lambda; \boldsymbol{\theta}_0)\}d\lambda \right], \end{aligned}$$

respectively.

The proof of Theorem 2.2 is given in Sect. 2.6. In what follows, we give several application examples of the theorem.

Example 2.1 (Autocorrelation) Denote the autocovariance and the autocorrelation of the process $\{X_i(t)\}$ (i -th component of the process $\{\mathbf{X}(t)\}$) with lag h by $\gamma_i(h)$ and $\rho_i(h)$, respectively. Suppose that we are interested in the joint estimation of $\rho_i(h)$ and $\rho_j(k)$. Take

$$\begin{aligned} \boldsymbol{\phi}_1(\lambda; \boldsymbol{\theta}) &= \begin{cases} \cos(h\lambda) - \theta_1 & (i, i)\text{-th component} \\ 0 & \text{otherwise} \end{cases}, \\ \boldsymbol{\phi}_2(\lambda; \boldsymbol{\theta}) &= \begin{cases} \cos(k\lambda) - \theta_2 & (j, j)\text{-th component} \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

Then (2.6) leads to $\theta_{10} = \gamma_i(h)/\gamma_i(0)$ and $\theta_{20} = \gamma_j(k)/\gamma_j(0)$, hence, then $\boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20})'$ corresponds to the desired autocorrelations $\boldsymbol{\rho} = (\rho_i(h), \rho_j(k))'$. ■

Example 2.2 (Portfolio selection) Let $X_i(t)$ be the log-return of i -th asset ($i = 1, \dots, m$) at time t and suppose that the process $\{\mathbf{X}(t) = (X_1(t), \dots, X_m(t))'\}$ is stationary with zero mean. Consider the portfolio $p(t) = \sum_{i=1}^m \theta_i X_i(t)$ where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$ is a vector of weights. The process $\{p(t)\}$ is a linear combination of the stationary process, hence $\{p(t)\}$ is also stationary and, from Herglotz's theorem, its variance is

$$\text{Var}\{p(t)\} = \boldsymbol{\theta}'\text{Var}\{\mathbf{X}(t)\}\boldsymbol{\theta} = \boldsymbol{\theta}'\left(\int_{-\pi}^{\pi} \mathbf{f}(\lambda)d\lambda\right)\boldsymbol{\theta}.$$

Our aim is to find the weights $\boldsymbol{\theta}_0 = (\theta_{10}, \dots, \theta_{m0})'$, that minimize the variance (the risk) of the portfolio $p(t)$. The first-order condition $(\partial/\partial\boldsymbol{\theta})_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\text{Var}\{p(t)\} = \mathbf{0}$ leads to

$$\left(\int_{-\pi}^{\pi} \mathbf{f}(\lambda) + \mathbf{f}(\lambda)'d\lambda\right)\boldsymbol{\theta}_0 = \mathbf{0}. \quad (2.9)$$

Now, for $j = 1, \dots, m$, consider to take

$$\phi_j(\lambda; \boldsymbol{\theta}) = \begin{cases} \theta_i & (j, i)\text{-th and } (i, j)\text{-th component, } (i = 1, \dots, m \text{ and } i \neq j) \\ 2\theta_j & (j, j)\text{-th component} \\ 0 & \text{otherwise} \end{cases}.$$

Then, (2.6) coincides with the condition (2.9), which implies that the best portfolio weights can be solved with the framework of the spectral moment condition. \blacksquare

Example 2.3 (Whittle estimation) Consider fitting a parametric spectral density model $\mathbf{f}_\theta(\lambda)$ to the true spectral density $\mathbf{f}(\lambda)$, and seeking the quasi-true value $\underline{\boldsymbol{\theta}}$ in (1.21). Assume that the spectral density model has the form described in Assumption (NGR) (iii). The Kolmogorov's formula (cf. p. 162 of Hannan (1970)) says

$$\det \mathbf{K} = \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \det \{2\pi \mathbf{f}_\theta(\lambda)\} d\lambda \right\}.$$

This implies that, if $\boldsymbol{\theta}$ is innovation-free, the quantity $\int_{-\pi}^{\pi} \log \det \{\mathbf{f}_\theta(\lambda)\} d\lambda$ is independent of $\boldsymbol{\theta}$ and (1.21) leads to

$$\left. \frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \text{tr} \{ \mathbf{f}_\theta(\lambda)^{-1} \mathbf{f}(\lambda) \} d\lambda \right|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}} = \mathbf{0}.$$

This corresponds to (2.6) when we set

$$\phi_j(\lambda; \boldsymbol{\theta}) = \frac{\partial \mathbf{f}_\theta(\lambda)^{-1}}{\partial \theta_j} \quad (j = 1, \dots, p),$$

so the quasi-true value can be expressed by the spectral moment condition. \blacksquare

2.4 Extensions of Empirical Likelihood

Since the classical empirical likelihood method was proposed, several extensions have been considered by many researchers. Among those, we introduce Cressie-Read power-divergence (CR) statistic and generalized empirical likelihood (GEL).

As an alternative of the empirical likelihood ratio, Baggerly (1998) made use of the CR statistic, which is used at χ^2 -goodness of fit test. It is defined as:

$$\frac{2}{\nu(\nu+1)} \sum_{i=1}^k N_i \left\{ \left(\frac{N_i}{np_i} \right)^\nu - 1 \right\}, \quad \nu \in \mathbb{R},$$

where k is a number of categories, N_i is a number of the observations which fell into the i -th category and p_i is the probability that the observation falls into the i -th category (see Read and Cressie (1988)). In the case of $\nu = -1, 0$, it is define by the continuous limits:

$$2 \sum_{i=1}^k (np_i) \log\left(\frac{np_i}{N_i}\right) \quad \text{and} \quad 2 \sum_{i=1}^k N_i \log\left(\frac{N_i}{np_i}\right).$$

The CR statistic contains the user-specified parameter $\nu \in \mathbb{R}$ and encompasses several commonly used test statistics, i.e., the Neyman-modified χ^2 -statistic ($\nu = -2$), the Kullback-Leibler divergence ($\nu = -1$), the Freeman-Tukey statistic ($\nu = -1/2$) and Pearson's χ^2 -statistic ($\nu = 1$).

In the empirical likelihood framework, we consider the distribution whose support is a set of observations. This means that $k = n$ and $N_i = 1$ for all i in the above setting. Therefore, CR statistic is reduced to

$$\text{CR}_\nu(\mathbf{p}) = \frac{2}{\nu(\nu+1)} \sum_{i=1}^n \{(np_i)^{-\nu} - 1\}, \quad \nu \in \mathbb{R}. \quad (2.10)$$

The empirical likelihood statistic corresponds to the specific case of $\nu = 0$, therefore, CR statistic can be considered as an extension of the empirical likelihood. Baggerly (1998) showed the CR statistic is asymptotically chi-square distributed for i.i.d. random vectors.

For the asymptotic distribution of CR statistic under the dependent process, we first consider the same estimating function as in (2.7). Following (2.8), we define the frequency domain CR statistic:

$$\widetilde{\text{CR}}_\nu(\boldsymbol{\theta}) = \min_{\mathbf{p}} \left\{ \text{CR}_\nu(\mathbf{p}) \left[\sum_{t=1}^n p_t \mathbf{m}(\lambda_t; \boldsymbol{\theta}) = \mathbf{0}, p_t \geq 0, \sum_{t=1}^n p_t = 1 \right] \right\}. \quad (2.11)$$

Then we obtain the following theorem.

Theorem 2.3 (Ogata and Taniguchi (2009)) *Let Assumption 2.1 hold. Then $\widetilde{\text{CR}}_\nu(\boldsymbol{\theta}_0)$ has the same asymptotic distribution as in Theorem 2.2 for any $\nu \in \mathbb{R}$.*

The proof of Theorem 2.3 is given in Sect. 2.6. In what follows, we give several application examples, in which we assume that $X(t)$ is scalar, i.e., $\mathbf{X}(t) = X(t)$, for simplicity.

Example 2.4 (Prediction) The h -step prediction problem as in Hannan (1970, Chapter III Section 2) is considered. We predict $X(t)$ using a linear combination of $X(t-j)$, $j \geq h$, that is, we use $\tilde{X}(t) = \sum_{j \geq h} a(j; \boldsymbol{\theta}) X(t-j)$ as a predictor where $a(j; \boldsymbol{\theta})$ s are constants. We measure the error of the predictor by $E[|X(t) - \tilde{X}(t)|^2]$ and find the best linear predictor which minimizes this error. The spectral representations of $X(t)$ and $\tilde{X}(t)$ are

$$X(t) = \int_{-\pi}^{\pi} \exp(-it\lambda) z(d\lambda), \quad \tilde{X}(t) = \int_{-\pi}^{\pi} \exp(-it\lambda) \left\{ \sum_{j \geq h} a(j; \boldsymbol{\theta}) \exp(ij\lambda) \right\} z(d\lambda)$$

where $E[|z(d\lambda)|^2] = f(\lambda) d\lambda$, $E[z(d\lambda_1) \overline{z(d\lambda_2)}] = 0$, $\lambda_1 \neq \lambda_2$. Therefore, the prediction error is written by

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j \geq h} a(j; \boldsymbol{\theta}) \exp(ij\lambda) \right|^2 f(\lambda) d\lambda.$$

We find the minimizer of this error, $\boldsymbol{\theta}_0$, by the equation

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \left| 1 - \sum_{j \geq h} a(j; \boldsymbol{\theta}) \exp(ij\lambda) \right|^2 f(\lambda) d\lambda \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}.$$

This is exactly our problem when we take

$$\phi_j(\lambda; \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} \left| 1 - \sum_{j \geq h} a(j; \boldsymbol{\theta}) \exp(ij\lambda) \right|^2$$

in (2.6).

Example 2.5 (Interpolation) Assume that the entire time series has been observed except for the time point $t = 0$. We would like to estimate $X(0)$ by a linear combination of the observed stochastic variables, that is, $\tilde{X}(0) = \sum_{j \neq 0} a(j; \boldsymbol{\theta}) X(j)$. Similar to the prediction problem, the error of interpolation becomes

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j \neq 0} a(j; \boldsymbol{\theta}) \exp(ij\lambda) \right|^2 f(\lambda) d\lambda.$$

We find the minimizer of this error, $\boldsymbol{\theta}_0$, by the equation

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \left| 1 - \sum_{j \neq 0} a(j; \boldsymbol{\theta}) \exp(ij\lambda) \right|^2 f(\lambda) d\lambda \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}.$$

This is exactly our problem when we take

$$\phi_j(\lambda; \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} \left| 1 - \sum_{j \neq 0} a(j; \boldsymbol{\theta}) \exp(ij\lambda) \right|^2$$

in (2.6).

Example 2.6 (Smoothing) Consider smoothing the trajectory of $X(t)$ by a linear combination of adjacent observations, $\sum_{j=-N}^N \theta_j X(t+j)$. Then, similar to the

previous problem, the error of this smoothing is expressed as:

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j=-N}^N \theta_j \exp(ij\lambda) \right|^2 f(\lambda) d\lambda.$$

Set $\boldsymbol{\theta} = (\theta_{-N}, \dots, \theta_N)'$ and we find the minimizer of this error, $\boldsymbol{\theta}_0$, by the equation

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \left| 1 - \sum_{j=-N}^N \theta_j \exp(ij\lambda) \right|^2 f(\lambda) d\lambda \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}.$$

This is exactly our problem when we take

$$\phi_j(\lambda; \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} \left| 1 - \sum_{j=-N}^N \theta_j \exp(ij\lambda) \right|^2$$

in (2.6).

Another extension of empirical likelihood is GEL in Newey and Smith (2004). GEL can be also considered as an alternative of generalized methods of moments (GMM) and it is known that its asymptotic bias does not grow with the number of moment restrictions, while the bias of GMM often does. Following the i.i.d. GEL described in Newey and Smith (2004), Ogata (2012) investigated the asymptotic properties of frequency domain GEL.

To describe GEL, let $\rho(y)$ be a function of a scalar y that is concave on its domain, an open interval \mathcal{Y} containing zero. Let $\hat{\Xi}_n(\boldsymbol{\theta}) = \{\boldsymbol{\xi} : \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}) \in \mathcal{Y}, t = 1, \dots, n\}$. The estimator is the solution to a saddle point problem

$$\hat{\boldsymbol{\theta}}_{\text{GEL}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \sup_{\boldsymbol{\xi} \in \hat{\Xi}_n(\boldsymbol{\theta})} \sum_{t=1}^n \rho(\boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta})).$$

The empirical likelihood (EL) estimator of Qin and Lawless (1994), the exponential tilting (ET) estimator of Kitamura and Stutzer (1997) and the continuous updating estimator (CUE) of Hansen et al. (1996) are special cases with $\rho(y) = \log(1 - y)$, $\rho(y) = -e^y$ and $\rho(y) = -(1 + y)^2/2$, respectively. Let $\boldsymbol{\Omega} = E[\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)']$. The following assumptions and theorems are frequency domain version of those in Newey and Smith (2004), and are found in Ogata (2012).

Assumption 2.2

- (i) $\boldsymbol{\theta}_0 \in \Theta$ is the unique solution to (2.6).
- (ii) Θ is compact.
- (iii) $\mathbf{m}(\lambda_t; \boldsymbol{\theta})$ is continuous at each $\boldsymbol{\theta} \in \Theta$ with probability one.

Table 2.1 Market

Index	Market
S&P 500	NYSE
Bovespa	São Paulo stock exchange
CAC40	Bourse de Paris
AEX	Amsterdam stock exchange
ATX	Wiener Börse
HKHSI	Hong Kong exchanges and clearing
Nikkei	Tokyo stock exchange

Taken from Ogata (2012). Published with the kind permission of © Hiroaki Ogata 2012. Published under the Creative Commons Attribution License

- (iv) $E[\sup_{\theta \in \Theta} \|\mathbf{m}(\lambda_t; \theta)\|^\alpha] < \infty$ for some $\alpha > 2$.
- (v) $\mathbf{\Omega}$ is nonsingular.
- (vi) $\rho(y)$ is twice continuously differentiable in a neighborhood of zero.

Theorem 2.4 *Let Assumption 2.2 hold. Then $\hat{\boldsymbol{\theta}}_{\text{GEL}} \xrightarrow{p} \boldsymbol{\theta}_0$.*

Furthermore, let $\mathbf{G} = E[\partial \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) / \partial \boldsymbol{\theta}]$.

Assumption 2.3

- (i) $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$.
- (ii) $\mathbf{m}(\lambda_t; \boldsymbol{\theta})$ is continuously differentiable in a neighborhood \mathcal{N} of $\boldsymbol{\theta}_0$ and

$$E \left[\sup_{\boldsymbol{\theta} \in \mathcal{N}} \left\| \frac{\partial \mathbf{m}(\lambda_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right\| \right] < \infty.$$

- (iii) $\text{rank}(\mathbf{G}) = p$.

Theorem 2.5 *Let Assumptions 2.2 and 2.3 hold. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{GEL}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, (\mathbf{G}' \mathbf{\Omega}^{-1} \mathbf{G})^{-1}).$$

Now we are applying the frequency domain GEL estimation method to the portfolio selection problem introduced in Example 2.2. The sample consists of 7 weekly market indices (S&P 500, Bovespa, CAC 40, AEX, ATX, HKHSI, and Nikkei) having 800 observations each: the initial date is April 30, 1993, and the ending date is August 22, 2008. The data are from Ogata (2012). Refer to Table 2.1 for the market of each index.

The log-return of the i -th ($i = 1, \dots, 7$) index at time t is denoted by $x_i(t)$ and each process is assumed to be stationary. We estimate the best portfolio weights $\boldsymbol{\theta}_0 = (\theta_1, \dots, \theta_7)'$ described in Example 2.2 by using three types of frequency domain GEL estimators (EL, ET, and CUE). The results are shown in Table 2.2, and explain that Bovespa and ATX contribute large part in the optimal portfolio.

Table 2.2 Estimated portfolio weights

	EL	ET	CUE
S&P 500	0.0759	0.0617	0.0001
Bovespa	0.6648	0.6487	0.6827
CAC40	0.0000	0.0000	0.0000
AEX	0.0000	0.0000	0.0000
ATX	0.2593	0.2558	0.3168
HKHSI	0.0000	0.0000	0.0000
Nikkei	0.0000	0.0338	0.0004

Taken from Ogata (2012). Published with the kind permission of © Hiroaki Ogata 2012. Published under the creative commons attribution license

2.5 GEL for Multivariate Stable Distributions

It is well known that financial data often show asymmetry and heavy-tails properties which cannot be captured by simple distributions such as normal distribution. One of the families of distributions which can be accommodated to such asymmetry and heavy-tails data is the stable family. Since Mandelbrot (1963) and Fama (1965) proposed the use of stable distributions to analyze financial data, these distributions have received considerable attention; see, for example, Embrechts et al. (1997), Belkacem et al. (2000), and Rachev and Han (2000). Moreover, the stable distributions arise from a generalization of the central limit theorem in which the assumption of finite variance is relaxed, and consequently, the stable distributions are closed under summation. This property is also a strong motivation to fit stable distributions to financial data, since low frequency financial returns can be regarded as the sum of high frequency data.

In order to deal with several financial assets simultaneously, we treat multivariate stable distributions. In fact, we have multivariate stable distributions that are included in an elliptical family. Elliptical stable distributions are easier to handle than nonelliptical stable distributions but we do not consider them in this section. The reason is in Fig. 2.1, that displays density plots of bivariate normal, elliptical stable, and nonelliptical stable distributions along with contour plots of the densities. The graphs for the nonelliptical stable distribution show skewness that cannot be handled by either normal or elliptical stable distributions. Therefore, we especially consider multivariate nonelliptical stable distributions and apply the i.i.d. GEL method to estimate their parameters, following Ogata (2013).

We briefly review the multivariate stable distributions. A random vector $\mathbf{X} = (X_1, \dots, X_d)' \in \mathbb{R}^d$ is said to be stable if its characteristic function is

$$\phi(\mathbf{v}) = E[\exp\{i\langle \mathbf{X}, \mathbf{v} \rangle\}] = \exp\{-I(\mathbf{v})\},$$

where the exponent function is

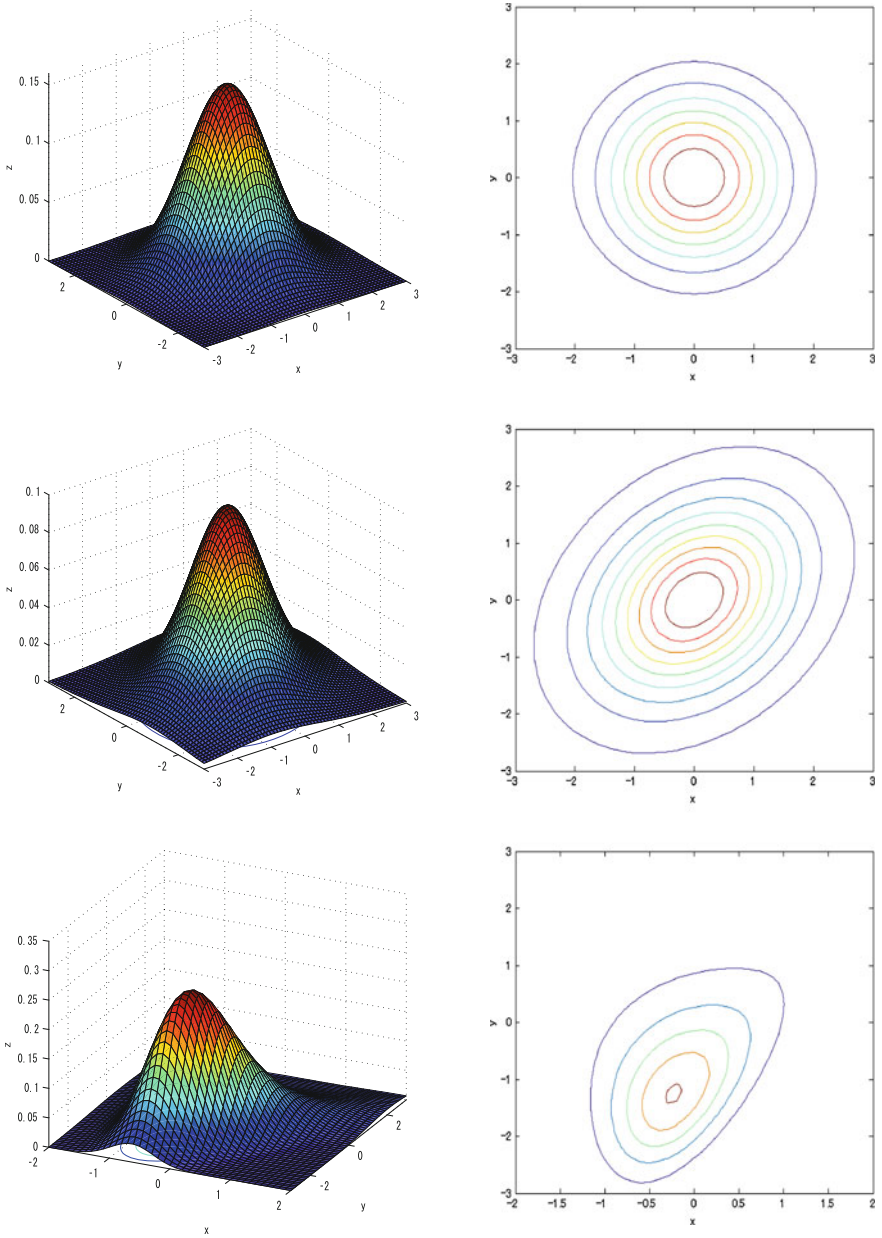


Fig. 2.1 The left three panels show bivariate density plots and right three panels show their contours. From top to bottom, the figures are for normal, elliptical stable, and nonelliptical stable distributions, respectively. Taken from Ogata (2013). Published with the kind permission of © Elsevier 2013. All Rights Reserved

$$I(\mathbf{v}) = \int_{S_d} \psi(\langle \mathbf{s}, \mathbf{v} \rangle) \Gamma(d\mathbf{s}) + i \langle \mathbf{v}, \boldsymbol{\mu} \rangle.$$

Here $S_d = \{\mathbf{s} : \|\mathbf{s}\| = 1\}$ is the unit sphere in \mathbb{R}^d , the symbol $\langle \cdot, \cdot \rangle$ denotes inner product, Γ is a finite spectral measure on S_d , $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$ is the location vector, and

$$\psi(u) = \begin{cases} |u|^\alpha \left(1 - i \operatorname{sign}(u) \tan \frac{\pi\alpha}{2}\right) & (\alpha \neq 1) \\ |u| \left(1 + i \frac{2}{\pi} \operatorname{sign}(u) \ln |u|\right) & (\alpha = 1), \end{cases}$$

where $\alpha \in (0, 2]$ is the characteristic exponent, which controls tail thickness. The d -dimensional stable distribution is denoted by $S^d(\alpha, \Gamma, \boldsymbol{\mu})$.

Our purpose is to estimate the spectral measure Γ , the location vector $\boldsymbol{\mu}$, and the characteristic exponent α , based on an i.i.d. sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ of d -dimensional random vectors drawn from this distribution. However, difficulties arise even in the univariate case, $d = 1$. Foremost is the complexity of the density function. Except for a few cases, no simple explicit form of the density exists, which is an obstacle to implementing the usual MLE method. Moreover, the stable distributions do not necessarily have second or even first moments, which impede using the ordinary method of moments.

One remedy for this difficulty is to use the empirical characteristic function

$$\hat{\phi}_n(\mathbf{v}) = \frac{1}{n} \sum_{j=1}^n \exp\{i \langle \mathbf{X}_j, \mathbf{v} \rangle\}.$$

Following Nolan et al. (2001), we first consider a discrete approximation of the spectral measure:

$$\Gamma^* = \sum_{\ell=1}^L \gamma_\ell \delta_{s_\ell}, \quad (2.12)$$

where $\gamma_\ell = \Gamma(A_\ell)$, $\ell = 1, \dots, L$, are weights and δ_{s_ℓ} is a point mass at $s_\ell \in S_d$. Here, A_ℓ , $\ell = 1, \dots, L$, are patches that partition the sphere S_d , where A_ℓ has some ‘‘center’’ s_ℓ . The estimation of the parameters is done via an estimating function based on theoretical and empirical characteristic functions. Many authors have investigated the use of empirical characteristic functions for estimation; for a summary and extensive references, see Yu (2004). In the context of multivariate stable distributions, suppose that $\{\mathbf{X}_j\}_{j=1}^n$ is an i.i.d. sequence from $S^d(\alpha, \Gamma^*, \boldsymbol{\mu})$, where Γ^* is a discrete spectral measure defined in (2.12). Denote the parameters by $\boldsymbol{\theta} = (\alpha, \boldsymbol{\gamma}', \boldsymbol{\mu}')' \in (0, 2] \times \mathbb{R}_+^L \times \mathbb{R}^d$, and define the estimating function as

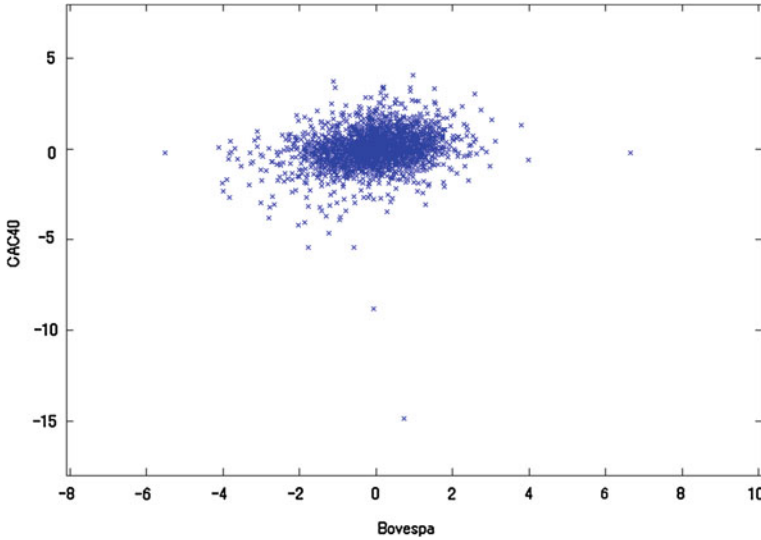


Fig. 2.2 Scatter plot of the pair of Bovespa and CAC40. Taken from Ogata (2013). Published with the kind permission of © Elsevier 2013. All Rights Reserved

$$h(\mathbf{v}, \mathbf{X}_j, \boldsymbol{\theta}) = \exp\{i\langle \mathbf{v}, \mathbf{X}_j \rangle\} - \phi_{\boldsymbol{\theta}}(\mathbf{v}),$$

where $\phi_{\boldsymbol{\theta}}$ is the theoretical characteristic function for model parameter $\boldsymbol{\theta}$. Denoting true parameter value by $\boldsymbol{\theta}_0$, we have $E[h(\mathbf{v}, \mathbf{X}_j, \boldsymbol{\theta}_0)] = 0$ for any frequency $\mathbf{v} \in \mathbb{R}^d$. After choosing some frequencies $\mathbf{v}_1, \dots, \mathbf{v}_K \in \mathbb{R}^d$, we redefine the estimating function as

$$\mathbf{g}(\mathbf{X}_j, \boldsymbol{\theta}) = (\Re\epsilon[h(\mathbf{v}_1, \mathbf{X}_j, \boldsymbol{\theta})], \dots, \Re\epsilon[h(\mathbf{v}_K, \mathbf{X}_j, \boldsymbol{\theta})], \Im\epsilon[h(\mathbf{v}_1, \mathbf{X}_j, \boldsymbol{\theta})], \dots, \Im\epsilon[h(\mathbf{v}_K, \mathbf{X}_j, \boldsymbol{\theta})])', \quad (2.13)$$

where $\Re\epsilon[\cdot]$ and $\Im\epsilon[\cdot]$ are the real and imaginary parts of a complex number. Obviously, we still have $E[\mathbf{g}(\mathbf{X}_j, \boldsymbol{\theta}_0)] = \mathbf{0}$. Using this estimating function, we estimate the parameters by the GEL method described in the previous section.

Now we are applying the GEL estimation method for multivariate stable distributions to real data. The sample consists of two log returns of daily market indexes, Bovespa (São Paulo Stock Exchange) and CAC40 (Bourse de Paris), having 2536 observations in each.¹ The initial date is January 4, 2000, and the ending date is September 22, 2009. Figure 2.2 shows a scatter plot of their returns. In this plot, several points are located far from the origin, and the data are skewed heavily downward. Thus, multivariate stable distributions are more appropriate than multivariate normal or elliptical distributions.

¹ This data is a part of the data set which was used in Dominicy and Veredas (2013).

Table 2.3 Estimated parameters and VaRs : σ_{BB} and σ_{CC} are the variances for Bovespa and CAC40, and σ_{BC} is the covariance

Model	Stable				Normal	
	α	1.757				
Estimated parameters	γ_1	0.409	γ_5	0.598		
	γ_2	0.000	γ_6	0.000	σ_{BB}	0.999
	γ_3	0.000	γ_7	0.000	σ_{BC}	0.219
	γ_4	0.226	γ_8	0.729	σ_{CC}	0.988
VaR _B (0.01)			-5.763			-2.960
VaR _C (0.01)			-3.778			-1.603

VaR_B(0.01) and VaR_C(0.01) stand for the 1 % VaRs for Bovespa and CAC40, respectively. Taken from Ogata (2013). Published with the kind permission of © Elsevier 2013. All Rights Reserved

We model the pair (Bovespa, CAC40) using the bivariate stable distribution $S^2(\alpha, \Gamma^*, \mathbf{0})$. Here, Γ^* is a discrete spectral measure defined in (2.12), and we take $L = 8, s_\ell = (\cos \theta_\ell, \sin \theta_\ell)'$ for $\theta_\ell = (\ell - 1)\pi/4, \ell = 1, \dots, 8$. For comparison, the bivariate normal distribution $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{BB} & \sigma_{BC} \\ \sigma_{BC} & \sigma_{CC} \end{pmatrix}\right)$ is also fitted to the data. We use ET estimators for the parameters of the stable distribution, $(\alpha, \gamma_1, \dots, \gamma_8)'$, and MLEs for the parameters of the normal distribution, $(\sigma_{BB}, \sigma_{BC}, \sigma_{CC})$. Based on the estimated parameters for both models, we calculate the 1 % VaRs for the Bovespa and the CAC40. Table 2.3 shows the estimation results and the VaRs.

Both VaRs from the fitted stable model are smaller than those from the fitted normal model, which means fitting the stable distribution evaluates the risk to be higher than that from fitting the normal distribution. The estimated marginal density functions for Bovespa and CAC40 are displayed in Fig. 2.3.

Last, we make a comment about the data. Perhaps, heavy tails may be caused by GARCH effects rather than by unconditional stable distributions. To remove the effect of the dynamic conditional volatility, Dominicy and Veredas (2013) adjusted the original data by using the AR(2)-GARCH(1,1) model. The analysis here is also based on this adjusted data.

2.6 Appendix

In this section, we provide the proofs of Theorems 2.2 and 2.3. For the simplicity of expression, we introduce

$$P_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n m(\lambda_t; \theta_0),$$

$$S_n = \frac{1}{n} \sum_{t=1}^n m(\lambda_t; \theta_0)m(\lambda_t; \theta_0)',$$

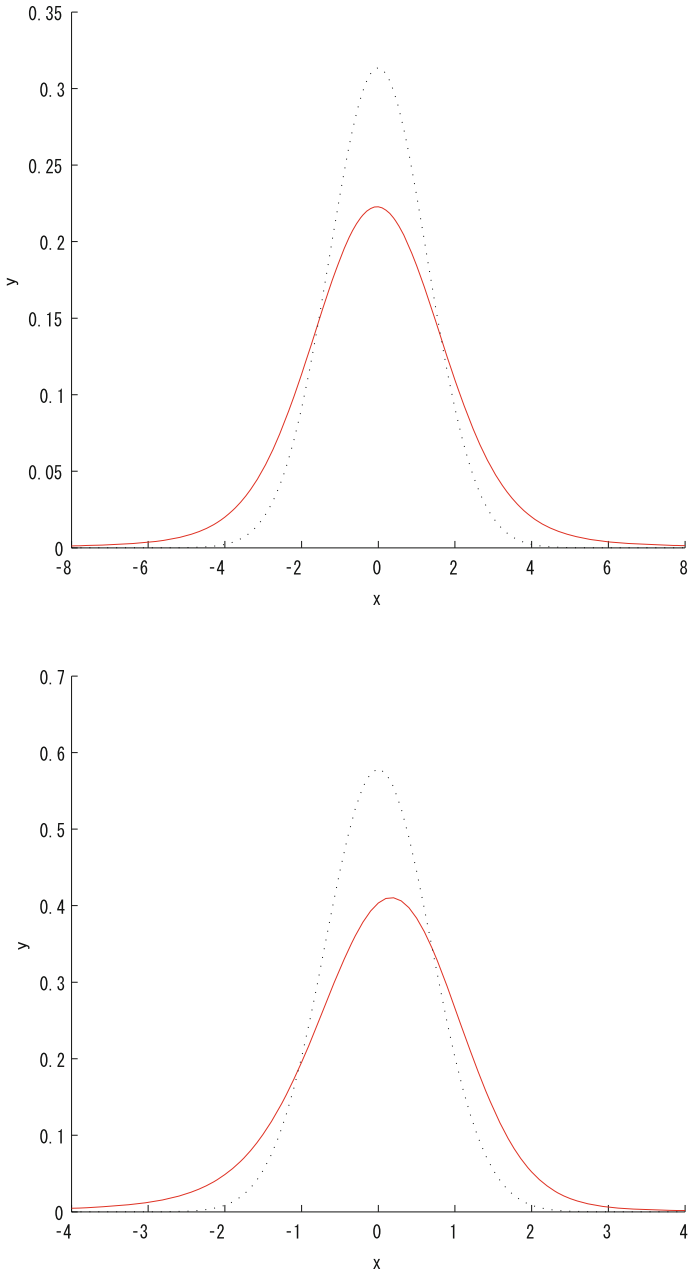


Fig. 2.3 The estimated probability density functions are displayed. The *upper figure* is for Bovespa while *lower* is for CAC40. The *real line* is for stable fitting and *dotted line* is for normal fitting. Taken from Ogata (2013). Published with the kind permission of © Elsevier 2013. All Rights Reserved

$$Z_n = \max_{1 \leq t \leq n} \|\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\|.$$

Before giving the proofs of the theorems, we prepare Lemmas 2.1, 2.2 and 2.3 whose proofs are omitted. Lemma 2.1 is proved by use of Theorem 5.10.2 and Lemma P5.1 of Brillinger (2001), and Proposition 1.2 in Chap. 1. Lemmas 2.2 and 2.3 are proved by Theorems 7.2.2 and 4.5.1 of Brillinger (2001), respectively.

Lemma 2.1 *Let Assumption 2.1 hold. Then*

$$\mathbf{P}_n \xrightarrow{d} N_q(\mathbf{0}, \boldsymbol{\Sigma}_1)$$

where $\boldsymbol{\Sigma}_1$ is the same matrix as in Theorem 2.2.

Lemma 2.2 *Let Assumption 2.1 hold. Then*

$$\mathbf{S}_n \xrightarrow{p} \boldsymbol{\Sigma}_2$$

where $\boldsymbol{\Sigma}_2$ is the same matrix as in Theorem 2.2.

Lemma 2.3 *Let Assumption 2.1 hold. Then*

$$Z_n = O(\log n).$$

2.6.1 Proof of Theorem 2.2

Following the way of leading to (2.5), we may write

$$\tilde{\mathcal{R}}(\boldsymbol{\theta}_0) = \prod_{t=1}^n \frac{1}{1 + \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)}$$

where the vector $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta}_0)$ satisfies q equations given by

$$\mathbf{J}(\boldsymbol{\xi}) \equiv \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)}{1 + \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)} = \mathbf{0}. \quad (2.14)$$

Put $Y_t = \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)$ and write $\boldsymbol{\xi} = \|\boldsymbol{\xi}\| \mathbf{u}$ where $\mathbf{u} \in U$ is a unit vector. By substituting $1/(1 + Y_t) = 1 - Y_t/(1 + Y_t)$ into $\mathbf{u}' \mathbf{J}(\boldsymbol{\xi}) = 0$, we have

$$\begin{aligned}
\mathbf{u}' \left\{ \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \left(1 - \frac{Y_t}{1 + Y_t} \right) \right\} &= 0 \\
\mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right) &= \mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \frac{\boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)}{1 + Y_t} \right) \\
\mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right) &= \|\boldsymbol{\xi}\| \mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \frac{\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)'}{1 + Y_t} \right) \mathbf{u}.
\end{aligned}$$

Then, we obtain

$$\begin{aligned}
\|\boldsymbol{\xi}\| \mathbf{u}' \mathbf{S}_n \mathbf{u} &\leq \|\boldsymbol{\xi}\| \mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \frac{\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)'}{1 + Y_t} \right) \mathbf{u} \cdot (1 + \max_t Y_t) \\
&\leq \|\boldsymbol{\xi}\| \mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^n \frac{\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)'}{1 + Y_t} \right) \mathbf{u} \cdot (1 + \|\boldsymbol{\xi}\| Z_n) \\
&= n^{-1/2} \mathbf{u}' \mathbf{P}_n (1 + \|\boldsymbol{\xi}\| Z_n)
\end{aligned}$$

and so

$$\|\boldsymbol{\xi}\| \left(\mathbf{u}' \mathbf{S}_n \mathbf{u} - n^{-1/2} Z_n \mathbf{u}' \mathbf{P}_n \right) \leq n^{-1/2} \mathbf{u}' \mathbf{P}_n.$$

From Lemmas 2.1–2.3, we have

$$\|\boldsymbol{\xi}\| \{O_p(1) - O(\log n) O_p(n^{-1/2})\} \leq O_p(n^{-1/2}),$$

and hence,

$$\|\boldsymbol{\xi}\| = O_p(n^{-1/2}). \quad (2.15)$$

After we have established an order of $\boldsymbol{\xi}$, we have from Lemma 2.3 and (2.15) that

$$\max_{1 \leq t \leq n} |Y_t| = O_p(n^{-1/2}) O(\log n). \quad (2.16)$$

Now, by (2.14),

$$\begin{aligned}
\mathbf{0} &= \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \left(1 - Y_t + \frac{Y_t^2}{1 + Y_t} \right) \\
&= n^{-1/2} \mathbf{P}_n - \mathbf{S}_n \boldsymbol{\xi} + \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \frac{Y_t^2}{1 + Y_t}
\end{aligned} \quad (2.17)$$

Noting that

$$\frac{1}{n} \sum_{t=1}^n \|\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\|^3 \leq Z_n \|\mathbf{S}_n\| = O(\log n),$$

the final term of (2.17) has a norm bounded by

$$\frac{1}{n} \sum_{t=1}^n \|\mathbf{m}(\lambda_t, \boldsymbol{\theta}_0)\|^3 \|\boldsymbol{\xi}\|^2 |1 + Y_t|^{-1} = O(\log n) O_p(n^{-1}) O_p(1) = O_p(n^{-1} \log n).$$

Hence we can write

$$\boldsymbol{\xi} = \frac{1}{\sqrt{n}} \mathbf{S}_n^{-1} \mathbf{P}_n + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} = O_p(n^{-1} \log n)$. By (2.16), we can write

$$\log(1 + Y_t) = Y_t - \frac{1}{2} Y_t^2 + \eta_t$$

where, for some finite constant $C > 0$,

$$\Pr(|\eta_t| \leq C|Y_t|^3, 1 \leq t \leq n) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Now we may write

$$\begin{aligned} -2 \log \tilde{\mathcal{R}}(\boldsymbol{\theta}_0) &= -2 \sum_{t=1}^n \log(nw_t) = -2 \sum_{t=1}^n \log(1 + Y_t) = 2 \sum_{t=1}^n Y_t - \sum_{t=1}^n Y_t^2 + 2 \sum_{t=1}^n \eta_t \\ &= \mathbf{P}'_n \mathbf{S}_n^{-1} \mathbf{P}_n - n \boldsymbol{\epsilon}' \mathbf{S}_n \boldsymbol{\epsilon} + 2 \sum_{t=1}^n \eta_t. \end{aligned}$$

Here it is seen that

$$\begin{aligned} n \boldsymbol{\epsilon}' \mathbf{S}_n \boldsymbol{\epsilon} &= n O_p(n^{-1} \log n) O_p(1) O_p(n^{-1} \log n) = O_p(n^{-1} (\log n)^2), \\ 2 \sum_{t=1}^n \eta_t &\leq C \|\boldsymbol{\xi}\|^3 \sum_{t=1}^n \|\mathbf{m}(\lambda_t, \boldsymbol{\theta}_0)\|^3 = O_p(n^{-3/2}) O(n \log n) = O_p(n^{-1/2} \log n), \end{aligned}$$

and finally, from Lemmas 2.1 and 2.2, we can show that

$$\mathbf{P}'_n \mathbf{S}_n^{-1} \mathbf{P}_n \xrightarrow{d} \left(\boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_1^{-1/2} \mathbf{P}_n \right)' \left(\boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_1^{-1/2} \mathbf{P}_n \right) \stackrel{d}{=} (\boldsymbol{\Sigma} \mathbf{N})' (\boldsymbol{\Sigma} \mathbf{N}). \quad \square$$

2.6.2 Proof of Theorem 2.3

Let $\nu \neq 0, -1$. To find the minimizing weights $\mathbf{p} = (p_1, \dots, p_n)'$ in (2.11), we proceed by the Lagrange multiplier method. Write

$$G = \text{CR}_\nu(\mathbf{p}) + \boldsymbol{\xi}' \sum_{t=1}^n p_t \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) + \gamma \left(\sum_{t=1}^n p_t - 1 \right),$$

where $\boldsymbol{\xi} \in \mathbb{R}^q$ and $\gamma \in \mathbb{R}$ are Lagrange multipliers. Since the objective function $\text{CR}_\nu(\mathbf{p})$ is convex with respect to \mathbf{p} , the solution of the first-order condition

$$(i) \quad \frac{\partial G}{\partial \mathbf{p}} = \mathbf{0} \quad (ii) \quad \frac{\partial G}{\partial \boldsymbol{\xi}} = \mathbf{0} \quad (iii) \quad \frac{\partial G}{\partial \gamma} = 0 \quad (2.18)$$

gives a global minimum point. The first condition of (2.18) gives

$$p_t = \frac{1}{n} \left\{ \frac{(\nu + 1)\gamma}{2n} \right\}^{-1/(1+\nu)} \left\{ 1 + \frac{\boldsymbol{\xi}'}{\gamma} \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right\}^{-1/(1+\nu)} \quad (t = 1, \dots, n).$$

With a simpler notation, we rewrite p_t as

$$p_t = C^* \left\{ 1 + \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right\}^{-1/(1+\nu)} \quad (t = 1, \dots, n) \quad (2.19)$$

where C^* and $\boldsymbol{\xi}$ are certain constant and vector, respectively. From the third condition of (2.18), the constant should be

$$C^* = \left[\sum_{t=1}^n \left\{ 1 + \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right\}^{-1/(1+\nu)} \right]^{-1}. \quad (2.20)$$

Put $Y_t = \boldsymbol{\xi}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)$ and write $\boldsymbol{\xi} = \|\boldsymbol{\xi}\| \mathbf{u}$ where $\mathbf{u} \in U$ is a unit vector. From the second condition of (2.18), we have

$$\frac{1}{n} \mathbf{u}' \sum_{t=1}^n p_t \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) = 0,$$

and by substituting (2.19) and (2.20) into this equation, we have

$$\begin{aligned} 0 &= \frac{1}{n} \mathbf{u}' \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) (1 + Y_t)^{-1/(1+\nu)} \\ &= \frac{1}{n} \mathbf{u}' \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \left\{ 1 - \frac{(1 + \eta_t Y_t)^{-1/(1+\nu)-1}}{1 + \nu} Y_t \right\} \quad (0 < \eta_t < 1). \end{aligned}$$

We rewrite this as

$$\begin{aligned} \|\xi\| &= \left[\mathbf{u}' \left\{ \frac{1}{n} \sum_{t=1}^n \{1 + \eta_t \|\xi\| \mathbf{u}' \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\}^{-1/(1+\nu)-1} \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)' \right\} \mathbf{u} \right]^{-1} \\ &\quad \times (1 + \nu) \mathbf{u}' \left\{ \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \right\}. \end{aligned} \quad (2.21)$$

From Lemmas 2.1–2.3, there exists a sufficiently large C and an integer n_0 such that if $n > n_0$ and $\|\xi\| < Cn^{-1/2}$, the right hand side of (2.21) is less than $Cn^{-1/2}$. Applying the Brouwer fixed point theorem to the right hand side of (2.21), we can see that a solution of Eq. (2.21) exists in the region $\|\xi\| < Cn^{-1/2}$. Hence it is shown that

$$\|\xi\| = O_p(n^{-1/2}). \quad (2.22)$$

Now, from the second condition of (2.18), we obtain

$$\begin{aligned} \mathbf{0} &= \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) (1 + Y_t)^{-1/(1+\nu)} \\ &= \frac{1}{n} \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \left\{ 1 - (1 + \nu)^{-1} Y_t + O_p(Y_t^2) \right\} \\ &= n^{-1/2} \mathbf{P}_n - (1 + \nu)^{-1} \mathbf{S}_n \xi + \boldsymbol{\beta}. \end{aligned}$$

With some constant C , the norm of the last term $\boldsymbol{\beta}$ is evaluated as

$$\|\boldsymbol{\beta}\| = \frac{C}{n} \sum_{t=1}^n \|\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\|^3 \|\xi\|^2 \leq CZ_n \mathbf{S}_n \|\xi\|^2 = O(\log n) O_p(n^{-1})$$

and we have

$$\xi = (1 + \nu) n^{-1/2} \mathbf{S}_n^{-1} \mathbf{P}_n + \boldsymbol{\beta}.$$

Now it is shown that

$$\sum_{t=1}^n Y_t = \xi' \sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) = (1 + \nu) \mathbf{P}_n' \mathbf{S}_n^{-1} \mathbf{P}_n + O(\log n) O_p(n^{-1/2}), \quad (2.23)$$

$$\begin{aligned} \sum_{t=1}^n Y_t^2 &= \xi' \left(\sum_{t=1}^n \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0) \mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)' \right) \xi \\ &= (1 + \nu)^2 \mathbf{P}_n' \mathbf{S}_n^{-1} \mathbf{P}_n + O(\log n) O_p(n^{-1/2}) \end{aligned} \quad (2.24)$$

$$\begin{aligned} \sum_{t=1}^n Y_t^3 &\leq \max_{1 \leq t \leq n} |Y_t| \sum_{t=1}^n Y_t^2 = \|\boldsymbol{\xi}\| \max_{1 \leq t \leq n} \|\mathbf{m}(\lambda_t; \boldsymbol{\theta}_0)\| \sum_{t=1}^n Y_t^2 \\ &= O(\log n) O_p(n^{-1/2}). \end{aligned} \quad (2.25)$$

From (2.19) and (2.20), we can write

$$\begin{aligned} CR_v(\mathbf{p}) &= \frac{2}{v(v+1)} \sum_{t=1}^n \{(np_t)^{-v} - 1\} \\ &= \frac{2}{v(v+1)} \sum_{t=1}^n \left[\left\{ \frac{n(1+Y_t)^{-1/(1+v)}}{\sum_{t=1}^n (1+Y_t)^{-1/(1+v)}} \right\}^{-v} - 1 \right] \\ &= \frac{2}{v(v+1)} \left[\left\{ \frac{1}{n} \sum_{t=1}^n (1+Y_t)^{-1/(1+v)} \right\}^v \left\{ \sum_{t=1}^n (1+Y_t)^{v/(1+v)} \right\} - n \right]. \end{aligned} \quad (2.26)$$

Taylor expansion leads to

$$\begin{aligned} &\left\{ \frac{1}{n} \sum_{t=1}^n (1+Y_t)^{-1/(1+v)} \right\}^v \\ &= \left\{ 1 - \frac{1}{n(1+v)} \sum_{t=1}^n Y_t + \frac{2+v}{2n(1+v)^2} \sum_{t=1}^n Y_t^2 + \frac{1}{n} \sum_{t=1}^n O_p(Y_t^3) \right\}^v \\ &= 1 - \frac{v}{n(1+v)} \sum_{t=1}^n Y_t + \frac{v(2+v)}{2n(1+v)^2} \sum_{t=1}^n Y_t^2 + \frac{1}{n} \sum_{t=1}^n O_p(Y_t^3) + O_p(n^{-2}) \end{aligned} \quad (2.27)$$

and

$$\sum_{t=1}^n (1+Y_t)^{v/(1+v)} = n + \frac{v}{1+v} \sum_{t=1}^n Y_t - \frac{v}{2(1+v)^2} \sum_{t=1}^n Y_t^2 + \sum_{t=1}^n O_p(Y_t^3). \quad (2.28)$$

Using (2.23)–(2.25), (2.27) and (2.28), we can show that

$$(A.13) = \frac{1}{(1+v)^2} \sum_{t=1}^n Y_t^2 + O(\log n) O_p(n^{-1/2}) = \mathbf{P}'_n \mathbf{S}_n^{-1} \mathbf{P}_n + O(\log n) O_p(n^{-1/2}).$$

Finally, from Lemmas 2.1 and 2.2, we can show that

$$\mathbf{P}'_n \mathbf{S}_n^{-1} \mathbf{P}_n \xrightarrow{d} \left(\boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_1^{-1/2} \mathbf{P}_n \right)' \left(\boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_1^{-1/2} \mathbf{P}_n \right) \stackrel{d}{=} (\boldsymbol{\Sigma} \mathbf{N})' (\boldsymbol{\Sigma} \mathbf{N})$$

This is the desired result.

When $\nu = 0$, Theorem 2.3 is reduced to Theorem 2.2. When $\nu = -1$, the proof is similar. \square

References

- Baggerly, K.A.: Empirical likelihood as a goodness-of-fit measure. *Biometrika* **85**, 535–547 (1998)
- Belkacem, L., Véhel, J.L., Walter, C.: Capm, risk and portfolio selection in alpha-stable markets. *Fractals* **8**, 99–115 (2000)
- Brillinger, D.R.: *Time Series: Data Analysis and Theory*, expanded edn. Holden-Day, San Francisco (2001)
- Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*, 2nd edn. Springer, Berlin (1991)
- Dominicy, Y., Veredas, D.: The method of simulated quantiles. *J. Econometrics* **172**, 235–247 (2013)
- Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling Extreme Events for Insurance and Finance*. Springer-Verlag, Berlin (1997)
- Fama, E.F.: The behavior of stock-market prices. *J. Bus.* **38**, 34–105 (1965)
- Hannan, E.J.: *Multiple Time Series*. John Wiley, New York (1970)
- Hansen, L.P., Heaton, J., Yaron, A.: Finite-sample properties of some alternative GMM estimators. *J. Bus. Econ. Statist.* **14**, 262–280 (1996)
- Kitamura, Y.: Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25**, 2084–2102 (1997)
- Kitamura, Y., Stutzer, M.: An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **65**, 861–874 (1997)
- Mandelbrot, B.: The variation of certain speculative prices. *J. Bus.* **36**, 394–419 (1963)
- Monti, A.C.: Empirical likelihood confidence regions in time series models. *Biometrika* **80**, 329–338 (1997)
- Nolan, J.P., Panorska, A.K., McCulloch, J.H.: Estimation of stable spectral measures. *Math. Comput. Model.* **34**, 1113–1122 (2001)
- Nordman, D.J., Lahiri, S.N.: A frequency domain empirical likelihood for short- and long-range dependence. *Ann. Statist.* **34**, 3019–3050 (2006)
- Newey, W.K., Smith, R.J.: Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**, 219–255 (2004)
- Ogata, H.: Optimal portfolio estimation for dependent financial returns with generalized empirical likelihood. *Adv. Decis. Sci.* **2012**, 8pages, Article ID 973173 (2012).
- Ogata, H.: Estimation for multivariate stable distributions with generalized empirical likelihood. *J. Econometrics* **172**, 248–254 (2013). doi:[10.1016/j.jeconom.2012.08.017](https://doi.org/10.1016/j.jeconom.2012.08.017)
- Ogata, H., Taniguchi, M.: Cressie-Read power-divergence statistics for non Gaussian vector stationary processes. *Scand. J. Stat.* **36**, 141–156 (2009)
- Ogata, H., Taniguchi, M.: An empirical likelihood approach for non-Gaussian vector stationary processes and its application to minimum contrast estimation. *Aust. N. Z. J. Stat.* **52**, 451–468 (2010)
- Owen, A.B.: Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249 (1988)
- Owen, A.B.: Empirical likelihood confidence regions. *Ann. Statist.* **18**, 90–120 (1990)
- Owen, A.B.: *Empirical likelihood*. Chapman & Hall/CRC, Boca Raton, London, New York, Washington, DC (2001)
- Qin, J., Lawless, J.: Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–325 (1994)
- Rachev, S.T., Han, S.: Portfolio management with stable distributions. *Math. Method. Oper. Res.* **51**, 341–352 (2000)

- Read, T.R.C., Cressie, N.A.C.: Goodness-of-fit statistics for discrete multivariate data. Springer-Verlag, New York (1988)
- Yu, J.: Empirical characteristic function estimation and its applications. *Econometric Rev.* **23**, 93–123 (2004)

Chapter 3

Various Methods for Financial Engineering

Abstract Various statistical methods have been introduced to many application fields. Such methods are often designed for standard settings, i.e., i.i.d. cases, regular model etc. However, financial data are usually dependent and have complicated features (see Chap. 1). In this chapter, we state various methods which are suitable for financial data. In Sect. 3.2, the control variate method is applied to time series models. Control variate method is the one to reduce the variance of estimators. However, this method has been developed mainly in i.i.d. cases. Because financial data are usually dependent, we extend this method to dependent case. In Sect. 3.3, we apply an instrumental variable method to a stochastic regression model. In stochastic regression models, a natural estimator for the regression coefficients is the ordinary least squares estimator (OLS). However, if the explanatory variable and the stochastic disturbance are correlated, this estimator is inconsistent. To overcome this difficulty, the instrumental variable method is used. In the CAPM model, it will be shown that the explanatory variable and the disturbance are fractionally cointegrated. Hence, we use the instrumental variable method to estimate the regression coefficients.

Keywords Control variate method · Instrumental variable method · Stochastic regression model · CAPM · Long memory process · Ordinary least squares estimator

3.1 Introduction

In this chapter, we state two important methods which are suitable for financial data.

In Sect. 3.2, the control variate method is applied to time series models. The sample mean is one of the most natural estimators of the population mean based on i.i.d. sample. However, if some control variables are available, it is known that the control variate method reduces the variance of the sample mean. This method has been discussed in the case when the sample and control variable are i.i.d.. Here we

assume that these variables are stationary processes. Then we propose an estimator of the mean of the stationary process of interest by using control variate method. It is shown that this estimator improves the sample mean in the sense of mean square error. Also this analysis is extended to the case when the mean dynamics is of the form of regression.

In Sect. 3.3, we apply an instrumental variable method to the stochastic regression model, which includes CAPM model with time dimension. In the CAPM, empirical studies suggest that the response process and the explanatory process are short memory dependent and long memory dependent, respectively. From this point of view, we have to assume that the error process is also long memory dependent and is correlated with the explanatory process. For the stochastic regression model, the most fundamental estimator is the ordinary least squares estimator. However, the dependence of the error process with the explanatory process makes this estimator to be inconsistent. To overcome this difficulty instrumental variable method was proposed. In this section, by using instrumental variable method, we propose the two-stage least squares (2SLS) estimator for the stochastic regression model in which the explanatory process and error process are long memory dependent and mutually correlated each other. Then we prove its consistency and CLT under some conditions.

3.2 Control Variate Methods for Financial Data

When we deal with financial time series data, estimation of dynamics mean is most fundamental, and is useful for the prediction of financial time series and construction of portfolio on financial assets etc.

For this purpose, the sample mean is the most natural one. However, it is often observed that the sample mean is not so good. In such a situation, if we can use some control variate information, we can introduce more effective method to estimate the dynamics mean, which is called the control variate method.

When some control variable vector is available (a random vector which is possibly correlated with the variable of interest), using the information about the control variate vector, it is known that the control variate method reduces the variance of the sample mean. That is, if μ_Y is an unknown mean of i.i.d. data $\{Y(1), Y(2), \dots, Y(n)\}$ and X is a control variable vector with known mean vector μ_X , then for any constant vector b , the mean of the control variate estimator $\hat{\mu}_Y(b) = \bar{Y}_n - b'(X - \mu_X)$ is μ_Y (\bar{Y}_n is the sample mean of $\{Y(1), Y(2), \dots, Y(n)\}$) and its variance is $Var\{\bar{Y}_n\} - 2b'Cov\{\bar{Y}_n, X\} + b'\sum_X b$, where \sum_X is the covariance matrix of X and $Cov\{\bar{Y}_n, X\}$ is the covariance vector between \bar{Y}_n and X . Hence if $2b'Cov\{\bar{Y}_n, X\} > b'\sum_X b$, then the variance of the control variate estimator is smaller than that of the sample mean.

This method has been discussed in the case when the sample and control variable are i.i.d.. Lavenberg and Welch (1981) reviews analyses of the control variate developed up to the date. In the paper, the value b_* of vector b which minimizes the variance of the control variate estimator is derived and the confidence interval of

$\hat{\mu}_Y(\mathbf{b}_*)$ is constructed. However, in practice, since the correlation between \bar{Y}_n and \mathbf{X} is unknown, this \mathbf{b}_* is not known and an estimator $\hat{\mathbf{b}}_*$ of \mathbf{b}_* is proposed. In general, the control variate estimator involving the estimator $\hat{\mathbf{b}}_*$ is not unbiased and the confidence interval cannot be constructed easily. They also discuss these problems. Rubinstein and Marcus (1985) extends the results to the case when the sample mean \bar{Y}_n is multidimensional vector and the multidimensional control variate estimator is represented as $\hat{\mu}_Y(\mathbf{B}) = \bar{Y}_n - \mathbf{B}(\mathbf{X} - \mu_X)$, where \mathbf{B} is an arbitrary matrix and \mathbf{X} is a control variate vector with mean vector μ_X . Giving the matrix \mathbf{B}_* which minimizes the determinant of $E\{\hat{\mu}_Y(\mathbf{B})'\hat{\mu}_Y(\mathbf{B})\}$, they introduce an estimator $\hat{\mu}_Y(\mathbf{B}_*)$. They also introduce an estimator of $\hat{\mathbf{B}}_*$ of \mathbf{B}_* and discuss the confidence ellipsoid. Nelson (1990) proves a central limit theorem of the control variate estimator. Since a lot of control variate theories have been discussed under a specific probability structure (usually normal distribution) for the sample and control variates, a number of authors introduced remedies for violations of these assumptions. Nelson (1990) gives a systematic analytical evaluation of them. In recent years, this method has been applied to financial engineering (e.g., Glasserman (2004), Chan and Wong (2006)).

As we saw in Chap. 1, empirical studies show that many financial data are dependent. However, the control variate theory is usually discussed under the assumption that the sample and control variates are i.i.d. Hence in this section, when the sample is generated from a stationary process and some control variable process is available, we propose an estimator $\hat{\theta}_C$ of the mean of the concerned process by using control variate method. Then it is shown that this estimator improves the sample mean in the sense of mean square error (MSE). The estimator $\hat{\theta}_C$ is expressed in terms of nonparametric estimators for spectra of the concerned process and the control variate process. We also apply this analysis to the case when the mean dynamics is of the form of regression. A control variate estimator for the regression coefficients is proposed and is shown to improve the LSE in the sense of MSE. Numerical studies show how our estimators behave. Our results have applications to various fields, including finance in particular.

Suppose that $\{Y(t); t \in \mathbb{Z}\}$ is a scalar-valued process with mean $E\{Y(t)\} = \theta$ and $\{\mathbf{X}(t); t \in \mathbb{Z}\}$ is an another m -dimensional process with the mean vector $E\{\mathbf{X}(t)\} = \mathbf{0}$, which is possibly correlated with $\{Y(t)\}$. We are now interested in estimation of θ . Let $\mathbf{Z}(t) \equiv (Y(t), \mathbf{X}'(t))'$ and we assume $\{\mathbf{Z}(t); t \in \mathbb{Z}\}$ is generated by the following linear process.

$$\mathbf{Z}(t) = \sum_{j=0}^{\infty} \mathbf{B}(j)\epsilon(t-j) + \boldsymbol{\theta} \quad (3.1)$$

where $\boldsymbol{\theta} = (\theta, 0, 0, \dots, 0)'$ is $m+1$ -dimensional vector and $\mathbf{B}(j)'$ s are $(m+1) \times (m+1)$ matrices, and $\{\epsilon(t)\}$ is a sequence of i.i.d. $(m+1)$ -dimensional random vectors with mean vector $\mathbf{0}$ and covariance matrix \mathbf{K} .

Henceforth $|U|$ denotes the sum of all the absolute values of elements of matrix U .

- Assumption 3.1** (i) $\det\{\sum_{u=0}^{\infty} \mathbf{B}(u)z^u\} = 0$ has no roots in the unit disc $\{z \in \mathbb{C}; |z| \leq 1\}$.
(ii) The coefficient matrices $\mathbf{B}(u)$ satisfy

$$\sum_{u=0}^{\infty} |u|^4 |\mathbf{B}(u)| < \infty. \quad (3.2)$$

Using the same k -th order cumulant notation as in Sect. 1.1, we set the following.

Assumption 3.2 For $k = 3, 4, \dots$,

$$C_k^\epsilon \equiv \sup_{a_1, \dots, a_k} |c_{a_1, \dots, a_k}^\epsilon(0, \dots, 0)| < \infty \quad (3.3)$$

and

$$\sum_{L=1}^{\infty} \left(\sum_{\nu} C_{n_1}^\epsilon \cdots C_{n_p}^\epsilon \right) z^L / L! < \infty, \quad (3.4)$$

for z in a neighborhood of 0, where the inner summation is over all indecomposable partitions (see p. 20 of Brillinger (2001)) $\nu = (\nu_1, \dots, \nu_p)$ of the table

$$\begin{array}{cc} 1 & 2 \\ 3 & 4 \\ \vdots & \vdots \\ 2L-1 & 2L \end{array} \quad (3.5)$$

with ν_p having $n_p > 1$ elements, $p = 1, \dots, P$.

Write

$$C_k \equiv \sup_{a_1, \dots, a_k} \sum_{t_1, \dots, t_{k-1} = -\infty}^{\infty} |c_{a_1, \dots, a_k}^z(t_1, \dots, t_{k-1})|. \quad (3.6)$$

Then Assumptions 3.1 and 3.2 imply Assumption **(B)** in Chap. 1 and

$$\sum_{L=1}^{\infty} \left(\sum_{\nu} C_{n_1} \cdots C_{n_p} \right) z^L / L! < \infty, \quad (3.7)$$

in a neighborhood of 0, where the summation \sum_{ν} is defined as in (3.4) (see p. 48 of Brillinger (2001)). From Assumption 3.1, it is seen that the process $\{\mathbf{Z}(t)\}$ becomes a stationary process with nonsingular spectral density matrix (e.g., Brillinger (2001)). We write the spectral density matrix by

$$\mathbf{f}(\lambda) = \begin{bmatrix} f_{YY}(\lambda) & \mathbf{f}_{YX}(\lambda) \\ \mathbf{f}_{XY}(\lambda) & f_{XX}(\lambda) \end{bmatrix}. \quad (3.8)$$

From Assumption 3.1, it follows that $\mathbf{R}(s) = \{c_{ij}^z(s); i, j = 1, \dots, m+1\}$ satisfies

$$\sum_{s=-\infty}^{\infty} |s|^4 |\mathbf{R}(s)| < \infty, \quad (3.9)$$

(e.g., p. 46 of Brillinger (2001)). Suppose that partial observations $\{Y(0), Y(1), \dots, Y(n-1)\}$ and $\{\mathbf{X}(-M_n), \mathbf{X}(-M_n+1), \dots, \mathbf{X}(0), \dots, \mathbf{X}(n-1)\}$ are available, where $M_n = O(n^\beta)$ ($\frac{1}{4} \leq \beta < \frac{1}{3}$).

Now we are interested in the estimation of θ . Based on the observations, we introduce the following estimator $\hat{\theta}_C$ of θ

$$\hat{\theta}_C \equiv \frac{1}{n} \sum_{t=0}^{n-1} \left\{ Y(t) - \sum_{u=0}^{M_n} \hat{\mathbf{a}}'_n(u) \mathbf{X}(t-u) \right\}, \quad (3.10)$$

where $\hat{\mathbf{a}}_n(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\mathbf{A}}_n(\lambda) \exp(iu\lambda) d\lambda$, $\hat{\mathbf{A}}_n(\lambda) = \hat{\mathbf{f}}_{XX}(\lambda)^{-1} \hat{\mathbf{f}}_{XY}(\lambda)$. Here $\hat{\mathbf{f}}_{XX}(\lambda)$ and $\hat{\mathbf{f}}_{XY}(\lambda)$ are, respectively, nonparametric estimators of $\mathbf{f}_{XX}(\lambda)$ and $\mathbf{f}_{XY}(\lambda)$ which are defined as,

$$\hat{\mathbf{f}}_{XY}(\lambda) \equiv \frac{2\pi}{n} \sum_{s=1}^{n-1} W_n \left(\lambda - \frac{2\pi s}{n} \right) \mathbf{I}_{XY} \left(\frac{2\pi s}{n} \right) \quad (3.11)$$

$$\hat{\mathbf{f}}_{XX}(\lambda) \equiv \frac{2\pi}{n} \sum_{s=1}^{n-1} W_n \left(\lambda - \frac{2\pi s}{n} \right) \mathbf{I}_{XX} \left(\frac{2\pi s}{n} \right) \quad (3.12)$$

where $\mathbf{I}_{XY}(\mu)$ and $\mathbf{I}_{XX}(\mu)$ are submatrices of the periodogram of $\mathbf{Z}(t)$

$$\mathbf{I}_n(\mu) = \begin{bmatrix} I_{YY}(\mu) & \mathbf{I}_{YX}(\mu) \\ \mathbf{I}_{XY}(\mu) & \mathbf{I}_{XX}(\mu) \end{bmatrix}, \quad (3.13)$$

and $\{W_n(\lambda)\}$ are weight functions which are described below. The $\hat{\mathbf{A}}_n(\lambda)$ and $\hat{\mathbf{a}}_n(u)$ are shown to be consistent estimators of $\mathbf{A}(\lambda) = \mathbf{f}_{XX}(\lambda)^{-1} \mathbf{f}_{XY}(\lambda)$, $\mathbf{a}(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{A}(\lambda) \exp(iu\lambda) d\lambda$, respectively.

Next, we will show that the proposed estimator $\hat{\theta}_C$ improves the sample mean in the sense of MSE.

Then we get the following theorem. For the proof, see Amano and Taniguchi (2011).

Theorem 3.1 *Suppose Assumptions 3.1 and 3.2, and that the spectral window $W_n(\lambda)$ satisfies Assumption 1.2. Then it holds that*

$$\lim_{n \rightarrow \infty} nE|\hat{\theta}_C - \theta|^2 = 2\pi \left(f_{YY}(0) - f_{YX}(0)f_{XX}(0)^{-1}f_{XY}(0) \right). \quad (3.14)$$

It is known that the asymptotic variance of the sample mean $\bar{Y}_n \equiv \frac{1}{n} \sum_{t=0}^{n-1} Y(t)$ is $2\pi f_{YY}(0)$ (e.g., Theorem 5.2.1 of Brillinger (2001)). Since

$$2\pi \left(f_{YY}(0) - f_{YX}(0)f_{XX}(0)^{-1}f_{XY}(0) \right) \leq 2\pi f_{YY}(0), \quad (3.15)$$

we observe that $\hat{\theta}_C$ improves \bar{Y}_n in the sense of MSE.

Next, we assume $\{Y(t); t \in \mathbb{Z}\}$ is a trend model whose mean $E[Y(t)] = \mu(t) = \boldsymbol{\phi}'(t)\boldsymbol{\theta}$ is a time-dependent function. Here $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_J(t))'$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J)'$. Let $\{\mathbf{X}(t); t \in \mathbb{Z}\}$ be another m -dimensional process with mean vector $E\{\mathbf{X}(t)\} = \mathbf{0}$, which is possibly correlated with $\{Y(t)\}$. Now we apply the control variate method to estimate the parameter $\boldsymbol{\theta}$. Let $\mathbf{Z}(t) \equiv (Y(t), \mathbf{X}'(t))'$, $t \in \mathbb{Z}$. We impose the following assumption.

Assumption 3.3 $\{\mathbf{Z}(t); t \in \mathbb{Z}\}$ is generated by the following linear process.

$$\mathbf{Z}(t) = \sum_{j=0}^{\infty} \mathbf{B}(j)\boldsymbol{\epsilon}(t-j) + \begin{pmatrix} \mu(t) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.16)$$

where $\mathbf{B}(j)$'s are $(m+1) \times (m+1)$ matrices satisfying Assumption 3.1 and $\boldsymbol{\epsilon}(t)$'s are i.i.d. random vectors with mean vector $\mathbf{0}$ and covariance matrix \mathbf{K} .

For convenience, we define $\eta(t)$ by $\sum_{j=0}^{\infty} \mathbf{B}(j)\boldsymbol{\epsilon}(t-j) = (\eta(t), \mathbf{X}'(t))'$, then as discussed before, $(\eta(t), \mathbf{X}'(t))'$ has the spectral density matrix,

$$\mathbf{f}(\lambda) = \begin{bmatrix} f_{\eta\eta}(\lambda) & f_{\eta\mathbf{X}}(\lambda) \\ f_{\mathbf{X}\eta}(\lambda) & f_{\mathbf{X}\mathbf{X}}(\lambda) \end{bmatrix}. \quad (3.17)$$

Suppose that partial observations $\{Y(0), Y(1), \dots, Y(n-1)\}$ and $\{\mathbf{X}(-M_n), \mathbf{X}(-M_n+1), \dots, \mathbf{X}(0), \dots, \mathbf{X}(n-1)\}$ are available.

We define nonparametric estimators $\hat{f}_{XX}(\lambda)$ and $\hat{f}_{X\hat{\eta}}(\lambda)$ for the spectral densities $f_{XX}(\lambda)$ and $f_{X\eta}(\lambda)$, respectively, as

$$\hat{f}_{XX}(\lambda) \equiv \frac{2\pi}{n} \sum_{s=1}^{n-1} W_n\left(\lambda - \frac{2\pi s}{n}\right) \mathbf{I}_{XX}\left(\frac{2\pi s}{n}\right) \quad (3.18)$$

$$\hat{f}_{X\hat{\eta}}(\lambda) \equiv \frac{2\pi}{n} \sum_{s=1}^{n-1} W_n\left(\lambda - \frac{2\pi s}{n}\right) \mathbf{I}_{X\hat{\eta}}\left(\frac{2\pi s}{n}\right) \quad (3.19)$$

where

$$\mathbf{I}_{XX}(\mu) \equiv \frac{1}{2\pi n} \left\{ \sum_{t=0}^{n-1} \mathbf{X}(t)e^{it\mu} \right\} \left\{ \sum_{t=0}^{n-1} \mathbf{X}(t)e^{it\mu} \right\}^* \quad (3.20)$$

$$\mathbf{I}_{X\hat{\eta}}(\mu) \equiv \frac{1}{2\pi n} \left\{ \sum_{t=0}^{n-1} \mathbf{X}(t)e^{it\mu} \right\} \left\{ \sum_{t=0}^{n-1} \hat{\eta}(t)e^{it\mu} \right\}^* \quad (3.21)$$

where $\hat{\eta}(t) = Y(t) - \boldsymbol{\phi}'(t)\hat{\boldsymbol{\theta}}_{LSE}$, $\hat{\boldsymbol{\theta}}_{LSE} = (\boldsymbol{\phi}'\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'\mathbf{Y}$ (the least squares estimator of $\boldsymbol{\theta}$), $\mathbf{Y} = (Y(1), \dots, Y(n))'$ and $\boldsymbol{\phi} = (\boldsymbol{\phi}(1), \dots, \boldsymbol{\phi}(n))'$. Let $\hat{A}(\lambda) = \hat{f}_{XX}(\lambda)^{-1}\hat{f}_{X\hat{\eta}}(\lambda)$ and $\hat{a}(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{A}(\lambda) \exp(iu\lambda) d\lambda$.

Now we propose an estimator $\hat{\boldsymbol{\theta}}_{LSE}^C$ of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_{LSE}^C = (\boldsymbol{\phi}'\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'(\mathbf{Y} - \hat{\mathbf{W}}_M) \quad (3.22)$$

where $\hat{\mathbf{W}}_M = (\hat{W}_M(1), \dots, \hat{W}_M(n))'$ with

$$\hat{W}_M(t) = \sum_{u=0}^{M_n} \hat{a}'(u)\mathbf{X}(t-u). \quad (3.23)$$

To describe the asymptotics of $\hat{\boldsymbol{\theta}}_{LSE}^C$, we impose the following Grenander's conditions.

Assumption 3.4 Let $c_{j,k}^n(h) = \sum_{t=1}^{n-h} \phi_j(t+h)\phi_k(t) = \sum_{t=1-h}^n \phi_j(t+h)\phi_k(t)$. $c_{j,k}^n(h)$'s satisfy

- (i) $c_{j,j}^n(0) = O(n^\gamma)$, $j = 1, \dots, J$ for some $\gamma > 0$.
- (ii) $\lim_{n \rightarrow \infty} \frac{\phi_j^2(n+1)}{c_{j,j}^n(0)} = 0$, $j = 1, \dots, J$.
- (iii)

$$\lim_{n \rightarrow \infty} \frac{c_{j,k}^n(h)}{\left\{ c_{j,j}^n(0)c_{k,k}^n(0) \right\}^{\frac{1}{2}}} = m_{jk}(h) \quad (3.24)$$

We may take $\phi_1(t) = 1$ (constant), which evidently satisfies Assumption 3.4, hence, the regression part $\boldsymbol{\phi}(t)$ of $\{Y(t)\}$ may include a constant.

We define the $J \times J$ matrix $\mathbf{m}_{\boldsymbol{\phi}\boldsymbol{\phi}}(u)$ by

$$\mathbf{m}_{\boldsymbol{\phi}\boldsymbol{\phi}}(u) = \{m_{jk}(u); j, k = 1, \dots, J\}. \quad (3.25)$$

From p. 175 of Brillinger (2001), there exists an $r \times r$ matrix-valued function $\mathbf{G}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\lambda)$, $-\pi < \lambda \leq \pi$, whose entries are of bounded variation, such that

$$\mathbf{m}_{\phi\phi}(u) = \int_{-\pi}^{\pi} \exp(iu\lambda) d\mathbf{G}_{\phi\phi}(\lambda) \quad (3.26)$$

for $u = 0, \pm 1, \dots$. Under these assumptions, we obtain the following theorem. For the proof, see Amano and Taniguchi (2011).

Theorem 3.2 *Suppose Assumptions 3.2–3.4, then*

$$\begin{aligned} & \lim_{n \rightarrow \infty} n^\gamma E\{(\hat{\boldsymbol{\theta}}_{LSE}^C - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{LSE}^C - \boldsymbol{\theta})'\} \\ & = 2\pi \mathbf{m}_{\phi\phi}(0)^{-1} \int_{-\pi}^{\pi} f_{\eta-V, \eta-V}(\lambda) d\mathbf{G}_{\phi\phi}(\lambda) \mathbf{m}_{\phi\phi}(0)^{-1}, \end{aligned} \quad (3.27)$$

where $f_{\eta-V, \eta-V}(\lambda) = f_{\eta, \eta}(\lambda) - \mathbf{f}_{\eta X}(\lambda) \mathbf{f}_{XX}(\lambda)^{-1} \mathbf{f}_{X\eta}(\lambda)$ is the spectral density of $\eta(t) - V(t)$. Here $V(t) = \sum_{u=0}^{\infty} \mathbf{a}'(u) \mathbf{X}(t-u)$, $\mathbf{a}(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{A}(\lambda) \exp(iu\lambda) d\lambda$, $\mathbf{A}(\lambda) = \mathbf{f}_{XX}(\lambda)^{-1} \mathbf{f}_{X\eta}(\lambda)$.

Note that the least squares estimator $\hat{\boldsymbol{\theta}}_{LSE}$ of $\boldsymbol{\theta}$ has the following asymptotic variance

$$\begin{aligned} & \lim_{n \rightarrow \infty} n^\gamma E\{(\hat{\boldsymbol{\theta}}_{LSE} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{LSE} - \boldsymbol{\theta})'\} \\ & = 2\pi \mathbf{m}_{\phi\phi}(0)^{-1} \int_{-\pi}^{\pi} f_{\eta, \eta}(\lambda) d\mathbf{G}_{\phi\phi}(\lambda) \mathbf{m}_{\phi\phi}(0)^{-1}, \end{aligned} \quad (3.28)$$

where $f_{\eta, \eta}(\lambda)$ is the spectral density of $\eta(t)$. It is seen that

$$f_{\eta-V, \eta-V}(\lambda) \equiv f_{\eta, \eta}(\lambda) - \mathbf{f}_{\eta X}(\lambda) \mathbf{f}_{XX}(\lambda)^{-1} \mathbf{f}_{X\eta}(\lambda) \leq f_{\eta, \eta}(\lambda), \quad (3.29)$$

which implies that the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{LSE}^C$ is smaller than that of $\hat{\boldsymbol{\theta}}_{LSE}$.

Next we examine the control variate estimators numerically. In Example 3.1, we compare the control variate estimators with sample means. Next the control variate estimators are compared with the least squares estimators in Example 3.2. In Example 3.3, we investigate its usefulness in real financial data.

Example 3.1 We consider the following model,

$$Y(t) = u(t) + v(t) \quad (3.30)$$

$$X(t) = a_1 u(t) + 0.4u(t-1) + a_2 v(t), \quad (3.31)$$

where $\{u(t)\}, \{v(t)\} \sim \text{i.i.d. } N(0,1)$ and they are mutually independent. The length of $Y(t)$ and $X(t)$ are set by 1,000 and based on 5,000 times simulation we report MSE of $\hat{\theta}_C$ with $M_n = 20$ and \bar{Y}_n . We set $a_1 = 0.1, 0.2, 0.3$ and $a_2 = 0.1, 0.2, 0.3, 4.0$ in Table 3.1.

Table 3.1 MSE of $\hat{\theta}_C$ and \bar{Y}_n

a_1	a_2	MSE $\hat{\theta}_C$	MSE \bar{Y}_n	MSE $\bar{Y}_n - \text{MSE } \hat{\theta}_C$
0.1	0.1	0.00062385	0.00205319	0.00142934
0.1	0.2	0.00030573	0.00196825	0.00166252
0.1	0.3	0.00012184	0.00201872	0.00189688
0.2	0.1	0.00068064	0.00193495	0.00125431
0.2	0.2	0.00042055	0.00206655	0.001646
0.2	0.3	0.0002056	0.00192683	0.00172123
0.3	0.1	0.00071185	0.0020196	0.00130775
0.3	0.2	0.00047111	0.00195619	0.00148508
0.3	0.3	0.00027794	0.00198016	0.00170222
0.3	4.0	0.00068665	0.00199143	0.00130478

Taken from Amano and Taniguchi (2011). Published with the kind permission of © Elsevier B.V. 2011. All Rights Reserved

Table 3.2 MSE of $\hat{\theta}_{LSE}^C$ and $\hat{\theta}_{LSE}$ for $\phi(t) = (1, t)'$

a_2	MSE $\hat{\theta}_{LSE}^C$	MSE $\hat{\theta}_{LSE}$	MSE $\hat{\theta}_{LSE} - \text{MSE } \hat{\theta}_{LSE}^C$
0.1	0.00786141	0.01047323	0.00261182
0.2	0.0041377	0.00543288	0.00129518
0.3	0.00493462	0.00533173	0.00039711

Taken from Amano and Taniguchi (2011). Published with the kind permission of © Elsevier B.V. 2011. All Rights Reserved

From Table 3.1, we can see $\text{MSE } \bar{Y}_n - \text{MSE } \hat{\theta}_C$ becomes larger as a_2 becomes large, which implies, if control variates are highly correlated with the disturbance, then $\hat{\theta}_C$ is better than \bar{Y}_n . However, excessive influence of the disturbance makes the performance of $\hat{\theta}_C$ worse.

Example 3.2 Consider the following model,

$$Y(t) = \mu(t) + u(t) + v(t) \tag{3.32}$$

$$X(t) = 0.3u(t) + 0.4u(t - 1) + a_2v(t), \tag{3.33}$$

where $\{u(t)\}, \{v(t)\} \sim \text{i.i.d. } N(0, 1)$ and they are mutually independent. Here $\mu(t) = \theta' \phi(t)$, $\phi(t)$ is a regression function and θ is a vector-valued parameter. The length of $Y(t)$ and $X(t)$ are set by 1,000 and based on 5,000 replications we report MSE of $\hat{\theta}_{LSE}^C$ with $M_n = 20$ and $\hat{\theta}_{LSE}$, that is MSE of $\hat{\theta}_{LSE}^C$ and $\hat{\theta}_{LSE}$ are $\frac{1}{5,000} \sum_{i=1}^{5,000} \|\hat{\theta}_{LSE}^C(i) - \theta\|^2$ and $\frac{1}{5,000} \sum_{i=1}^{5,000} \|\hat{\theta}_{LSE}(i) - \theta\|^2$ ($\hat{\theta}_{LSE}^C(i)$ is i th control variate estimator and $\hat{\theta}_{LSE}(i)$ is i th least squares estimator). We set $\theta = (1, 1)'$ and $a_2 = 0.1, 0.2, 0.3$. Table 3.2 shows MSE of $\hat{\theta}_{LSE}^C$ and $\hat{\theta}_{LSE}$ for $\phi(t) = (1, t)'$, and Table 3.3 shows those for $\phi(t) = (1, \cos(\frac{\pi}{4}t))'$.

From Tables 3.2 and 3.3, we observe that $\text{MSE } \hat{\theta}_{LSE}^C$ are smaller than $\text{MSE } \hat{\theta}_{LSE}$. That is, control variate estimator also improves the least squares estimator.

Table 3.3 MSE of $\hat{\theta}_{LSE}^C$ and $\hat{\theta}_{LSE}$ for $\phi(t) = (1, \cos(\frac{\pi}{4}t))'$

a_2	MSE $\hat{\theta}_{LSE}^C$	MSE $\hat{\theta}_{LSE}$	MSE $\hat{\theta}_{LSE} - \text{MSE } \hat{\theta}_{LSE}^C$
0.1	0.00594558	0.00837971	0.00243413
0.2	0.00747352	0.00749953	0.00002601
0.3	0.00913403	0.00940378	0.00026975

Taken from Amano and Taniguchi (2011). Published with the kind permission of © Elsevier B.V. 2011. All Rights Reserved

Table 3.4 Prediction of the stock price $S(N + H)$ by $\hat{\theta}_{H,C}$ and $\bar{Y}_{H,n}$

H	$\hat{S}^C(N + H)$	$\hat{S}(N + H)$	$S(N + H)$
1	891.0968	891.4768	891
2	888.3209	888.7471	876

Modified from Amano and Taniguchi (2011). Published with the kind permission of © Elsevier B.V. 2011. All Rights Reserved

Example 3.3 We calculate the control variate estimator $\hat{\theta}_{H,C}$ and the sample mean $\bar{Y}_{H,n}$ of $\{Y_H(t)\}$ where $Y_H(t) = \log S(t) - \log S(t - H)$ and $\{S(t)\}$ are NIPPON OIL CORPORATION’s stock prices (7/20/2007 ~ 12/12/2007). We set the difference between Yen–Euro’s exchange rate and its sample mean (7/4/2007 ~ 12/11/2007) as the control variate process. Then by use of $\hat{\theta}_{H,C}$ and $\bar{Y}_{H,n}$, we forecast NIPPON OIL CORPORATION’s stock $S(N + H)$ at $N = 12/12/2007$ by $\hat{S}^C(N + H) \equiv e^{\hat{\theta}_{H,C} + \log S(N)}$ and $\hat{S}(N + H) \equiv e^{\bar{Y}_{H,n} + \log S(N)}$ in Table 3.4.

From Table 3.4, the prediction values $\hat{S}^C(N + H)$ are nearer to the true values $S(N + H)$ than $\hat{S}(N + H)$, which implies the prediction by the control variate estimator is better than that by the sample mean.

There are many cases in finance where we should estimate the statistical models for data of interest under the circumstance that we can use some related variables. In such situations, the estimators $\hat{\theta}_C$ and $\hat{\theta}_{LSE}^C$ can be used, and are more efficient than the usual estimators.

3.3 Statistical Estimation for Stochastic Regression Models with Long Memory Dependence

The CAPM is one of the typical models of risk asset’s price on equilibrium market and has been used for pricing individual stocks and portfolios. At first, Markowitz (1991) worked the groundwork of this model. In his research, he cast the investor’s portfolio selection problem in terms of expected return and variance. Sharpe (1964) and Lintner (1965) developed Markowitz’s idea for economical implication. Black (1972) derived a more general version of the CAPM. In their version, the CAPM is constructed based on the excess of the return of the asset over zero-beta return $E\{R_i\} = E\{R_{0m}\} + \beta_{im}(E\{R_m\} - E\{R_{0m}\})$, where R_i and R_m are the return of the i th asset and the market portfolio, and R_{0m} is the return of zero-beta portfolio of the

market portfolio. Campbell et al. (1997) discussed the estimation of CAPM, but in their work they did not discuss the time dimension. However, in the econometric analysis, CAPM with the time dimension should be investigated, that is, the model should be represented as $R_i(t) = \alpha_{im} + \beta_{im}R_m(t) + \epsilon_i(t)$. These models are included in the stochastic regression model, $R_i(t) = \alpha_i + \beta_i'Z(t) + \epsilon_i(t)$, where $Z(t)$ is the risk factors. In the CAPM with the time dimension, the risk is only the return of the market portfolio. However, there are some evidences of other common risk factors besides the return of the market portfolio and stochastic regression models allow more factors than the return of the market portfolio as the risk. Hence here, we investigate the stochastic regression models. Recently from the empirical analysis, it is known that the return of asset follows a short memory process and if we assume $Z(t)$ follows short memory process, such a stochastic regression model has been investigated enough. Besides, if we consider the volatility of the return of the market portfolio as the risk factor, $Z(t)$ follows long memory dependent. From this point of view, we observe that $Z(t)$ and the error process $\epsilon(t)$ are long memory dependent and correlated with each other.

For the stochastic regression model, the most fundamental estimator is the ordinary least squares estimator. However, the dependence of the error process with the explanatory process makes this estimator to be inconsistent. To overcome this difficulty, the instrumental variable method is proposed by use of the instrumental variables which are uncorrelated with the error process and correlated with the explanatory process. This method was first used by Wright (1928) and many researchers developed this method (see, Reiersøl (1945) and Geary (1949) etc.). Comprehensive reviews are seen in White (2001). However, the instrumental variable method has been discussed in the case where the error process does not follow the long memory process and this makes the estimation difficult.

For the analysis of long memory process, Robinson and Hidalgo (1997) considered a stochastic regression model defined by $y(t) = \alpha + \beta'x(t) + u(t)$, where α , $\beta = (\beta_1, \dots, \beta_K)'$ are unknown parameters and the K -vector processes $\{x(t)\}$ and $\{u(t)\}$ are long memory dependent with $E\{x(t)\} = \mathbf{0}$, $E\{u(t)\} = 0$. Furthermore, in Choy and Taniguchi (2001), they consider the stochastic regression model $y(t) = \beta x(t) + u(t)$, where $\{x(t)\}$ and $\{u(t)\}$ are stationary process with $E\{x(t)\} = \mu \neq 0$, and Choy and Taniguchi (2001) introduced a ratio estimator, the least squares estimator, and the best linear unbiased estimator for β . However, Robinson and Hidalgo (1997) and Choy and Taniguchi (2001) assume that the explanatory process $\{x(t)\}$ and the error process $\{u(t)\}$ are independent.

In this section, by using instrumental variable method we propose the two-stage least squares (2SLS) estimator for the stochastic regression models in which stochastic explanatory process and error process are long memory dependent and mutually correlated each other. Then we prove its consistency and CLT under some conditions. Also, some numerical studies are provided.

For Sharpe and Lintner version of CAPM (see Sharpe (1964) and Lintner (1965)), the expected return of asset i is given by

$$E\{R_i\} = R_f + \beta_{im} (E\{R_m - R_f\}), \quad (3.34)$$

where

$$\beta_{im} = \frac{\text{Cov}\{R_i, R_m\}}{\text{Var}\{R_m\}},$$

R_m is the return of the market portfolio and R_f is the return of the risk-free asset. Another Sharpe–Lintner CAPM (see Sharpe (1964) and Lintner (1965)) is defined for $Z_i \equiv R_i - R_f$,

$$E\{Z_i\} = \beta_{im}E\{Z_m\}$$

where

$$\beta_{im} = \frac{\text{Cov}\{Z_i, Z_m\}}{\text{Var}\{Z_m\}}$$

and $Z_m = R_m - R_f$.

Black (1972) derived a more general version of CAPM, which is written as

$$E\{R_i\} = \alpha_{im} + \beta_{im}E\{R_m\}, \quad (3.35)$$

where $\alpha_{im} = E\{R_{0m}\}(1 - \beta_{im})$ and R_{0m} is the return on the *zero-beta portfolio*.

Since CAPM is single-period model, (3.34) and (3.35) do not have a time dimension. However for econometric analysis of the model, it is necessary to add assumptions concerning the time dimension. Hence it is natural to consider the model with the time dimension

$$R_i(t) = \alpha_{im} + \beta_{im}R_m(t) + \epsilon_i(t)$$

where i denotes the asset, t denotes the period, and $R_i(t)$ and $R_m(t)$, $i = 1, \dots, q$; $t = 1, \dots, n$ are, respectively, considered as the returns of the asset i and the market portfolio at t .

Indeed, there are some evidences of other common risk factors besides the return of market portfolio. Stochastic regression models generalize CAPM with the time dimension by allowing more factors than simply the return of market portfolio $R_m(t)$

$$R_i(t) = \alpha_i + \beta'_i \mathbf{Z}(t) + \epsilon_i(t)$$

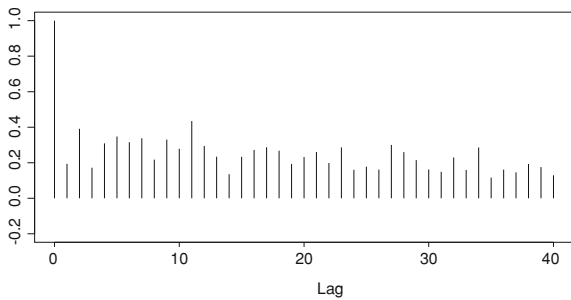
where $\beta'_i = (\beta_{i,1}, \dots, \beta_{i,p})'$ is a p -dimensional vector and $\{\mathbf{Z}(t) = (Z_1(t), \dots, Z_p(t))'\}$ is factors process.

Empirical features of the realized returns for assets and market portfolios are well-known. SACF, which was defined in Chap. 1, of returns of IBM stock was plot in Fig. 1.1 (Sect. 1.1) and SACF of returns of S&P500 (squared transformed) is plot in Fig. 3.1.

From Figs. 1.1 and 3.1, we observe that the return of stock (i.e., IBM) shows the short memory dependence, and that squared transformed of a market index (i.e., S&P500) shows the long memory dependence.

Suppose that an q -dimensional process $\{\mathbf{Y}(t) = (Y_1(t), \dots, Y_q(t))'\}$ is generated by

Fig. 3.1 SACF of return of S&P500 (square transformed). Taken from Amano et al. (2012). Published with the kind permission of © Tomoyuki Amano et al. (2012). Published under the creative commons attribution license



$$\mathbf{Y}(t) = \boldsymbol{\alpha} + \mathbf{B}'\mathbf{Z}(t) + \boldsymbol{\epsilon}(t), \quad (t = 1, 2, \dots, n) \quad (3.36)$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)'$ and $\mathbf{B} = \{\beta_{ij}; i = 1, \dots, p, j = 1, \dots, q\}$ are unknown vector and matrix, respectively, $\{\mathbf{Z}(t) = (Z_1(t), \dots, Z_p(t))'\}$ is a stochastic explanatory process and $\{\boldsymbol{\epsilon}(t) = (\epsilon_1(t), \dots, \epsilon_q(t))'\}$ is a sequence of disturbance process. The i th component is written as

$$Y_i(t) = \alpha_i + \boldsymbol{\beta}'_i \mathbf{Z}(t) + \epsilon_i(t)$$

where $\boldsymbol{\beta}'_i = (\beta_{i,1}, \dots, \beta_{i,p})$.

In the stochastic regression model, $\mathbf{Y}(t)$ is the return of assets, and $\mathbf{Z}(t)$ is the factors process. As we saw, empirical studies suggest that the return of assets $\{\mathbf{Y}(t)\}$ is short memory dependent and that the squared transformed return of the market portfolio is long memory dependent. Furthermore, as a market risk, the volatility of the return of the market portfolio should be considered and the stochastic regression model (3.36) in the case $\mathbf{Z}(t)$ is short memory dependent has been investigated enough. Then we assume that $\mathbf{Z}(t)$ is long memory dependent. On this ground, we investigate the conditions that the stochastic regression model (3.36) are well-defined. It is seen that if the model (3.36) is valid, we have to assume that $\{\boldsymbol{\epsilon}(t)\}$ is also long memory dependent and is correlated with $\{\mathbf{Z}(t)\}$.

Hence, we suppose that $\{\mathbf{Z}(t)\}$ and $\{\boldsymbol{\epsilon}(t)\}$ are defined by

$$\begin{aligned} \mathbf{Z}(t) &= \sum_{j=0}^{\infty} \boldsymbol{\gamma}(j)\mathbf{a}(t-j) + \sum_{j=0}^{\infty} \boldsymbol{\rho}(j)\mathbf{b}(t-j), \\ \boldsymbol{\epsilon}(t) &= \sum_{j=0}^{\infty} \boldsymbol{\eta}(j)\mathbf{e}(t-j) + \sum_{j=0}^{\infty} \boldsymbol{\xi}(j)\mathbf{b}(t-j), \end{aligned} \quad (3.37)$$

where $\{\mathbf{a}(t)\}$, $\{\mathbf{b}(t)\}$, and $\{\mathbf{e}(t)\}$ are p -dimensional zero-mean uncorrelated processes, and they are mutually independent. Here the coefficients $\{\boldsymbol{\gamma}(j)\}$ and $\{\boldsymbol{\rho}(j)\}$ are $p \times p$ -matrices, and all the components of $\boldsymbol{\gamma}(j)$ are ℓ^1 -summable, (for short, $\boldsymbol{\gamma}(j) \in \ell^1$), and those of $\boldsymbol{\rho}(j)$ are ℓ^2 -summable (for short, $\boldsymbol{\rho}(j) \in \ell^2$). The coefficients $\{\boldsymbol{\eta}(j)\}$ and $\{\boldsymbol{\xi}(j)\}$ are $q \times p$ -matrices, and $\boldsymbol{\eta}(j) \in \ell^1$ and $\boldsymbol{\xi}(j) \in \ell^2$. From (3.37) it follows that

$$Y(t) = \alpha + \sum_{j=0}^{\infty} (\mathbf{B}'\boldsymbol{\gamma}(j)\mathbf{a}(t-j) + \eta(j)\mathbf{e}(t-j)) + \sum_{j=0}^{\infty} (\mathbf{B}'\boldsymbol{\rho}(j) + \boldsymbol{\xi}(j))\mathbf{b}(t-j).$$

Although $(\mathbf{B}'\boldsymbol{\rho}(j) + \boldsymbol{\xi}(j)) \in \ell^2$ generally, if $\mathbf{B}'\boldsymbol{\rho}(j) + \boldsymbol{\xi}(j) = O\left(\frac{1}{j^\alpha}\right)$, $\alpha > 1$, then $(\mathbf{B}'\boldsymbol{\rho}(j) + \boldsymbol{\xi}(j)) \in \ell^1$, which leads to

Proposition 3.1 *If $\mathbf{B}'\boldsymbol{\rho}(j) + \boldsymbol{\xi}(j) = O(j^{-\alpha})$, $\alpha > 1$, then the process $\{Y(t)\}$ is short memory dependent.*

Proposition 3.1 provides an important view for the stochastic regression model, i.e., if we assume natural conditions on (3.36) based on the empirical studies, then they impose a sort of “curved structure”: $\mathbf{B}'\boldsymbol{\rho}(j) + \boldsymbol{\xi}(j) = O(j^{-\alpha})$ on the regressor and disturbance. More important view is the statement implies that the process $\{\boldsymbol{\beta}'_i\mathbf{Z}(t) + \epsilon_i(t)\}$ is fractionally cointegrated. Here $\boldsymbol{\beta}_i$ and $\epsilon_i(t)$ are called the cointegrating vector and error, respectively (see Robinson and Yajima (2002)).

We discuss estimation of (3.36) satisfying Proposition 3.1. Since $E(\mathbf{Z}(t)\boldsymbol{\epsilon}(t)') \neq \mathbf{0}$, the least squares estimator for \mathbf{B} is known to be inconsistent. In what follows we assume that $\boldsymbol{\alpha} = \mathbf{0}$ in (3.36), because it can be estimated consistently by the sample mean. However, by use of econometric theory, it is often possible to find other variables that are uncorrelated with the errors $\boldsymbol{\epsilon}(t)$, which we call instrumental variables, and to overcome this difficulty. Without instrumental variables, correlations between the observables $\{\mathbf{Z}(t)\}$ and unobservables $\{\boldsymbol{\epsilon}(t)\}$ persistently contaminate our estimator for \mathbf{B} . Hence, instrumental variables are useful in allowing us to estimate \mathbf{B} .

Let $\{\mathbf{X}(t)\}$ be r -dimensional vector ($p \leq r$) instrumental variables with $E\{\mathbf{X}(t)\} = \mathbf{0}$, $Cov\{\mathbf{X}(t), \mathbf{Z}(t)\} \neq \mathbf{0}$ and $Cov\{\mathbf{X}(t), \boldsymbol{\epsilon}(t)\} = \mathbf{0}$. Consider the OLS regression of $\mathbf{Z}(t)$ on $\mathbf{X}(t)$. If $\mathbf{Z}(t)$ can be represented as

$$\mathbf{Z}(t) = \boldsymbol{\delta}'\mathbf{X}(t) + \mathbf{u}(t), \quad (3.38)$$

where $\boldsymbol{\delta}$ is a $r \times p$ matrix and $\{\mathbf{u}(t)\}$ is a p -dimensional vector process which is independent of $\{\mathbf{X}(t)\}$, $\boldsymbol{\delta}$ can be estimated by the OLS estimator

$$\hat{\boldsymbol{\delta}} = \left[\sum_{t=1}^n \mathbf{X}(t)\mathbf{X}(t)' \right]^{-1} \left[\sum_{t=1}^n \mathbf{X}(t)\mathbf{Z}(t)' \right]. \quad (3.39)$$

From (3.36) with $\boldsymbol{\alpha} = \mathbf{0}$ and (3.38), $Y(t)$ has the form

$$Y(t) = \mathbf{B}'\boldsymbol{\delta}'\mathbf{X}(t) + \mathbf{B}'\mathbf{u}(t) + \boldsymbol{\epsilon}(t)$$

and $\boldsymbol{\delta}'\mathbf{X}(t)$ is uncorrelated with $\mathbf{B}'\mathbf{u}(t) + \boldsymbol{\epsilon}(t)$, hence \mathbf{B} can be estimated by the OLS estimator

$$\hat{\mathbf{B}}_{OLS} = \left[\sum_{t=1}^n (\delta' \mathbf{X}(t)) (\delta' \mathbf{X}(t))' \right]^{-1} \left[\sum_{t=1}^n (\delta' \mathbf{X}(t)) \mathbf{Y}'(t) \right]. \quad (3.40)$$

Using (3.39) and (3.40), we can propose the 2SLS estimator

$$\hat{\mathbf{B}}_{2SLS} = \left[\sum_{t=1}^n (\hat{\delta}' \mathbf{X}(t)) (\hat{\delta}' \mathbf{X}(t))' \right]^{-1} \left[\sum_{t=1}^n (\hat{\delta}' \mathbf{X}(t)) \mathbf{Y}'(t) \right]. \quad (3.41)$$

Now, we aim at proving the consistency and asymptotic normality of the 2SLS estimator $\hat{\mathbf{B}}_{2SLS}$. For this we assume that $\{\epsilon(t)\}$ and $\{\mathbf{X}(t)\}$ jointly constitute the following linear process.

$$\begin{pmatrix} \epsilon(t) \\ \mathbf{X}(t) \end{pmatrix} = \sum_{j=0}^{\infty} \mathbf{G}(j) \boldsymbol{\Gamma}(t-j) = \mathbf{A}(t) \quad (\text{say}),$$

where $\{\boldsymbol{\Gamma}(t)\}$ is uncorrelated $(q+r)$ -dimensional vector process with

$$\begin{aligned} E\{\boldsymbol{\Gamma}(t)\} &= \mathbf{0} \\ E\{\boldsymbol{\Gamma}(t) \boldsymbol{\Gamma}(s)^*\} &= \delta(t, s) \mathbf{K} \\ \delta(t, s) &= \begin{cases} 1, & t = s \\ 0, & t \neq s \end{cases} \end{aligned}$$

and $\mathbf{G}(j)$'s are $(q+r) \times (q+r)$ matrices which satisfy $\sum_{j=0}^{\infty} \text{tr}\{\mathbf{G}(j) \mathbf{K} \mathbf{G}(j)^*\} < \infty$. Then $\{\mathbf{A}(t)\}$ has the spectral density matrix

$$\mathbf{f}(\lambda) = \frac{1}{2\pi} \mathbf{k}(\lambda) \mathbf{K} \mathbf{k}(\lambda)^* = \{f_{ab}(\lambda); 1 \leq a, b \leq (q+r)\}, \quad (-\pi < \lambda \leq \pi),$$

where

$$\mathbf{k}(\lambda) = \sum_{j=0}^{\infty} \mathbf{G}(j) e^{i\lambda j} = \{k_{ab}(\lambda); 1 \leq a, b \leq (q+r)\}, \quad (-\pi < \lambda \leq \pi).$$

Further, we assume that $\int_{-\pi}^{\pi} \log \det \mathbf{f}(\lambda) d\lambda > -\infty$, so that the process $\{\mathbf{A}(t)\}$ is non-deterministic. For the asymptotics of $\hat{\mathbf{B}}_{2SLS}$, from page 108–109 of Hosoya (1997), we impose Assumption (HT) (ii), (iii), and (iv) in Chap. 1 on $\boldsymbol{\Gamma}(t)$ and the following Assumption 3.5.

Assumption 3.5 Each $f_{ab}(\lambda)$ is square-integrable.

Under above assumptions, we can establish the following theorem. For the proof, see Amano et al. (2012).

Theorem 3.3 *If $\Gamma(t)$ satisfies Assumption (HT) (ii), (iii), and (iv) in Chap. 1, under Assumption 3.5 it holds that*

(i)

$$\hat{\mathbf{B}}_{2SLS} \xrightarrow{P} \mathbf{B},$$

(ii)

$$\sqrt{n} \left(\hat{\mathbf{B}}_{2SLS} - \mathbf{B} \right) \xrightarrow{d} \mathbf{Q}^{-1} E \left[\mathbf{Z}(t) \mathbf{X}'(t) \right] E \left[\mathbf{X}(t) \mathbf{X}(t)'\right]^{-1} \mathbf{U}$$

where

$$\mathbf{Q} = \left[E(\mathbf{Z}(t) \mathbf{X}(t)') \right] \left[E(\mathbf{X}(t) \mathbf{X}(t)') \right]^{-1} \left[E(\mathbf{X}(t) \mathbf{Z}(t)') \right]$$

and $\mathbf{U} = \{U_{i,j}; 1 \leq i \leq r, 1 \leq j \leq q\}$ is a random matrix whose elements follow normal distributions with mean 0 and

$$\begin{aligned} \text{Cov}[U_{i,j}, U_{k,l}] &= 2\pi \int_{-\pi}^{\pi} [f_{q+i,q+k}(\lambda) \bar{f}_{j,l}(\lambda) + f_{q+i,l}(\lambda) \bar{f}_{j,q+k}(\lambda)] d\lambda \\ &+ 2\pi \sum_{\beta_1, \dots, \beta_4=1}^p \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \kappa_{q+i,\beta_1}(\lambda_1) \kappa_{j,\beta_2}(-\lambda_1) \\ &\times \kappa_{q+k,\beta_3}(\lambda_2) \kappa_{l,\beta_4}(-\lambda_2) \mathcal{Q}_{\beta_1, \dots, \beta_4}^{\Gamma}(\lambda_1, -\lambda_2, \lambda_2) d\lambda_1 d\lambda_2, \quad (3.42) \end{aligned}$$

where $\mathcal{Q}_{\beta_1, \dots, \beta_4}^{\Gamma}(\lambda_1, \lambda_2, \lambda_3)$ is the fourth order spectral density of $\Gamma(t)$, which was defined in Chap. 1.

Next example prepares the asymptotic variance formula of $\hat{\mathbf{B}}_{2SLS}$ to investigate its features in simulation study.

Example 3.4 Let $\{Z(t)\}$ and $\{X(t)\}$ be scalar long memory processes, with spectral densities $\{2\pi |1 - e^{i\lambda}|^{2d_Z}\}^{-1}$ and $\{2\pi |1 - e^{i\lambda}|^{2d_X}\}^{-1}$, respectively, and cross spectral density $\frac{1}{2\pi} (1 - e^{i\lambda})^{-d_X} (1 - e^{-i\lambda})^{-d_Z}$, where $0 < d_Z < 1/2$ and $0 < d_X < 1/2$. Then

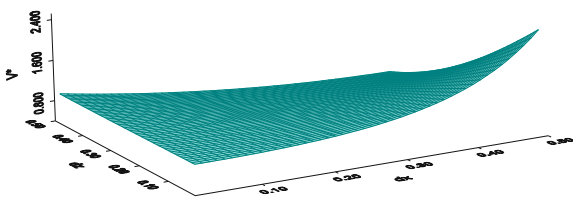
$$E(X(t)Z(t)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{(1 - e^{i\lambda})^{d_X}} \frac{1}{(1 - e^{-i\lambda})^{d_Z}} d\lambda.$$

Suppose that $\{\epsilon(t)\}$ is a scalar uncorrelated process with $\sigma_{\epsilon}^2 \equiv E\{\epsilon(t)^2\}$. Assuming Gaussianity of $\{\mathbf{A}(t)\}$, it is seen that the right hand of (3.42) is

$$2\pi \int_{-\pi}^{\pi} \frac{1}{2\pi |1 - e^{i\lambda}|^{2d_X}} \frac{\sigma_{\epsilon}^2}{2\pi} d\lambda,$$

which entails

Fig. 3.2 $V_*(d_x, d_z)$ in Example 3.5. Taken from Amano et al. (2012). Published with the kind permission of © Tomoyuki Amano et al. 2012. Published under the creative commons attribution license



$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Var} \left\{ \sqrt{n} \left(\hat{B}_{2SLS} - B \right) \right\} &= \frac{2\pi \int_{-\pi}^{\pi} \frac{1}{2\pi |1 - e^{i\lambda}|^{2d_X}} \frac{\sigma_\epsilon^2}{2\pi} d\lambda}{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{(1 - e^{i\lambda})^{d_X}} \frac{1}{(1 - e^{-i\lambda})^{d_Z}} d\lambda \right)^2} \\ &= \sigma_\epsilon^2 \left(2\pi \frac{\int_{-\pi}^{\pi} \frac{1}{|1 - e^{i\lambda}|^{2d_X}} d\lambda}{\left(\int_{-\pi}^{\pi} \frac{1}{(1 - e^{i\lambda})^{d_X}} \frac{1}{(1 - e^{-i\lambda})^{d_Z}} d\lambda \right)^2} \right) \\ &= \sigma_\epsilon^2 \times V_*(d_X, d_Z). \end{aligned}$$

Now, we evaluate the behavior of \hat{B}_{2SLS} in the case $p = 1$ in (3.36) numerically.

Example 3.5 Under the condition of Example 3.4, we investigate the asymptotic variance behavior of \hat{B}_{2SLS} by simulation. Figure 3.2 plots $V_*(d_X, d_Z)$ for $0 < d_X < \frac{1}{2}$ and $0 < d_Z < \frac{1}{2}$.

From Fig. 3.2, we observe that if $d_Z \searrow 0$ and if $d_X \nearrow 1/2$, then V_* becomes large and otherwise V_* is small. This result implies only in the case that the long memory behavior of $Z(t)$ is weak and the long memory behavior of $X(t)$ is strong, V_* is large. Note that long memory behavior of $Z(t)$ makes the asymptotic variance of the 2SLS estimator small but one of $X(t)$ makes it large.

Example 3.6 We consider the following model,

$$\begin{aligned} Y(t) &= Z(t) + \epsilon(t), \\ Z(t) &= X(t) + u(t), \\ \epsilon(t) &= w(t) + u(t), \end{aligned}$$

where $X(t)$, $w(t)$, and $u(t)$ are scalar long memory processes which follow FARIMA(0, d_1 , 0), FARIMA(0, d_2 , 0), and FARIMA(0, 0.1, 0), respectively. Note that $Z(t)$ and $\epsilon(t)$ are correlated, $X(t)$ and $Z(t)$ are correlated, but $X(t)$ and $\epsilon(t)$ are independent. Under this model, we compare \hat{B}_{2SLS} with the ordinary least squares estimator \tilde{B}_{OLS} for B , which is defined as

$$\tilde{B}_{OLS} = \left[\sum_{t=1}^n Z^2(t) \right]^{-1} \left[\sum_{t=1}^n Z(t)Y(t) \right].$$

Table 3.5 MSE of \hat{B}_{2SLS} and \tilde{B}_{OLS}

d_2	0.1	0.2	0.3
$\hat{B}_{2SLS} (d_1 = 0.1)$	0.03	0.052	0.189
$\tilde{B}_{OLS} (d_1 = 0.1)$	0.259	0.271	0.34
$\hat{B}_{2SLS} (d_1 = 0.2)$	0.03	0.075	0.342
$\tilde{B}_{OLS} (d_1 = 0.2)$	0.178	0.193	0.307
$\hat{B}_{2SLS} (d_1 = 0.3)$	0.019	0.052	0.267
$\tilde{B}_{OLS} (d_1 = 0.3)$	0.069	0.089	0.23

Taken from Amano et al. (2012). Published with the kind permission of © Tomoyuki Amano et al. 2012. Published under the creative commons attribution license

Table 3.6 \hat{B}_{2SLS} based on the actual financial data

Stock	IBM	Nike	Amazon	American Express	Ford
\hat{B}_{2SLS}	0.75	1.39	1.71	2.61	-1.89

Taken from Amano et al. (2012). Published with the kind permission of © Tomoyuki Amano et al. 2012. Published under the creative commons attribution license

The length of $X(t)$, $Y(t)$, and $Z(t)$ are set by 100, and based on 5,000 times simulation we report MSE of \hat{B}_{2SLS} and \tilde{B}_{OLS} . We set $d_1, d_2 = 0.1, 0.2, 0.3$ in Table 3.5.

In most cases of d_1 and d_2 in Table 3.5, MSE of \hat{B}_{2SLS} is smaller than that of \tilde{B}_{OLS} . Hence from this Example we can see our estimator \hat{B}_{2SLS} is better than \tilde{B}_{OLS} in the sense of MSE. Furthermore from Table 3.5, we can see that MSE of \hat{B}_{2SLS} and \tilde{B}_{OLS} increases as d_2 becomes large, that is, long memory behavior of $w(t)$ makes the asymptotic variances of \hat{B}_{2SLS} and \tilde{B}_{OLS} large.

Example 3.7 In this example, we calculate \hat{B}_{2SLS} based on the actual financial data. We choose S&P500 (square transformed) as $Z(t)$ and Nikkei Average as an instrumental variable $X(t)$. Assuming that $Y(t)$ (5×1) consists of the return of IBM, Nike, Amazon, American Expresses, and Ford, the 2SLS estimates for $B_i, i = 1, \dots, 5$ are recorded in Table 3.6. We chose Nikkei Stock Average as the instrumental variable, because we got the following correlation analysis between the residual processes of returns and Nikkei:

Correlation of IBM's residual and Nikkei Average's return: -0.000311

Correlation of Nike's residual and Nikkei Average's return: -0.00015

Correlation of Amazon's residual and Nikkei Average's return: -0.000622

Correlation of American Express's residual and Nikkei Average's return: 0.000147

Correlation of Ford's residual and Nikkei Average's return: -0.000536 ,

which supports the assumption $Cov(X(t), \epsilon(t)) = \mathbf{0}$.

From Table 3.6, we observe that the return of American Express is strongly correlated with that of S&P500 and the return of the auto industry stock (Ford) is negatively correlated with that of S&P500.

References

- Amano, T., Kato, T., Taniguchi, M.: Statistical estimation for CAPM with long-memory dependence. *Adv. Decis. Sci.* **2012**, Article ID 571034 (2012)
- Amano, T., Taniguchi, M.: Control variate method for stationary processes. *J. Econometrics* **165**, 20–29 (2011)
- Black, F.: Capital market equilibrium with restricted borrowing. *J. Bus.* **45**, 444–455 (1972)
- Brillinger, D.R.: *Time Series: Data Analysis and Theory*, Expanded edn. Holden-Day, San Francisco (2001)
- Campbell, J.Y., Lo, A.W., Mackinlay, A.C.: *The Econometrics of Financial Markets*. Princeton University Press, New Jersey (1997)
- Chan, N.H., Wong, H.Y.: *Simulation Techniques in Financial Risk Management*. Wiley, New York (2006)
- Choy, K., Taniguchi, M.: Stochastic regression model with dependent disturbances. *J. Time Ser. Anal.* **22**, 175–196 (2001)
- Geary, R.C.: Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown. *Econometrica* **17**, 30–58 (1949)
- Glasserman, P.: *Monte Carlo Methods in Financial Engineering*. Springer, New York (2004)
- Hosoya, Y.: A limit theory for long-range dependence and statistical inference on related models. *Ann. Stat.* **25**, 105–137 (1997)
- Lavenberg, S.S., Welch, P.D.: A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Manag. Sci.* **27**, 322–335 (1981)
- Lintner, J.: The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev. Econ. Stat.* **47**, 13–37 (1965)
- Markowitz, H.: *Portfolio Selection: Efficient Diversification of Investments*. Wiley, New York (1991)
- Nelson, B.L.: Control variate remedies. *Oper. Res.* **38**, 974–992 (1990)
- Reiersöl, O.: Confluence analysis by means of instrumental sets of variables. *Aktiv för Matematik, Astronomi och Fysik.* **32A**, 1–119 (1945)
- Robinson, P.M., Hidalgo, F.J.: Time series regression with long-range dependence. *Ann. Stat.* **25**, 77–104 (1997)
- Robinson, P.M., Yajima, Y.: Determination of cointegrating rank in fractional system. *J. Econometrics* **106**, 217–241 (2002)
- Rubinstein, R.Y., Marcus, R.: Efficiency of multivariate control variates in Monte Carlo simulation. *Oper. Res.* **33**, 661–677 (1985)
- Sharpe, W.: Capital asset prices: a theory of market equilibrium under conditions of risk. *J. Finance* **19**, 425–442 (1964)
- White, H.: *Asymptotic Theory for Econometricians*. Academic Press, New York (2001)
- Wright, P.G.: *The Tariff on Animal and Vegetable Oils*. MacMillan, New York (1928)

Chapter 4

Some Techniques for ARCH Financial Time Series

Abstract This chapter introduces two techniques which can be utilized in study of financial risks. The first one is the method called Quantile Regression (QR), which can be used to analyze the conditional quantile of financial assets. There, by means of rank-based semiparametrics, we provide the statistically efficient inference under the autoregressive conditional heteroskedasticity (ARCH). The second technique, the realized volatility, estimates the conditional variance, or “volatility” of financial assets. Revealing the fact that its inference can be greatly affected by the existence of additional noise called market microstructure, we introduce and study the asymptotics of some appropriate estimator under the microstructure with ARCH-dependent structure.

Keywords ARCH model · Quantile regression · Rank-based semiparametrics · Realized volatility · Market microstructure

4.1 Introduction

In this chapter, two techniques for analysis of financial risks will be introduced. For both techniques, namely, the Quantile Regression (QR) with its semiparametrics and Realized Volatility (RV) with dependent microstructure, here we extend their applicability to the AutoRegressive Conditional Heteroskedasticity (ARCH) models. As is well known, ARCH structure is one of the most particular characteristics of financial asset’s prices. (See, e.g., Gouriéroux (1997).)

The construction of this chapter is as follows. In Sect. 4.2, we explain how QR method can be used for ARCH time series and provide the semiparametrically efficient inference as well. This method allows us to estimate the conditional quantiles of financial assets, such as some variants of Value-at-Risk. Section 4.3 is concerned with estimation of the conditional variance, which is called “volatility” in finance. There, the existence of additional noise called market microstructure affects our inference greatly. By introducing some appropriate estimator, we show that this estimator works well in the case where microstructure exhibits the ARCH property.

4.2 Quantile Regression and Its Semiparametric Efficiency for ARCH Series

Motivated by squared ARCH models, we investigate a Quantile AutoRegression (QAR) model for non-negative series (Koenker and Bassett (1978); Koenker and Xiao (2006); Koenker (2005)). This QAR can be interpreted as a Random Coefficient AutoRegression (RCAR) model for which we see the Local Asymptotic Normality (LAN). More precisely, we decompose the random coefficient as a product of two components, a quantile coefficient $\boldsymbol{\gamma}(\tau)$ for fixed τ and a “standardized random coefficient” with τ -quantile one, and think of the former as a parameter of interest while the density g_1 of the latter is treated as a nuisance. Having the LAN result with respect to $\boldsymbol{\gamma}(\tau)$ for each τ -th quantile restriction (identification constraint) given on the nuisance, we follow the one-step estimation method of Le Cam. Consequently, we provide a semiparametrically efficient (at g_1) version of QAR estimators.

4.2.1 Model, Estimators, and Some Asymptotics

Suppose that we have observation $\{y_t^2\}_{t=1}^n$ from some experiment $(\mathbb{R}_+^n, \mathcal{B}^{(n)}, \mathcal{P}^{(n)})$. Also, let us consider a sequence of probability measures $\mathcal{P}^{(n)} = \{P_{\boldsymbol{\beta}, g}^{(n)} | \boldsymbol{\beta} \in \boldsymbol{\Theta}, g \in \mathcal{G}\}$ which is described by the model known as ARCH(1), that is,

$$Y_t^2 = (\beta_0 + \beta_1 Y_{t-1}^2) Z_t^2, \quad (4.1)$$

where $\boldsymbol{\beta} := (\beta_0, \beta_1)'$, and $\{Z_t^2\}_{t=1}^n$ are i.i.d. with density g and distribution function G . Here $\boldsymbol{\Theta} \subset \mathbb{R}^2$ and \mathcal{G} are some parameter space and family of probability density functions, respectively. Note that we introduced the innovation density g as that of the squared process $\{Z_t^2\}_{t=1}^n$, not that of $\{Z_t\}_{t=1}^n$ itself. The conditional τ -quantile of Y_t^2 (conditioned by the realized value y_{t-1}^2) in model (4.1) is

$$F_{Y_t^2}^{-1}(\tau | y_{t-1}^2) := F_{Y_t^2 | Y_{t-1}^2 = y_{t-1}^2}^{-1}(\tau) = (\beta_0 + \beta_1 y_{t-1}^2) G^{-1}(\tau),$$

where $F_{X|S}^{-1}(\cdot)$ denotes the (conditional) quantile function, i.e., $F_{X|S}^{-1}(\tau) := \inf\{x : P(X \leq x | S) \geq \tau\}$.

Motivated by the squared series of ARCH(1) above, let $X_t \geq 0, t \in \mathbb{Z}$, be a process which takes non-negative values. Then we define a Quantile AutoRegression (QAR) model, for given $\tau \in (0, 1)$, by

$$F_{X_t}^{-1}(\tau | x_{t-1}) = \gamma_0(\tau) + \gamma_1(\tau) x_{t-1}, \quad (4.2)$$

where x_{t-1} is the realized value of X_{t-1} and $\boldsymbol{\gamma}(\tau) := (\gamma_0(\tau), \gamma_1(\tau))'$ is the parameter of interest. So, we see that the model (4.1) is a special case of (4.2) with

$\boldsymbol{\gamma}(\tau) \equiv \boldsymbol{\beta}G^{-1}(\tau)$. In the following, although we discuss this quantile regression model of order 1 for simplicity, our method can be easily extended to the case of higher dependency so that it can be applicable to ARCH(p), $p \geq 1$.

Now, we define the QR estimator $\hat{\boldsymbol{\gamma}}_n(\tau)$ as

$$\hat{\boldsymbol{\gamma}}_n(\tau) := \underset{b_0, b_1}{\operatorname{argmin}} \sum_{t=1}^n \rho_\tau(X_t - (b_0 + b_1 X_{t-1})), \quad (4.3)$$

where $\rho_\tau(u) := \tau|u| \cdot 1\{u \geq 0\} + (1 - \tau)|u| \cdot 1\{u < 0\} = u(\tau - 1\{u < 0\})$, $u \in \mathbb{R}$ is called the *check function* (see Koenker (2005)), and $1A$ is the indicator function defined by $1A = 1_A(\omega) := 1$ if $\omega \in A$, $:= 0$ if $\omega \notin A$. Also, defining

$$D_0 := E[X_{t-1}X'_{t-1}], \quad \text{and} \\ D_{\boldsymbol{\gamma}(\tau)} := E[f_{X_t}(X'_{t-1}\boldsymbol{\gamma}(\tau)|X_{t-1})X_{t-1}X'_{t-1}]$$

with $X_{t-1} := (1, X_{t-1})'$, some regularity conditions are now in order:

Assumption 4.1

- (i) There exists a unique strictly stationary solution $\{X_t\}$ of model (4.2) with $E[X_t^2] < \infty$;
- (ii) The matrices D_0 and $D_{\boldsymbol{\gamma}(\tau)}$ are nonsingular;
- (iii) For $\tau = \tau_0$ of our interest, the density of X_t conditional on X_{t-1} , denoted by $f_{X_t}(x|X_{t-1})$, is continuous and positive at the points $x = F_{X_t}^{-1}(\tau_0|X_{t-1})$, $t = 1, 2, \dots$

Proposition 4.1 *Let Assumption 4.1 hold. Then, for any $\tau \in (0, 1)$, as $n \rightarrow \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}_n(\tau) - \boldsymbol{\gamma}(\tau)) \xrightarrow{d} N\left(\mathbf{0}, \tau(1 - \tau)D_{\boldsymbol{\gamma}(\tau)}^{-1}D_0D_{\boldsymbol{\gamma}(\tau)}^{-1}\right), \quad (4.4)$$

Proof See e.g., Koenker (2005, Chap. 4), and Koul (1992, Eq. (7.3b.6)). \square

Now, observe that, for each fixed τ , the QR coefficient $\boldsymbol{\gamma}(\tau)$ can be characterized as the parameter $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)'$ of some model such as

$$X_t = (\gamma_0 + \gamma_1 X_{t-1})\xi_t, \quad (4.5)$$

where $\{\xi_t \geq 0, t \in \mathbb{Z}\}$ is i.i.d. with distribution function G_1 and density g_1 ,

$$g_1 \in \mathcal{F}^\tau := \left\{ f : [0, \infty) \rightarrow [0, \infty) \mid \int_0^1 f(x)dx = \tau = 1 - \int_1^\infty f(x)dx \right\}.$$

This “quantile-restricted ARCH model” is a fixed- τ submodel of general QAR model (4.2). There we find that

$$f_{X_t}(X'_{t-1}\boldsymbol{\gamma}(\tau)|X_{t-1}) = \frac{f_{X_t/X'_{t-1}\boldsymbol{\gamma}(1)}}{X'_{t-1}\boldsymbol{\gamma}} \left(= \frac{g_1(G_1^{-1}(\tau))}{X'_{t-1}\boldsymbol{\gamma}} \right),$$

so that the following corollary holds from Proposition 4.1.

Corollary 4.1 *Let Assumption 4.1 hold. Then, for any $\boldsymbol{\gamma} \in \Theta \subset \mathbb{R}_+^2$ and any $g_1 \in \mathcal{F}^\tau$, as $n \rightarrow \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}_n(\tau) - \boldsymbol{\gamma}) \xrightarrow{d} N\left(\mathbf{0}, \frac{\tau(1-\tau)}{g_1^2(1)} D_{1,\boldsymbol{\gamma}}^{-1} D_0 D_{1,\boldsymbol{\gamma}}^{-1}\right), \quad (4.6)$$

where $D_{1,\boldsymbol{\gamma}} := E[X_{t-1}X'_{t-1}/X'_{t-1}\boldsymbol{\gamma}]$.

Remark 4.1 Note here that, for the submodel (4.5), Assumption 4.1-(i) is satisfied if the following conditions are fulfilled

- (i) The “true” values γ_0 and γ_1 are positive;
- (ii) $E[\xi_0] < \infty$, $E[\xi_0^2] < \infty$ and $(E[\xi_0^2])^{1/2} \gamma_1 < 1$.

The latter is a sufficient condition due to Giraitis et al. (2000), which may be easy to verify. Necessary and sufficient conditions can also be found in He and Teräsvirta (1999). ■

Now, as a preparation for Sect. 4.2.2, let us write $\mathcal{P}_{\boldsymbol{\gamma}}^{(n)} := \{P_{\boldsymbol{\gamma},g_1}^{(n)} | g_1 \in \mathcal{F}^\tau\}$ for the fixed- $\boldsymbol{\gamma}$ subfamily of $\mathcal{P}^{(n)}$. In order to apply the general results of Hallin et al. (2006b), we first establish the existence of a generating group of $\mathcal{P}_{\boldsymbol{\gamma}}^{(n)}$. Denoting by \mathcal{H}_1 the set of all continuous and strictly increasing functions h from \mathbb{R}_+ to \mathbb{R}_+ satisfying $h(0) = 0$ and $h(1) = 1$, define

$$a_h^{(n)}(\xi_1, \dots, \xi_n) := (h(\xi_1), \dots, h(\xi_n))$$

and consider the transformation group $A_{\boldsymbol{\gamma}}^{(n)}$ (acting on \mathbb{R}_+^n)

$$A_{\boldsymbol{\gamma}}^{(n)} := \left\{ (\boldsymbol{\xi}_{\boldsymbol{\gamma}}^{(n)})^{-1} \circ a_h^{(n)} \circ \boldsymbol{\xi}_{\boldsymbol{\gamma}}^{(n)}, h \in \mathcal{H}_1 \right\}, \quad (4.7)$$

where $\boldsymbol{\xi}_{\boldsymbol{\gamma}}^{(n)}$ is a residual function defined by

$$\boldsymbol{\xi}_{\boldsymbol{\gamma}}^{(n)}(X_1, \dots, X_n) := \left(\frac{X_1}{\gamma_0 + \gamma_1 X_0}, \dots, \frac{X_n}{\gamma_0 + \gamma_1 X_{n-1}} \right) = (\xi_{\boldsymbol{\gamma},1}, \dots, \xi_{\boldsymbol{\gamma},n}).$$

Then, $A_{\boldsymbol{\gamma}}^{(n)}$ is a generating group for $\mathcal{P}_{\boldsymbol{\gamma}}^{(n)}$, in the sense that for all $P_1, P_2 \in \mathcal{P}_{\boldsymbol{\gamma}}^{(n)}$, there exists $a \in A_{\boldsymbol{\gamma}}^{(n)}$ such that $(X_1, \dots, X_n) \stackrel{d}{\sim} P_1$ iff $a(X_1, \dots, X_n) \stackrel{d}{\sim} P_2$. So, defining $S_{\boldsymbol{\gamma},t}$ and $R_{\boldsymbol{\gamma},t}^{(n)}$ as the sign and the rank (among n copies) of $(\xi_{\boldsymbol{\gamma},t} - 1)$ respectively, we see that the map

$$T(X_1, \dots, X_n) = (S_{\boldsymbol{\gamma},1}, \dots, S_{\boldsymbol{\gamma},n}; R_{\boldsymbol{\gamma},1}^{(n)}, \dots, R_{\boldsymbol{\gamma},n}^{(n)}), \quad (4.8)$$

is maximal invariant w.r.t. $A_{\boldsymbol{\gamma}}^{(n)}$ (cf. Schmitterer (1974, Sections 7.4 and 7.6)). Maximal invariance implies that, for any invariant map T' (i.e., $T'(\mathbf{x}) = T'(a\mathbf{x})$, $a \in A_{\boldsymbol{\gamma}}^{(n)}$), there exists a measurable function χ such that $T' = \chi \circ T$. Consequently, we have that any test which is invariant w.r.t. $A_{\boldsymbol{\gamma}}^{(n)}$ is $\mathcal{B}_{\boldsymbol{\gamma}}^{(n)}$ -measurable, where

$$\mathcal{B}_{\boldsymbol{\gamma}}^{(n)} := \sigma(S_{\boldsymbol{\gamma},1}, \dots, S_{\boldsymbol{\gamma},n}; R_{\boldsymbol{\gamma},1}^{(n)}, \dots, R_{\boldsymbol{\gamma},n}^{(n)}) \quad (4.9)$$

which is the σ -field generated by $\{S_{\boldsymbol{\gamma},t}\}_{t=1}^n$ and $\{R_{\boldsymbol{\gamma},t}^{(n)}\}_{t=1}^n$. This $\mathcal{B}_{\boldsymbol{\gamma}}^{(n)}$ will play a key role in Sect. 4.2.2.

4.2.2 Semiparametrically Efficient Inference

In this section, we improve the efficiency of $\hat{\boldsymbol{\gamma}}_n(\tau)$ with regard to the quantile-restricted ARCH model (4.5). Since here we do not have any knowledge about the true density g_1 , except for the fact that it belongs to \mathcal{F}^τ , we arbitrarily choose a “reference density” f from \mathcal{F}^τ , and correspondingly define a “reference model”

$$X_t = (\gamma_0 + \gamma_1 X_{t-1})\xi_t, \quad (4.10)$$

where $X_t \geq 0$, $t \in \mathbb{Z}$, and $\{\xi_t \geq 0, t \in \mathbb{Z}\}$ is i.i.d. with density $f \in \mathcal{F}^\tau$. The goal is to construct an asymptotically efficient version of $\hat{\boldsymbol{\gamma}}_n(\tau)$ based on some feasible $f \in \mathcal{F}^\tau$, that is, attaining the semi-parametric lower bound at correctly specified density $f = g_1$ that nevertheless remains \sqrt{n} -consistent under misspecified density ($f \neq g_1$).

First, let us assume stationarity by assuming, e.g., that the conditions of Remark 4.1 hold. Also, in order to ensure the regularity of the reference model (4.10), we require the following assumption, which is essentially same as that for standard (i.e., nonsquared) ARCH models since $\boldsymbol{\gamma}$ has no effect on the sign of Y_t .

Assumption 4.2

- (i) The distribution of the initial value, $\mathcal{L}(X_0|\cdot)$, say, is continuous in probability with respect to $\boldsymbol{\gamma}$, i.e., for any $\boldsymbol{\gamma}_n \rightarrow \boldsymbol{\gamma}$,

$$\mathcal{L}(X_0|\boldsymbol{\gamma}_n) \xrightarrow{P} \mathcal{L}(X_0|\boldsymbol{\gamma});$$

- (ii) the density $f \in \mathcal{F}^\tau$ is absolutely continuous with derivative f' and has finite Fisher information for scale, i.e.,

$$0 < \mathcal{I}_f := \int_0^\infty \varphi_f^2(x) f(x) dx = \int_0^1 \varphi_f^2(F^{-1}(u)) du < \infty,$$

$$\text{where } \varphi_f(x) := -1 - \frac{xf'(x)}{f(x)}.$$

Then, by Drost et al. (1997, Theorem 2.1), the model (4.10) satisfies the uniform LAN (ULAN) condition under $f \in \mathcal{F}^\tau$ and for any $\boldsymbol{\gamma}_n$ of the form $\boldsymbol{\gamma} + O(n^{-1/2})$, with the central sequence

$$\mathbf{A}_{\boldsymbol{\gamma}_n, f}^{(n)} := \frac{1}{\sqrt{n}} \sum_{t=1}^n \varphi_f(\xi_{\boldsymbol{\gamma}_n, t}) \mathbf{W}_t(\boldsymbol{\gamma}_n), \quad \mathbf{W}_t(\boldsymbol{\gamma}_n) \equiv \frac{X_{t-1}}{X'_{t-1} \boldsymbol{\gamma}_n}, \quad (4.11)$$

and Fisher information $\mathcal{I}_f(\boldsymbol{\gamma}) \equiv \mathcal{I}_f \cdot E[\mathbf{W}_t(\boldsymbol{\gamma}) \mathbf{W}'_t(\boldsymbol{\gamma})]$, where $\xi_{\boldsymbol{\gamma}_n, t} := X_t / (\gamma_{n,0} + \gamma_{n,1} X_{t-1})$ denotes the residuals. The reason why we have stated ULAN, rather than LAN at single $\boldsymbol{\gamma}$, is due to the one-step improvement, which will be discussed below.

Here we observe that, by recursively replacing X_{t-i} by $(\gamma_0 + \gamma_1 X_{t-i-1}) \xi_{\boldsymbol{\gamma}, t-i}$, recovering the exact X_{t-i} requires observing the infinitely many lagged residuals $\{\xi_{\boldsymbol{\gamma}, t-i}\}_{i=1}^\infty$. However, this can be overcome by re-expressing (up to $o_P(1)$ terms) the central sequence (4.11) by another central sequence which involves finitely lagged residuals as follows. Since $X_{t-1} = \gamma_0 \sum_{i=1}^{t-1} \gamma_1^{i-1} \prod_{j=1}^i \xi_{\boldsymbol{\gamma}, t-j}$, we have for the second component (first component can be treated similarly) that

$$\begin{aligned} E \left| \frac{X_{t-1}}{\gamma_0 + \gamma_1 X_{t-1}} - \frac{\gamma_0 \sum_{i=1}^p \gamma_1^{i-1} \prod_{j=1}^i \xi_{\boldsymbol{\gamma}, t-j}}{\gamma_0 + \gamma_1 \gamma_0 \sum_{i=1}^p \gamma_1^{i-1} \prod_{j=1}^i \xi_{\boldsymbol{\gamma}, t-j}} \right| \\ \leq E \left[\sum_{i=p+1}^{t-1} \gamma_1^{i-1} \prod_{j=1}^i \xi_{\boldsymbol{\gamma}, t-j} \right] \leq \frac{\sigma_Z^2}{1 - \gamma_1 \sigma_Z^2} (\gamma_1 \sigma_Z^2)^p, \end{aligned} \quad (4.12)$$

where $\sigma_Z^2 := E[\xi_{\boldsymbol{\gamma}, t}]$. By the stationarity, we know that (4.12) converges to zero at a geometric rate as $p = p(n) \rightarrow \infty$, where this convergence $p \rightarrow \infty$ can be made arbitrarily slow. Hence, (4.12) is $o_P(1)$ as $n \rightarrow \infty$ and so we do not lose anything through this re-expression.

Then, following Hallin and Werker (2003), a semi-parametrically efficient procedure can be obtained by projecting (4.11) on the maximal invariant σ -field, which is $\mathcal{B}_{\boldsymbol{\gamma}}^{(n)}$ in our case. So, together with (4.12), “good” inference should be based on

$$\begin{aligned} \mathbf{A}_{\boldsymbol{\gamma}, f}^{(n)} &:= E_{\boldsymbol{\gamma}, f}^{(n)}[\mathbf{A}_{\boldsymbol{\gamma}, f}^{(n)} | \mathcal{B}_{\boldsymbol{\gamma}}^{(n)}] \\ &= \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^n E_{\boldsymbol{\gamma}, f}^{(n)} \left[\varphi_f(\xi_{\boldsymbol{\gamma}, t}) \frac{[1, \gamma_0 \sum_{i=1}^p \gamma_1^{i-1} \prod_{j=1}^i \xi_{\boldsymbol{\gamma}, t-j}]'}{\gamma_0 + \gamma_1 \gamma_0 \sum_{i=1}^p \gamma_1^{i-1} \prod_{j=1}^i \xi_{\boldsymbol{\gamma}, t-j}} \middle| \mathcal{B}_{\boldsymbol{\gamma}}^{(n)} \right] + o_P(1) \\ &= \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^n E_{\boldsymbol{\gamma}, f}^{(n)} \left[\varphi_f[F^{-1}(U_{\boldsymbol{\gamma}, t})] \right] \end{aligned}$$

$$\begin{aligned}
& \times \frac{[1, \gamma_0 \sum_{i=1}^p \gamma_1^{i-1} \prod_{j=1}^i F^{-1}(U_{\mathcal{Y},t-j})]'}{\gamma_0 + \gamma_1 \gamma_0 \sum_{i=1}^p \gamma_1^{i-1} \prod_{j=1}^i F^{-1}(U_{\mathcal{Y},t-j})} \Big| \mathcal{B}_{\mathcal{Y}}^{(n)} \Big] + o_P(1) \\
\stackrel{(\text{say})}{=} & \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^n E_{\mathcal{Y},f}^{(n)}[\boldsymbol{\psi}_{f,p}(U_{\mathcal{Y},t}, \dots, U_{\mathcal{Y},t-p}) | \mathcal{B}_{\mathcal{Y}}^{(n)}] + o_P(1), \tag{4.13}
\end{aligned}$$

where $U_{\mathcal{Y},t} := F(\xi_{\mathcal{Y},t})$ which is i.i.d. uniform on $[0, 1]$ under $P_{\mathcal{Y},f}^{(n)}$. The quantity (4.13), which involves the expectation of $\boldsymbol{\psi}_{f,p}$, is said to be made of “exact scores”. A more convenient version, i.e., those made of “approximate scores”, is

$$\frac{1}{\sqrt{n-p}} \sum_{t=p+1}^n \boldsymbol{\psi}_{f,p}(V_{\mathcal{Y},t}^{(n)}, \dots, V_{\mathcal{Y},t-p}^{(n)}) + o_P(1), \tag{4.14}$$

where

$$V_{\mathcal{Y},t}^{(n)} := \begin{cases} \tau \cdot \frac{R_{\mathcal{Y},t}^{(n)}}{N_{\mathcal{Y},L}^{(n)} + 1} & \text{if } R_{\mathcal{Y},t}^{(n)} \leq N_{\mathcal{Y},L}^{(n)}, \\ \tau + (1 - \tau) \cdot \frac{R_{\mathcal{Y},t}^{(n)} - N_{\mathcal{Y},L}^{(n)}}{n - N_{\mathcal{Y},L}^{(n)} + 1} & \text{otherwise,} \end{cases}$$

with $N_{\mathcal{Y},L}^{(n)} := \#\{t \in \{1, \dots, n\} | S_{\mathcal{Y},t} = -1\}$. In short, we are first rewriting the residual $\xi_{\mathcal{Y},t}$ as $F^{-1}(U_{\mathcal{Y},t})$ with realization $U_{\mathcal{Y},t}$ of $[0, 1]$ -uniform r.v., then approximating those $U_{\mathcal{Y},t}$ by $V_{\mathcal{Y},t}^{(n)}$ with $\{N_{\mathcal{Y},L}^{(n)}; R_{\mathcal{Y},1}^{(n)}, \dots, R_{\mathcal{Y},n}^{(n)}\}$ given. Since this version is known to be equivalent to the original (4.13), in the remainder of this chapter, $\boldsymbol{\Delta}_{\mathcal{Y},f}^{(n)}$ will be meant to be (4.14), which is computable from the sample. Finally letting

$$\mathbf{W}_t(\mathbf{b}, f) := \frac{[1, b_0 \sum_{i=1}^p b_1^{i-1} \prod_{j=1}^i F^{-1}(V_{\mathbf{b},t-j}^{(n)})]'}{b_0 + b_1 b_0 \sum_{i=1}^p b_1^{i-1} \prod_{j=1}^i F^{-1}(V_{\mathbf{b},t-j}^{(n)})},$$

we have that

$$\begin{aligned}
\overline{\mathbf{W}}^{(n)}(\mathcal{Y}, f) &:= \frac{1}{n-p} \sum_{t=p+1}^n \mathbf{W}_t(\mathcal{Y}, f) \xrightarrow{P} \boldsymbol{\mu}_{\mathbf{W}(\mathcal{Y},f)}, \quad \text{and} \\
\frac{1}{n-p} \sum_{t=p+1}^n \{\mathbf{W}_t(\mathcal{Y}, f) - \overline{\mathbf{W}}^{(n)}(\mathcal{Y}, f)\} \{\mathbf{W}_t(\mathcal{Y}, f) - \overline{\mathbf{W}}^{(n)}(\mathcal{Y}, f)\}' &\xrightarrow{P} \boldsymbol{\Sigma}_{\mathbf{W}(\mathcal{Y},f)},
\end{aligned}$$

for some $\boldsymbol{\mu}_{\mathbf{W}(\mathcal{Y},f)}$ and $\boldsymbol{\Sigma}_{\mathbf{W}(\mathcal{Y},f)}$. We are now ready to state the asymptotic representation of the serial linear sign-and-rank statistics $\boldsymbol{\Delta}_{\mathcal{Y},f}^{(n)}$ of the form (4.14) as follows.

Theorem 4.1 *Let Assumptions 4.1 and 4.2 hold. Then, for any $f \in \mathcal{F}^\tau$ and some suitably large p , we have, under $\boldsymbol{\gamma}$ and g_1 ,*

$$\begin{aligned} \mathbf{A}_{\boldsymbol{\gamma},f}^{(n)} = & \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^n \left[\boldsymbol{\psi}_{f,p}[G_1(\xi_{\boldsymbol{\gamma},t}), \dots, G_1(\xi_{\boldsymbol{\gamma},t-p})] \right. \\ & - E[\boldsymbol{\psi}_{f,p}(U_0, \dots, U_p) | U_0 = G_1(\xi_{\boldsymbol{\gamma},t})] \\ & \left. - \frac{f(1)}{\tau(1-\tau)} (1\{\xi_{\boldsymbol{\gamma},t} \leq 1\} - \tau) \boldsymbol{\mu}_{\mathbf{W}(\boldsymbol{\gamma},f)} \right] + o_P(1). \end{aligned} \quad (4.15)$$

If moreover $\mathbf{0} < \int_{[0,1]^{p+1}} |\boldsymbol{\psi}_{f,p}(u_0, u_1, \dots, u_p)|^{2+\delta} du < \infty$ for some $\delta > 0$, then, for any $f \in \mathcal{F}^\tau$, we have, under $\boldsymbol{\gamma}$ and g_1 ,

$$\mathbf{A}_{\boldsymbol{\gamma},f}^{(n)} \xrightarrow{d} N\left(\mathbf{0}, \Sigma_f(\boldsymbol{\gamma})\right), \quad (4.16)$$

$$\text{where } \Sigma_f(\boldsymbol{\gamma}) := \mathcal{I}_f \Sigma_{\mathbf{W}(\boldsymbol{\gamma},f)} + \frac{f^2(1)}{\tau(1-\tau)} \boldsymbol{\mu}_{\mathbf{W}(\boldsymbol{\gamma},f)} \boldsymbol{\mu}'_{\mathbf{W}(\boldsymbol{\gamma},f)}.$$

Proof See Sect. 4.4.1. □

Remark 4.2 Note that the behavior of $\mathbf{A}_{\boldsymbol{\gamma},f}^{(n)}$, which is based on the reference density f , is now studied under the true density g_1 . Also, due to the invariance properties, the limiting distribution in (4.16) only depends on the reference density f , and not on the true density g_1 . ■

To state the “one-step estimator” of $\boldsymbol{\gamma}$, the asymptotic behavior under $\boldsymbol{\gamma}$ and g of $\mathbf{A}_{\tilde{\boldsymbol{\gamma}}_n, f}^{(n)}$, for some local alternative $\tilde{\boldsymbol{\gamma}}_n$ defined below, is needed. For this purpose, let us denote, in regard to (4.15),

$$\mathbf{T}_{\mathbf{b},f}^{(n)} := \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^n \left(\varphi_f[F^{-1}(G_1(\xi_{\mathbf{b},t}))] (\mathbf{W}_t(\mathbf{b}, f) - \boldsymbol{\mu}_{\mathbf{W}(\mathbf{b},f)}) \right).$$

Now, consider a local alternative, with respect to $\boldsymbol{\gamma}$, of the form $\boldsymbol{\gamma}_n = \boldsymbol{\gamma} + \mathbf{h}_n/\sqrt{n} + o(n^{-1/2})$, with $\mathbf{h}_n \rightarrow \mathbf{h}$ as $n \rightarrow \infty$. Note that this is still within our ULAN framework. Then, under the following assumption, we have the asymptotic linearity.

Assumption 4.3

- (i) $\sup_{x \in \mathbb{R}_+} |xg_1(x)| < \infty$ and $\lim_{x \rightarrow \infty} xg_1(x) = 0$;
- (ii) $\varphi_f^* := \varphi_f \circ F^{-1}$ is nondecreasing and differentiable with its derivative $\dot{\varphi}_f^*$ being uniformly continuous on $[0, 1]$.

Lemma 4.1 *Let Assumptions 4.1 (with Remark 4.1) and 3 hold. Then, for an arbitrary sequence $\boldsymbol{y}_n = \boldsymbol{y} + \boldsymbol{h}_n/\sqrt{n} + o(n^{-1/2})$ with $\boldsymbol{h}_n \rightarrow \boldsymbol{h}$, we have*

$$\boldsymbol{T}_{\boldsymbol{y}_n, f}^{(n)} - \boldsymbol{T}_{\boldsymbol{y}, f}^{(n)} + \begin{pmatrix} \mathcal{I}_{fg_1} \boldsymbol{\Sigma}_{1, \boldsymbol{W}(\boldsymbol{y}, f)} \\ -g_1(1) \boldsymbol{\mu}'_{\boldsymbol{W}(\boldsymbol{y})} \end{pmatrix} \boldsymbol{h} = o_P(1), \quad (4.17)$$

where $\mathcal{I}_{fg_1} := \int_0^1 \varphi_f[F^{-1}(u)] \varphi_{g_1}[G_1^{-1}(u)] du$, $\boldsymbol{\mu}_{\boldsymbol{W}(\boldsymbol{y})} := E[\boldsymbol{W}_t(\boldsymbol{y})]$, and $\boldsymbol{\Sigma}_{1, \boldsymbol{W}(\boldsymbol{y}, f)}$ is such that

$$\boldsymbol{S}_{1, \boldsymbol{W}(\boldsymbol{y}, f)}^{(n)} := \frac{1}{n-p} \sum_{t=p+1}^n \boldsymbol{W}_t(\boldsymbol{y}, f) \boldsymbol{W}_t(\boldsymbol{y})' \xrightarrow{P} \boldsymbol{\Sigma}_{1, \boldsymbol{W}(\boldsymbol{y}, f)}. \quad (4.18)$$

Proof The proof for the first component of (4.17) is similar to that of Koul (1992, Theorem 7.3b.1). So we omit the details (See Taniai (2009)) and only remark that the quantity Q therein corresponds here to

$$\begin{aligned} Q &\equiv \int_0^1 G_1^{-1}(u) g_1(G_1^{-1}(u)) \dot{\varphi}_f^*(u) du \\ &= \left[G_1^{-1}(u) g_1(G_1^{-1}(u)) \varphi_f^*(u) \right]_0^1 - \int_0^1 \frac{d\{G_1^{-1}(u) g_1(G_1^{-1}(u))\}}{du} \varphi_f^*(u) du \\ &= \int_0^1 \left\{ -1 - \frac{G_1^{-1}(u) g_1'(G_1^{-1}(u))}{g_1(G_1^{-1}(u))} \right\} \varphi_f^*(u) du = \mathcal{I}_{fg_1}, \end{aligned}$$

by Assumption 4.3-(i). Also, the second component of (4.17) follows from an application of Koul (1992, Corollary 7.2.2). \square

Finally in preparing for the one-step estimator, a “discretization trick” is to be introduced.

Definition 4.1 For any sequence of estimators $\hat{\theta}_n$, the discretized estimator $\bar{\theta}$ is defined to be the nearest vertex of $\{\theta : \theta = \frac{1}{\sqrt{n}}(i_1, i_2, \dots, i_k)', i_j : \text{integers}\}$.

We denote by $\bar{\boldsymbol{y}}_n (= \bar{\boldsymbol{y}}_n(\tau))$ a locally discretized version of $\hat{\boldsymbol{y}}_n(\tau)$ which is defined at (4.3). Recall here that Corollary 4.1 ensures the \sqrt{n} -consistency of $\hat{\boldsymbol{y}}_n(\tau)$. So, if we write as $\hat{\boldsymbol{y}}_n(\tau) = \boldsymbol{y} + \boldsymbol{k}_n/\sqrt{n}$ then \boldsymbol{k}_n is a sequence which is bounded in probability. The great advantage of the discretization is that it allows us to regard \boldsymbol{k}_n further as a nonrandom sequence (See Kreiss (1987, Lemma 4.4), or Linton (1993) etc.), i.e., $\bar{\boldsymbol{y}}_n = \boldsymbol{y} + \boldsymbol{h}_n/\sqrt{n}$ as in Lemma 4.1.

Now, using the rank-based central sequence $\boldsymbol{\Delta}_{\bar{\boldsymbol{y}}_n, f}^{(n)}$ based on a \sqrt{n} -consistent discretized estimator $\bar{\boldsymbol{y}}_n$, we construct some $\tilde{\boldsymbol{y}}_n$ for which $\sqrt{n}(\tilde{\boldsymbol{y}}_n - \boldsymbol{y})$ behaves well. If $\sqrt{n}(\tilde{\boldsymbol{y}}_n - \bar{\boldsymbol{y}}_n)$ is of the form $A_n^{-1} \boldsymbol{\Delta}_{\bar{\boldsymbol{y}}_n, f}^{(n)}$ for some matrix A_n , then we have by (4.17), under \boldsymbol{y} and g_1 ,

$$\begin{aligned}
\sqrt{n}(\tilde{\boldsymbol{y}}_n - \boldsymbol{\gamma}) &= \sqrt{n}(\tilde{\boldsymbol{y}}_n - \bar{\boldsymbol{y}}_n) + \sqrt{n}(\bar{\boldsymbol{y}}_n - \boldsymbol{\gamma}) = A_n^{-1} \mathbf{\Delta}_{\tilde{\boldsymbol{y}}_n, f}^{(n)} + \boldsymbol{h}_n \\
&= A_n^{-1} \left(\mathbf{\Delta}_{\boldsymbol{\gamma}, f}^{(n)} - \mathcal{I}_{fg_1} \Sigma_{1, \mathbf{W}(\boldsymbol{\gamma}, f)} \boldsymbol{h} - \frac{f(1)g_1(1)}{\tau(1-\tau)} \boldsymbol{\mu}_{\mathbf{W}(\boldsymbol{\gamma}, f)} \boldsymbol{\mu}'_{\mathbf{W}(\boldsymbol{\gamma})} \boldsymbol{h} \right) + \boldsymbol{h}_n + o_P(1).
\end{aligned} \tag{4.19}$$

This, together with $\boldsymbol{h}_n \rightarrow \boldsymbol{h}$, suggests us to choose A_n to be consistent for

$$\Sigma_{fg_1}(\boldsymbol{\gamma}) := \mathcal{I}_{fg_1} \Sigma_{1, \mathbf{W}(\boldsymbol{\gamma}, f)} + \frac{f(1)g_1(1)}{\tau(1-\tau)} \boldsymbol{\mu}_{\mathbf{W}(\boldsymbol{\gamma}, f)} \boldsymbol{\mu}'_{\mathbf{W}(\boldsymbol{\gamma})}, \tag{4.20}$$

so that (4.19) becomes asymptotically equivalent to $A_n^{-1} \mathbf{\Delta}_{\boldsymbol{\gamma}, f}^{(n)}$. For this purpose, let us remind (4.18) and define

$$M_{1, \mathbf{W}(\boldsymbol{\gamma}, f)}^{(n)} := \left(\frac{1}{n-p} \sum_{t=p+1}^n \mathbf{W}_t(\boldsymbol{\gamma}, f) \right) \left(\frac{1}{n-p} \sum_{t=p+1}^n \mathbf{W}_t(\boldsymbol{\gamma}) \right)'.$$

The consistencies $S_{1, \mathbf{W}(\tilde{\boldsymbol{y}}_n, f)}^{(n)} \xrightarrow{P} \Sigma_{1, \mathbf{W}(\boldsymbol{\gamma}, f)}$ and $M_{1, \mathbf{W}(\tilde{\boldsymbol{y}}_n, f)}^{(n)} \xrightarrow{P} \boldsymbol{\mu}_{\mathbf{W}(\boldsymbol{\gamma}, f)} \boldsymbol{\mu}'_{\mathbf{W}(\boldsymbol{\gamma})}$ can be verified in the manner of Linton (1993, Theorem 2). Now, we define the one-step estimator $\tilde{\boldsymbol{y}}_n$ as follows.

Definition 4.2 The rank-based one-step estimator of $\boldsymbol{\gamma}$ starting from the \sqrt{n} -consistent and locally discrete estimator $\bar{\boldsymbol{y}}_n$ and based on reference density $f \in \mathcal{F}^\tau$ is defined as

$$\begin{aligned}
\tilde{\boldsymbol{y}}_{n, f} &:= \bar{\boldsymbol{y}}_n + \hat{\Sigma}_{fg_1}(\bar{\boldsymbol{y}}_n)^{-1} \frac{\mathbf{\Delta}_{\bar{\boldsymbol{y}}_n, f}^{(n)}}{\sqrt{n}}, \quad \text{with} \\
\hat{\Sigma}_{fg_1}(\bar{\boldsymbol{y}}_n) &:= \hat{\mathcal{I}}_{fg_1} S_{1, \mathbf{W}(\bar{\boldsymbol{y}}_n, f)}^{(n)} + \frac{\tau}{1-\tau} \cdot \frac{f(1)}{-\tau} \cdot \hat{\boldsymbol{\mu}}_{\varphi_{g_1}, L} M_{1, \mathbf{W}(\bar{\boldsymbol{y}}_n, f)}^{(n)},
\end{aligned} \tag{4.21}$$

where $\hat{\mathcal{I}}_{fg_1}$ and $\hat{\boldsymbol{\mu}}_{\varphi_{g_1}, L}$ are consistent estimates of \mathcal{I}_{fg_1} and

$$\boldsymbol{\mu}_{\varphi_{g_1}, L} := E[\varphi_{g_1}[G_1^{-1}(U)] | U \leq \tau] \quad (= -g_1(1)/\tau),$$

respectively.

The consistent estimators $\hat{\mathcal{I}}_{fg_1}$ and $\hat{\boldsymbol{\mu}}_{\varphi_{g_1}, L}$ can be obtained in the manner of Hallin et al. (2006a, Sect. 4.2), which will be done without the kernel estimation of g_1 . (See Taniai (2009) for the details.)

If Assumptions 4.1 and 4.2 hold then we have, under $\boldsymbol{\gamma}$ and g_1 ,

$$\sqrt{n}(\tilde{\boldsymbol{y}}_{n, f} - \boldsymbol{\gamma}) \xrightarrow{d} N\left(\mathbf{0}, \Sigma_{fg_1}^{-1}(\boldsymbol{\gamma}) \Sigma_f(\boldsymbol{\gamma}) \Sigma_{fg_1}^{-1}(\boldsymbol{\gamma})\right), \tag{4.22}$$

as a consequence of Theorem 4.1. Also, similarly to Hallin et al. (2008), this $\tilde{\boldsymbol{\gamma}}_{n,f}$ is shown to attain the semiparametric lower bound.

Proposition 4.2 *Under Assumptions 4.1–4.3, the one-step estimator $\tilde{\boldsymbol{\gamma}}_{n,f}$ defined by (4.21) for $\boldsymbol{\gamma}$ is semi-parametrically efficient at $f = g_1$.*

Proof The proof is similar to that of Hallin et al. (2006b, Lemma 2), i.e., the parametric submodels are constructed as in Bickel et al. (1998, Example 3.2.1). We omit the details. \square

4.2.3 Numerical Studies

Based on the result of Lemma 4.1 and Le Cam’s third lemma (See van der Vaart (1998)), we can have the convergence-in-law of $\boldsymbol{T}_{\boldsymbol{\gamma}_0}^{(n)}$ under the local alternatives $\boldsymbol{\gamma}_n = \boldsymbol{\gamma}_0 + \boldsymbol{h}/\sqrt{n}$ (and the “true” density g_1). Namely, we have, under the local alternatives $\boldsymbol{\gamma}_n$,

$$\underline{\boldsymbol{\Delta}}_{\boldsymbol{\gamma}_0,f}^{(n)} \xrightarrow{d} N\left(\Sigma_{f,g_1}(\boldsymbol{\gamma}_0)\boldsymbol{h}, \Sigma_f(\boldsymbol{\gamma}_0)\right), \quad \text{under } P_{\boldsymbol{\gamma}_n,g_1} \quad (4.23)$$

and hence the distribution (under $\boldsymbol{\gamma}_n$) of Lagrange multiplier (LM) statistics

$$LM_f^{(n)} := \underline{\boldsymbol{\Delta}}_{\boldsymbol{\gamma}_0,f}^{(n)'} \Sigma_f^{-1}(\boldsymbol{\gamma}_0) \underline{\boldsymbol{\Delta}}_{\boldsymbol{\gamma}_0,f}^{(n)} \quad (4.24)$$

is asymptotically a noncentral chi-square with 2 degrees of freedom and noncentrality parameter

$$\boldsymbol{h}' \Sigma'_{f,g_1}(\boldsymbol{\gamma}_0) \Sigma_f^{-1}(\boldsymbol{\gamma}_0) \Sigma_{f,g_1}(\boldsymbol{\gamma}_0) \boldsymbol{h}. \quad (4.25)$$

Note that the distribution of $LM_f^{(n)}$ under the null $\boldsymbol{\gamma}_0$ is asymptotically a (central) chi-square with 2 degrees of freedom, χ_2^2 . Thus, the tests that reject the null hypothesis $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ if $LM_f^{(n)}$ exceeds $F_{\chi_2^2}^{-1}(1 - \alpha)$ are asymptotically of level α .

Assume now that we have two such level α tests $A^{(1)}$ and $A^{(2)}$ based on the statistics $LM_{f_1}^{(n_1)}$ and $LM_{f_2}^{(n_2)}$. Also, let us assume that their noncentrality parameter (under g_1) can be expressed in a quadratic form of \boldsymbol{h} , and denote them as $\lambda_{g_1}(f_i, \boldsymbol{h}) = \boldsymbol{h}' B_i \boldsymbol{h}$, $i = 1, 2$. Then the two tests have the same limiting power against the same sequence of alternatives if

$$\frac{\boldsymbol{h}_1}{\sqrt{n_1}} = \frac{\boldsymbol{h}_2}{\sqrt{n_2}} \quad (4.26)$$

and

$$\lambda_{g_1}(f_1, \boldsymbol{h}_1) = \lambda_{g_1}(f_2, \boldsymbol{h}_2). \quad (4.27)$$

So, from (4.26) and (4.27), we see that the asymptotic relative efficiency (ARE, Pitman efficiency) of $A^{(2)}$ with respect to $A^{(1)}$ under g_1 should be

$$\text{ARE}_{g_1}(A^{(2)}, A^{(1)}) := \lim \frac{n_1}{n_2} = \frac{\lambda_{g_1}(f_2, \mathbf{h}_1)}{\lambda_{g_1}(f_1, \mathbf{h}_1)}. \quad (4.28)$$

Since this involves $\mathbf{h}_1 = (h_{11}, h_{12})'$, the unique answer regarding ARE of $A^{(2)}$ w.r.t. $A^{(1)}$ may not be possible. However, as suggested in Puri and Sen (1971, Sect. 3.8.3), the supremum and the infimum of $\text{ARE}_{g_1}(A^{(2)}, A^{(1)})$ can be obtained as the largest and smallest eigenvalues of $B_2 B_1^{-1}$ by the Courant-Fischer minimax theorem. We will show these supremum and infimum later. Still, it is worthwhile, in fact, to observe also the values of (4.28) for $\mathbf{h}_1 = (1, 0)'$ and for $\mathbf{h}_1 = (0, 1)'$. This is because they would correspond to the testing, not both but, one of the parameters while keeping the other parameter fixed. In order to make those displays, further description of our setting is now explained as follows.

Let us first define the reference and/or underlying densities we consider. Below, we see some numerical results concerning ARE of LM tests for a family of models, which contains the squared series of ARCH model (4.1). As ARCH models require the (unsquared) innovation Z_t to be i.i.d.(0, 1), let us think of generating the density of squared innovations from such i.i.d.(0, 1) densities. So, let us suppose that $\{Z_t\}$ is a sequence of i.i.d.(0, 1) random variables with continuous distribution function F_Z of the symmetric density f_Z . Then

$$F_{Z^2}(z) = P\{Z_t^2 \leq z\} = \begin{cases} 0, & z \leq 0, \\ 2F_Z(\sqrt{z}) - 1, & z > 0. \end{cases} \quad (4.29)$$

We shall now compute (4.29) in the following particular choices of the innovation distribution function F_Z , and define families of densities, which contain such particular choices within.

(i) F_Z being Normal:

$$\begin{aligned} F_N(z) &= \Phi(z) = \int_{-\infty}^z \phi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt, \\ \Rightarrow F_{N^2}(z) &= 2\Phi(\sqrt{z}) - 1, \quad f_{N^2}(z) = \frac{\phi(\sqrt{z})}{\sqrt{z}}. \end{aligned}$$

In this case, we have $f_{N^2} \in \mathcal{F}^{0.6827}$ so that the commonly performed analysis of β corresponds to the situation $\tau = 0.6827$. But our analysis of γ ($= \gamma(\tau)$) concerns more generally the class of the following densities for $\tau \in (0, 1)$:

$$f_{N^2}^\tau(z) := f_{N^2}(z \cdot F_{N^2}^{-1}(\tau)) \cdot F_{N^2}^{-1}(\tau) \left(= \frac{dF_{N^2}(z \cdot F_{N^2}^{-1}(\tau))}{dz} \right) \in \mathcal{F}^\tau.$$

Table 4.1 The values of $(E[\xi_0^2])^{-1/2}$: In order to maintain the stationarity of the process, γ_1 ($= \gamma_1(\tau)$) should be smaller than these values

		τ								
		0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
g_1	$f_{N^2}^\tau$	0.02	0.06	0.12	0.20	0.33	0.50	0.76	1.20	2.22
	$f_{DE^2}^\tau$	0.005	0.02	0.04	0.07	0.13	0.22	0.39	0.73	1.83
	$f_{LGT^2}^\tau$	0.01	0.04	0.08	0.14	0.23	0.36	0.56	0.94	1.99

(ii) F_Z being Double Exponential (bilateral exponential, Laplace):

$$F_{DE}(z) = DE\left(z; 0, \frac{1}{\sqrt{2}}\right) = \frac{\sqrt{2}}{2} \int_{-\infty}^z e^{-\sqrt{2}|t|} dt,$$

$$\Rightarrow f_{DE^2}(z) = \frac{1}{\sqrt{2z}} \exp(-\sqrt{2z}).$$

In this case, while $f_{DE^2} \in \mathcal{F}^{0.7569}$, we apply the class of densities

$$f_{DE^2}^\tau(z) := f_{DE^2}(z \cdot F_{DE^2}^{-1}(\tau)) \cdot F_{DE^2}^{-1}(\tau) \in \mathcal{F}^\tau,$$

(iii) F_Z being Logistic:

$$F_{LGT}(z) = \text{Logistic}\left(z; 0, \frac{\sqrt{3}}{\pi}\right) = \left\{1 + \exp\left(-\frac{\pi}{\sqrt{3}}z\right)\right\}^{-1},$$

$$\Rightarrow f_{LGT^2}(z) = \frac{\pi}{\sqrt{3z}} \frac{\exp(-\sqrt{\pi^2 z/3})}{\{1 + \exp(-\sqrt{\pi^2 z/3})\}^2}.$$

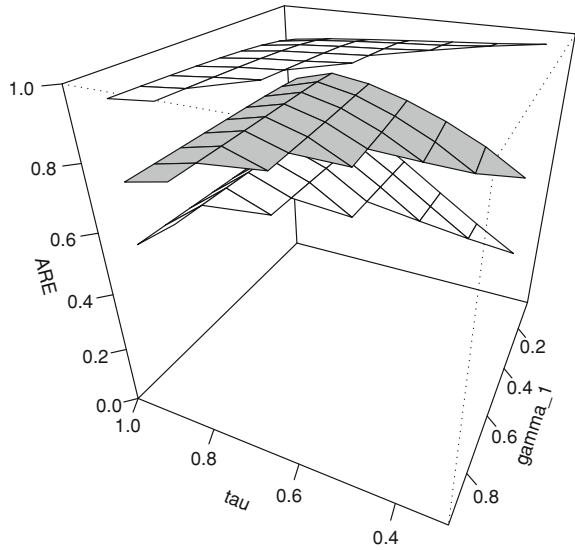
Again, while $f_{LGT^2} \in \mathcal{F}^{0.7196}$, we apply

$$f_{LGT^2}^\tau(z) := f_{LGT^2}(z \cdot F_{LGT^2}^{-1}(\tau)) \cdot F_{LGT^2}^{-1}(\tau) \in \mathcal{F}^\tau.$$

Being provided the exact form of $g_1 \in \mathcal{F}^\tau$, we have the collection of γ_1 , which makes the process stationary. According to the sufficient condition of Remark 4.1, the following Table 4.1 tells that the availability of γ_1 are limited especially for small τ .

Now, we start with calculating ARE of rank based LM statistics (4.24) with respect to a LM statistic which is based on the parametric ‘‘Gaussian’’ (: more accurately, squared Gaussian $f_{N^2}^\tau$) score function. In order to ensure the asymptotic results for the latter quantity, the Gaussian parametric score, here we suppose that the underlying density is being correctly specified, i.e., $g_1 = f_{N^2}^\tau$. Consequently, our rank-based tests, or any tests, can not outperform such a correctly specified parametric score test. But still we use this as a benchmark, and observe the performances of several

Fig. 4.1 The *top* and the *bottom* surfaces are sup and inf of $e_{g_1}(\text{RB}(f_{N_2}^\tau), P(g_1))$, the ARE of rank-based test (with reference density $f_{N_2}^\tau$) w.r.t. the correctly-specified parametric score test under the true density g_1 . Those sup and inf are taken among all the direction \mathbf{h}_1 of the local perturbation. (The values for a specific direction $\mathbf{h}_1 = (0, 1)'$ is provided as the shadowed surface.) In this case, our setting $g_1 = f_{N_2}^\tau$ means the equality of the reference density and the true density. Still, we may observe that the supremum of ARE seems to be not attaining 1



reference densities against this benchmark. To summarize, the quantity which we first concern is

$$\begin{aligned}
 & \text{ARE}_{g_1}(\text{Rank-Based with } f, \text{Parametric with } g_1) \\
 &= \frac{\mathbf{h}' \Sigma'_{f_{g_1}}(\boldsymbol{\gamma}_0) \Sigma_f^{-1}(\boldsymbol{\gamma}_0) \Sigma_{f_{g_1}}(\boldsymbol{\gamma}_0) \mathbf{h}}{\mathbf{h}' (\mathcal{I}_{g_1}(\Sigma_{\mathbf{W}}(\boldsymbol{\gamma}, g_1) + M_{\mathbf{W}}(\boldsymbol{\gamma}, g_1))) \mathbf{h}} \\
 &\stackrel{(\text{say})}{=} e_{g_1}(\text{RB}(f), P(g_1)), \tag{4.30}
 \end{aligned}$$

with $g_1 = f_{N_2}^\tau$ and several f are chosen for comparison. As discussed in (4.28), we can calculate the supremum and infimum of $e_{g_1}(\text{RB}(f), P(g_1))$. Figure 4.1 shows the case of $f \equiv f_{N_2}^\tau$, and there the values of $e_{g_1}(\text{RB}(f), P(g_1))$ for $\mathbf{h}_1 = (0, 1)'$ are also shown. There, the sample size is $n = 500$ and the finite approximation is done with $p = 50$. Also, there we set γ_0 to be $1 - \gamma_1$ in order to keep the unconditional variance being the same regardless of the changes of ARCH parameters (this does not make any substantial effect on the result).

In fact, the cases where $f \equiv f_{\text{DE}^2}^\tau$ and $f \equiv f_{\text{LGT}^2}^\tau$ yield the similar result as shown in Fig. 4.1 (See Taniai (2009)). According to this observation, it seems that the value of the parameter γ_1 does not have substantial effect on the result. On the other hand, the dependence on τ is evident. Those τ 's being too small or too big would makes our rank-based inference difficult comparing to the parametric method. Also, even for those supremum values, they seem to be not attaining 1. This means that there is a definite gap between the semiparametric efficiency (the best feasible solution), and the correctly specified parametric efficiency (the best possible but infeasible solution).

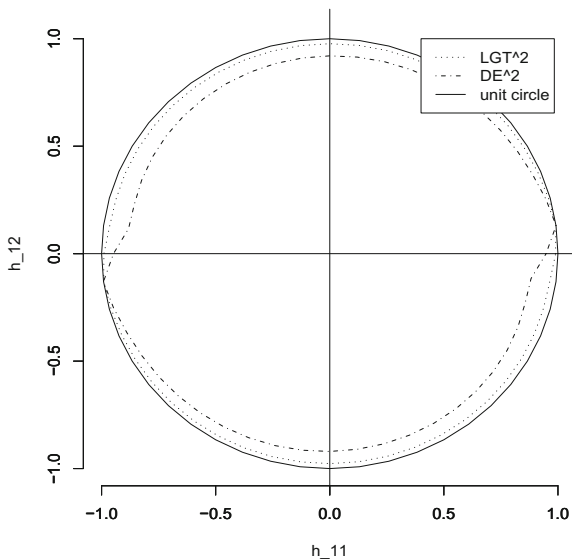


Fig. 4.2 The ARE among rank based tests, $e_{g_1}(\text{RB}(f), \text{RB}(g_1))$ with $g_1 = f_{N_2}^\tau$ for $\tau = 0.35$ and $\gamma_1 = 0.1$. Since here the underlying density is $g_1 = f_{N_2}^\tau$, the unit circle represents the rank based test with the reference density based on Gaussian

Next, let us have a close look to the ARE among those rank based procedures. That is, we observe the value of

$$\begin{aligned} & \text{ARE}_{g_1}(\text{Rank-Based with } f, \text{Rank-Based with } g_1) \\ &= \frac{\mathbf{h}' \Sigma'_{fg_1}(\boldsymbol{\gamma}_0) \Sigma_f^{-1}(\boldsymbol{\gamma}_0) \Sigma_{fg_1}(\boldsymbol{\gamma}_0) \mathbf{h}}{\mathbf{h}' \Sigma_{g_1}(\boldsymbol{\gamma}_0) \mathbf{h}} \\ &\stackrel{(\text{say})}{=} e_{g_1}(\text{RB}(f), \text{RB}(g_1)), \end{aligned} \tag{4.31}$$

with several choices of f and g_1 for comparison. Figures 4.2 and 4.3 show $e_{g_1}(\text{RB}(f), \text{RB}(g_1))$ for each direction $\mathbf{h}_1 = (h_{11}, h_{12})'$ with $g_1 = f_{N_2}^\tau$ for different choices of τ and γ_1 . This explain how much of efficiency will be lost according to our choices for the reference densities, under the case where true density $g_1 = f_{N_2}^\tau$ (i.e., Gaussian-based case). As for the other cases, i.e., Double Exponential based case ($g_1 \equiv f_{\text{DE}^2}^\tau$) and Logistic based case ($g_1 \equiv f_{\text{LGT}^2}^\tau$), we refer to Taniai (2009).

According to the figures, it can be seen that the rank-based tests are robust against the misspecification of the underlying density. So, we may conclude as follows: Once we enter the semi-parametrics, the rank-based inference in this chapter, the price for misspecification can be reduced. Also, the reason why we may need to enter is due to the fact that we accepted, in the first place, our ignorance or indifference about the underlying densities.

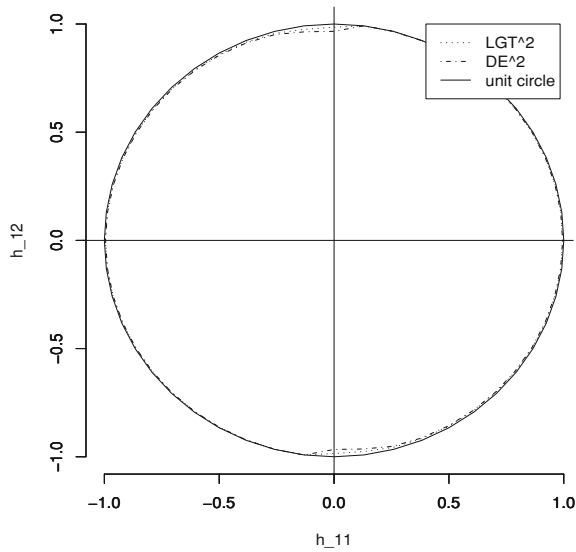


Fig. 4.3 (The same as Figure 4.2, but for $\tau = 0.85$ and $\gamma_1 = 0.9$.)

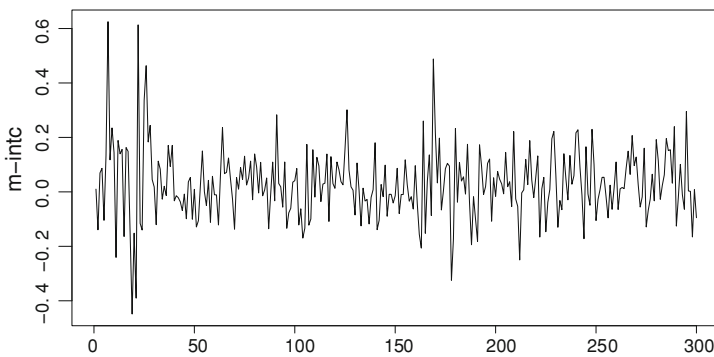


Fig. 4.4 Time plot of the monthly log stock returns of Intel from January 1973 to December 1997

Finally, let us see the real data application with the monthly log stock returns of Intel Corporation from January 1973 to December 1997, which consists of $n = 300$ data points (Fig. 4.4). This data can be found in Tsay (2002, Example 3.1), and studied to behave as ARCH(1) under the Gaussian assumption. For example, denoting X_t for the squared value of log stock returns at time t , ordinary Quantile Regression analysis for $\tau = 0.7$ becomes,

$$\hat{F}_{X_t}^{-1}(0.7|x_{t-1}) = 0.0137838 + 0.1379812x_{t-1}.$$

Table 4.2 MSE of conditional quantiles: This table shows, for each cases of g_1 , how the ordinary QR estimator (the row “QR”) and our rank-based estimators (the row “OS $_{N^2}$ ”, “OS $_{DE^2}$ ”, “OS $_{LGT^2}$ ”) performs in estimation

		τ				
		0.1	0.3	0.5	0.7	0.9
$g_1 = f_{N^2}^\tau$	QR	5.00×10^{-6}	1.49×10^{-4}	2.11×10^{-3}	7.41×10^{-3}	5.73×10^{-2}
	OS $_{N^2}$	14.9×10^{-6}	1.44×10^{-4}	1.12×10^{-3}	4.67×10^{-3}	4.89×10^{-2}
	OS $_{DE^2}$	14.9×10^{-6}	1.45×10^{-4}	1.13×10^{-3}	4.77×10^{-3}	5.12×10^{-2}
	OS $_{LGT^2}$	14.9×10^{-6}	1.46×10^{-4}	1.14×10^{-3}	4.77×10^{-3}	5.01×10^{-2}
$g_1 = f_{DE^2}^\tau$	QR	9.91×10^{-6}	2.95×10^{-4}	4.19×10^{-3}	14.6×10^{-3}	11.3×10^{-2}
	OS $_{N^2}$	15.2×10^{-6}	1.18×10^{-4}	1.10×10^{-3}	5.46×10^{-3}	11.0×10^{-2}
	OS $_{DE^2}$	15.2×10^{-6}	1.19×10^{-4}	1.10×10^{-3}	5.38×10^{-3}	10.5×10^{-2}
	OS $_{LGT^2}$	15.1×10^{-6}	1.18×10^{-4}	1.09×10^{-3}	5.39×10^{-3}	10.6×10^{-2}

QR corresponds to ordinary quantile regression estimator (4.3), and OS $_f$ corresponds to our rank-based one-step estimator (4.21) with reference density f and $p = 5$. N Normal, DE Double Exponential, LGT Logistic. (The case $g_1 = f_{LGT^2}^\tau$ is similar to that of $f_{DE^2}^\tau$ so we omitted.) For example, the category “ $g_1 = f_{N^2}^\tau$ ” exhibits the case where the simulation is generated from the density $f_{N^2}^\tau$, and hence the estimators “OS $_{DE^2}$ ” and “OS $_{LGT^2}$ ” must be misspecifying in fact. As a consequence their performances are worse than “OS $_{N^2}$,” but we note that even these misspecifying estimators can perform better than the ordinary QR estimator

In order to check the efficiency gain of our one-step estimator, we compute the mean squared errors (MSE) of the conditional quantile of the squared series. That is, we compute

$$\begin{aligned} \text{MSE}^{(\text{QR})} &:= E\sqrt{n}\{\hat{F}_{X_t}^{-1}(\tau|x_{t-1}) - F_{X_t}^{-1}(\tau|x_{t-1})\}^2 \\ &= E\{[\sqrt{n}(X'_{t-1}\hat{\gamma}_n(\tau) - \gamma(\tau))\}^2] \\ &\approx \frac{1}{n} \sum_{t=1}^n \text{tr}[X_{t-1}X'_{t-1} \frac{\tau(1-\tau)}{g_1^2(1)} D_{1,\hat{\gamma}_n(\tau)}^{-1} D_0 D_{1,\hat{\gamma}_n(\tau)}^{-1}] \end{aligned}$$

and

$$\begin{aligned} \text{MSE}_f^{(\text{OS})} &= E\{[\sqrt{n}(X'_{t-1}\tilde{\gamma}_{n,f}(\tau) - \gamma(\tau))\}^2] \\ &\approx \frac{1}{n} \sum_{t=1}^n \text{tr}[X_{t-1}X'_{t-1} \Sigma_{fg_1}^{-1}(\tilde{\gamma}_{n,f}(\tau)) \Sigma_f(\tilde{\gamma}_{n,f}(\tau)) \Sigma_{fg_1}^{-1}(\tilde{\gamma}_{n,f}(\tau))]. \end{aligned}$$

The computed MSEs for several choice of reference density f and the true density g_1 are shown in Table 4.2. Note that, there is no way to find the true density g_1 since this is the real data.

From Table 4.2, we find that our rank-based one-step estimator improves efficiency for moderate choice of τ . Unfortunately, for very small τ (e.g., $\tau = 0.1$) our estimator failed to improve efficiency. This may be related with the remark from Table 4.1 since the estimated value is found to be $\hat{\gamma}_{n,1}(0.1) = 0.012$. That is, our efficiency gain is realized only when the model (4.5) maintains the stationarity, not only the model

(4.2). Still, it is remarkable that, for moderate value of τ , the improvement method of ours is achieved even without the knowledge of true density g_1 . In addition, the robustness against the misspecification (as noted in Figs. 4.2 and 4.3) is observed in this analysis as well.

4.3 Asymptotics of Realized Volatility with Non-Gaussian ARCH(∞) Microstructure Noise

In order to control the risk of financial assets, good estimation of the conditional variance, volatility, is indispensable. For this purpose, the quantity known as RV is having attention, especially when the frequency of the sampling is high (cf. Dacorogna et al. (2001)). Namely, let X_t denote the log-price of an asset, and follow some process such as $dX_t = \mu_t dt + \sigma_t dW_t$ which we will discuss later in (4.35) of Sect. 4.3.1. Then the RV (or sometimes called also as Realized Variance) over time $[0, T]$ is defined as the sum of the frequently sampled squared returns (see Andersen et al. (2003); McAleer and Medeiros (2008)):

$$[X, X]_T := \sum_{t=1}^T (X_t - X_{t-1})^2. \quad (4.32)$$

This quantity estimates the Integrated Volatility (IV) defined as

$$\langle X, X \rangle_T := \int_0^T \sigma_t^2 dt, \quad (4.33)$$

which corresponds to the quadratic variation in the Stochastic Volatility model (see Barndorff-Nielsen and Shephard (2002)).

But here we need to pay attention also to the existence of market microstructure. That is, it is known that the estimator $[X, X]_T$ does not converge as sampling frequency increases when there is market microstructure dynamics behind. Here, the market microstructure can arise through the bid-ask bounce, asynchronous trading, infrequent trading, and price discreteness, among other factors (see Madhavan (2000); O'Hara (1997)). Still, Zhang et al. (2005) proposed some subsample-based estimator to estimate the integrated volatility and proved its consistency in the presence of i.i.d. market microstructure noise. Their estimator is defined by adjusting the bias from the following preliminary estimator:

$$[X, X]_T^{(\text{avg})} := \frac{1}{K} \sum_{k=1}^K [X, X]_T^{(k)} \quad (4.34)$$

where

$$[X, X]_T^{(k)} := \sum_{t=1}^{T_k} (X_{k-1+tK} - X_{k-1+(t-1)K})^2, \quad \text{with}$$

$$T_k := \lfloor \frac{T - k + 1}{K} \rfloor. \quad (\lfloor x \rfloor \text{ is the integer part of a real number } x)$$

Here, K is prespecified and the asymptotics are studied under suitably chosen T and K , as we will see also in Theorem 4.3 below. Namely, this estimator $[X, X]_T^{(\text{avg})}$ firstly uses the time scale, for example, 10 min ($K = 10$) to measure from 9:30 with the intervals 9:30–9:40, 9:40–9:50, ..., then 9:31–9:41, 9:41–9:51, and so on. Then those will be averaged so that we make use of the full data whose time scale is, say, 1 min. By employing a bias-adjusting to this $[X, X]_T^{(\text{avg})}$, the resulting consistent estimator is called the Two Time Scales estimator of the integrated volatility, and later renamed as Two Scales Realized Volatility (TSRV) in Aït-Sahalia et al. (2011). This latter reference, Aït-Sahalia et al. (2011), provides the results for dependent microstructure noise which is stationary and strong mixing with the mixing coefficients decaying exponentially, together with a moment condition for the microstructure. Also, they include explicit example of the AR(1) microstructure.

Extending these serial-dependent microstructure noise, in this section, we investigate the asymptotic distribution of realized volatility and its subsample estimator under the case that microstructure noise follows an ARCH model of order ∞ (ARCH(∞)) (Engle (1982); Giraitis et al. (2000)). In fact, empirical evidence shows that the process of market microstructure usually has the conditional heteroskedasticity (see Madhavan (2000), and we will discuss on this issue in Remark 4.4 below). The usual ARCH model itself may be included by the situation of Aït-Sahalia et al. (2011), since some conditions make it satisfy the strong mixing condition with geometric rate (see Francq and Zakoïan (2006); Meitz and Saikkonen (2008)). The distinction between our setting and that of Aït-Sahalia et al. (2011) will be described in Remark 4.3. Also, there is a merit when we specify the model for the microstructure, such as forecasting, etc. We also provide the numerical results, which show the MSE of lag averaging realized volatility is smaller than that of usual RV even if noise follows ARCH(∞).

4.3.1 Model, Estimators, and Main Results

In this section, first, we describe stochastic models for price, and explain the integrated and realized volatilities. Let (Ω, \mathcal{F}, P) be a probability space, and let $\{\mathcal{F}_t : 0 \leq t \leq \tau\}$ be a nondecreasing family of σ -algebra satisfying $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}, s \leq t$. We suppose that the logarithmic price of a given asset $X = \{X_t, \mathcal{F}_t : 0 \leq t \leq \tau\}$ follows a $\{\mathcal{F}_t\}$ -adapted continuous time diffusion process, that is, Ito's process:

$$dX_t = \mu_t dt + \sigma_t dW_t \quad (4.35)$$

where μ_t is the drift component, σ_t is the instantaneous volatility and $\{W_t, \mathcal{F}_t\}$ is the $\{\mathcal{F}_t\}$ -adapted standard Brownian motion process.

Assumption 4.4 μ_t and σ_t are $\{\mathcal{F}_t\}$ -adapted for each $0 \leq t \leq \tau$, and satisfy

$$P \left\{ \int_0^\tau |\mu_t| dt < \infty \right\} = 1, \quad (4.36)$$

$$P \left\{ \int_0^\tau \sigma_t^2 dt < \infty \right\} = 1. \quad (4.37)$$

In real trading, we can only use discretized data of assets though price process is assumed to follow continuous process (4.35). In what follows, we denote the t -th observation of i -th day by $X_{i+\frac{t}{M}}$, and write the true (latent) return as

$$r_{i+\frac{t}{M}} := X_{i+\frac{t}{M}} - X_{i+\frac{t-1}{M}},$$

where $1/M$ is the length of sampling interval. The Integrated Volatility was introduced to describe the true volatility of continuous latent process. The Integrated Volatility of the i -th day, $\langle X, X \rangle_{i,i+1}$, is defined as

$$\langle X, X \rangle_{i,i+1} := \int_i^{i+1} \sigma_t^2 dt. \quad (4.38)$$

A good estimator for $\langle X, X \rangle_{i,i+1}$ is the Realized Volatility of i -th day defined by

$$[X, X]_{i,i+1} := \sum_{t=1}^M r_{i+\frac{t}{M}}^2. \quad (4.39)$$

Then, the following results are given by Bandi and Russell (2008).

Proposition 4.3 (Bandi and Russell (2008)) *Under Assumption 4.4,*

$$p\text{-}\lim_{M \rightarrow \infty} ([X, X]_{i,i+1} - \langle X, X \rangle_{i,i+1}) = 0, \quad (4.40)$$

$$\sqrt{M}([X, X]_{i,i+1} - \langle X, X \rangle_{i,i+1}) \xrightarrow{d} \mathcal{MN}(0, 2Q_{i,i+1}), \quad M \rightarrow \infty, \quad (4.41)$$

where $\mathcal{MN}(0, 2Q_{i,i+1})$ is the mixed normal distribution with mean 0 and random variance $2Q_{i,i+1}$, where

$$Q_{i,i+1} := \int_i^{i+1} \sigma_t^4 dt. \quad (4.42)$$

Now, let us consider the case that $Y_{i+\frac{t}{M}}$ is a perturbed observation with microstructure noise corresponding to the true (latent) log price $X_{i+\frac{t}{M}}$. The relation can be modeled as

$$Y_{i+\frac{t}{M}} = X_{i+\frac{t}{M}} + \epsilon_{i+\frac{t}{M}}, \quad (4.43)$$

where $\{\epsilon_t\}$ is a sequence of microstructure noise. In this section we assume that the series $\{\epsilon_{i+\frac{t}{M}}, t \in \mathbf{Z}\}$ follows an autoregressive conditionally heteroskedastic process of order ∞ (ARCH(∞)), i.e.,

$$\epsilon_{i+\frac{t}{M}} = \left(a_0 + \sum_{j=1}^{\infty} a_j \epsilon_{i+\frac{t-j}{M}}^2 \right)^{1/2} z_{i+\frac{t}{M}}, \quad (4.44)$$

where $a_0 > 0$, $a_j \geq 0$ ($j \geq 1$), and z_t 's are i.i.d. random variables.

Remark 4.3 Zhang et al. (2005) dealt with the case when $\{\epsilon_t\} \sim$ i.i.d. Further, Aït-Sahalia et al. (2011) discussed the case when $\{\epsilon_t\}$ is a strong mixing process with the mixing coefficient decaying exponentially (i.e., j -lag autocovariance function is of order ρ^j , $|\rho| < 1$.) In this section we deal with ARCH(∞) in (4.44), which is uncorrelated but dependent. Here, we assume that the coefficients $\{a_j\}$ of our ARCH(∞) satisfy Assumption 4.5 in Theorem 4.2, and that $a_j \sim O(j^{-r})$, $r > 1$ in Theorem 4.3. Hence our model and that of Aït-Sahalia et al. (2011) show essentially different dependence structure. It may be noted that, the ARCH(∞) setting is convenient for description of heteroskedasticity in the volatility. \square

Remark 4.4 This assumption of heteroskedastic microstructure is natural when considering that the market microstructure is caused by artificial operation such as bid-ask bounce and asynchronous trading. Empirical evidence also shows the process of market microstructure usually has the conditional heteroskedasticity (see Madhavan (2000)). For example, automated limit order books systems of the type used by the Toronto Stock Exchange and Paris Bourse offer continuous trading with high degrees of transparency (i.e., public display of current and away limit orders) without reliance on dealers. Also, this heteroskedastic noise may have a reasoning in view of the informational asymmetries (see e.g., O'Hara (1997)). That is, by this persistence of ARCH(∞) we are trying to describe how quickly the information is assimilated into the latent process of asset price. Here, the existence of new information is governed by the uncorrelated z_i 's but once if they realized then its direction and impact of concerning information are believed to persist since they are not yet assimilated. There, the duration how long it takes to attain the full-information price (i.e., the latent price) is unspecified but described in terms of the significant order of ARCH coefficients. \square

To guarantee the stationarity and geometrical ergodicity (see Giraitis et al. (2000)), we assume the following.

Assumption 4.5 The coefficients in (4.44) satisfy the conditions below:

$$a_0 > 0, \quad a_j \geq 0, \quad \sum_{j=1}^{\infty} a_j < 1, \quad (4.45)$$

$$E[z_0^2]^{1/2} \sum_{j=1}^{\infty} a_j < 1. \quad (4.46)$$

Perturbed version of the return is

$$\tilde{r}_{i+\frac{t}{M}} := Y_{i+\frac{t}{M}} - Y_{i+\frac{t-1}{M}} = r_{i+\frac{t}{M}} + \eta_{i+\frac{t}{M}}, \quad (4.47)$$

where

$$\eta_{i+\frac{t}{M}} := \epsilon_{i+\frac{t}{M}} - \epsilon_{i+\frac{t-1}{M}}.$$

The observed RV is

$$[Y, Y]_{i,i+1} = \sum_{t=1}^M \tilde{r}_{i+\frac{t}{M}}^2. \quad (4.48)$$

We can rewrite the realized volatility (4.48) as the sum of three components,

$$[Y, Y]_{i,i+1} = [X, X]_{i,i+1} + \sum_{t=1}^M \eta_{i+\frac{t}{M}}^2 + 2 \sum_{t=1}^M r_{i+\frac{t}{M}} \eta_{i+\frac{t}{M}} \quad (4.49)$$

$$= \mathcal{I}_1 + \mathcal{I}_2 + 2\mathcal{I}_3. \quad (\text{say}) \quad (4.50)$$

If the true price process was observable, only the term \mathcal{I}_1 would drive the limiting properties of $[Y, Y]_{i,i+1}$. The presence of microstructure noise introduces two additional components \mathcal{I}_2 and \mathcal{I}_3 . We will show that it is mainly the term \mathcal{I}_2 that makes standard consistency arguments fail. Intuitively, \mathcal{I}_2 diverge to infinity almost surely as the number of observations increases (or equivalently, as the frequency of observations increases) since more and more noise will get accumulated. Under ARCH(∞) disturbances, we have the following theorem.

Theorem 4.2 *Suppose that Y follows the model (4.44) with X of (4.35), and that Assumptions 4.4 and 4.5 hold. Then, as $M \rightarrow \infty$, conditionally on the X process, we have*

$$\frac{1}{\sqrt{M}}([Y, Y]_{i,i+1} - [X, X]_{i,i+1} - 2Mv) \xrightarrow{d} N(0, \rho_1 + \rho_2),$$

where

$$\begin{aligned}\rho_1 &:= 4 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f^\epsilon(\lambda_1, \lambda_2, -\lambda_2) d\lambda_1 d\lambda_2 d\lambda_3 + 12v^2, \\ \rho_2 &:= 4 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left\{ e^{i(\lambda_2 + \lambda_3)} f^\epsilon(\lambda_1, \lambda_2, \lambda_3) - 2e^{-i\lambda_2} f^\epsilon(\lambda_1, \lambda_2, -\lambda_2) \right\} d\lambda_1 d\lambda_2 d\lambda_3.\end{aligned}$$

Here,

$$\begin{aligned}v &:= E\left[\epsilon_{i+\frac{t}{M}}^2\right], \\ f^\epsilon(\lambda_1, \lambda_2, \lambda_3) &:= \frac{1}{(2\pi)^3} \sum_{l_1, l_2, l_3=-\infty}^{\infty} Q^\epsilon(l_1, l_2, l_3) e^{-i(\lambda_1 l_1 + \lambda_2 l_2 + \lambda_3 l_3)}, \\ Q^\epsilon(l_1, l_2, l_3) &:= \text{cum}\left(\epsilon_i, \epsilon_{i+\frac{l_1}{M}}, \epsilon_{i+\frac{l_2}{M}}, \epsilon_{i+\frac{l_3}{M}}\right).\end{aligned}$$

Proof See Taniai et al. (2012, Theorem 1). \square

This theorem implies that the observed RV, $[Y, Y]_{i,i+1}$, has a diverging bias $2Mv$ around the true RV $[X, X]_{i,i+1}$ as in the case of i.i.d. noise. However, the asymptotics are affected by non-Gaussian dependent structure of the noise process. Indeed, when the ϵ is i.i.d. (see Bandi and Russell (2008)), our $f^\epsilon(\lambda_1, \lambda_2, \lambda_3)$ becomes a constant, hence, $\rho_2 = 0$. Thus, ρ_2 represents a deviation from i.i.d. assumption on $\{\epsilon_t\}$. Since the fourth-order cumulant spectral density $f^\epsilon(\lambda_1, \lambda_2, \lambda_3)$ shows the degree of dependence and non-Gaussianity of $\{\epsilon_t\}$, we can examine their influences on the asymptotics of $[Y, Y]_{i,i+1}$.

As for the estimator of v above, Zhang et al. (2005) introduced \tilde{v} defined by

$$\tilde{v} := \frac{\sum_{t=1}^M \tilde{r}_{i+\frac{t}{M}}^2}{2M} = \frac{[Y, Y]_{i,i+1}}{2M}.$$

Accordingly, they proposed a bias-adjusted estimator for $\langle X, X \rangle_{i,i+1}$ by

$$[Y, Y]_{i,i+1}^{(\text{tsrv})} := [Y, Y]_{i,i+1}^{(\text{avg})} - \frac{\bar{M}}{M} [Y, Y]_{i,i+1}, \quad (:\text{ Two Scales Realized Volatility (tsrv)}) \quad (4.51)$$

where $[Y, Y]_{i,i+1}^{(\text{avg})}$ is defined, in the manner of (4.34), as

$$\begin{aligned}[Y, Y]_{i,i+1}^{(\text{avg})} &:= \frac{1}{K} \sum_{k=1}^K [Y, Y]_{i,i+1}^{(k)}, \\ [Y, Y]_{i,i+1}^{(k)} &:= \sum_{t=1}^{n_k} \left(Y_{i+\frac{k-1+tK}{M}} - Y_{i+\frac{k-1+(t-1)K}{M}} \right)^2 \quad \text{with } n_k := \left\lfloor \frac{M-k+1}{K} \right\rfloor,\end{aligned} \quad (4.52)$$

and \bar{M} is

$$\bar{M} := \frac{1}{K} \sum_{k=1}^K n_k \approx \frac{M+1}{K} - \frac{K+1}{2} \approx \frac{M}{K},$$

where the last approximation is used when M is much larger relatively to K . The following theorem of ours states the result for ARCH(∞) disturbances case.

Theorem 4.3 *Suppose that Y follows the model (4.43) with X of (4.35), and that Assumptions 4.4 and 4.5 hold. Further, if the coefficients a_j 's satisfy $a_j < b_j^{-r}$ for some $r > 1$ and $b > 0$, and if $K = cM^{2/3}$, for some $c > 0$, then it holds that*

$$\begin{aligned} M^{1/6}([Y, Y]_{i,i+1}^{(\text{tsrv})} - \langle X, X \rangle_{i,i+1}) &\xrightarrow{d} \mathcal{MN}(0, c^{-2}(4\gamma + 12v^2)) + \mathcal{MN}(0, c(\eta^2)) \\ &= \mathcal{MN}(0, c^{-2}(4\gamma + 12v^2) + c(\eta^2)), \end{aligned}$$

where

$$\begin{aligned} \eta &:= \frac{4}{3} \int_i^{i+1} \sigma_t^4 dt, \\ \gamma &:= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i(\lambda_2 + \lambda_3)} f^\epsilon(\lambda_1, \lambda_2, \lambda_3) d\lambda_1 d\lambda_2 d\lambda_3. \end{aligned}$$

Proof See Taniai et al. (2012, Theorem 2). □

Theorem 4.3 implies the modified estimator $[Y, Y]_{i,i+1}^{(\text{tsrv})}$ is consistent for $\langle X, X \rangle_{i,i+1}$ even if the noise process is ARCH(∞). This is because the subsampling frequency K also diverges to infinity when the original sampling frequency M goes to infinity. It may be noted that the asymptotics of $[Y, Y]_{i,i+1}^{(\text{tsrv})}$ depend on non-Gaussianity and dependence of the noise through the integral of fourth-order cumulant spectra, which describes the asymptotics of the Whittle estimator for non-Gaussian dependent processes in unified manner (see, Hosoya and Taniguchi (1982), Sect. 3.1 of Taniguchi and Kakizawa (2000)).

Remark 4.5 Estimation for the asymptotic variances γ and v^2 seems important. It is difficult to develop the general theory. But, a partial solution is given as follows. Suppose that $\{X_t\}$ is generated by a stationary continuous time ARMA process (See Brockwell (1994), and also Taniguchi and Kakizawa (2000, p. 149)), which includes a simple diffusion process whose discretized version is an autoregressive model. Consider

$$Y_{t_j} = X_{t_j} + \epsilon_{t_j}, \quad j = 1, 2, \dots, n. \quad (4.53)$$

Based on the observation $\{Y_{t_j}\}$ we can estimate the spectral density $f_Y(\lambda)$ of $\{Y_{t_j}\}$. By using a method due to Hosoya and Taniguchi (1982, pp. 139–140), we can estimate the spectral density $f_X(\lambda)$ of $\{X_{t_j}\}$, and derive an estimator \hat{X}_{t_j} of X_{t_j} . Then we can calculate $\hat{\epsilon}_{t_j} \equiv Y_{t_j} - \hat{X}_{t_j}$. Based on $\hat{\epsilon}_{t_1}, \dots, \hat{\epsilon}_{t_n}$, we can estimate the integral

$$\gamma = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i(\lambda_2 + \lambda_3)} f^\epsilon(\lambda_1, \lambda_2, \lambda_3) d\lambda_1 d\lambda_2 d\lambda_3.$$

by a method proposed by Taniguchi (1982). Hence we can construct a consistent estimator $\hat{\gamma}$ of γ .

4.3.2 Numerical Studies

In this section, we calculate two estimators $[Y, Y]_{i,i+1}$, $[Y, Y]_{i,i+1}^{(\text{tsrv})}$ to estimate the integrated volatility $\langle X, X \rangle_{i,i+1}$. The observed frequency M is chosen as 25,000, which corresponds to the time interval about 1 s. In order to generate the X process, here we use the Euler-Maruyama Discretization (e.g., Kloeden and Platen (1992); Gouriéroux and Jasiak (2001)):

$$X_{i+\frac{t}{M}} = X_{i+\frac{t-1}{M}} + \mu_{i+\frac{t-1}{M}} \frac{1}{M} + \sigma_{i+\frac{t-1}{M}} (W_{i+\frac{t}{M}} - W_{i+\frac{t-1}{M}}).$$

So we may express the log-return process as

$$r_{i+\frac{t}{M}} = X_{i+\frac{t}{M}} - X_{i+\frac{t-1}{M}} = \mu_{i+\frac{t-1}{M}} \frac{1}{M} + \sigma_{i+\frac{t-1}{M}} \sqrt{\frac{1}{M}} u_{i+\frac{t}{M}}, \quad (4.54)$$

with $\{u_t\}$ being a sequence of i.i.d. $N(0, 1)$ random variables. Consequently, provided the condition (4.37), a discretized version of Integrated Volatility (4.38), which here we define as

$$\langle X, X \rangle_{i,i+1}^{(d)} := \frac{1}{M} \sum_{t=1}^M \sigma_{i+\frac{t-1}{M}}^2, \quad (4.55)$$

is known to converge to $\langle X, X \rangle_{i,i+1}$ as $M \rightarrow \infty$. Now, let us assume that $\mu_{i+\frac{t}{M}}$ and $\sigma_{i+\frac{t}{M}}$ take the form as

$$\mu_{i+\frac{t}{M}} = \phi_0 + \phi_1 r_{i+\frac{t-1}{M}}, \quad \sigma_{i+\frac{t}{M}}^2 = s_0 + s_1 (r_{i+\frac{t-1}{M}} - \mu_{i+\frac{t-1}{M}})^2,$$

so that the log-return $r_{i+\frac{t}{M}}$ follows an AR(1)-ARCH(1) model characterized by

$$\begin{aligned} r_{i+\frac{t}{M}} &= \frac{\phi_0}{M} + \frac{\phi_1}{M} r_{i+\frac{t-1}{M}} + e_{i+\frac{t}{M}}, \quad e_{i+\frac{t}{M}} := h_{i+\frac{t}{M}} u_{i+\frac{t}{M}}, \\ h_{i+\frac{t}{M}}^2 &= \frac{s_0}{M} + \frac{s_1}{M} e_{i+\frac{t-1}{M}}^2. \end{aligned} \quad (4.56)$$

Then, we generate the process $\{X_{i+\frac{t}{M}}\}$ through the above AR-ARCH model (4.56) with a choice of parameters as follows:

$$\begin{aligned}
r_{i+\frac{t}{M}} &= 10^{-5} + 0.3r_{i+\frac{t-1}{M}} + e_{i+\frac{t}{M}}, \quad e_{i+\frac{t}{M}} := h_{i+\frac{t}{M}} u_{i+\frac{t}{M}}, \\
h_{i+\frac{t}{M}}^2 &= \frac{s_0}{M} + 0.5e_{i+\frac{t-1}{M}}^2.
\end{aligned} \tag{4.57}$$

Here, according to our choice $s_1/M \equiv 0.5$, the value of s_0 is set to be $s_0 \equiv 1 - s_1/M$ so that the unconditional standard deviation over 1 period ($1/M$) becomes $(\text{Var}[r_{i+\frac{t}{M}}])^{1/2} \equiv 1/\sqrt{M}$. Also, we assume the market microstructure noise $\{\epsilon_t\}$ follows an ARCH(1) process

$$\epsilon_{i+\frac{t}{M}} = (a_0 + a_1\epsilon_{i+\frac{t-1}{M}}^2)^{1/2} \sqrt{\frac{1}{M}} z_{i+\frac{t}{M}} = \left(\frac{a_0}{M} + \frac{a_1}{M} \epsilon_{i+\frac{t-1}{M}}^2 \right)^{1/2} z_{i+\frac{t}{M}}. \tag{4.58}$$

In our first observation below, we set $a_0 \equiv p^2 - a_1/M$ for varying values of a_1 to keep the unconditional standard deviation to be $(\text{Var}[\epsilon_{i+\frac{t}{M}}])^{1/2} \equiv p/\sqrt{M}$. That is, the microstructure noise is set to always have $100p\%$ magnitude with respect to the asset. Substituting $u_{i+\frac{t}{M}}$ in (4.54) and $z_{i+\frac{t}{M}}$ in (4.58) with realizations of $N(0, 1)$, we generate $20 \times M$ (which amounts to 1-month data points) samples of $\{X_{i+\frac{t}{M}}\}$.

Let $\text{MSE}^{(\text{rv})}$ and $\text{MSE}^{(\text{tsrv})}$ denote the mean squared error (MSE) of $[Y, Y]_{i,i+1}$ and $[Y, Y]_{i,i+1}^{(\text{tsrv})}$, i.e.,

$$\begin{aligned}
\text{MSE}^{(\text{rv})} &:= \frac{1}{20} \sum_{i=1}^{20} ([Y, Y]_{i,i+1} - \langle X, X \rangle_{i,i+1}^{(d)})^2, \\
\text{MSE}^{(\text{tsrv})} &:= \frac{1}{20} \sum_{i=1}^{20} ([Y, Y]_{i,i+1}^{(\text{tsrv})} - \langle X, X \rangle_{i,i+1}^{(d)})^2.
\end{aligned}$$

Figure 4.5 shows the values of $\text{MSE}^{(\text{tsrv})}$ with $p = 1$ for several choices of sampling interval K and ARCH parameter a_1 .

As we will see in Table 4.3, for this set of parameters the value of $\text{MSE}^{(\text{rv})}$ is about 4.42763. So, it can be seen that the TSRV estimator yields smaller estimation error with any choice of the sampling interval from $K = 2$ (2s) to $K = 600$ (10 min.). Also, it should be noted that $\text{MSE}^{(\text{tsrv})}$ is almost robust against the choice of the ARCH parameter. This feature of TSRV is contrast to those of RV, so we study the following Table 4.3 with this in our mind.

Table 4.3 shows the values of $\text{MSE}^{(\text{rv})}$ and $\text{MSE}^{(\text{tsrv})}$ for $K = 10$, along the choices of the noise magnitude p and ARCH parameter a_1 . As mentioned, there we see that $\text{MSE}^{(\text{rv})}$ is more sensitive to the change of ARCH parameter than $\text{MSE}^{(\text{tsrv})}$. From this Table 4.3, we may find two features. Firstly, the difference between $\text{MSE}^{(\text{rv})}$ and $\text{MSE}^{(\text{tsrv})}$ varies along the choice of the size of noise p . In fact, if the relative size of $\epsilon_{i+\frac{t}{M}}$ is small enough with respect to the asset $X_{i+\frac{t}{M}}$, we even observe $\text{MSE}^{(\text{rv})}$ outperforming $\text{MSE}^{(\text{tsrv})}$ uniformly in a_1 (and in K also). This may be related to our second remark of Table 4.3, which is the fact that those MSEs exhibit a decreasing

Fig. 4.5 $MSE^{(tsrv)}$ with $p = 1$. Sampling interval $\log K$ are $\log(2, 3, 5, 10, 30, 60, 300, 600)$. As a comparison, $MSE^{(rv)}$ is larger than 2.5 for $a_1/M \in [0.1, 0.9]$, as shown in Table 4.3. Taken from Taniai et al. (2012). Published with the kind permission of © Oxford University Press 2012. All Rights Reserved

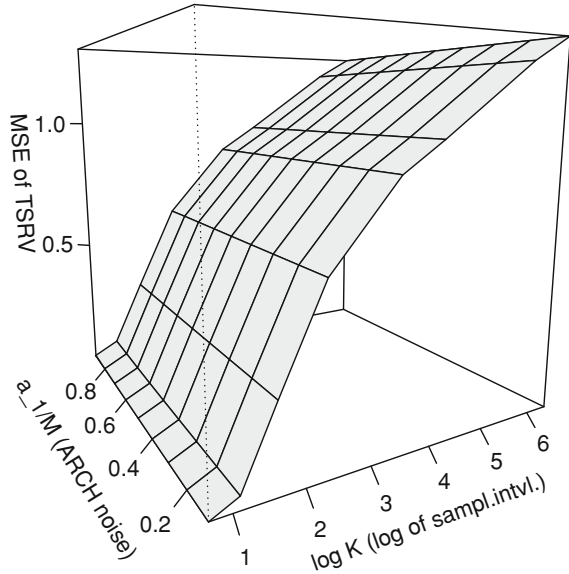


Table 4.3 The values of $MSE^{(rv)}$ (upper) and $MSE^{(tsrv)}$ (lower) when $K = 10$

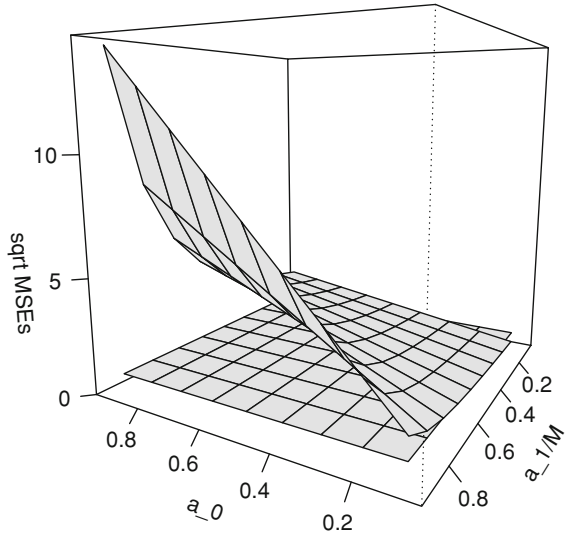
		ARCH parameter a_1/M				
		0.1	0.3	0.5	0.7	0.9
Noise magnitude	0	0.00987	0.00987	0.00987	0.00987	0.00987
	p	0.61247	0.61247	0.61247	0.61247	0.61247
	0.2	0.03210	0.03208	0.03199	0.03153	0.02579
		0.61271	0.61272	0.61280	0.61295	0.61298
	0.4	0.17595	0.17577	0.17496	0.17085	0.12040
		0.61302	0.61305	0.61318	0.61348	0.61346
	0.6	0.67318	0.67237	0.66891	0.65122	0.43442
		0.61340	0.61344	0.61363	0.61405	0.61394
	0.8	1.90997	1.90755	1.89733	1.84492	1.20249
		0.61384	0.61391	0.61414	0.61468	0.61440
	1	4.42692	4.42117	4.39710	4.27319	2.75318
		0.61435	0.61444	0.61471	0.61535	0.61486

Taken from Taniai et al. (2012). Published with the kind permission of © Oxford University Press 2012. All Rights Reserved

feature in a_1 . Suspecting that this stems from our request $a_0 \equiv p^2 - a_1/M$, which was to keep the unconditional variance, we investigate the behavior of MSEs along the choice of ARCH parameters both a_0 and a_1 below.

By releasing the constraint $a_0 = p^2 - a_1/M$, Fig. 4.6 exhibits the changes of MSEs in ARCH parameters a_0 and a_1 . Again, as for the sampling interval of TSRV estimator, we applied $K = 10$. We observe that $MSE^{(rv)}$, the upper surface, is more sensitive than $MSE^{(tsrv)}$, the lower surface, to both parameters a_0 and a_1 . In fact, the

Fig. 4.6 $\sqrt{\text{MSEs}}$ for ARCH parameters. The *upper* and *lower* surfaces correspond to $\sqrt{\text{MSE}^{(\text{rv})}}$ and $\sqrt{\text{MSE}^{(\text{tsrv})}}$ ($K = 10$) respectively. Taken from Taniai et al. (2012). Published with the kind permission of © Oxford University Press 2012. All Rights Reserved



lower surface $\text{MSE}^{(\text{tsrv})}$ is almost a horizontal plane, which shows the robustness of TSRV estimator against the ARCH affection. This, in turn, suggest the importance of recognition of ARCH behavior in microstructure noise when applying the commonly used RV estimator (4.32).

4.4 Appendix

4.4.1 Proof of Theorem 4.1

To begin with, define, in regard to (4.13),

$$D_{\gamma,f}^{(n)} := \frac{1}{n-p} \sum_{t=p+1}^n E_{\gamma,f}^{(n)}[\psi_{f,p}(U_{\gamma,t}, \dots, U_{\gamma,t-p}) | \mathcal{B}_{\gamma}^{(n)}].$$

Then, as in Hallin et al. (2008, Equation (A.4)), it can be shown that

$$D_{\gamma,f}^{(n)} - E_{\gamma,f}^{(n)}[D_{\gamma,f}^{(n)} | N_{\gamma,L}^{(n)}] = \frac{1}{n-p} \sum_{t=p+1}^n \left[\psi_{f,p}(U_{\gamma,t}, \dots, U_{\gamma,t-p}) - E[\psi_{f,p}(U_0, \dots, U_p) | U_0 = U_{\gamma,t}] \right] + o_P(n^{-1/2}). \tag{4.59}$$

More precisely, first, with help of results for Linear Serial Rank Statistics (See Hallin et al. (2003, Sect. 4.2)), it can be shown that

$$\mathbf{D}_{\gamma,f}^{(n)} - E_{\gamma,f}^{(n)}[\mathbf{D}_{\gamma,f}^{(n)} | N_{\gamma,L}^{(n)}] = \mathbf{T}_{\gamma,f}^{(n)} - E[\mathbf{T}_{\gamma,f}^{(n)} | \boldsymbol{\xi}_{\gamma,(\cdot)}] + o_P(n^{-1/2}), \quad (4.60)$$

where

$$\mathbf{T}_{\gamma,f}^{(n)} := \frac{1}{n-p} \sum_{t=p+1}^n \boldsymbol{\psi}_{f,p}(U_{\gamma,t}, \dots, U_{\gamma,t-p}).$$

and $\boldsymbol{\xi}_{\gamma,(\cdot)}$ denotes the vector of Order Statistics. Then, by van der Vaart (1998, Theorem 12.3) or Serfling (1980, Sect. 5.3), the second term of the r.h.s. of (4.60) becomes into

$$\begin{aligned} E[\mathbf{T}_{\gamma,f}^{(n)} | \boldsymbol{\xi}_{\gamma,(\cdot)}] &= E[\mathbf{T}_{\gamma,f}^{(n)} | U_{\gamma,(\cdot)}] \\ &= \frac{1}{n-p} \sum_{t=p+1}^n \sum_{\ell=0}^p E[\boldsymbol{\psi}_{f,p}(U_0, \dots, U_p) | U_t = U_{\gamma,t-\ell}] \\ &\quad + (p+1)E[\boldsymbol{\psi}_{f,p}(U_0, \dots, U_p)] + o_P(n^{-1/2}). \end{aligned}$$

So, together with the fact that $\boldsymbol{\psi}_{f,p}$ forms a martingale difference, that is, for all $(u_1, \dots, u_p) \in [0, 1]^p$,

$$\int_0^1 \boldsymbol{\psi}_{f,p}(u_0, u_1, \dots, u_p) du_0 = \mathbf{0}, \quad (4.61)$$

we have (4.59).

Evaluation of $E_{\gamma,f}^{(n)}[\mathbf{D}_{\gamma,f}^{(n)} | N_{\gamma,L}^{(n)}]$ is also a modification of Hallin et al. (2008). That is, firstly observe that, by the tower property of conditional expectations and WLLN,

$$\begin{aligned} E_{\gamma,f}^{(n)}[\mathbf{D}_{\gamma,f}^{(n)} | N_{\gamma,L}^{(n)}] &= \frac{1}{n-p} \sum_{t=p+1}^n E_{\gamma,f}^{(n)}[\boldsymbol{\psi}_{f,p}(U_{\gamma,t}, \dots, U_{\gamma,t-p}) | N_{\gamma,L}^{(n)}] \\ &= E[\boldsymbol{\psi}_{f,p}^{(s)}(s) | N_{\gamma,L}^{(n)}] + o_P(n^{-1/2}), \end{aligned}$$

where $\mathbf{s} = (s_0, \dots, s_p) \in \{-1, 1\}^{p+1}$ and

$$\boldsymbol{\psi}_{f,p}^{(s)}(\mathbf{s}) := E_{\gamma,f}^{(n)}[\boldsymbol{\psi}_{f,p}(U_{\gamma,t}, \dots, U_{\gamma,t-p}) | S_{\gamma,t} = s_0, \dots, S_{\gamma,t-p} = s_p].$$

The number of -1 's in its argument is distributed as

$$\#\{i = 0, 1, \dots, p : S_{\mathcal{Y}, t-i} = -1\} \stackrel{d}{\sim} \text{HyperGeometric}(n, N_{\mathcal{Y}, L}^{(n)}, p),$$

so we approximate this with binomial distribution. Then, noting that $E_{\mathcal{Y}, f}^{(n)}[\mathbf{D}_{\mathcal{Y}, f}^{(n)}] = 0$ by (4.61), we have

$$\begin{aligned} & E_{\mathcal{Y}, f}^{(n)}[\mathbf{D}_{\mathcal{Y}, f}^{(n)} | N_{\mathcal{Y}, L}^{(n)}] - E_{\mathcal{Y}, f}^{(n)}[\mathbf{D}_{\mathcal{Y}, f}^{(n)}] \\ &= \sum_{s \in \{-1, 1\}^{p+1}} \boldsymbol{\psi}_{f, p}^{(s)}(s) \cdot B_p\left(\frac{N_{\mathcal{Y}, L}^{(n)}}{n}, \#\{i : s_i = -1\}\right) \\ &\quad - \sum_{s \in \{-1, 1\}^{p+1}} \boldsymbol{\psi}_{f, p}^{(s)}(s) \cdot B_p(\tau, \#\{i : s_i = -1\}) + o_P(n^{-1/2}) \\ &= \sum \boldsymbol{\psi}_{f, p}^{(s)}(s) \cdot \frac{\partial}{\partial x} B_p(x, \#\{i : s_i = -1\}) \Big|_{x=\tau} \left(\frac{N_{\mathcal{Y}, L}^{(n)}}{n} - \tau\right) + o_P(n^{-1/2}) \\ &= \sum \boldsymbol{\psi}_{f, p}^{(s)}(s) \cdot B_p(\tau, \#\{s_i = -1\}) \frac{\#\{i : s_i = -1\} - \tau(p+1)}{\tau(1-\tau)} \\ &\quad \times \left(\frac{N_{\mathcal{Y}, L}^{(n)}}{n} - \tau\right) + o_P(n^{-1/2}) \\ &= \frac{E[\boldsymbol{\psi}_{f, p}^{(s)}(s)(\#\{i : s_i = -1\} - \tau(p+1))]}{\tau(1-\tau)} \times \left(\frac{N_{\mathcal{Y}, L}^{(n)}}{n} - \tau\right) + o_P(n^{-1/2}), \end{aligned} \tag{4.62}$$

where $B_p(x, w) := x^w(1-x)^{p+1-w}$. Now, observe that

$$\begin{aligned} & E[\boldsymbol{\psi}_{f, p}^{(s)}(s)(\#\{i = 0, 1, \dots, p : s_i = -1\} - \tau(p+1)) | s_1, \dots, s_p] \\ &= P\{s_0 = -1\} \boldsymbol{\psi}_{f, p}^{(s)}(-1, s_1, \dots, s_p)(1 + \#\{i = 1, \dots, p : s_i = -1\} - \tau(p+1)) \\ &\quad + P\{s_0 = 1\} \boldsymbol{\psi}_{f, p}^{(s)}(1, s_1, \dots, s_p)(0 + \#\{i = 1, \dots, p : s_i = -1\} - \tau(p+1)) \\ &= \tau \cdot \boldsymbol{\psi}_{f, p}^{(s)}(-1, s_1, \dots, s_p), \end{aligned}$$

where the last equality follows from $P\{s_0 = -1\} = P\{S_{\mathcal{Y}, t} = -1\} = \tau$ and

$$P\{s_0 = -1\} \boldsymbol{\psi}_{f, p}^{(s)}(-1, s_1, \dots, s_p) + P\{s_0 = 1\} \boldsymbol{\psi}_{f, p}^{(s)}(1, s_1, \dots, s_p) = 0,$$

by (4.61). The tower property yields that

$$\begin{aligned}
& E[\boldsymbol{\psi}_{f,p}^{(s)}(s)(\#\{i : s_i = -1\} - \tau(p+1))] \\
&= \tau \cdot E[\boldsymbol{\psi}_{f,p}^{(s)}(s) | s_0 = -1] = \tau \cdot E[\boldsymbol{\psi}_{f,p}(U_0, \dots, U_p) | U_0 \leq \tau] \\
&= \tau \cdot \frac{\int_0^\tau \varphi_f[F^{-1}(u)] du}{\int_0^\tau du} \cdot \boldsymbol{\mu}_{\mathbf{W}(\boldsymbol{\gamma}, f)} = -f(1) \cdot \boldsymbol{\mu}_{\mathbf{W}(\boldsymbol{\gamma}, f)}.
\end{aligned}$$

Finally, the rest of (4.62) is $N_{\boldsymbol{\gamma}, L}^{(n)}/n - \tau = n^{-1} \sum_{t=1}^n (1\{\xi_{\boldsymbol{\gamma}, t} \leq 1\} - \tau)$.

The asymptotic normality (4.16) follows from Yoshihara's CLT for U-statistics, as shown in Hallin et al. (1985). The proof is now completed. \square

References

- Aït-Sahalia, Y., Mykland, P.A., Zhang, L.: Ultra high frequency volatility estimation with dependent microstructure noise. *J. Econometric* **160**(1), 160–175 (2011)
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: Modeling and forecasting realized volatility. *Econometrica* **71**(2), 579–625 (2003)
- Bandi, F.M., Russell, J.R.: Microstructure noise, realized variance, and optimal sampling. *Rev. Econom. Stud.* **75**(2), 339–369 (2008)
- Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**(2), 253–280 (2002)
- Bickel, P.J., Klaassen, C.A.J., Ritov, Y.: Efficient and Adaptive Estimation for Semiparametric Models. Springer, New York (1998). Reprint of the 1993 original
- Brockwell, P.J.: On continuous-time threshold ARMA processes. *J. Statist. Plann. Infer.* **39**(2), 291–303 (1994)
- Dacorogna, M., Gençay, R., Muller, U.A., Olsen, R., Pictet, O.: An Introduction to High-Frequency Finance. Academic, San Diego (2001)
- Drost, F.C., Klaassen, C.A.J., Werker, B.J.M.: Adaptive estimation in time-series models. *Ann. Statist.* **25**(2), 786–817 (1997)
- Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4), 987–1007 (1982)
- Francq, C., Zakoïan, J.M.: Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theor.* **22**(5), 815–834 (2006)
- Giraitis, L., Kokoszka, P., Leipus, R.: Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theor.* **16**(1), 3–22 (2000)
- Gouriéroux, C.: ARCH Models and Financial Applications. Springer, New York (1997). Springer Series in Statistics
- Gouriéroux, C., Jasiak, J.: Financial Econometrics. Princeton University Press, Princeton (2001). Princeton-Series in Finance
- Hallin, M., Ingenbleek, J.F., Puri, M.L.: Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.* **13**(3), 1156–1181 (1985)
- Hallin, M., Oja, H., Paindaveine, D.: Semiparametrically efficient rank-based inference for shape. II. Optimal R -estimation of shape. *Ann. Statist.* **34**(6), 2757–2789 (2006a)
- Hallin, M., Vermandele, C., Werker, B.J.M.: Serial and nonserial sign-and-rank statistics: asymptotic representation and asymptotic normality. *Ann. Statist.* **34**(1), 254–289 (2006b)
- Hallin, M., Vermandele, C., Werker, B.J.M.: Semiparametrically efficient inference based on signs and ranks for median-restricted models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70**(2), 389–412 (2008)

- Hallin, M., Werker, B.J.M.: Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9**(1), 137–165 (2003)
- He, C., Teräsvirta, T.: Fourth moment structure of the GARCH(p, q) process. *Econometric Theory* **15**(6), 824–846 (1999)
- Hosoya, Y., Taniguchi, M.: A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Statist.* **10**(1), 132–153 (1982)
- Kloeden, P.E., Platen, E.: Numerical solution of stochastic differential equations, Applications of Mathematics (New York), vol. 23. Springer, Berlin (1992)
- Koenker, R.: Quantile regression, *Econometric Society Monographs*, vol. 38. Cambridge University Press, Cambridge (2005)
- Koenker, R., Bassett Jr, G.: Regression quantiles. *Econometrica* **46**(1), 33–50 (1978)
- Koenker, R., Xiao, Z.: Quantile autoregression. *J. Amer. Statist. Assoc.* **101**(475), 980–990 (2006)
- Koul, H.L.: Weighted Empirical and Linear Models. Institute of Mathematical Statistics Lecture Notes-Monograph Series, 21. Institute of Mathematical Statistics, Hayward, CA (1992).
- Kreiss, J.P.: On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15**(1), 112–133 (1987)
- Linton, O.: Adaptive estimation in ARCH models. *Econometric Theor.* **9**(4), 539–569 (1993)
- Madhavan, A.: Market microstructure: A survey. *J. Finan. Markets.* **3**(3), 205–258 (2000)
- McAleer, M., Medeiros, M.: Realized volatility: a review. *Econometric Rev.* **27**(1—3), 10–45 (2008)
- Meitz, M., Saikkonen, P.: Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theor.* **24**(5), 1291–1320 (2008)
- O'Hara, M.: *Market Microstructure Theory*. Blackwell, Cambridge, MA (1997)
- Puri, M.L., Sen, P.K.: *Nonparametric Methods in Multivariate Analysis*. Wiley, New York-London-Sydney (1971)
- Schmetterer, L.: *Introduction to Mathematical Statistics*. Springer, Berlin (1974). Translated from the second German edition by Kenneth Wickwire, *Die Grundlehren der mathematischen Wissenschaften*, Band 202
- Serfling, R.J.: *Approximation Theorems of Mathematical Statistics*. Wiley, New York (1980). Wiley Series in Probability and Mathematical Statistics
- Taniai, H.: Inference for the quantiles of ARCH processes. The doctoral dissertation. Université libre de Bruxelles, Belgium (2009)
- Taniai, H., Usami, T., Suto, N., Taniguchi, M.: Asymptotics of realized volatility with non-gaussian arch(∞) microstructure noise. *J. Finan. Econometrics* **10**(4), 617–636 (2012)
- Taniguchi, M.: On estimation of the integrals of the fourth order cumulant spectral density. *Biometrika* **69**(1), 117–122 (1982)
- Taniguchi, M., Kakizawa, Y.: *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York (2000). Springer Series in Statistics
- Tsay, R.S.: *Analysis of Financial Time Series*. Wiley, New York (2002). Wiley series in probability and mathematical statistics
- van der Vaart, A.W.: *Asymptotic Statistics*. Cambridge University Press, Cambridge (1998). Cambridge Series in Statistical and Probabilistic Mathematics
- Zhang, L., Mykland, P.A., Ait-Sahalia, Y.: A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100**(472), 1394–1411 (2005)

Index

A

American option, 29
Arbitrage opportunity, 28
Arbitrage-free, 28
ARCH model, 85, 105
ARCH(∞) model, 19
ARCH(∞)-SM model, 19
ARCH(q), 3
Asian call option, 30
Asset, 27
Asymptotic relative efficiency (ARE), 96
Asymptotically centering, 23
Asymptotically efficient, 23, 27

B

Black-Scholes formula, 32

C

Call option, 29
CAPM, 74
Central sequence, 22
Characteristic exponent, 53
CHARN, 5
Classification, 35
Complete, 29
Conditional LSE, 24
Conditional quantile, 86
Contiguous alternative, 25
Contingent claim, 29
Control variate estimator, 66
Control variate method, 66
CR statistic, 46
 frequency domain—, 47
Curved structure, 78

D

Dendrogram, 37
Discretization, 93
Disparity measure, 18

E

Edgeworth expansion, 32
EGARCH(p, q), 4
Empirical likelihood ratio function, 42
 —for frequency domain, 44
Equivalent martingale measure, 29
Estimating function, 42
European option, 29

F

Fourth-order cumulant, 6
Fourth-order stationary, 6
Fractionally cointegrated, 78

G

GARCH(p, q), 4
Generalized empirical likelihood (GEL), 46,
 49
Geometric Brownian motion, 31
Grenander's condition, 71

H

h -step ahead prediction, 12

I

Innovation-free, 12
Instrumental variable, 78

Instrumental variable method, 75
 Integrated volatility (IV), 102

K

Kolmogorov's formula, 46

L

LAN, 19
 Locally stationary, 26
 Log-likelihood ratio, 25
 Long memory dependence, 74
 Long memory dependent, 77

M

Market microstructure, 102
 Martingale, 28
 Maturity, 29
 Maximal invariant, 89
 Maximum likelihood estimator, 25
 Mean squared error (MSE), 101
 Measure of instantaneous linear feedback, 16
 Measure of linear dependence, 17
 Measure of linear feedback, 16
 Misclassification, 35

N

Non-Gaussian robustness, 36
 Nonparametric likelihood, 42
 Nonparametric spectral estimator, 14

O

One-step estimator, 92
 Option, 29
 Ordinary least squares (OLS) estimator, 75, 78

P

Portfolio, 13, 27
 Put option, 29

Q

Quantile regression (QR), 85
 Quasi-Gaussian maximum likelihood estimator, 9
 Quasi-maximum likelihood estimator, 27

R

Realized volatility (RV), 85, 102

S

Sample autocorrelation function (SACF), 2
 Sample mean, 66, 70
 Second-order stationary, 5
 Self-financing portfolio, 29
 Semiparametric efficiency, 95
 Short memory dependent, 77
 Short memory process, 75
 Spectral measure, 53
 Spectral moment condition, 43
 Spectral window, 13
 Stable distribution, 51
 multivariate—, 51
 Stochastic regression model, 74, 76
 Strike price, 29
 $SV(m)$, 4

T

The least squares estimator (LSE), 67, 72
 Time varying spectral density, 26
 Trend model, 70
 Two scales realized volatility (TSRV), 103
 Two-stage least squares (2SLS) estimator, 75, 79

U

ULAN, 90

W

Wold decomposition theorem, 5