Advances in

Decision Sciences

Special Issue
Statistical Estimation of Portfolios for Dependent Financial Returns

Guest Editors: Masanobu Taniguchi, Cathy W. S. Chen, Junichi Hirukawa, Hiroshi Shiraishi, Kenichiro Tamaki, and David Veredas
Statistical Estimation of Portfolios for Dependent Financial Returns
Statistical Estimation of Portfolios for Dependent Financial Returns

Guest Editors: Masanobu Taniguchi, Cathy W. S. Chen, Junichi Hirukawa, Hiroshi Shiraishi, Kenichiro Tamaki, and David Veredas
Editorial Board

Mahyar A. Amouzegar, USA
Fernando Beltran, New Zealand
Omer S. Benli, USA
David Bulger, Australia
Raymond Chan, Hong Kong
Wai Ki Ching, Hong Kong
Stefanka Chukova, New Zealand

S. Dempe, Germany
C. D. Lai, New Zealand
YanXia Lin, Canada
Chenghu Ma, China
Khosrow Moshirvaziri, USA
Shelton Peiris, Australia
Jack Penm, Australia

Roger Z. Rios-Mercado, Mexico
Henry Schellhorn, USA
Andreas Soteriou, Cyprus
Olivier Thas, Belgium
WingKeung Wong, Hong Kong
Graham Raymond Wood, Australia
## Contents

**Statistical Estimation of Portfolios for Dependent Financial Returns**, Masanobu Taniguchi, Cathy W. S. Chen, Junichi Hirukawa, Hiroshi Shiraishi, Kenichiro Tamaki, and David Veredas  
Volume 2012, Article ID 681490, 3 pages

**Large-Deviation Results for Discriminant Statistics of Gaussian Locally Stationary Processes**, Junichi Hirukawa  
Volume 2012, Article ID 572919, 15 pages

**Asymptotic Optimality of Estimating Function Estimator for CHARN Model**, Tomoyuki Amano  
Volume 2012, Article ID 515494, 11 pages

**Optimal Portfolio Estimation for Dependent Financial Returns with Generalized Empirical Likelihood**, Hiroaki Ogata  
Volume 2012, Article ID 973173, 8 pages

**Statistically Efficient Construction of α-Risk-Minimizing Portfolio**, Hiroyuki Tanai and Takayuki Shiohama  
Volume 2012, Article ID 980294, 17 pages

**Estimation for Non-Gaussian Locally Stationary Processes with Empirical Likelihood Method**, Hiroaki Ogata  
Volume 2012, Article ID 704693, 22 pages

**A Simulation Approach to Statistical Estimation of Multiperiod Optimal Portfolios**, Hiroshi Shiraishi  
Volume 2012, Article ID 341476, 13 pages

**On the Causality between Multiple Locally Stationary Processes**, Junichi Hirukawa  
Volume 2012, Article ID 261707, 15 pages

**Optimal Portfolios with End-of-Period Target**, Hiroshi Shiraishi, Hiroaki Ogata, Tomoyuki Amano, Valentin Patilea, David Veredas, and Masanobu Taniguchi  
Volume 2012, Article ID 703465, 13 pages

**Least Squares Estimators for Unit Root Processes with Locally Stationary Disturbance**, Junichi Hirukawa and Mako Sadakata  
Volume 2012, Article ID 893497, 16 pages

**Statistical Portfolio Estimation under the Utility Function Depending on Exogenous Variables**, Kenta Hamada, Dong Wei Ye, and Masanobu Taniguchi  
Volume 2012, Article ID 127571, 15 pages

**Statistical Estimation for CAPM with Long-Memory Dependence**, Tomoyuki Amano, Tsuyoshi Kato, and Masanobu Taniguchi  
Volume 2012, Article ID 571034, 12 pages
Editorial

Statistical Estimation of Portfolios for Dependent Financial Returns

Masanobu Taniguchi,1 Cathy W. S. Chen,2 Junichi Hirukawa,3 Hiroshi Shiraishi,4 Kenichiro Tamaki,1 and David Veredas5

1 Department of Applied Mathematics, Waseda University, Tokyo 169-8555, Japan
2 Department of Statistics/Graduate Institute of Statistics & Actuarial Science, Feng Chia University Taichung 407, Taiwan
3 Department of Mathematics, Faculty of Science, Niigata University, Niigata 950-2181, Japan
4 The Jikei University School of Medicine, Tokyo 105-8461, Japan
5 ECARES-Solvay Brussels School of Economics and Management, Universite Libre de Bruxelles, 1050 Brussels, Belgium

Correspondence should be addressed to Masanobu Taniguchi, taniguchi@waseda.jp

Received 22 August 2012; Accepted 22 August 2012

Copyright © 2012 Masanobu Taniguchi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The field of financial engineering has developed as a huge integration of economics, mathematics, probability theory, statistics, time series analysis, operation research, and so forth, over the last decade. The construction of portfolios for financial assets is one of the most important issues in financial engineering. It is empirically observed that financial returns are non-Gaussian and dependent, and it is shown that the classical mean-variance portfolio estimator is not statistically optimal. Knowledge and understanding of these have led to the development of general time series modeling for financial returns, sophisticated optimal estimation theory, robust estimation methods, empirical likelihood for time series, nonstationary time series analysis, prediction of time series, and various numerical approaches for portfolios.

As the contents, this special volume includes the following topics in financial time series analysis and financial engineering.

The paper titled “Large deviation results for discriminant statistics of Gaussian locally stationary processes” by J. Hirukawa discusses the large deviation principle of discriminant statistics for Gaussian locally stationary processes. The large deviation theorems for quadratic forms and the log-likelihood ratio for a Gaussian locally stationary process with a mean function are proved. Their asymptotics are described by the large deviation rate functions. Next, the situation where processes are misspecified to be stationary is considered. In these
misspecified cases, the log-likelihood ratio discriminant statistics are formally constructed and the large deviation theorems of them are derived. Since they are mathematically complicated, they are evaluated and illustrated by numerical examples. We see that the misspecification of the process to be stationary seriously affects these discrimination.

The paper by T. Amano is titled "Asymptotic optimality of estimating function estimator for CHARN model." CHARN model is a famous and important model in the finance, which includes many financial time series models and can be used to model the return processes of assets. One of the most fundamental estimators for financial time series models is the conditional least squares CL estimator. However, recently, it was shown that the optimal estimating function estimator (G estimator) is better than CL estimator for some time series models in the sense of efficiency. This paper examines efficiencies of CL and G estimators for CHARN model and derives the condition that G estimator is asymptotically optimal.

The next paper titled “Optimal portfolio estimation for dependent financial returns with generalized empirical likelihood” by H. Ogata proposes to use the method of generalized empirical likelihood to find the optimal portfolio weights. The log-returns of assets are modeled by multivariate stationary processes rather than i.i.d. sequences. The variance of the portfolio is written by the spectral density matrix, and we seek the portfolio weights minimizing it. The illustration of this method to the real market index data is also given.

The paper titled “Statistically efficient construction of α-risk-minimizing portfolio” by Hiroyuki Taniai and Takayuki Shiohama proposes a semiparametrically efficient estimator for α-risk-minimizing portfolio weights. The optimal portfolio whose α-risk being minimized is formulated in a linear quantile regression problem. The authors apply the rank-based semiparametric method, using the signs and ranks of residual, to provide the efficient construction of the optimal portfolios. This efficiency gain is verified by Monte Carlo simulations and Empirical applications.

The paper titled “Estimation for non-Gaussian locally stationary processes with empirical likelihood method” by H. Ogata considered an estimation problem for non-Gaussian locally stationary processes by empirical likelihood. The parameter of interest is specified by the time-varying spectral moment condition, and it can express various important indices for locally stationary processes such as an autocorrelation function. The asymptotic distribution of maximum empirical likelihood estimator and empirical likelihood ratio test statistic are given based on the central limit theorem for locally stationary processes.

The paper “A simulation approach to statistical estimation of multiperiod optimal portfolios” by H. Shiraishi discusses a simulation-based method for solving discrete-time multiperiod portfolio choice problems under AR(1) return process. Based on the AR bootstrap, first, simulation sample paths of the random returns are generated. Then, for each sample path and each investment time, an optimal portfolio estimator, which optimizes a constant relative risk aversion (CRRA) utility function, is obtained.

The paper by J. Hirukawa entitled “On the causality between multiple locally stationary processes” is concerned with the concepts of dependence and causality which can describe the relations between multivariate time series. These concepts also appear to be useful when one is describing the properties of an engineering or econometric model. Although the measures of dependence and causality under stationary assumption are well established, empirical studies show that these measures are not constant in time. In this paper, the generalized measures of linear dependence and causality to multiple locally stationary processes are proposed. The measures of linear dependence, linear causality from one series to the other, and instantaneous linear feedback, at each time and each frequency, are given.
The paper titled “Optimal portfolios with end-of-period target” by H. Shiraishi et al. studies the estimation of optimal portfolios for a Reserve Fund with an end-of-period target, when the returns of the assets constituting the Reserve Fund portfolio follow two specifications. They focus the case when assets are split into short memory bonds and long memory equities or when returns of the distribution are heavy-tailed stable.

The next paper is by J. Hirukawa and M. Sadakata entitled “Least squares estimators for unit root processes with locally stationary disturbance.” It contains a discussion of various properties of the least squares estimators for unit root processes with locally stationary innovation processes. Since locally stationary process is not a stationary process, these models include different two types of nonstationarity, namely, unit root and locally stationarity. The locally stationary innovation has time-varying spectral structure, hence, it is suitable for describing the empirical financial time series data. Due to its nonstationarity, the least squares estimators of these models do not satisfy asymptotic normality. In this paper, the limiting distributions of the least squares estimators of unit root, near unit root, and general integrated processes with LSP innovation are derived.

The paper titled “Statistical portfolio estimation under the utility function depending on exogenous variables” by K. Hamada et al. develops the portfolio estimation under the situation. To estimate the optimal portfolio, a function of sample moments of the return process and sample cumulant between the return processes and exogenous variables is introduced. Then, its asymptotic distribution is derived, and the influence of exogenous variable on the return process is illuminated.

The paper titled “Statistical estimation for CAPM with long-memory dependence” by T. Amano et al. investigates the Capital Asset Pricing Model (CAPM) with time dimension. In view of time series analysis, authors describe the model of CAPM such as a regression model that market portfolio and the error process are long-memory process and correlated each other. They give a sufficient condition for the return of assets in the CAPM to be short memory. In this setting, they propose a two-stage least squares estimator for the regression coefficient and derive the asymptotic distribution. Some numerical studies are given.

This issue develops a modern, high-level statistical optimal estimation theory for portfolio coefficients, assuming that the financial returns are “dependent and non-Gaussian,” which opens up a new horizon in the field of portfolio estimation.

Masanobu Taniguchi  
Cathy W. S. Chen  
Junichi Hirukawa  
Hiroshi Shiraishi  
Kenichiro Tamaki  
David Veredas
Research Article

Large-Deviation Results for Discriminant Statistics of Gaussian Locally Stationary Processes

Junichi Hirukawa

Faculty of science, Niigata University, 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan

Correspondence should be addressed to Junichi Hirukawa, hirukawa@math.sc.niigata-u.ac.jp

Received 16 February 2012; Accepted 9 April 2012

Academic Editor: Kenichiro Tamaki

Copyright © 2012 Junichi Hirukawa. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper discusses the large-deviation principle of discriminant statistics for Gaussian locally stationary processes. First, large-deviation theorems for quadratic forms and the log-likelihood ratio for a Gaussian locally stationary process with a mean function are proved. Their asymptotics are described by the large deviation rate functions. Second, we consider the situations where processes are misspecified to be stationary. In these misspecified cases, we formally make the log-likelihood ratio discriminant statistics and derive the large deviation theorems of them. Since they are complicated, they are evaluated and illustrated by numerical examples. We realize the misspecification of the process to be stationary seriously affecting our discrimination.

1. Introduction

Consider a sequence of random variables $S_1, S_2, \ldots$ converging (in probability) to a real constant $c$. By this we mean that $\Pr(|S_T - c| > \varepsilon) \to 0$ as $T \to \infty$ for all $\varepsilon > 0$. The simplest setting in which to obtain large-deviation results is that considering sums of independent identically distributed (iid) random variables on the real line. For example, we would like to consider the large excursion probabilities of sums as the sample average:

$$S_T = T^{-1} \sum_{i=1}^{T} X_i$$

(1.1)

where the $X_i, i = 1, 2, \ldots,$ are i.i.d., and $T$ approaches infinity. Suppose that $E(X_i) = m$ exists and is finite. By the law of large numbers, we know that $S_T$ should be converging to $m$. Hence,
c is merely the expected value of the random process. It is often the case that not only does \( \Pr\{|S_T - c| > \varepsilon\} \) go to zero, but it does so exponentially fast. That is,

\[
\Pr\{|S_T - c| > \varepsilon\} \approx K(\varepsilon, c, T) \exp\{-TI(\varepsilon, c)\},
\]

where \( K(\varepsilon, c, T) \) is a slowly varying function of \( T \) (relative to the exponential), and \( I(\varepsilon, c) \) is a positive quantity. Loosely, if such a relationship is satisfied, we will say that the sequence \( \{S_n\} \) satisfies a large-deviation principle. Large-deviation theory is concerned primarily with determining the quantities \( I(\varepsilon, c) \) and (to a lesser extent) \( K(\varepsilon, c, T) \). The reason for the nomenclature is that for a fixed \( \varepsilon > 0 \) and a large index \( T \), a large-deviation from the nominal value occurs if \( |S_T - c| > \varepsilon \). Large-deviation theory can rightly be considered as a generalization or extension of the law of large numbers. The law of large numbers says that certain probabilities converge to zero. Large-deviation theory is concerned with the rate of convergence. Bucklew [1] describes the historical statements of large-deviation in detail.

There have been a few works on the large-deviation theory for time series data. Sato et al. [2] discussed the large-deviation theory of several statistics for short- and long-memory stationary processes. However, it is still hard to find the large-deviation results for nonstationary processes. Recently, Dahlhaus [3, 4] has formulated an important class of nonstationary processes with a rigorous asymptotic theory, which he calls locally stationary. A locally stationary process has a time-varying spectral density whose spectral structure changes smoothly with time. There are several papers which discuss discriminant analysis for locally stationary processes (e.g., Chandler and Polonik [5], Sakiyama and Taniguchi [6], and Hirukawa [7]). In this paper, we discuss the large-deviation theory of discriminant statistics of Gaussian locally stationary processes. In Section 2 we present the Gärtner-Ellis theorem which establishes a large-deviation principle of random variables based only upon convergence properties of the associated sequence of cumulant generating functions. Since no assumptions are made about the dependency structure of random variables, we can apply this theorem to non-stationary time series data. In Section 3, we deal with a Gaussian locally stationary process with a mean function. First, we prove the large-deviation principle for a general quadratic form of the observed stretch. We also give the large-deviation principle for the log-likelihood ratio and the misspecified log-likelihood ratio between two hypotheses. These fundamental statistics are important not only in statistical estimation and testing theory but in discriminant problems. The above asymptotics are described by the large-deviation rate functions. In our stochastic models, the rate functions are very complicated. Thus, in Section 4, we evaluate them numerically. They demonstrate that the misspecifications of non-stationary has serious effects. All the proofs of the theorems presented in Section 3 are given in the Appendix.

2. Gärtner-Ellis Theorem

Cramér’s theorem (e.g., Bucklew [1]) is usually credited with being the first large-deviation result. It gives the large-deviation principle for sums of independent identically distributed random variables. One of the most useful and surprising generalizations of this theorem is the one due to Gärtner [8] and, more recently, Ellis [9]. These authors established a large-deviation principle of random variables based only upon convergence properties of the associated sequence of moment generating functions \( \Phi(\omega) \). Their methods thus allow large-deviation results to be derived for dependent random processes such as Markov chains and
functionals of Gaussian random processes. Gärtner [8] assumed throughout that $\Phi(\omega) < \infty$ for all $\omega$. By extensive use of convexity theory, Ellis [9] relaxed this fairly stringent condition.

Suppose that we are given an infinite sequence of random variables $\{Y_T, T \in \mathbb{N}\}$. No assumptions are made about the dependency structure of this sequence. Define

$$
\psi_T(\omega) \equiv T^{-1} \log E\{\exp(\omega Y_T)\}.
$$

(2.1)

Now let us list two assumptions.

Assumption 2.1. $\psi(\omega) \equiv \lim_{T \to \infty} \psi_T(\omega)$ exists for all $\omega \in \mathbb{R}$, where we allow $\infty$ both as a limit value and as an element of the sequence $\{\psi_T(\omega)\}$.

Assumption 2.2. $\psi(\omega)$ is differentiable on $D_{\psi} \equiv \{\omega : \psi(\omega) < \infty\}$.

Define the large-deviation rate function by

$$
I(x) \equiv \sup_{\omega} \{\omega x - \psi(\omega)\};
$$

(2.2)

this function plays a crucial role in the development of the theory. Furthermore, define

$$
\psi'(D_{\psi}) \equiv \{\psi'(\omega) : \omega \in D_{\psi}\},
$$

(2.3)

where $\psi'$ indicates the derivative of $\psi$. Before proceeding to the main theorem, we first state some properties of this rate function.

Property 1. $I(x)$ is convex.

We remark that a convex function $I(\cdot)$ on the real line is continuous everywhere on $D_I \equiv \{x : I(x) < \infty\}$, the domain of $I(\cdot)$.

Property 2. $I(x)$ has its minimum value at $m = \lim_{T \to \infty} T^{-1}E(Y_T)$, and $I(m) = 0$.

We now state a simple form of a general large-deviation theorem which is known as the Gärtner and Ellis theorem (e.g., Bucklew [1]).

Lemma 2.3 (Gärtner-Ellis). Let $(a, b)$ be an interval with $[a, b] \cap D_I \neq \emptyset$. If Assumption 2.1 holds and $a < b$, then

$$
\limsup_{T \to \infty} T^{-1} \log \Pr\{T^{-1}Y_T \in [a, b]\} \leq -\inf_{x \in [a, b]} I(x).
$$

(2.4)

If Assumptions 2.1 and 2.2 hold and $(a, b) \subset \psi'(D_{\psi})$, then

$$
\liminf_{T \to \infty} T^{-1} \log \Pr\{T^{-1}Y_T \in (a, b)\} \geq -\inf_{x \in (a, b)} I(x).
$$

(2.5)

Large-deviation theorems are usually expressed as two separate limit theorem: an upper bound for closed sets and a lower bound for open sets. In the case of interval subsets
of $R$, it can be guaranteed that the upper bound equals the lower bound by the continuity of $I(\cdot)$. For the applications that we have in mind, the interval subsets will be sufficient.

3. Large-Deviation Results for Locally Stationary Processes

In this section, using the Gartner-Ellis theorem, we develop the large-deviation principle for some non-stationary time series statistics. When we deal with non-stationary processes, one of the difficult problems to solve is how to set up an adequate asymptotic theory. To overcome this problem, an important class of non-stationary process has been formulated in rigorous asymptotic framework by Dahlhaus [3, 4], called locally stationary processes. Locally stationary processes have time-varying densities, whose spectral structures smoothly change in time. We give the precise definition of locally stationary processes which is due to Dahlhaus [3, 4].

**Definition 3.1.** A sequence of stochastic processes $X_{t,T} (t = 1, \ldots, T; T \geq 1)$ is called locally stationary with transfer function $A^\circ$ and trend $\mu$ if there exists a representation:

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t) A^\circ_{t,T}(\lambda) d\xi(\lambda),$$

where

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

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

where $\text{cum}\{\ldots\}$ denotes the cumulant of $k$-th order, $\nu_1 = 0$, $\nu_2(\lambda) = 1$, $|\nu_k(\lambda_1, \ldots, \lambda_{k-1})| \leq \text{const}_k$ for all $k$ and $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period $2\pi$ extension of the Dirac delta function. To simplify the problem, we assume in this paper that the process $X_{t,T}$ is Gaussian, namely, we assume that $\nu_k(\lambda) = 0$ for all $k \geq 3$;

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

$$\sup_{t,\lambda} \left| A^\circ_{t,T}(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right| \leq KT^{-1},$$

for all $T$. $A(u, \lambda)$ and $\mu(u)$ are assumed to be continuous in $u$.

The function $f(u, \lambda) := |A(u, \lambda)|^2$ is called the time-varying spectral density of the process. In the following, we will always denote by $s$ and $t$ time points in the interval $[1, T]$, while $u$ and $v$ will denote time points in the rescaled interval $[0, 1]$, that is $u = t/T$.

We discuss the asymptotics away from the expectation of some statistics used for the problem of discriminating between two Gaussian locally stationary processes with specified
mean functions. Suppose that \( \{X_{i,T}, t = 1, \ldots, T; T \geq 1\} \) is a Gaussian locally stationary process which under the hypothesis \( \Pi_i \) has mean function \( \mu^{(i)}(u) \) and time-varying spectral density \( f^{(i)}(u, \lambda) \) for \( j = 1, 2 \). Let \( X_T = (X_{1,T}, \ldots, X_{T,T})' \) be a stretch of the series \( \{X_{i,T}\} \), and let \( p^{(j)}(\cdot) \) be the probability density function of \( X_T \) under \( \Pi_j (j = 1, 2) \). The problem is to classify \( X_T \) into one of two categories \( \Pi_1 \) and \( \Pi_2 \) in the case that we do not have any information on the prior probabilities of \( \Pi_1 \) and \( \Pi_2 \).

Set \( \mu_T^{(j)} = [\mu^{(j)}(1/T), \ldots, \mu^{(j)}(T/T)]' \) and \( \Sigma_T^{(j)} = \Sigma_T (A^{(j)}, A^{(j)}) \), where

\[
\Sigma_T (A, B) = \left\{ \int_0^\pi A_{s,T}^0 (\lambda) B_{s,T}^0 (-\lambda) \exp(i\lambda(s-t)) d\lambda \right\}_{s,t=1,\ldots,T}.
\]  

Initially, we make the following assumption.

**Assumption 3.2.** (i) We observe a realisation \( X_{1,T}, \ldots, X_{T,T} \) of a Gaussian locally stationary process with mean function \( \mu^{(i)} \) and transfer function \( A^{(i)} \), under \( \Pi_j, j = 1, 2 \);

(ii) the \( A^{(j)}(u, \lambda) \) are uniformly bounded from above and below, and are differentiable in \( u \) and \( \lambda \) with uniformly continuous derivatives \( (\partial/\partial u)(\partial/\partial \lambda) A^{(j)} \);

(iii) the \( \mu^{(j)}(u) \) are differentiable in \( u \) with uniformly continuous derivatives.

In time series analysis, the class of statistics which are quadratic forms of \( X_T \) is fundamental and important. This class of statistics includes the first-order terms (in the expansion with respect to \( T \)) of quasi-Gaussian maximum likelihood estimator (QMLE), tests and discriminant statistics, and so forth. Assume that \( G^* \) is the transfer function of a locally stationary process, where the corresponding \( G \) satisfies Assumption 3.2 (ii) and \( g(u) \) is a continuous function of \( u \) which satisfies Assumption 3.2 (iii), if we replace \( A^{(i)} \) by \( G \) and \( \mu^{(j)}(u) \) by \( g(u) \), respectively. And set \( G_T \equiv \Sigma_T(G, G), f_G(u, \lambda) \equiv |G(u, \lambda)|^2, g_T \equiv \{ g(1/T), \ldots, g(T/T) \}' \) and \( Q_T \equiv X^T_T X_T + g_T' X_T \). Henceforth, \( E^{(j)}(\cdot) \) stands for the expectation with respect to \( p^{(j)}(\cdot) \).

Set \( S_T^{(j)}(Q) = Q_T - E^{(j)}(Q_T) \) for \( j = 1, 2 \). We first prove the large-deviation theorem for this quadratic form \( Q_T \) of \( X_T \). All the proofs of the theorems are in the Appendix.

**Theorem 3.3.** Let Assumption 3.2 hold. Then under \( \Pi_1 \),

\[
\lim_{T \to \infty} T^{-1} \log \Pr \left\{ T^{-1} S_T^{(1)}(Q) > x \right\} = \inf_{\omega} \left\{ \psi_Q (\omega; f^{(1)}) - \omega \max(x, 0) \right\},
\]  

and under \( \Pi_2 \),

\[
\lim_{T \to \infty} T^{-1} \log \Pr \left\{ T^{-1} S_T^{(2)}(Q) < x \right\} = \inf_{\omega} \left\{ \psi_Q (\omega; f^{(2)}) - \omega \min(x, 0) \right\},
\]

where for \( j = 1, 2 \), \( \psi_Q (\omega; f^{(j)}) \) equals

\[
\frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \left[ \log \frac{f_G(u, \lambda)}{f_G(u, \lambda) - 2\omega f^{(j)}(u, \lambda)} - \frac{2\omega f^{(j)}(u, \lambda)}{f_G(u, \lambda)} \right] d\lambda du.
\]  

\[ \psi_Q (\omega; f^{(j)}) \] equals
Next, one considers the log-likelihood ratio statistics. It is well known that the log-likelihood ratio criterion:

$$\Lambda_T \equiv \log \frac{p^{(2)}(X_T)}{p^{(1)}(X_T)} \quad (3.8)$$

gives the optimal discrimination rule in the sense that it minimizes the probability of misdiscrimination (Anderson [10]). Set $S_T^{(j)}(\Lambda) \equiv \Lambda_T - E^{(j)}(\Lambda_T)$ for $j = 1, 2$. For discrimination problem one gives the large-deviation principle for $\Lambda_T$.

**Theorem 3.4.** Let Assumption 3.2 hold. Then under $\Pi_1$,

$$\lim_{T \to \infty} T^{-1} \log \Pr_1 \{ T^{-1} S_T^{(1)}(\Lambda) > x \} = \inf_{\omega} \left\{ \varphi_L(\omega; f^{(1)}, f^{(2)}) - \omega \max(x, 0) \right\}, \quad (3.9)$$

where $\varphi_L(\omega; f^{(1)}, f^{(2)})$ equals

$$\frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^{2\pi} \left[ \log \frac{f^{(2)}(u, \lambda)}{(1 - \omega)f^{(2)}(u, \lambda) + \omega f^{(1)}(u, \lambda)} + \omega \left\{ \frac{f^{(1)}(u, \lambda)}{f^{(2)}(u, \lambda)} - 1 \right\} \right.$$

$$\left. + \frac{\omega^2 \left( \mu^{(1)}(u) - \mu^{(2)}(u) \right)^2 f^{(1)}(u, 0)}{2\pi \left[ (1 - \omega)f^{(2)}(u, 0) + \omega f^{(1)}(u, 0) \right] f^{(2)}(u, 0)} \right] d\lambda du. \quad (3.10)$$

Similarly, under $\Pi_2$,

$$\lim_{T \to \infty} T^{-1} \log \Pr_2 \{ T^{-1} S_T^{(2)}(\Lambda) < x \} = \inf_{\omega} \left\{ \varphi_L(-\omega; f^{(2)}, f^{(1)}) - \omega \min(x, 0) \right\}. \quad (3.11)$$

In practice, misspecification occurs in many statistical problems. We consider the following three situations. Although actually $\{X_{t,T}\}$ has the time-varying mean functions $\mu^{(j)}(u)$ and the time-varying spectral densities $f^{(j)}(u, \lambda)$, under $\Pi_j$, $j = 1, 2$, respectively,

(i) the mean functions are misspecified to $\mu^{(j)}(u) \equiv 0$, $j = 1, 2$;

(ii) the spectral densities are misspecified to $f^{(j)}(u, \lambda) \equiv f^{(j)}(0, \lambda)$, $j = 1, 2$;

(iii) the mean functions and the spectral densities are misspecified to $\mu^{(j)}(u) \equiv 0$ and $f^{(j)}(u, \lambda) \equiv f^{(j)}(0, \lambda)$, $j = 1, 2$. Namely, $X_T$ is misspecified to stationary.
In each misspecified case, one can formally make the log-likelihood ratio in the form:

\[
M_{1,T} = \frac{1}{2} \left[ \log \left( \frac{\Sigma^{(1)}}{\Sigma^{(2)}} \right) + X_T' \left( \Sigma_t^{(1)} - \Sigma_t^{(2)} \right) X_T \right],
\]

\[
M_{2,T} = \frac{1}{2} \left[ \log \left( \frac{\Sigma^{(1)}}{\Sigma^{(2)}} \right) + \left( X_T - \mu_T^{(1)} \right)' \Sigma_t^{(1)} \left( X_T - \mu_T^{(1)} \right) \right. \\
\left. - \left( X_T - \mu_T^{(2)} \right)' \Sigma_t^{(2)} \left( X_T - \mu_T^{(2)} \right) \right],
\]

\[
M_{3,T} = \frac{1}{2} \left[ \log \left( \frac{\Sigma^{(1)}}{\Sigma^{(2)}} \right) + X_T' \left( \Sigma_t^{(1)} - \Sigma_t^{(2)} \right) X_T \right],
\]

where

\[
\Sigma_t^{(j)} = \left\{ \int_{-\pi}^{\pi} \exp(i\lambda(t-s)) f^{(j)}(0,\lambda) d\lambda \right\}_{s,t=1,...,T}.
\]

Set \( S_T^{(j)}(M_k) \equiv M_{k,T} - E^{(j)}(M_{k,T}) \) for \( j = 1,2 \) and \( k = 1,2,3 \). The next result is a large-deviation theorem for the misspecified log-likelihood ratios \( M_{k,T} \). It is useful in investigating the effect of misspecification.

**Theorem 3.5.** Let Assumption 3.2 hold. Then under \( \Pi_1 \),

\[
\lim_{T \to \infty} T^{-1} \log \Pr \left\{ T^{-1} S_T^{(1)}(M_k) > x \right\} = \inf_{D} \left\{ \eta_{M_k}(\omega; f^{(1)}, f^{(2)}, \mu^{(1)}, \mu^{(2)}) - \omega \max(x,0) \right\},
\]

where \( \eta_{M}(\omega; f^{(1)}, f^{(2)}, \mu^{(1)}, \mu^{(2)}) \) equals

\[
\frac{1}{4\pi} \int_{-\pi}^{\pi} \left[ \log \left( \frac{f^{(2)}(u,\lambda)}{1-\omega f^{(2)}(u,\lambda) + \omega f^{(1)}(u,\lambda)} + \omega \left( \frac{f^{(1)}(u,\lambda)}{f^{(2)}(u,\lambda)} - 1 \right) \right) \right. \\
\left. \frac{\omega \mu^{(1)}(u) \left( f^{(1)}(u,0) - f^{(2)}(u,0) \right)^2}{2\pi \left( 1-\omega f^{(2)}(u,0) + \omega f^{(1)}(u,0) \right) f^{(1)}(u,0) f^{(2)}(u,0)} \right] d\lambda du,
\]
\[ q_{M_1}(\omega; f^{(1)}, f^{(2)}, \mu^{(1)}, \mu^{(2)}) \text{ equals} \]

\[
\frac{1}{4\pi} \int_{0}^{1} \int_{-\pi}^{\pi} \left[ \log \frac{f^{(1)}(0, \lambda) f^{(2)}(0, \lambda)}{f^{(1)}(0, \lambda) f^{(2)}(0, \lambda) - \omega f^{(1)}(u, \lambda) \{ f^{(2)}(0, \lambda) - f^{(1)}(0, \lambda) \}} \\
+ \omega \left\{ \frac{f^{(1)}(u, \lambda)}{f^{(2)}(0, \lambda)} - \frac{f^{(1)}(u, \lambda)}{f^{(1)}(0, \lambda)} \right\} \\
+ \frac{\omega^2 \mu^{(2)}(u) - \mu^{(1)}(u) \}^2 f^{(1)}(u, 0) f^{(2)}(0, 0) / f^{(1)}(0, 0)}{2\pi [f^{(1)}(0, 0) f^{(2)}(0, 0) - \omega f^{(1)}(u, 0) \{ f^{(2)}(0, 0) - f^{(1)}(0, 0) \}]} \right] d\lambda du \]

and \( q_{M_2}(\omega; f^{(1)}, f^{(2)}, \mu^{(1)}, \mu^{(2)}) \) equals

\[
\frac{1}{4\pi} \int_{0}^{1} \int_{-\pi}^{\pi} \left[ \log \frac{f^{(1)}(0, \lambda) f^{(2)}(0, \lambda)}{f^{(1)}(0, \lambda) f^{(2)}(0, \lambda) - \omega f^{(1)}(u, \lambda) \{ f^{(2)}(0, \lambda) - f^{(1)}(0, \lambda) \}} \\
+ \omega \left\{ \frac{f^{(1)}(u, \lambda)}{f^{(2)}(0, \lambda)} - \frac{f^{(1)}(u, \lambda)}{f^{(1)}(0, \lambda)} \right\} \\
+ \frac{\omega^2 \mu^{(1)}(u) \{ f^{(1)}(0, 0) - f^{(2)}(0, 0) \}^2 / \{ f^{(1)}(0, 0) f^{(2)}(0, 0) \}}{2\pi [f^{(1)}(0, 0) f^{(2)}(0, 0) - \omega f^{(1)}(u, 0) \{ f^{(2)}(0, 0) - f^{(1)}(0, 0) \}]} \right] d\lambda du. \]

Similarly, under \( \Pi_2 \),

\[
\lim_{T \to \infty} T^{-1} \log \Pr \{ T^{-1} S_T^2(M_k) < x \} = \inf_{\omega} \{ q_{M_1}(\omega; f^{(1)}, f^{(2)}, \mu^{(1)}, \mu^{(2)}) - \omega \min(x, 0) \}. \]

Now, we turn to the discussion of our discriminant problem of classifying \( X_T \) into one of two categories described by two hypotheses as follows:

\[ \Pi_1 : \mu^{(1)}(u), f^{(1)}(u, \lambda), \quad \Pi_2 : \mu^{(2)}(u), f^{(2)}(u, \lambda). \]
We use $\Lambda_T$ as the discriminant statistic for the problem (3.19), namely, if $\Lambda_T > 0$ we assign $X_T$ into $\Pi_2$, and otherwise into $\Pi_1$. Taking $x = -\lim_{T \to \infty} T^{-1} E_1^2(\Lambda_T)$ in (3.9), we can evaluate the probability of misdiscrimination of $X_T$ from $\Pi_1$ into $\Pi_2$ as follows:

$$P(2 | 1) \equiv \Pr_1 \{ \Lambda_T > 0 \} \approx \exp \left[ T \inf_{\omega} \left\{ \frac{1}{4\pi} \int_{0}^{1} \int_{-\pi}^{\pi} \log \frac{f^{(1)}(u, \lambda) \omega f^{(2)}(u, \lambda)^{1-\omega}}{(1-\omega) f^{(2)}(u, \lambda) + \omega f^{(1)}(u, \lambda)} + \frac{\omega(\omega-1)\{\mu^{(1)}(u) - \mu^{(2)}(u)\}^2}{2\pi (1-\omega)f^{(2)}(u,0) + \omega f^{(1)}(u,0)} \right\} du d\lambda \right].$$  

(3.20)

Thus, we see that the rate functions play an important role in the discriminant problem.

4. Numerical Illustration for Nonstationary Processes

We illustrate the implications of Theorems 3.4 and 3.5 by numerically evaluating the large-deviation probabilities of the statistics $\Lambda_T$ and $M_{k,T}$, $k = 1, 2, 3$ for the following hypotheses:

- (Stationary white noise) $\Pi_1 : \mu^{(1)}(u) \equiv 0, \quad f^{(1)}(u, \lambda) \equiv 1,$
- (Time-varying AR(1)) $\Pi_2 : \mu^{(2)}(u) = \mu(u), \quad f^{(2)}(u, \lambda) = \frac{\sigma(u)^2}{[1 - a(u)e^{i\lambda}]^2}$

where $\mu(u) = (1/2) \exp(-u^2)$, $\sigma(u) = (1/2) \exp(-u-1)^2$ and $a(u) = (1/2) \exp(-4u-1/2)^2$, $u \in [0, 1]$, respectively. Figure 1 plots the mean function $\mu(u)$ (the solid line), the coefficient functions $\sigma(u)$ (the dashed line), and $a(u)$ (the dotted line). The time-varying spectral density $f^{(2)}(u, \lambda)$ is plotted in Figure 2.

From these figures, we see that the magnitude of the mean function is large at $u$ close to 0, while the magnitude of the time-varying spectral density is large at $u$ close to 1.

Specifically, we use the formulae in those theorems concerning $\Pi_2$ to evaluate the limits of the large-deviation probabilities:

$$\text{LDP}(\Lambda) = \lim_{T \to \infty} T^{-1} \log \Pr_2 \left\{ T^{-1} S^{(2)}_T(\Lambda) < x \right\},$$

$$\text{LDP}(M_k) = \lim_{T \to \infty} T^{-1} \log \Pr_2 \left\{ T^{-1} S^{(2)}_T(M_k) < x \right\}, \quad k = 1, 2, 3.$$  

(4.2)

Though the result is an asymptotic theory, we perform the simulation with a limited sample size. Therefore, we use some levels of $x$ to expect fairness, that is, we take $x = -0.1, -1, -10$. The results are listed in Table 1.

For each value $x$, the large-deviation rate of $\Lambda_T$ is the largest and that of $M_{3,T}$ is the smallest. Namely, we see that the correctly specified case is the best, and on the other hand the misspecified to stationary case is the worst. Furthermore, the large-deviation rates $-\text{LDP}(M_2)$...
Figure 1: The mean function $\mu(u)$ (the solid line), the coefficient functions $\sigma(u)$ (the dashed line), and $\alpha(u)$ (the dotted line).

Figure 2: The time-varying spectral density $f^{(2)}(u, \lambda)$.

Table 1: The limits of the large-deviation probabilities of $\Lambda_T$ and $M_{k,T}$, $k = 1, 2, 3$.

|       | $x = -0.1$ | $x = -1$   | $x = -10$  |
|-------|------------|------------|------------|
| LDP($\Lambda$) | -0.012078  | -0.562867  | -9.460066  |
| LDP($M_1$)     | -0.009895  | -0.486088  | -8.857859  |
| LDP($M_2$)     | -0.000348  | -0.026540  | -0.703449  |
| LDP($M_3$)     | -0.000290  | -0.022313  | -0.629251  |

and $-\text{LDP}(M_2)$ are significantly small, comparing with $-\text{LDP}(M_1)$. This fact implies that the misspecification of the spectral density to be constant in the time seriously affects the large-deviation rate.

Figures 3, 4, 5, and 6 show the large-deviation probabilities of $\Lambda_T$ and $M_{k,T}$, $k = 1, 2, 3$, for $x = -1$, at each time $u$ and frequency $\lambda$.

We see that the large-deviation rate of $\Lambda_T$ keeps the almost constant value at all the time $u$ and frequency $\lambda$. On the other hand, that of $M_{1,T}$ is small at $u$ close to 0 and those of $M_{2,T}$ and $M_{3,T}$ are small at $u$ close to 1 and $\lambda$ close to 0. That is, the large-deviation probability of $M_{1,T}$ is violated by the large magnitude of the mean function, while those of $M_{2,T}$ and
Figure 3: The time-frequency plot of the large-deviation probabilities of $\Lambda_T$.

Figure 4: The time-frequency plot of the large-deviation probabilities of $M_{1,T}$.

$M_{3,T}$ are violated by that of the time-varying spectral density. Hence, we can conclude the misspecifications seriously affect our discrimination.

Appendix

We sketch the proofs of Theorems 3.3–3.5. First, we summarize the assumptions used in this paper.

Assumption A.1. (i) Suppose that $A : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}$ is a $2\pi$-periodic function with $A(u, \lambda) = \overline{A(u, -\lambda)}$ which is differentiable in $u$ and $\lambda$ with uniformly bounded derivative $(\partial / \partial u)(\partial / \partial \lambda)A$. $f_A(u, \lambda) \equiv |A(u, \lambda)|^2$ denotes the time-varying spectral density. $A_{t,T}^\circ : \mathbb{R} \rightarrow \mathbb{C}$ are $2\pi$-periodic functions with

$$
\sup_{t,\lambda} \left| A_{t,T}^\circ (\lambda) - A\left( \frac{t}{T}, \lambda \right) \right| \leq KT^{-1}, \quad (A.1)
$$

(ii) suppose that $\mu : [0, 1] \rightarrow \mathbb{R}$ is differentiable with uniformly bounded derivative.
We introduce the following matrices (see Dahlhaus [4] p.154 for the detailed definition):

$$W_T(\phi) = \frac{S}{N} \sum_{j=1}^{M} K_T^{(j)} W_T^{(j)} (\phi) K_T^{(j)},$$  \hspace{1cm} (A.2)

where

$$W_T^{(j)} (\phi) = \left\{ \int_{-\pi}^{\pi} \phi(u_j, \lambda) \exp(i\lambda(k-l)) d\lambda \right\}_{k,l=1,\ldots,L_j},$$  \hspace{1cm} (A.3)

and $K_T^{(j)} = (0_{j1}, I_{j2}, 0_{j2})$. According to Lemmata 4.4 and 4.7 of Dahlhaus [4], we can see that

$$\|\Sigma_T(A, A)\| \leq C + o(1), \hspace{1cm} \left\| \Sigma_T(A, A)^{-1} \right\| \leq C + o(1),$$  \hspace{1cm} (A.4)

and $W_T(f_A)$ and $W_T\{4\pi^2 f_A\}^{-1}$ are the approximations of $\Sigma_T(A, A)$ and $\Sigma_T(A, A)^{-1}$, respectively. We need the following lemmata which are due to Dahlhaus [3, 4]. Lemma A.2 is Lemma A.5 of Dahlhaus [3] and Lemma A.3 is Theorem 3.2 (ii) of Dahlhaus [4].
**Lemma A.2.** Let $k \in \mathbb{N}$, $A_i$, $B_i$ fulfill Assumption A.1 (i) and $\mu_1$, $\mu_2$ fulfill Assumption A.1 (ii). Let $\Sigma_i = \Sigma_T(A_i, A_i)$ or $W_T(f_A)$. Furthermore, let $\Gamma_i = \Sigma_T(B_i, B_i)$, $W_T((4\pi^2)^{-1} f_B)$ or $\Gamma_i^{-1} = W_T((4\pi^2 f_B)^{-1})$. Then we have

\[
T^{-1} \text{tr} \left[ \prod_{i=1}^{k} \Gamma_i^{-1} \Sigma_i \right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \prod_{i=1}^{k} \frac{f_A(u, \lambda)}{f_B(u, \lambda)} \right\} d\lambda du + O\left(T^{-1/2} \log^{2k+2} T\right),
\]

\[
T^{-1} \mu_1^{T} \left[ \prod_{i=1}^{k-1} \Gamma_i^{-1} \Sigma_i \right] \Gamma_k^{-1} \mu_2^{T} = \frac{1}{2\pi} \int_{0}^{1} \left\{ \prod_{i=1}^{k-1} \frac{f_A(u, 0)}{f_B(u, 0)} \right\} f_B(u, 0)^{-1} \mu_1(u) \mu_2(u) du + O\left(T^{-1/2} \log^{2k+2} T\right). \tag{A.5}
\]

**Lemma A.3.** Let $D^\ast$ be the transfer function of a locally stationary process $\{Z_t, T\}$, where the corresponding $D$ is bounded from below and has uniformly bounded derivative $(\partial/\partial u)(\partial/\partial \lambda)D$. $f_D(u, \lambda) \equiv |D(u, \lambda)|^2$ denotes the time-varying spectral density of $Z_t$. Then, for $\Sigma_T(d) \equiv \Sigma_T(D, D)$, we have

\[
\lim_{T \to \infty} T^{-1} \log |\Sigma_T(d)| = \frac{1}{2\pi} \int_{0}^{1} \int_{-\pi}^{\pi} \log 2\pi f_D(u, \lambda) d\lambda du. \tag{A.6}
\]

We also remark that if $U_T$ and $V_T$ are real nonnegative symmetric matrices, then

\[
\text{tr}\{U_T V_T\} \leq \text{tr}\{U_T\} \|V_T\|. \tag{A.7}
\]

**Proof of Theorems 3.3–3.5.** We need the cumulant generating function of the quadratic form in normal variables $X_T \sim \mathcal{N}(\nu_T, \Sigma_T)$. It is known that the quadratic form $S_T^{(j)} = X_T^\prime H_T X_T + h_T^\prime X_T - E^{(j)}(X_T^\prime H_T X_T + h_T^\prime X_T)$ has cumulant generating function $\log E^{(j)}(e^{\omega S_T^{(j)}})$ equals to

\[
\begin{align*}
&-\frac{1}{2} \log |\Sigma_T^{(j)}| - \frac{1}{2} \log |\Sigma_T^{(j)} - 2\omega H_T| - \omega \text{tr} \left[ H_T \Sigma_T^{(j)} \right] \\
&+ \frac{1}{2} \omega \left\{ h_T + 2H_T \mu_T^{(j)} \right\}^\prime \left( \Sigma_T^{(j)} - 2\omega H_T \right)^{-1} \left\{ h_T + 2H_T \mu_T^{(j)} \right\} \tag{A.8}
\end{align*}
\]
We prove Theorem 3.5 for respective cases:

\[ H_T(Q) = G_T^{-1}, \quad h_T(Q) = g_T; \]

\[ H_T(\Lambda) = \frac{1}{2} \left( \Sigma_T^{(1)} - \Sigma_T^{(2)} \right), \quad h_T(\Lambda) = -\Sigma_T^{(1)} \mu_T^{(1)} + \Sigma_T^{(2)} \mu_T^{(2)}; \]

\[ H_T(M_1) = \frac{1}{2} \left( \Sigma_T^{(1)} - \Sigma_T^{(2)} \right), \quad h_T(M_1) = 0; \]

\[ H_T(M_2) = \frac{1}{2} \left( \Sigma_T^{(1)} - \Sigma_T^{(2)} \right), \quad h_T(M_2) = -\Sigma_T^{(1)} \mu_T^{(1)} + \Sigma_T^{(2)} \mu_T^{(2)}; \]

\[ H_T(M_3) = \frac{1}{2} \left( \Sigma_T^{(1)} - \Sigma_T^{(2)} \right), \quad h_T(M_3) = 0. \]

We prove Theorem 3.5 for \( M_{3,T} \) (under \( \Pi_1 \)) only. Theorems 3.3 and 3.4 are similarly obtained. In order to use the Gärtner-Ellis theorem, consider

\[ \varphi_T^{(j)}(\omega) \equiv T^{-1} \log \left[ \mathbb{E}^{(j)} \left\{ \exp \left( \omega S_T^{(j)}(M_3) \right) \right\} \right]. \]  

(A.10)

Setting \( H_T = H_T(M_3) \) and \( h_T = h_T(M_3) \) in (A.8), we have under \( \Pi_1 \) the following:

\[ \varphi_T^{(1)}(\omega) = -(2T)^{-1} \left[ \log |\Sigma_T^{(1)}| - \log |\Sigma_T^{(2)}| - \omega \left( \Sigma_T^{(1)} - \Sigma_T^{(2)} \right) \right] \]

\[ - \omega \text{ tr} \left\{ \left( \Sigma_T^{(1)} - \Sigma_T^{(2)} \right) \Sigma_T^{(1)} \right\} + \omega^2 \mu_T^{(1)} \left\{ \Sigma_T^{(1)} - \Sigma_T^{(2)} \right\} \]

\[ \times \left\{ \Sigma_T^{(1)} - \omega \left( \Sigma_T^{(1)} - \Sigma_T^{(2)} \right) \right\}^{-1} \left\{ \Sigma_T^{(1)} - \Sigma_T^{(2)} \right\} \mu_T^{(1)}. \]  

(A.11)

Using the inequality (A.7), we then replace \( \left\{ \Sigma_T^{(1)} - \omega \left( \Sigma_T^{(1)} - \Sigma_T^{(2)} \right) \right\} \) by \( W_T(f_0^{(1)} f_0^{(2)} - \omega f_0^{(1)} f_0^{(2)} f_0^{(1)} f_0^{(2)}) \), where \( f_0^{(j)} \) denote \( f^{(j)}(0, \lambda), j = 1, 2 \), that is, we obtain the approximation

\[ - (2T)^{-1} \left[ \log |\Sigma_T^{(1)}| - \log W_T \left( \frac{f_0^{(1)} f_0^{(2)} - \omega f_0^{(1)} f_0^{(2)} f_0^{(1)} f_0^{(2)}}{4 \pi^2 f_0^{(1)} f_0^{(2)}} \right) \right] - \omega \text{ tr} \left\{ \left( \Sigma_T^{(1)} - \Sigma_T^{(2)} \right) \Sigma_T^{(1)} \right\} \]

\[ + \omega^2 \mu_T^{(1)} \left\{ \Sigma_T^{(1)} - \Sigma_T^{(2)} \right\} \left\{ W_T \left( \frac{f_0^{(1)} f_0^{(2)} - \omega f_0^{(1)} f_0^{(2)} f_0^{(1)} f_0^{(2)}}{4 \pi^2 f_0^{(1)} f_0^{(2)}} \right) \right\}^{-1} \]

\[ \times \left\{ \Sigma_T^{(1)} - \Sigma_T^{(2)} \right\} \mu_T^{(1)}. \]  

(A.12)
In view of Lemmas A.2 and A.3, the above $\psi_T^{(1)}(\omega)$ converges to $\psi_{M_3}$, given in Theorem 3.5. Clearly, $\psi_{M_3}$ exists for $\omega \in D_{\psi_{M_3}} = \{ \omega : 1 - \omega^* \{ (f^{(1)}(u, \lambda)/f^{(1)}(0, \lambda)) - (f^{(1)}(u, \lambda)/f^{(2)}(0, \lambda)) \} > 0 \}$ and is convex and continuously differentiable with respect to $\omega$. For a sequence $\{\omega_m\} \to \omega_0 \in \partial D_{\psi_{M_3}}$ as $m \to \infty$, we can show that

$$\frac{\partial \psi_{M_3}(\omega_m; f^{(1)}, f^{(2)}, \mu^{(1)}, \mu^{(2)})}{\partial \omega} \to \infty, \quad \frac{\partial \psi_{M_3}(0; f^{(1)}, f^{(2)}, \mu^{(1)}, \mu^{(2)})}{\partial \omega} = 0. \quad (A.13)$$

Hence, $\psi_{M_3}'(D_{\psi_{M_3}}) \supset (x, \infty)$ for every $x > 0$. Application of the Gärtner-Ellis theorem completes the proof.

**Acknowledgments**

The author would like to thank the referees for their many insightful comments, which improved the original version of this paper. The author would also like to thank Professor Masanobu Taniguchi who is the lead guest editor of this special issue for his efforts and celebrate his sixtieth birthday.

**References**

[1] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*, Wiley, New York, NY, USA, 1990.

[2] T. Sato, Y. Kakizawa, and M. Taniguchi, “Large deviation results for statistics of short- and long-memory Gaussian processes,” *Australian & New Zealand Journal of Statistics*, vol. 40, no. 1, pp. 17–29, 1998.

[3] R. Dahlhaus, “Maximum likelihood estimation and model selection for locally stationary processes,” *Journal of Nonparametric Statistics*, vol. 6, no. 2-3, pp. 171–191, 1996.

[4] R. Dahlhaus, “On the Kullback-Leibler information divergence of locally stationary processes,” *Stochastic Processes and their Applications*, vol. 62, no. 1, pp. 139–168, 1996.

[5] G. Chandler and W. Polonik, “Discrimination of locally stationary time series based on the excess mass functional,” *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 240–253, 2006.

[6] K. Sakiyama and M. Taniguchi, “Discriminant analysis for locally stationary processes,” *Journal of Multivariate Analysis*, vol. 90, no. 2, pp. 282–300, 2004.

[7] J. Hirukawa, “Discriminant analysis for multivariate non-Gaussian locally stationary processes,” *Scientiae Mathematicae Japonicae*, vol. 60, no. 2, pp. 357–380, 2004.

[8] J. Gärtner, “On large deviations from an invariant measure,” *Theory of Probability and its Applications*, vol. 22, no. 1, pp. 24–39, 1977.

[9] R. S. Ellis, “Large deviations for a general class of random vectors,” *The Annals of Probability*, vol. 12, no. 1, pp. 1–12, 1984.

[10] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, NY, USA, 2nd edition, 1984.

[11] A. M. Mathai and S. B. Provost, *Quadratic Forms in Random Variables: Theory and Applications*, Marcel Dekker, New York, NY, USA, 1992.
Asymptotic Optimality of Estimating Function Estimator for CHARN Model

Tomoyuki Amano

Faculty of Economics, Wakayama University, 930 Sakaedani, Wakayama 640-8510, Japan

Correspondence should be addressed to Tomoyuki Amano, tomchami@center.wakayama-u.ac.jp

Received 15 February 2012; Accepted 9 April 2012

Academic Editor: Hiroshi Shiraishi

CHARN model is a famous and important model in the finance, which includes many financial time series models and can be assumed as the return processes of assets. One of the most fundamental estimators for financial time series models is the conditional least squares estimator. However, recently, it was shown that the optimal estimating function estimator (estimator) is better than CL estimator for some time series models in the sense of efficiency. In this paper, we examine efficiencies of CL and G estimators for CHARN model and derive the condition that G estimator is asymptotically optimal.

1. Introduction

The conditional least squares (CL) estimator is one of the most fundamental estimators for financial time series models. It has the two advantages which can be calculated with ease and does not need the knowledge about the innovation process (i.e., error term). Hence this convenient estimator has been widely used for many financial time series models. However, Amano and Taniguchi [1] proved it is not good in the sense of the efficiency for ARCH model, which is the most famous financial time series model.

The estimating function estimator was introduced by Godambe ([2, 3]) and Hansen [4]. Recently, Chandra and Taniguchi [5] constructed the optimal estimating function estimator (G estimator) for the parameters of the random coefficient autoregressive (RCA) model, which was introduced to describe occasional sharp spikes exhibited in many fields and ARCH model based on Godambe's asymptotically optimal estimating function. In Chandra and Taniguchi [5], it was shown that G estimator is better than CL estimator by simulation. Furthermore, Amano [6] applied CL and G estimators to some important time series models (RCA, GARCH, and nonlinear AR models) and proved that G estimator is
better than CL estimator in the sense of the efficiency theoretically. Amano [6] also derived the conditions that G estimator becomes asymptotically optimal, which are not strict and natural.

However, in Amano [6], G estimator was not applied to a conditional heteroscedastic autoregressive nonlinear model (denoted by CHARN model). CHARN model was proposed by Härdle and Tsybakov [7] and Härdle et al. [8], which includes many financial time series models and is used widely in the finance. Kanai et al. [9] applied G estimator to CHARN model and proved its asymptotic normality. However, Kanai et al. [9] did not compare efficiencies of CL and G estimators and discuss the asymptotic optimality of G estimator theoretically. Since CHARN model is the important and rich model, which includes many financial time series models and can be assumed as return processes of assets, more investigation of CL and G estimators for this model are needed. Hence, in this paper, we compare efficiencies of CL and G estimators and investigate the asymptotic optimality of G estimator for this model.

This paper is organized as follows. Section 2 describes definitions of CL and G estimators. In Section 3, CL and G estimators are applied to CHARN model, and efficiencies of these estimators are compared. Furthermore, we derive the condition of asymptotic optimality of G estimator. We also compare the mean squared errors of $\hat{\theta}_{CL}$ and $\hat{\theta}_{G}$ by simulation in Section 4. Proofs of Theorems are relegated to Section 5. Throughout this paper, we use the following notation: $|A|$: Sum of the absolute values of all entries of A.

### 2. Definitions of CL and G Estimators

One of the most fundamental estimators for parameters of the financial time series models is the conditional least squares (CL) estimator $\hat{\theta}_{CL}$ introduced by Tjøstheim [10], and it has been widely used in the finance. $\hat{\theta}_{CL}$ for a time series model $\{X_t\}$ is obtained by minimizing the penalty function

$$Q_n(\theta) \equiv \sum_{t=m+1}^{n} (X_t - E[X_t | F_t(m)])^2,$$

where $F_t(m)$ is the $\sigma$-algebra generated by $\{X_s : t - m \leq s \leq t - 1\}$, and $m$ is an appropriate positive integer (e.g., if $\{X_t\}$ follows $k$th-order nonlinear autoregressive model, we can take $m = k$). CL estimator generally has a simple expression. However, it is not asymptotically optimal in general (see Amano and Taniguchi [1]).

| $a$  | $\theta_{CL}(n = 100)$ | $\theta_{CL}(n = 100)$ | $\theta_{CL}(n = 200)$ | $\theta_{CL}(n = 300)$ | $\theta_{CL}(n = 300)$ |
|------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 0.1  | 0.01103311             | 0.01135393             | 0.01035005             | 0.00596135             | 0.00586717             |
| 0.2  | 0.01096804             | 0.01113519             | 0.01006857             | 0.00565094             | 0.00555699             |
| 0.3  | 0.00371269             | 0.00376673             | 0.00351314             | 0.00356603             | 0.00359798             |

Table 1: MSE of $\hat{\theta}_{CL}$ and $\hat{\theta}_{G}$ for the parameter $a$ in (4).
In this section, we discuss the asymptotics of \( \hat{\theta}_G \) constructed \( \hat{\theta}_G \) based on Godambe’s asymptotically optimal estimating function for RCA and ARCH models. For the definition of \( \hat{\theta}_G \), we prepare the following estimating function \( G(\theta) \). Let \( \{X_i\} \) be a stochastic process which is depending on the \( k \)-dimensional parameter \( \theta \). Let \( h_t = X_t - E[X_t | F_{t-1}] \), and \( F_{t-1} \) is the \( \sigma \)-field generated by \( \{X_s, s \leq t - 1\} \). The estimating function estimator \( \hat{\theta}_F \) for the parameter \( \theta_0 \) is defined as

\[
G(\hat{\theta}_F) = 0.
\]

Chandra and Taniguchi [5] derived the asymptotic variance of \( \sqrt{n}(\hat{\theta}_F - \theta_0) \) as

\[
\left( \frac{1}{n} E \left[ \frac{\partial}{\partial \theta} G(\theta_0) \right] \right)^{-1} E \left[ \frac{1}{n} E \left[ \frac{\partial}{\partial \theta} G(\theta_0) \right] \right]^{-1}.
\]

and gave the following lemma by extending the result of Godambe [3].

**Lemma 2.1.** The asymptotic variance (2.4) is minimized if \( G(\theta) = G^*(\theta) \) where

\[
G^*(\theta) = \sum_{t=1}^{n} \alpha_{t-1}^* h_t,
\]

\[
\alpha_{t-1}^* = E \left[ \frac{\partial h_t}{\partial \theta} | F_{t-1} \right] E \left[ h_t^2 | F_{t-1} \right]^{-1}.
\]

Based on the estimating function \( G^*(\theta) \) in Lemma 2.1, Chandra and Taniguchi [5] constructed \( \hat{\theta}_G \) for the parameters of RCA and ARCH models and showed that \( \hat{\theta}_G \) is better than \( \hat{\theta}_{CL} \) by simulation. Furthermore, Amano [6] applied \( \hat{\theta}_G \) to some important financial time series models (RCA, GARCH, and nonlinear AR models) and showed that \( \hat{\theta}_G \) is better than \( \hat{\theta}_{CL} \) in the sense of the efficiency theoretically. Amano [6] also derived conditions that \( \hat{\theta}_G \) becomes asymptotically optimal. However, in Amano [6], \( \hat{\theta}_{CL} \) and \( \hat{\theta}_G \) were not applied to CHARN model, which includes many important financial time series models. Hence, in the next section, we apply \( \hat{\theta}_{CL} \) and \( \hat{\theta}_G \) to this model and prove \( \hat{\theta}_G \) is better than \( \hat{\theta}_{CL} \) in the sense of the efficiency for this model. Furthermore, conditions of asymptotical optimality of \( \hat{\theta}_G \) are also derived.

**3. CL and G Estimators for CHARN Model**

In this section, we discuss the asymptotics of \( \hat{\theta}_{CL} \) and \( \hat{\theta}_G \) for CHARN model.
CHARN model of order \( m \) is defined as

\[
X_t = F_\theta(X_{t-1}, \ldots, X_{t-m}) + H_\theta(X_{t-1}, \ldots, X_{t-m})u_t,
\tag{3.1}
\]

where \( F_\theta, H_\theta : \mathbb{R}^m \to \mathbb{R} \) are measurable functions, and \( \{u_t\} \) is a sequence of i.i.d. random variables with \( E[u_t] = 0, E[u_t^2] = 1 \) and independent of \( \{X_s; s < t\} \). Here, the parameter vector \( \theta = (\theta_1, \ldots, \theta_k)' \) is assumed to be lying in an open set \( \Theta \subseteq \mathbb{R}^k \). Its true value is denoted by \( \theta_0 \).

First we estimate the true parameter \( \theta_0 \) of (3.1) by use of \( \hat{\theta}_{CL} \), which is obtained by minimizing the penalty function

\[
Q_n(\theta) = \sum_{t=m+1}^{n} (X_t - E[X_t | F_t(m)])^2
\tag{3.2}
= \sum_{t=m+1}^{n} (X_t - F_\theta)^2.
\]

For the asymptotics of \( \hat{\theta}_{CL} \), we impose the following assumptions.

Assumption 3.1. (1) \( u_t \) has the probability density function \( f(u) > 0 \) a.e. \( u \in \mathbb{R} \).

(2) There exist constants \( a_i \geq 0, b_i \geq 0, 1 \leq i \leq m \), such that for \( x \in \mathbb{R}^m \) with \( |x| \to \infty \),

\[
|F_\theta(x)| \leq \sum_{i=1}^{m} a_i |x_i| + o(|x|),
\tag{3.3}
\]

\[
|H_\theta(x)| \leq \sum_{i=1}^{m} b_i |x_i| + o(|x|).
\]

(3) \( H_\theta(x) \) is continuous and symmetric on \( \mathbb{R}^m \), and there exists a positive constant \( \lambda \) such that

\[
H_\theta(x) \geq \lambda \quad \text{for} \ \forall x \in \mathbb{R}^m.
\tag{3.4}
\]

(4) Consider the following

\[
\left\{ \sum_{i=1}^{m} a_i + E[u_1] \sum_{i=1}^{m} b_i \right\} < 1.
\tag{3.5}
\]

Assumption 3.1 makes \( \{X_t\} \) strict stationary and ergodic (see [11]). We further impose the following.

Assumption 3.2. Consider the following

\[
E_\theta[F_\theta(X_{t-1}, \ldots, X_{t-m})]^2 < \infty,
\]

\[
E_\theta[H_\theta(X_{t-1}, \ldots, X_{t-m})]^2 < \infty,
\tag{3.6}
\]

for all \( \theta \in \Theta \).
Assumption 3.3. (1) $F_\theta$ and $H_\theta$ are almost surely twice continuously differentiable in $\Theta$, and their derivatives $\partial F_\theta / \partial \theta_j$ and $\partial H_\theta / \partial \theta_j$, $j = 1, \ldots, k$, satisfy the condition that there exist square-integrable functions $A_j$ and $B_j$ such that

$$\left| \frac{\partial F_\theta}{\partial \theta_j} \right| \leq A_j$$

$$\left| \frac{\partial H_\theta}{\partial \theta_j} \right| \leq B_j,$$

for all $\theta \in \Theta$.

(2) $f(u)$ satisfies

$$\lim_{|u| \to \infty} |u| f(u) = 0,$$

$$\int u^2 f(u) du = 1.$$

(3) The continuous derivative $f'(u) \equiv \delta f(u) / \delta u$ exists on $R$ and satisfies

$$\int \left( \frac{f'}{f} \right)^4 f(u) du < \infty,$$

$$\int u^2 \left( \frac{f'}{f} \right)^2 f(u) du < \infty.$$

From Tjøstheim [10], the following lemma holds.

Lemma 3.4. Under Assumptions 3.1, 3.2, and 3.3, $\hat{\theta}_{CL}$ has the following asymptotic normality:

$$\sqrt{n} \left( \hat{\theta}_{CL} - \theta_0 \right) \xrightarrow{d} U^{-1} W U^{-1},$$

where

$$W = E \left[ H_{\theta_0}^2 \frac{\partial}{\partial \theta} F_{\theta_0} \frac{\partial}{\partial \theta} F_{\theta_0} \right],$$

$$U = E \left[ \frac{\partial}{\partial \theta} F_{\theta_0} \frac{\partial}{\partial \theta} F_{\theta_0} \right].$$

Next, we apply $\hat{\theta}_C$ to CHARN model. From Lemma 2.1, $\hat{\theta}_C$ is obtained by solving the equation

$$\sum_{t=m+1}^n \frac{1}{H_\theta^2} \frac{\partial}{\partial \theta} F_\theta (X_t - F_\theta) = 0.$$
For the asymptotic of \( \hat{\theta}_C \), we impose the following Assumptions.

**Assumption 3.5.** (1) Consider the following

\[
E_\theta \left\| \frac{\partial a_{\theta, t-1}}{\partial \theta_j} \right\|^2 < \infty
\]  
(3.13)

for all \( \theta \in \Theta \).

(2) \( a_{\theta, t-1} \) is almost surely twice continuously differentiable in \( \Theta \), and for the derivatives \( \frac{\partial a_{\theta, t-1}}{\partial \theta_j} \), \( j = 1, \ldots, k \), there exist square-integrable functions \( C_j \) such that

\[
\left| \frac{\partial a_{\theta, t-1}}{\partial \theta_j} \right| \leq C_j
\]  
(3.14)

for all \( \theta \in \Theta \).

(3) \( V = E[1/\theta] \) is \( k \times k \)-positive definite matrix and satisfies

\[
|V| < \infty.
\]  
(3.15)

(4) For \( \theta \in \mathcal{B} \) (a neighborhood of \( \theta_0 \) in \( \Theta \)), there exist integrable functions \( P_{ij}^{\theta_l}(X_{t-1}) \), \( Q_{ij}^{\theta_l}(X_{t-1}) \), and \( R_{ij}^{\theta_l}(X_{t-1}) \) such that

\[
\left| \frac{\partial^2}{\partial \theta_j \partial \theta_l} \left( a_{\theta, t-1} \right)_{ij} \right| \leq P_{ij}^{\theta_l}(X_{t-1}),
\]  
(3.16)

\[
\left| \frac{\partial}{\partial \theta_j} \left( a_{\theta, t-1} \right)_{ij} \frac{\partial}{\partial \theta_l} \right| \leq Q_{ij}^{\theta_l}(X_{t-1}),
\]  
(3.16)

\[
\left| \left( a_{\theta, t-1} \right)_{ij} \frac{\partial^2}{\partial \theta_j \partial \theta_l} \right| \leq R_{ij}^{\theta_l}(X_{t-1}),
\]  
(3.16)

for \( i, j, l = 1, \ldots, k \), where \( X_{t-1} = (X_1, \ldots, X_{t-1}) \).

From Kanai et al. [9], the following lemma holds.

**Lemma 3.6.** Under Assumptions 3.1, 3.2, 3.3, and 3.5, the following statement holds:

\[
\sqrt{n} \left( \hat{\theta}_C - \theta_0 \right) \xrightarrow{d} N \left( 0, V^{-1} \right).
\]  
(3.17)

Finally we compare efficiencies of \( \hat{\theta}_{CL} \) and \( \hat{\theta}_C \). We give the following theorem.

**Theorem 3.7.** Under Assumptions 3.1, 3.2, 3.3, and 3.5, the following inequality holds:

\[
U^{-1}WU^{-1} \geq V^{-1},
\]  
(3.18)
and equality holds if and only if $H_{\theta_0}$ is constant or $\partial F_{\theta_0} / \partial \theta = 0$ (for matrices $A$ and $B$, $A \geq B$ means $A - B$ is positive definite).

This theorem is proved by use of Kholevo inequality (see Kholevo [12]). From this theorem, we can see that the magnitude of the asymptotic variance of $\hat{\theta}_G$ is smaller than that of $\hat{\theta}_{CL}$, and the condition that these asymptotic variances coincide is strict. Therefore, $\hat{\theta}_G$ is better than $\hat{\theta}_{CL}$ in the sense of the efficiency. Hence, we evaluate the condition that $\hat{\theta}_G$ is asymptotically optimal based on local asymptotic normality (LAN). LAN is the concept of local asymptotic normality for the likelihood ratio of general statistical models, which was established by Le Cam [13]. Once LAN is established, the asymptotic optimality of estimators and tests can be described in terms of the LAN property. Hence, its Fisher information matrix $\Gamma$ is described in terms of LAN, and the asymptotic variance of an estimator has the lower bound $\Gamma^{-1}$. Now, we prepare the following Lemma, which is due to Kato et al. [14].

**Lemma 3.8.** Under Assumptions 3.1, 3.2, and 3.3, CHARN model has LAN, and its Fisher information matrix $\Gamma$ is

\[
E \left[ \frac{1}{H_{\theta_0}^2} \left( -\frac{\partial H_{\theta_0}}{\partial \theta}, \frac{\partial F_{\theta_0}}{\partial \theta} \right) \left( \begin{array}{cc} a & c \\ c & b \end{array} \right) \left( -\frac{\partial H_{\theta_0}}{\partial \theta}, \frac{\partial F_{\theta_0}}{\partial \theta} \right) \right],
\]

where

\[
a_t = u_t \left( f'(u_t)/f(u_t) \right) + 1, \quad b_t = -(f'(u_t)/f(u_t)),
\]

\[
a = E[a_t^2], \quad b = E[b_t^2], \quad c = E[a_t b_t].
\]

From this Lemma, the asymptotic variance of $\hat{\theta}_G V^{-1}$ has the lower bound $\Gamma^{-1}$, that is,

\[
V^{-1} \geq \Gamma^{-1}.
\]

The next theorem gives the condition that $V^{-1}$ equals $\Gamma^{-1}$, that is $\hat{\theta}_G$ becomes asymptotically optimal.

**Theorem 3.9.** Under Assumptions 3.1, 3.2, 3.3, and 3.5, if $\partial H_{\theta_0} / \partial \theta = 0$ and $u_t$ is Gaussian, then $\hat{\theta}_G$ is asymptotically optimal, that is,

\[
V^{-1} = \Gamma^{-1}.
\]

Finally, we give the following example which satisfies the assumptions in Theorems 3.7 and 3.9.

**Example 3.10.** CHARN model includes the following nonlinear AR model:

\[
X_t = F_\theta(X_{t-1}, \ldots, X_{t-m}) + u_t,
\]
where \( F_\theta : \mathbb{R}^m \to \mathbb{R} \) is a measurable function, and \( \{ u_t \} \) is a sequence of i.i.d. random variables with \( E u_t = 0, E[u_t^2] = 1 \) and independent of \( \{ X_s; s < t \} \), and we assume Assumptions 3.1, 3.2, 3.3, and 3.5 (for example, we define \( F_\theta = \sqrt{a_0 + a_1 X_{t-1}^2 + \cdots + a_m X_{t-m}^2} \) where \( a_0 > 0, a_j \geq 0, j = 1, \ldots, m, \sum_{j=1}^m a_j < 1 \)). In Amano [6], it was shown that the asymptotic variance of \( \hat{\theta}_{CL} \) attains that of \( \hat{\theta}_G \). Amano [6] also showed under the condition that \( u_t \) is Gaussian, \( \hat{\theta}_G \) is asymptotically optimal.

4. Numerical Studies

In this section, we evaluate accuracies of \( \hat{\theta}_{CL} \) and \( \hat{\theta}_G \) for the parameter of CHARN model by simulation. Throughout this section, we assume the following model:

\[
X_t = a X_{t-1} + \sqrt{0.2 + 0.1 X_{t-1}^2} u_t, \tag{4.1}
\]

where \( \{ u_t \} \sim \text{i.i.d.} \, N(0, 1) \). Mean squared errors (MSEs) of \( \hat{\theta}_{CL} \) and \( \hat{\theta}_G \) for the parameter \( a \) are reported in the following Table 1. The simulations are based on 1000 realizations, and we set the parameter value \( a \) and the length of observations \( n \) as \( a = 0.1, 0.2, 0.3 \) and \( n = 100, 200, 300 \).

From Table 1, we can see that MSE of \( \hat{\theta}_G \) is smaller than that of \( \hat{\theta}_{CL} \). Furthermore it is seen that MSE of \( \hat{\theta}_G \) decreases as the length of observations \( n \) increases.

5. Proofs

This section provides the proofs of the theorems. First, we prepare the following lemma to compare the asymptotic variances of CL and G estimators (see Kholevo [12]).

**Lemma 5.1.** We define \( \psi(\omega) \) and \( \phi(\omega) \) as \( r \times s \) and \( t \times s \) random matrices, respectively, and \( h(\omega) \) as a random variable that is positive everywhere. If the matrix \( E[\phi \phi' / h]^{-1} \) exists, then the following inequality holds:

\[
E[\psi \psi' h] \geq E[\psi \phi'] E \left[ \frac{\phi \phi'}{h} \right]^{-1} E[\phi \phi']'. \tag{5.1}
\]

The equality holds if and only if there exists a constant \( r \times t \) matrix \( C \) such that

\[
h \psi + C \phi = 0 \quad \text{a.e.} \tag{5.2}
\]

Now we proceed to prove Theorem 3.7.
Proof of Theorem 3.7. Let \( \psi = \frac{\partial}{\partial \theta} F_{\theta^0} \), \( \phi = \frac{\partial}{\partial \theta} F_{\theta^0} \), and \( h = H_{\theta^0}^2 \), then from the definitions of matrices \( U \), \( W \) and \( V \), it can be represented as
\[
U = E[\psi \phi'], \\
W = E[\psi \psi'h], \\
V = E[\phi \phi'/h].
\]

(5.3)

Hence from Lemma 5.1, we can see that
\[
W \geq UV^{-1}U.
\]

(5.4)

From this inequality, we can see that
\[
U^{-1}WU^{-1} \geq V^{-1}.
\]

(5.5)

\[\square\]

Proof of Theorem 3.9. Fisher information matrix of CHARN model based on LAN \( \Gamma \) can be represented as
\[
\Gamma = E \left[ \frac{1}{H_{\theta_0}^2} \left( \begin{array}{cc} \frac{\partial H_{\theta_0}}{\partial \theta'} & \frac{\partial F_{\theta_0}}{\partial \theta'} \\ \frac{\partial F_{\theta_0}}{\partial \theta'} & \frac{\partial F_{\theta_0}}{\partial \theta'} \end{array} \right) \right]
\]
\[
+ E \left[ \frac{f'(u)}{f(u)} \right]^2 E \left[ \frac{1}{H_{\theta_0}^2} \frac{\partial F_{\theta_0}}{\partial \theta'} \frac{\partial F_{\theta_0}}{\partial \theta'} \right].
\]

(5.6)

From (5.6), if \( \partial H_{\theta_0}/\partial \theta = 0 \), \( \Gamma \) becomes
\[
\Gamma = E \left[ \frac{f'(u)}{f(u)} \right]^2 E \left[ \frac{1}{H_{\theta_0}^2} \frac{\partial F_{\theta_0}}{\partial \theta'} \frac{\partial F_{\theta_0}}{\partial \theta'} \right].
\]

(5.7)
Next, we show under the Gaussianity of \( u_t \) that \( E[(f'(u_t)/f(u_t))^2] = 1 \). From the Schwarz inequality, it can be obtained that

\[
E\left[ \left( \frac{f'(u_t)}{f(u_t)} \right)^2 \right] = E[u_t^2] E\left[ \left( \frac{f'(u_t)}{f(u_t)} \right)^2 \right] \\
\geq \left( E\left[ u_t \frac{f'(u_t)}{f(u_t)} \right] \right)^2 \\
= \left( \int_{-\infty}^{\infty} xf'(x)dx \right)^2 \\
= \left( \int_{-\infty}^{\infty} xf(x)dx - \int_{-\infty}^{\infty} f(x)dx \right)^2 \\
= 1. 
\]

(5.8)

The equality holds if and only if there exists some constant \( c \) such that

\[
 cx = \frac{f'(x)}{f(x)}. 
\]

(5.9)

Equation (5.9) becomes, for some constant \( k \),

\[
 cx = (\log f(x))', \\
\frac{c}{2}x^2 + k = \log f(x), \\
f(x) = e^{(c/2)x^2+k} = e^ke^{(c/2)x^2}. 
\]

(5.10)

Hence, \( c \) is \(-1\), and \( f(x) \) becomes the density function of the normal distribution.

Acknowledgment

The author would like to thank referees for their comments, which improved the original version of this paper.

References

[1] T. Amano and M. Taniguchi, “Asymptotic efficiency of conditional least squares estimators for ARCH models,” Statistics & Probability Letters, vol. 78, no. 2, pp. 179–185, 2008.
[2] V. P. Godambe, “An optimum property of regular maximum likelihood estimation,” Annals of Mathematical Statistics, vol. 31, pp. 1208–1211, 1960.
[3] V. P. Godambe, “The foundations of finite sample estimation in stochastic processes,” Biometrika, vol. 72, no. 2, pp. 419–428, 1985.
[4] L. P. Hansen, “Large sample properties of generalized method of moments estimators,” Econometrica, vol. 50, no. 4, pp. 1029–1054, 1982.
[5] S. A. Chandra and M. Taniguchi, “Estimating functions for nonlinear time series models,” *Annals of the Institute of Statistical Mathematics*, vol. 53, no. 1, pp. 125–141, 2001.

[6] T. Amano, “Asymptotic efficiency of estimating function estimators for nonlinear time series models,” *Journal of the Japan Statistical Society*, vol. 39, no. 2, pp. 209–231, 2009.

[7] W. Hardle and A. Tsybakov, “Local polynomial estimators of the volatility function in nonparametric autoregression,” *Journal of Econometrics*, vol. 81, no. 1, pp. 223–242, 1997.

[8] W. Härdle, A. Tsybakov, and L. Yang, “Nonparametric vector autoregression,” *Journal of Statistical Planning and Inference*, vol. 68, no. 2, pp. 221–245, 1998.

[9] H. Kanai, H. Ogata, and M. Taniguchi, “Estimating function approach for CHARN models,” *Metron*, vol. 68, pp. 1–21, 2010.

[10] D. Tjøstheim, “Estimation in nonlinear time series models,” *Stochastic Processes and Their Applications*, vol. 21, no. 2, pp. 251–273, 1986.

[11] Z. Lu and Z. Jiang, “$L_1$ geometric ergodicity of a multivariate nonlinear AR model with an ARCH term,” *Statistics & Probability Letters*, vol. 51, no. 2, pp. 121–130, 2001.

[12] A. S. Kholevo, “On estimates of regression coefficients,” *Theory of Probability and its Applications*, vol. 14, pp. 79–104, 1969.

[13] L. Le Cam, “Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses,” vol. 3, pp. 37–98, 1960.

[14] H. Kato, M. Taniguchi, and M. Honda, “Statistical analysis for multiplicatively modulated nonlinear autoregressive model and its applications to electrophysiological signal analysis in humans,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 3414–3425, 2006.
Research Article

Optimal Portfolio Estimation for Dependent Financial Returns with Generalized Empirical Likelihood

Hiroaki Ogata

School of International Liberal Studies, Waseda University, Tokyo 169-8050, Japan

Correspondence should be addressed to Hiroaki Ogata, hiroakiogata@aoni.waseda.jp

Received 16 February 2012; Accepted 10 April 2012

Academic Editor: Junichi Hirukawa

Copyright © 2012 Hiroaki Ogata. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes to use the method of generalized empirical likelihood to find the optimal portfolio weights. The log-returns of assets are modeled by multivariate stationary processes rather than i.i.d. sequences. The variance of the portfolio is written by the spectral density matrix, and we seek the portfolio weights which minimize it.

1. Introduction

The modern portfolio theory has been developed since circa the 1950s. It is common knowledge that Markowitz [1, 2] is a pioneer in this field. He introduced the so-called mean-variance theory, in which we try to maximize the expected return (minimize the variance) under the constant variance (the constant expected return). After that, many researchers followed, and portfolio theory has been greatly improved. For a comprehensive survey of this field, refer to Elton et al. [3], for example.

Despite its sophisticated paradigm, we admit there exists several criticisms against the early portfolio theory. One of them is that it blindly assumes that the asset returns are normally distributed. As Mandelbrot [4] pointed out, the price changes in the financial market do not seem to be normally distributed. Therefore, it is appropriate to use the nonparametric estimation method to find the optimal portfolio. Furthermore, it is empirically observed that financial returns are dependent. Therefore, it is unreasonable to fit the independent model to it.

One of the nonparametric techniques which has been capturing the spotlight recently is the empirical likelihood method. It was originally proposed by Owen [5, 6] as a method of
inference based on a data-driven likelihood ratio function. Smith [7] and Newey and Smith [8] extended it to the generalized empirical likelihood (GEL). 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.

From the above point of view, we consider to find the optimal portfolio weights by using the GEL method under the multivariate stationary processes. The optimal portfolio weights are defined as the weights which minimize the variance of the return process with constant mean. The analysis is done in the frequency domain.

This paper is organized as follows. Section 2 explains about a frequency domain estimating function. In Section 3, we review the GEL method and mention the related asymptotic theory. Monte Carlo simulations and a real-data example are given in Section 4. Throughout this paper, \(A'\) and \(A^*\) indicate the transposition and adjoint of a matrix \(A\), respectively.

2. Frequency Domain Estimating Function

Here, we are concerned with the \(m\)-dimensional stationary process \(\{X(t)\}_{t \in \mathbb{Z}}\) with mean vector \(\theta\), the autocovariance matrix

\[
\Gamma(h) = \text{Cov}[X(t + h), X(t)] = E[X(t + h)X'(t)],
\]

and spectral density matrix

\[
f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma(h)e^{-ih\lambda}, \quad -\pi \leq \lambda < \pi.
\]

Suppose that information of an interested \(p\)-dimensional parameter \(\theta \in \Theta \subset \mathbb{R}^p\) exists through a system of general estimating equations in frequency domain as follows. Let \(\phi_j(\lambda; \theta), (j = 1, \ldots, q, q \geq p)\) be \(m \times m\) matrix-valued continuous functions on \([-\pi, \pi)\) satisfying \(\phi_j(\lambda; \theta) = \phi_j(\lambda; \theta)^*\) and \(\phi_j(-\lambda; \theta) = \phi_j(\lambda; \theta)^*\). We assume that each \(\phi_j(\lambda; \theta)\) satisfies the spectral moment condition

\[
\int_{-\pi}^{\pi} \text{tr}\{\phi_j(\lambda; \theta_0) f(\lambda)\} d\lambda = 0 \quad (j = 1, \ldots, q),
\]

where \(\theta_0 = (\theta_1, \ldots, \theta_p)'\) is the true value of the parameter. By taking an appropriate function for \(\phi_j(\lambda; \theta)\), (2.3) can express the best portfolio weights as shown in Example 2.1.

Example 2.1 (portfolio selection). In this example, we set \(m = p = q\). Let \(x_i(t)\) be the log-return of \(i\)th asset \((i = 1, \ldots, m)\) at time \(t\) and suppose that the process \(\{X(t) = (x_1(t), \ldots, x_m(t))^T\}\) is stationary with zero mean. Consider the portfolio \(p(t) = \sum_{i=1}^{m} \theta_i x_i(t)\), where \(\theta = (\theta_1, \ldots, \theta_m)'\) is a vector of weights, satisfying \(\sum_{i=1}^{m} \theta_i = 1\). The process \(\{p(t)\}\) is a linear combination of the
stationary process, hence \( \{ p(t) \} \) is still stationary, and, from Herglotz’s theorem, its variance is

\[
\text{Var}\{ p(t) \} = \theta' \text{Var}\{ X(t) \} \theta = \theta' \left\{ \int_{-\pi}^{\pi} f(\lambda) d\lambda \right\} \theta. \tag{2.4}
\]

Our aim is to find the weights \( \theta_0 = (\theta_{10}, \ldots, \theta_{m0})' \) that minimize the variance (the risk) of the portfolio \( p(t) \) under the constrain of \( \sum_{i=1}^{m} \theta_i = 1 \). The Lagrange function is given by

\[
L(\theta, \lambda) = \theta' \left\{ \int_{-\pi}^{\pi} f(\lambda) d\lambda \right\} \theta + \xi (\theta' e - 1), \tag{2.5}
\]

where \( e = (1, 1, \ldots, 1)' \) and \( \xi \) is Lagrange multiplier. The first order condition leads to

\[
(I - e\theta_0) \left\{ \int_{-\pi}^{\pi} \{ f(\lambda) + f(\lambda)' \} d\lambda \right\} \theta_0 = 0, \tag{2.6}
\]

where \( I \) is an identity matrix. Now, for fixed \( j = 1, \ldots, m \), consider to take

\[
\phi_j(\lambda; \theta) = \begin{cases} 
2\theta_j (1 - \theta_j) & \text{\( (j, j) \)th component} \\
1 - 2\theta_j \theta_\ell & \text{\( (j, \ell) \)th and \( (\ell, j) \)th component with \( \ell = 1, \ldots, m \) and \( \ell \neq j \)} \\
-2\theta_k \theta_\ell & \text{\( (k, \ell) \)th component with \( k, \ell = 1, \ldots, m \) and \( k \neq j, \ell \neq j \)}.
\end{cases} \tag{2.7}
\]

Then, (2.3) coincides with the first order condition (2.6), which implies that the best portfolio weights can be solved with the framework of the spectral moment condition.

Besides, we can express other important indices for time series. In what follows, several examples are given.

Example 2.2 (autocorrelation). Denote the autocovariance and the autocorrelation of the process \( \{ x_i(t) \} \) (ith component of the process \( \{ X(t) \} \)) 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

\[
\phi_1(\lambda; \theta) = \begin{cases} 
\cos(h\lambda) - \theta_1 & \text{\( (i, i) \)th component} \\
0 & \text{otherwise,}
\end{cases} \tag{2.8}
\]

\[
\phi_2(\lambda; \theta) = \begin{cases} 
\cos(k\lambda) - \theta_2 & \text{\( (j, j) \)th component} \\
0 & \text{otherwise.}
\end{cases}
\]
Then, (2.3) leads to

\[
\begin{align*}
\theta_{10} & = \left\{ \int_{-\pi}^{\pi} \exp(ih\lambda) f_{ii}(\lambda) d\lambda \right\} \left\{ \int_{-\pi}^{\pi} f_{ii}(\lambda) d\lambda \right\}^{-1}, \\
\theta_{20} & = \left\{ \int_{-\pi}^{\pi} \exp(ik\lambda) f_{jj}(\lambda) d\lambda \right\} \left\{ \int_{-\pi}^{\pi} f_{jj}(\lambda) d\lambda \right\}^{-1}.
\end{align*}
\] (2.9)

From Herglotz’s theorem, \(\theta_{10} = \gamma_i(h)/\gamma_i(0)\), and \(\theta_{20} = \gamma_j(k)/\gamma_j(0)\). Then, \(\theta_0 = (\theta_{10}, \theta_{20})'\) corresponds to the desired autocorrelations \(\rho = (\rho_i(h), \rho_j(k))'\). The idea can be directly extended to more than two autocorrelations.

**Example 2.3** (Whittle estimation). In this example, we set \(p = q\). Consider fitting a parametric spectral density model \(f_0(\lambda)\) to the true spectral density \(f(\lambda)\). The disparity between \(f_0(\lambda)\) and \(f(\lambda)\) is measured by the following criterion:

\[
D(f_\theta, f) \equiv \int_{-\pi}^{\pi} \left[ \log \det f_\theta(\lambda) + \text{tr}\left\{ f_\theta(\lambda)^{-1} f(\lambda) \right\} \right] d\lambda,
\] (2.10)

which is based on Whittle likelihood. The purpose here is to seek the quasi-true value \(\overline{\theta}\) defined by

\[
\overline{\theta} = \arg \min_{\theta \in \Theta} D(f_\theta, f).
\] (2.11)

Assume that the spectral density model is expressed by the following form:

\[
f_\theta(\lambda) = \left\{ \sum_{j=0}^{\infty} B_\theta(j) \exp(ij\lambda) \right\} K \left\{ \sum_{j=0}^{\infty} B_\theta(j) \exp(ij\lambda) \right\}^*,
\] (2.12)

where each \(B_\theta(j)\) is an \(m \times m\) matrix (\(B_\theta(0)\) is defined as identity matrix), and \(K\) is an \(m \times m\) symmetric matrix. The general linear process has this spectral form, so this assumption is not so restrictive. The key of this assumption is that the elements of the parameter \(\theta\) do not depend on \(K\). We call such a parameter innovation-free. Imagine that you fit the VARMA process, for example. Innovation-free implies that the elements of the parameter \(\theta\) depend on only AR or MA coefficients and not on the covariance matrix of the innovation process. Now, let us consider the equation:

\[
\frac{\partial}{\partial \theta} D(f_\theta, f) \bigg|_{\theta = \overline{\theta}} = 0,
\] (2.13)

to find quasi-true value. The Kolmogorov’s formula says

\[
\det K = \exp \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \det \{2\pi f_\theta(\lambda)\} d\lambda \right].
\] (2.14)
This implies that if $\theta$ is innovation-free, the quantity $\int_{-\infty}^{\infty} \log \det \{ f_\theta(\lambda) \} d\lambda$ is independent of $\theta$ and (2.13) leads to

$$\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} \operatorname{tr} \left\{ f_\theta(\lambda)^{-1} f(\lambda) d\lambda \right\} \bigg|_{\theta=\hat{\theta}} = 0.$$  \hspace{1cm} (2.15)

This corresponds to (2.3), when we set

$$\phi_j(\lambda; \theta) = \frac{\partial f_\theta(\lambda)}{\partial \theta_j} \big( \lambda = 1, \ldots, p \big),$$  \hspace{1cm} (2.16)

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

Based on the form of (2.3), we set the estimating function for $\theta$ as

$$m(\lambda; \theta) = \left( \operatorname{tr} \{ \phi_1(\lambda; \theta) I_n(\lambda) \}, \ldots, \operatorname{tr} \{ \phi_p(\lambda; \theta) I_n(\lambda) \} \right)'$$  \hspace{1cm} (2.17)

where $I_n(\lambda)$ is the periodogram, defined by

$$I_n(\lambda) = (2\pi n)^{-1} \left( \sum_{t=1}^{n} X(t) \exp(it\lambda) \right) \left( \sum_{t=1}^{n} X(t) \exp(it\lambda) \right)'$$  \hspace{1cm} (2.18)

where $\lambda_t = (2\pi t)/n$, $t = -(n-1)/2, \ldots, [n/2]$. Then, we have

$$\frac{2\pi}{n} \sum_{t=-(n-1)/2}^{[n/2]} \mathbb{E} \left[ m(\lambda_t; \theta) \right] \longrightarrow \left[ \int_{-\infty}^{\infty} \operatorname{tr} \left\{ \phi_j(\lambda; \theta_0) f(\lambda) \right\} d\lambda \right]_{j=1,\ldots,p} = 0.$$  \hspace{1cm} (2.19)

### 3. Generalized Empirical Likelihood

Once we construct the estimating function, we can make use of the method of GEL as in the work by Smith [7] and Newey and Smith [8]. GEL is introduced as an alternative to GMM and it is pointed out that its asymptotic bias does not grow with the number of moment restrictions, while the bias of GMM often does.

To describe GEL let $\rho(v)$ be a function of a scalar $v$ which is concave on its domain, an open interval $\mathcal{U}$ containing zero. Let $\hat{\Lambda}_n(\theta) = \{ \lambda : \lambda' m(\lambda; \theta) \in \mathcal{U}, \ t = 1, \ldots, n \}$. The estimator is the solution to a saddle point problem

$$\hat{\theta}_{GEL} = \arg \min_{\theta \in \mathcal{O}} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{t=1}^{n} \rho(\lambda' \theta).$$  \hspace{1cm} (3.1)

The empirical likelihood (EL) estimator (cf. [9]), the exponential tilting (ET) estimator (cf. [10]), and the continuous updating estimator (CUE) (cf. [11]) are special cases with $\rho(v) = \log(1-v)$, $\rho(v) = -e^v$ and $\rho(v) = -(1+v)^2/2$, respectively. Let $\Omega = \mathbb{E} \left[ m(\lambda_t; \theta_0) m(\lambda_t; \theta_0)' \right]$,
Theorem 3.4. Let Assumptions 3.1 and 3.3 hold. Then

\[ \theta_0 = \hat{\theta}_{GEL} \rightarrow N(\theta_0). \]

**Table 1:** Estimated autocorrelations for two-dimensional-AR(1) model.

|         | EL Mean | EL s.d. | ET Mean | ET s.d. | CUE Mean | CUE s.d. |
|---------|---------|---------|---------|---------|-----------|----------|
| \( \rho_1 \) (1) | 0.3779  | 0.0890  | 0.3760  | 0.0900  | 0.3797    | 0.0911   |
| \( \rho_2 \) (1) | 0.4680  | 0.0832  | 0.4660  | 0.0817  | 0.4650    | 0.0855   |

\[ G = E[\hat{m}(\lambda; \theta_0) / \hat{\theta}], \] and \( \Sigma = (G^T \Omega^{-1} G)^{-1}. \) The following assumptions and theorems are due to Newey and Smith [8].

**Assumption 3.1.** (i) \( \theta_0 \in \Theta \) is the unique solution to (2.3).
(ii) \( \Theta \) is compact.
(iii) \( m(\lambda; \theta) \) is continuous at each \( \theta \in \Theta \) with probability one.
(iv) \( E[\sup_{\theta \in \Theta} |m(\lambda; \theta)|^a] < \infty \) for some \( a > 2. \)
(v) \( \Omega \) is nonsingular.
(vi) \( \rho(v) \) is twice continuously differentiable in a neighborhood of zero.

**Theorem 3.2.** Let Assumption 3.1 hold. Then \( \hat{\theta}_{GEL} \rightarrow N(\theta_0). \)

**Assumption 3.3.** (i) \( \theta_0 \in \text{int}(\Theta). \)
(ii) \( m(\lambda; \theta) \) is continuously differentiable in a neighborhood \( \mathcal{N} \) of \( \theta_0 \) and \( E[|\sup_{\theta \in \mathcal{N}} |\hat{m}(\lambda; \theta) / \hat{\theta}|^2] < \infty. \)
(iii) \( \text{rank}(G) = p. \)

**Theorem 3.4.** Let Assumptions 3.1 and 3.3 hold. Then \( \sqrt{n}(\hat{\theta}_{GEL} - \theta_0) \rightarrow N(0, \Sigma). \)

4. Monte Carlo Studies and Illustration

In the first part of this section, we summarize the estimation results of Monte Carlo studies of the GEL method. We generate 200 observations from the following two-dimensional-AR(1) model

\[
\begin{bmatrix}
X_1(t) \\
X_2(t)
\end{bmatrix} = \begin{bmatrix}
0.2 & -0.5 \\
-0.5 & 0.3
\end{bmatrix} \begin{bmatrix}
X_1(t-1) \\
X_2(t-2)
\end{bmatrix} + \begin{bmatrix}
U_1(t) \\
U_2(t)
\end{bmatrix},
\]

where \{\( U(t) = (U_1(t), U_2(t)) \)\}_t \in \mathbb{Z} is an i.i.d. innovation process, distributed to two-dimensional \( t \)-distribution whose correlation matrix is identity, and degree of freedom is 5. The true autocorrelations with lag 1 of this process are \( \rho_1(1) = 0.3894 \) and \( \rho_2(1) = 0.4761, \) respectively. As described in Example 2.2, we estimate \( \rho_1(1) \) and \( \rho_2(1) \) by using three types of frequency domain GEL method (EL, ET, and CUE). Table 1 shows the mean and standard deviation of the estimation results whose repetition time is 1000. All types work properly.

Next, we apply the proposed method to the returns of market index data. The sample consists of 7 weekly indices (S&Ps 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. Refer to Table 2 for the market of each index.
Table 2: Market.

| Index    | Market                      |
|----------|-----------------------------|
| S&P 500  | NYSE                        |
| Bovespa  | São Paulo Stock Exchange    |
| CAC 40   | Bourse de Paris             |
| AEX      | Amsterdam Stock Exchange    |
| ATX      | Wiener Börse                |
| HKHSI    | Hong Kong Exchanges and Clearing |
| Nikkei   | Tokyo Stock Exchange        |

Table 3: Estimated portfolio weights.

|          | EL   | ET   | CUE  |
|----------|------|------|------|
| S&P 500  | 0.0759| 0.0617| 0.0001|
| Bovespa  | 0.6648| 0.6487| 0.6827|
| CAC 40   | 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|

As shown in Example 2.1, we use frequency domain GEL method to estimate the optimal portfolio weights, and the results are shown in Table 3. Bovespa and ATX account for large part in the optimal portfolio.

Acknowledgment

This work was supported by Grant-in-Aid for Young Scientists (B) (22700291).

References

[1] H. Markowitz, “Portfolio selection,” *Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.
[2] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons, New York, NY, USA, 1959.
[3] E. J. Elton, M. J. Gruber, S. J. Brown, and W. N. Goetzmann, *Modern Portfolio Theory and Investment Analysis*, John Wiley & Sons, New York, NY, USA, 7th edition, 2007.
[4] B. Mandelbrot, “The variation of certain speculative prices,” *Journal of Business*, vol. 36, pp. 394–419, 1963.
[5] A. B. Owen, “Empirical likelihood ratio confidence intervals for a single functional,” *Biometrika*, vol. 75, no. 2, pp. 237–249, 1988.
[6] A. Owen, “Empirical likelihood ratio confidence regions,” *The Annals of Statistics*, vol. 18, no. 1, pp. 90–120, 1990.
[7] R. J. Smith, GEL criteria for moment condition models. Working paper, 2004, http://cemmap .ifs.org.uk/wps/cwp0419.pdf.
[8] W. K. Newey and R. J. Smith, “Higher order properties of GMM and generalized empirical likelihood estimators,” *Econometrica*, vol. 72, no. 1, pp. 219–255, 2004.
[9] J. Qin and J. Lawless, “Empirical likelihood and general estimating equations,” *The Annals of Statistics*, vol. 22, no. 1, pp. 300–325, 1994.
[10] Y. Kitamura and M. Stutzer, “An information-theoretic alternative to generalized method of moments estimation,” *Econometrica*, vol. 65, no. 4, pp. 861–874, 1997.

[11] L. P. Hansen, J. Heaton, and A. Yaron, “Finite-sample properties of some alternative GMM estimators,” *Journal of Business and Economic Statistics*, vol. 14, no. 3, pp. 262–280, 1996.
Research Article

Statistically Efficient Construction of $\alpha$-Risk-Minimizing Portfolio

Hiroyuki Taniai$^1$ and Takayuki Shiohama$^2$

$^1$ School of International Liberal Studies, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku, Tokyo 169-8050, Japan
$^2$ Department of Management Science, Faculty of Engineering, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku, Tokyo 162-8601, Japan

Correspondence should be addressed to Hiroyuki Taniai, htaniai@aoni.waseda.jp

Received 21 March 2012; Accepted 19 April 2012

Academic Editor: Masanobu Taniguchi

Copyright © 2012 H. Taniai and T. Shiohama. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a semiparametrically efficient estimator for $\alpha$-risk-minimizing portfolio weights. Based on the work of Bassett et al. (2004), an $\alpha$-risk-minimizing portfolio optimization is formulated as a linear quantile regression problem. The quantile regression method uses a pseudolikelihood based on an asymmetric Laplace reference density, and asymptotic properties such as consistency and asymptotic normality are obtained. We apply the results of Hallin et al. (2008) to the problem of constructing $\alpha$-risk-minimizing portfolios using residual signs and ranks and a general reference density. Monte Carlo simulations assess the performance of the proposed method. Empirical applications are also investigated.

1. Introduction

Since the first formation of Markowitz’s mean-variance model, portfolio optimization and construction have been a critical part of asset and fund management. At the same time, portfolio risk assessment has become an essential tool in risk management. Yet there are well-known shortcomings of variance as a risk measure for the purposes of portfolio optimization; namely, variance is a good risk measure only for elliptical and symmetric return distributions.

The proper mathematical characterization of risk is of central importance in finance. The choice of an adequate risk measure is a complex task that, in principle, involves deep consideration of the attitudes of market players and the structure of markets. Recently, value at risk (VaR) has gained widespread use, in practice as well as in regulation. VaR has been criticized, however, because as a quantile is no reason to be convex, and indeed, it is easy to construct portfolios for which VaR seriously violates convexity. The shortcomings of VaR led to the introduction of coherent risk measures. Artzner et al. [1] and Föllmer and Schied [2]
question whether VaR qualifies as such a measure, and both find that VaR is not an adequate measure of risk. Unlike VaR, expected shortfall (or tail VaR), which is defined as the expected portfolio tail return, has been shown to have all necessary characteristics of a coherent risk measure. In this paper, we use $\alpha$-risk as a risk measure that satisfies the conditions of coherent risk measure (see [3]). Variants of the $\alpha$-risk measure include expected shortfall and tail VaR. The $\alpha$-risk-minimizing portfolio, introduced as a pessimistic portfolio in Bassett et al. [3], can be formulated as a problem of linear quantile regression.

Since the seminal work by Koenker and Bassett [4], quantile regression (QR) has become more widely used to describe the conditional distribution of a random variable given a set of covariates. One common finding in the extant literature is that the quantile regression estimator has good asymptotic properties under various data dependence structures, and for a wide variety of conditional quantile models and data structures. A comprehensive guide to quantile regression is provided by Koenker [5].

Quantile regression methods use a pseudolikelihood based on an asymmetric Laplace reference density (see [6]). Komunjer [7] introduced a class of “tick-exponential” distribution, which includes an asymmetric Laplace density as a particular case, and showed that the tick-exponential QMLE reduces to the standard quantile regression estimator of Koenker and Bassett [4].

In quantile regression, one must know the conditional error density at zero, and incorrect specification of the conditional error density leads to inefficient estimators. Yet correct specification is difficult, because reliable shape information may be scarce. Zhao [8], Whang [9], and Komunjer and Vuong [10] propose efficiency corrections for the univariate quantile regression model.

This paper describes a semiparametrically efficient estimation of an $\alpha$-risk-minimizing portfolio in place of an asymmetric Laplace reference density (a standard quantile regression estimator), by using any other $\alpha$-quantile zero reference density $f$, based on residual ranks and signs. A $\sqrt{n}$-consistent and asymptotically normal one-step estimator is proposed. Like all semiparametric estimators in the literature, our method relies on the availability of a $\sqrt{n}$-consistent first-round estimator, a natural choice being the standard quantile regression estimator. Under correct specifications, they attain the semiparametric efficiency bound associated with $f$.

The remainder of this paper is organized as follows. In Section 2, we introduce the setup and definition of an $\alpha$-risk-minimizing portfolio and present its equivalent formation under quantile regression settings. Section 3 contains theoretical results for our one-step estimator, and Section 4 describes its computation and performance. Section 5 gives empirical applications, and Section 6 our conclusions.

2. $\alpha$-Risk-Minimizing Portfolio Formulation

"$\alpha$-risk" can be considered a coherent measure of risk as discussed in Artzner et al. [1]. The $\alpha$-risk of $X$, say $\rho_{\alpha}(X)$, is defined as

$$
\rho_{\alpha}(X) := -\int_{0}^{1} F_X^{-}(t)d\nu_{\alpha}(t) = -\frac{1}{\alpha}\int_{0}^{\alpha} F_X^{-}(t)dt, \quad \alpha \in (0, 1),
$$

where $\nu_{\alpha}(t) := \min\{t/\alpha, 1\}$ and $F_X^{-}(\alpha) := \inf\{x : F_X(x) \geq \alpha\}$ denote the quantile function of a random variable $X$ with distribution function $F_X$. Here, we recall the definition of expected
shortfall and the relationship among the tail risk measures in finance. The $\alpha$-expected shortfall defined for $\alpha \in (0, 1)$ as

$$\text{ES}_{(\alpha)}(X) = -\frac{1}{\alpha} \int_0^\alpha F_X^{-1}(p) \, dp \quad (2.2)$$

can be shown to be a risk measure that satisfies the axioms of a coherent measure of risk. It is worth mentioning that the expected shortfall is closely related but not coincident to the notion of conditional value at risk (CVaR) defined in Uryasev [11] and Pflug [12]. We note that expected shortfall and conditional VaR or tail conditional expectations are identical "extreme" risk measures only for continuous distributions, that is,

$$\text{CVaR}_{(\alpha)}(X) = \text{TCE}_{(\alpha)}(X) = -E[X \mid X > F_X^{-1}(\alpha)]. \quad (2.3)$$

To avoid confusion, in this paper, we use the term "$\alpha$-risk measure" instead of terms like expected shortfall, CVaR, or tail conditional expectation.

Bassett et al. [3] showed that a portfolio with minimized $\alpha$-risk can be constructed via the quantile regression (QR) methods of Koenker and Bassett [4]. QR is based on the fact that a quantile can be characterized as the minimizer of some expected asymmetric absolute loss function, namely,

$$F_X^{-1}(\alpha) = \arg \min_{\theta} \mathbb{E}[(\alpha 1_{X - \theta \geq 0} + (1 - \alpha) 1_{X - \theta < 0})|X - \theta|]$$

$$= \arg \min_{\theta} \mathbb{E}[\rho_\alpha(X - \theta)], \quad (2.4)$$

where $\rho_\alpha(u) := u(\alpha - 1\{u < 0\}), u \in \mathbb{R}$ is called the check function (see [5]), and $1_A$ is the indicator function defined by $1_A = 1_A(\omega) := 1$ if $\omega \in A, := 0$ if $\omega \notin A$. To construct the optimal (i.e., $\alpha$-risk minimized) portfolio, the following lemma is needed.

**Lemma 2.1** (Theorem 2 of [3]). Let $X$ be a real-valued random variable with $EX = \mu < \infty$, then

$$\min_{\theta \in \mathbb{R}} \mathbb{E}[\rho_\alpha(X - \theta)] = \alpha (\mu + \rho_\alpha(X)). \quad (2.5)$$

Then, $Y = Y(\pi) = X^T \pi$ denotes a portfolio consisting of $d$ different assets $X := (X_1, \ldots, X_d)$ with allocation weights $\pi := (\pi_1, \ldots, \pi_d)'$ (subject to $\sum_{j=1}^d \pi_j = 1$), and the optimization problem under study is, for some prespecified expected return $\mu_0$,

$$\min_{\pi} \rho_\alpha(X^T \pi) \quad \text{subject to } E(X^T \pi) = \mu_0, \quad 1_d \pi = 1. \quad (2.6)$$

The sample or empirical analogue of this problem can be expressed as

$$\min \sum_{b \in \mathbb{R}^d} \rho_\alpha(Z_i - (A_w b)), \quad (2.7)$$
where $X_{ij}$ denotes the $j$th sample value of asset $i$, $\overline{X}_i := n^{-1} \sum_{j=1}^{n} X_{ij}$,

$$Z = (Z_1, \ldots, Z_n, Z_{n+1})' := (X_{11}, \ldots, X_{n1}, \kappa (\overline{X}_1 - \mu_0))'$$

with some $\kappa$ sufficiently large. The minimizer of (2.7), namely,

$$\hat{\beta}^{(n)}(\alpha) := \arg\min_{b \in \mathbb{R}^d} \sum_{i=1}^{n+1} \rho_\alpha(Z_i - (A_n b)_i) = \arg\min_{b \in \mathbb{R}^d} \sum_{i=1}^{n+1} \rho_\alpha(Z_i - W_i b)$$

and $\hat{\pi}^{(n)}(\alpha) := 1 - \sum_{i=2}^{d} \hat{p}_i^{(n)}(\alpha)$, provides the optimal weights yielding the minimal $\alpha$-risk.

The large sample properties of $\hat{\beta}^{(n)}(\alpha)$, especially its $\sqrt{n}$-consistency, can be implied from the standard arguments and assumptions in the QR context (see [5]).

Let $\overline{W}^{(n)}$ and $\hat{\Sigma}_W^{(n)}$ be the mean vector and the covariance matrix of $W_i$ which are given by

$$\overline{W}^{(n)} := \frac{1}{n+1} \sum_{i=1}^{n+1} W_i, \quad \hat{\Sigma}_W^{(n)} := \frac{1}{n+1} \sum_{i=1}^{n+1} (W_i - \overline{W}^{(n)})(W_i - \overline{W}^{(n)})'$$

respectively. Here, the $(p, q)$th element of $\hat{\Sigma}_W^{(n)}$ is

$$\hat{\sigma}_{W, pq}^{(n)} = \begin{cases} \frac{n}{n+1} & \text{if } p = q = 1, \\ \frac{n(1-\kappa)}{(n+1)^2} (\overline{X}_1 - \overline{X}_p) & \text{if } q = 1, \ p = 2, \ldots, d, \\ \frac{n(1-\kappa)}{(n+1)^2} (\overline{X}_1 - \overline{X}_q) & \text{if } p = 1, \ q = 2, \ldots, d, \\ \hat{\sigma}_{pq} + \frac{n(\kappa-1)^2}{(n+1)^2} (\overline{X}_1 - \overline{X}_p)(\overline{X}_1 - \overline{X}_q) & \text{if } p, q = 2, \ldots, d, \end{cases}$$

where

$$\hat{\sigma}_{pq} := \frac{1}{n+1} \sum_{i=1}^{n} (X_{1i} - \overline{X}_p)(X_{1i} - \overline{X}_q) - \frac{n}{n+1} (\overline{X}_1 - \overline{X}_p)(\overline{X}_1 - \overline{X}_q).$$
Let $\mathbf{D}_{\Sigma_w} := \text{diag}(\hat{\sigma}_{W,11}, \ldots, \hat{\sigma}_{W,dd})$. Then the correlation matrix of $\{W_i\}_{i=1,\ldots,n+1}$ becomes $\mathbf{R} := \mathbf{D}_{\Sigma_w}^{-1/2} \Sigma_w^{-1/2} \mathbf{D}_{\Sigma_w}^{-1/2}$, and the $(p, q)$th element of $\mathbf{R}$ is given by

$$r_{pq} = \frac{1 + \left(\frac{(n+1)^2}{n(\kappa-1)^2} \cdot \left(\hat{\sigma}_{pp}/(\bar{X}_1 - \bar{X}_p)\right)\left(\bar{X}_1 - \bar{X}_q\right)\right)}{\left(1 + \left(\frac{(n+1)^2}{n(\kappa-1)^2} \cdot \left(\hat{\sigma}_{pp}/(\bar{X}_1 - \bar{X}_p)\right)\left(\bar{X}_1 - \bar{X}_q\right)\right)^2\right)^{1/2}}$$

for $p, q = 2, \ldots, d$. The above correlation coefficient can take values close to 1 when $n/\kappa^2$ is close to 0 with $(\bar{X}_1 - \bar{X}_p) \neq 0$ and $(\bar{X}_1 - \bar{X}_q) \neq 0$. Hence, the correlation of the estimated portfolio weights is possibly highly correlated among assets whose sample means differ from $\bar{X}_1$, while these problems are ignorable in an asymptotic inference problem if we take $\kappa = O(n^{1/2})$.

Thus far, we have seen that the $\alpha$-risk-minimizing portfolio can be obtained by (2.9), which was the result of Bassett et al. [3]. In what follows, we show that semiparametrically efficient inference of the optimal weights $\hat{\beta}^{(n)}(\alpha)$ is feasible. The quantity estimated by (2.9) can be regarded as a QR coefficient $\hat{\beta}(\alpha)$, defined by

$$F_{Z_i}^{-1}(\alpha \mid W_i) := F_{Z_i|W_i=w_i}(\alpha) =: W_i' \beta(\alpha), \tag{2.14}$$

where $F_{X|S}(\cdot)$ denotes a conditional quantile function, that is, $F_{X|S}(\alpha) := \inf\{x : P(X \leq x \mid S) \geq \alpha\}$. Note that here the QR model (2.14) has a random coefficient regression (RCR) interpretation of the form $Z_i = W_i' \beta(U_i)$ with componentwise monotone increasing function $\beta$ and random variables $U_i$ that are uniformly distributed over $[0,1]$, that is, $U_i \sim \text{Uniform}[0,1]$ (see [5]). Here, a choice such that $\beta(u) = [\beta_1(u), \beta_2(u), \ldots, \beta_d(u)]' := [b_1 + F_z^{-1}(u), b_2, \ldots, b_d]'$ with $F_z$ the distribution function of some independent and identically distributed (i.i.d.) $n$-tuple $(\xi_1, \ldots, \xi_n)$ yields

$$Z_i = W_i' \beta(U_i) = W_i' \beta(F_z(\xi_i)) = W_i' \begin{bmatrix} b_1 + \xi_i \\ b_2 \\ \vdots \\ b_d \end{bmatrix}, \tag{2.15}$$

Hence, recalling that the first component of $W_i$ is 1, it follows that, for any fixed $\alpha \in [0,1]$, the QR coefficient $\beta(\alpha)$ can be characterized as the parameter $b \in \mathbb{R}^d$ of a model such as

$$Z_i = W_i' b + \xi_i, \quad \xi_i \overset{\text{iid}}{\sim} G, \tag{2.16}$$

where the density $g$ of $G$ is subject to

$$g \in \mathcal{F}^\alpha := \left\{ f : \int_{-\infty}^0 f(x)dx = \alpha = 1 - \int_0^\infty f(x)dx \right\}, \tag{2.17}$$
The procedure that we will apply here to achieve semiparametric efficiency is based on the invariance principle, as introduced by Hallin and Werker [13]. To this end, first we should have locally asymptotic normality (LAN; see, e.g., van der Vaart [14]) for a parametric submodel \( P_{b,g}^{(n)} \), namely,

\[
\log \frac{P_{b+h/\sqrt{n},g}^{(n)}}{P_{b,g}^{(n)}} = h' \Delta_{b,g}^{(n)} + \frac{1}{2} h' I_{b,g} h + o_P(1), \quad h \in \mathbb{R}^d,
\]

where all the stochastic convergences are taken under \( P_{b,g} := P_{b,g}^{(\infty)} \). Here, the random vector \( \Delta_{b,g}^{(n)} \) is called the central sequence, and the positive definite matrix \( I_{b,g} \) is the information matrix. To ensure the LAN condition for model (2.18), the following assumption is required.

**Assumption 3.1.** The reference density \( f \) has finite Fisher information for location:

\[
0 < \mathcal{I}_f := \int_{-\infty}^{\infty} \varphi_f(x)^2 f(x) dx = \int_{0}^{1} \varphi_f(F^{-1}(u))^2 du < \infty, \quad \text{where } \varphi_f(x) := \frac{-f'(x)}{f(x)}. \quad \text{(3.2)}
\]

**Assumption 3.2.** The regression vectors \( W_i \) satisfy, under \( P_{b,g} \),

\[
\overline{W}_i^{(n)} \xrightarrow{P} \mu_W, \quad \bar{\Sigma}_W^{(n)} \xrightarrow{P} \Sigma_W, \quad \text{(3.3)}
\]

for some vector \( \mu_W \) and positive definite \( \Sigma_W \), where \( \overline{W}_i^{(n)} \) and \( \bar{\Sigma}_W^{(n)} \) are defined by (2.10).
Then, by Theorem 2.1 and Example 4.1 of Drost et al. [15], model (2.18) satisfies the uniform LAN condition for any $b_n$ of the form $b + O(n^{-1/2})$, with central sequence and information matrix

$$
\Delta_{b_0, f}^{(n)} = \frac{1}{\sqrt{n} + 1} \sum_{i=1}^{n+1} \varphi_f(e_{b_0,i}) W_i, \quad I_{b,f} = \mathcal{O}(\Sigma W + \mu W \mu'), \tag{3.4}
$$

where $e_{b_0,i}$ denotes the residual (i.e., $e_{b_0,i} := Z_i - W_i^* b_n$). Consequently, we have the contiguity $P_{b_0,h}^{(n)} \triangleright P_{b_0,f}^{(n)}$, and of course $P_{b_0,f}^{(n)} \triangleright P_{b_0}^{(n)}$ as well. Recall that the contiguity $P_{b_0}^{(n)} \triangleright Q_{b_0}^{(n)}$ means that for any sequence $S_{b_0}^{(n)}$, if $P_{b_0}^{(n)}(S_{b_0}^{(n)}) \to 0$, then $Q_{b_0}^{(n)}(S_{b_0}^{(n)}) \to 0$ also. The reason why we have specified uniform LAN, rather than LAN at single $b$, is the one-step improvement, which will be discussed later.

By following Hallin and Werker [13], a semiparametrically efficient procedure can be obtained by projecting $\Delta_{b_0, f}^{(n)}$ on some $\sigma$-field to which the generating group for $\{P_{b_0,f}^{(n)} | f \in \mathcal{F}^a\}$ becomes maximal invariant (see, e.g., Schmetterer [16]). For the quantile-restricted regression model (2.16), such a $\sigma$-field is studied by Hallin et al. [6] and found to be generated by signs and ranks of the residuals. Here, let us denote the sign of a residual as $S_{b_0,i}$, the rank of a residual as $R_{b_0,i}$, and the $\sigma$-field generated by them as

$$
\mathcal{B}_{b_0} := \sigma\left(S_{b_0,1}, \ldots, S_{b_0,n}; R_{b_0,1}, \ldots, R_{b_0,n}\right). \tag{3.5}
$$

Then, “good” inference should be based on

$$
\Delta_{b_0, f}^{(n)} := E_{b_0, f}^{(n)} \left[ \Delta_{b_0, f}^{(n)} \mid \mathcal{B}_{b_0} \right] = \frac{1}{\sqrt{n} + 1} \sum_{i=1}^{n+1} E_{b_0, f}^{(n)} \left[ \varphi_f(e_{b_0,i}) \mid \mathcal{B}_{b_0} \right] W_i

= \frac{1}{\sqrt{n} + 1} \sum_{i=1}^{n+1} E_{b_0, f}^{(n)} \left[ \varphi_f(F^{-1}(U_{b_0,i})) \mid \mathcal{B}_{b_0} \right] W_i

= \frac{1}{\sqrt{n} + 1} \sum_{i=1}^{n+1} \varphi_f(F^{-1}(V_{b_0,i})) W_i + o_P(1), \tag{3.6}
$$

where $U_{b_0,i} := F(e_{b_0,i})$ is i.i.d. uniform on $[0, 1]$ under $P_{b_0,f}^{(n)}$ and hence approximated by

$$
V_{b_0,i}^{(n)} := \begin{cases} 
\alpha \cdot \frac{R_{b_0,i}^{(n)}}{N_{b_0,L}^{(n)} + 1} & \text{if } R_{b_0,i}^{(n)} \leq N_{b_0,L}^{(n)}, \\
\alpha + (1 - \alpha) \cdot \frac{R_{b_0,i}^{(n)} - N_{b_0,L}^{(n)}}{n - N_{b_0,L}^{(n)} + 1} & \text{otherwise},
\end{cases} \tag{3.7}
$$

with $N_{b_0,L}^{(n)} := \# \{i \in \{1, \ldots, n\} \mid S_{b_0,i} \leq 0\}$. In short, we are first rewriting the residual $e_{b_0,i}$ as $F^{-1}(U_{b_0,i})$ with realizations $U_{b_0,i}$ of a $[0, 1]$-uniform random variable, and then approximating those $U_{b_0,i}$ as $V_{b_0,i}$ given $\{N_{b_0,L}^{(n)}, R_{b_0,1}^{(n)}, \ldots, R_{b_0,n}^{(n)}\}$.
Using this rank-based central sequence, we can construct the one-step estimator (see, e.g., Bickel [17]; Bickel et al. [18]) as follows.

**Definition 3.3.** For any sequence of estimators \( \hat{\theta}_n \), the discretized estimator \( \bar{\theta}_n \) is defined to be the nearest vertex of \( \{ \theta : \theta = (1/\sqrt{n})(i_1,i_2,\ldots,i_d) : i_i : \text{integers} \} \).

**Definition 3.4.** Let \( \tilde{\beta}^{(n)}(a) \) be the discretized version of \( \hat{\beta}^{(n)}(a) \) defined at (2.9). We define the (rank-based) one-step estimator of \( b \) based on reference density \( f \in \mathcal{F}^s \) as

\[
\hat{b}_f^{(n)} := \tilde{\beta}^{(n)}(a) + \frac{\Delta^{(n)}_{fg}}{\sqrt{n}}, \quad \text{with} \quad \Delta^{(n)}_{fg} := \hat{\beta}^{(n)} - \frac{f^{(0)}}{1 - a} \hat{\mu}_g \hat{\mu}_W.
\]  
(3.8)

where \( \hat{\beta}^{(n)} \) and \( \hat{\mu}_g \) are consistent estimates of

\[
\Delta^{(n)}_{fg} := \int_0^1 \varphi_f[F^{-1}(u)] \varphi_g[G^{-1}(u)] du,
\]  
(3.9)

\[
\hat{\mu}_g := \mathbb{E}[\varphi_g[G^{-1}(U)] \mid U \leq a] = \frac{-g^{(0)}}{a},
\]  
(3.10)

respectively.

Consistent estimates \( \hat{\beta}^{(n)} \) and \( \hat{\mu}_g \) can be obtained in the manner of Hallin et al. [19], which is done without the kernel estimation of \( g \), though here we omit the details.

**Lemma 3.5** (Section 4.1 of [6]). Under \( P_{b,g} \) with \( g \in \mathcal{F}^s \),

\[
\sqrt{n}(\hat{b}_f^{(n)} - b) \xrightarrow{d} \mathcal{N}(0, \Sigma_{fg}^{-1} \Sigma_{gf}^{-1}).
\]  
(3.11)

Therefore, the one-step estimator \( \hat{b}_f^{(n)} \) defined by (3.8) for \( b \) is semiparametrically efficient at \( f = g \).

In our original notation, the above statement can be rewritten as, for some \( a \in (0,1) \) fixed, \( \sqrt{n}(\hat{B}_f^{(n)} - \beta(a)) \xrightarrow{d} \mathcal{N}(0, \Sigma_{fg}^{-1} \Sigma_{gf}^{-1}) \).

Recall that the standard QR estimator, defined at (2.9), is asymptotically normal (see Koenker [5]):

\[
\sqrt{n}(\hat{\beta}^{(n)}(a) - \beta(a)) \xrightarrow{d} \mathcal{N}(0, D^{-1}),
\]  
(3.12)

where

\[
D := \frac{g^{2}(0)}{a(1 - a)} D_0, \quad \text{with} \quad D_0 = \lim_{n \to \infty} \frac{1}{n + 1} \sum_{i=1}^{n+1} W_i W_i',
\]  
(3.13)
Denote the true portfolio weight with respect to risk probability $\alpha$ by $\pi = (1 - \mathbf{1}_{d-1}^{\prime} \pi_2(\alpha), \pi_2(\alpha))^{\prime}$, where $\pi_2(\alpha) = (\pi_2(a), \ldots, \pi_d(a))^{\prime}$, and its standard quantile regression and our one-step estimators by $\tilde{\pi}^{(QR)} := (1 - \mathbf{1}_{d-1}^{\prime} \tilde{p}_2^{(n)}(\alpha), \tilde{p}_2^{(n)}(\alpha))^{\prime}$ and $\tilde{\pi}^{(OS)} := (1 - \mathbf{1}_{d-1}^{\prime} \tilde{b}_2^{(n)}, \tilde{b}_2^{(n)})^{\prime}$, respectively. Denote the block matrix of the covariance matrix of standard quantile and one-step estimators by

$$
D^{-1} = \begin{pmatrix} D_{11} & D_{12} \\ D_{12}^{\prime} & D_{22} \end{pmatrix}, \quad \Sigma_{fg}^{-1} \Sigma_{ff} \Sigma_{fg}^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^{\prime} & \Sigma_{22} \end{pmatrix},
$$

where submatrices $D_{22}$ and $\Sigma_{22}$ are $(d - 1) \times (d - 1)$ symmetric matrices for the covariance of portfolio weights $\pi_2$. Then we obtain the variances of the $\alpha$-risk-minimizing portfolio constructed by the standard quantile, and the one-step estimators are stated in the following proposition. Since direct evaluation gives the following statement, we skip its proof.

**Proposition 3.6.** The asymptotic conditional variances of an $\alpha$-risk-minimizing portfolio using the standard quantile regression and one-step estimators given at $X = x$ are, respectively,

$$
\text{Var}(X' \pi^{(QR)} | X = x = [x_1, x_2]) = x_1^2 \cdot \text{tr}[\mathbf{1}_{d-1}^{\prime} \mathbf{1}_{d-1} D_{22}] + 2x_1 \cdot \text{tr}[\mathbf{1}_{d-1}^{\prime} x_1 D_{22} + x_2 x_2^{\prime} D_{22}], \\
\text{Var}(X' \pi^{(OS)} | X = x = [x_1, x_2]) = x_1^2 \cdot \text{tr}[\mathbf{1}_{d-1}^{\prime} \mathbf{1}_{d-1} \Sigma_{22}] + 2x_1 \cdot \text{tr}[\mathbf{1}_{d-1}^{\prime} x_1^{\prime} \Sigma_{22} + x_2 x_2^{\prime} \Sigma_{22}],
$$

where $x_2 = (x_2, \ldots, x_d)^{\prime}$.

For any positive definite matrices $A$ and $B$, we say $A \leq B$ if $B - A$ is nonnegative definite. To compare the efficiency of the standard quantile regression estimator and the one-step estimator, we need to show that $\Sigma_{fg}^{-1} \Sigma_{ff} \Sigma_{fg}^{-1} \leq D^{-1}$. To see this, as in Section 3 of Koenker and Zhao [20], let us consider

$$
\Sigma := \begin{pmatrix} \Sigma_{fg} & \Sigma_{fg}^{-1} \\ \Sigma_{fg} & \Sigma_{fg}^{-1} D^{-1} \Sigma_{fg} \end{pmatrix}.
$$

Note that $\Sigma$ is a nonnegative definite matrix. If $\Sigma_{fg}^{-1} D \Sigma_{fg}^{-1}$ is a positive definite, then there exists orthogonal matrix $P$, such that

$$
P' \Sigma P = \begin{pmatrix} \Sigma_{fg} & \Sigma_{fg}^{-1} \\ 0 & \Sigma_{fg}^{-1} D^{-1} \Sigma_{fg} - \Sigma_{ff} \end{pmatrix},
$$
so $\Sigma_{fg}D^{-1}\Sigma_{ff} - \Sigma_{ff}$ is nonnegative definite. Hence, $D^{-1} - \Sigma_{fg}^{-1}\Sigma_{ff}\Sigma_{fg}^{-1}$ is nonnegative definite if $\Sigma_{fg}$ is nonsingular. This result assures that the one-step estimator is asymptotically more efficient than the standard quantile regression estimator. From this result, it is easy to see that

$$\text{Var}(X'\hat{\pi}_f^{(OS)}|X = x) \leq \text{Var}(X'\hat{\pi}_f^{(QR)}|X = x).$$

(3.18)

Also, by taking expectation on both sides, the same inequality holds for unconditional variances.

4. Numerical Studies

In this section, we examine the finite sample properties of the proposed one-step estimator described in Section 3 for the cases where $\alpha = 0.1$ and $0.5$. Our simulations are performed with two data generating processes to focus on the underlying true density $g$ and how the choice of the reference density $f$ might affect the finite sample performances.

The first data-generating process (DGP1) is the same as that investigated by Bassett et al. [3]. For DGP1, we consider the construction of an $\alpha$-minimizing portfolio from four independently distributed assets, that is, asset 1 is normally distributed with mean $0.05$ and standard deviation $0.02$. Asset 2 is a reversed $\chi^2_3$ density with location and scale chosen so that its mean and variance are identical to those of asset 1. Asset 3 is normally distributed with mean $0.09$ and standard deviation $0.05$. Finally, asset 4 has a $\chi^2_3$ density with identical mean and standard deviation to asset 3. DGP2 is a four-dimensional normal distribution with mean vectors the same as those of DGP1, and covariance matrix $\Sigma = [\sigma_{ij}]_{i,j=1,...,4}$ with $\text{diag} \Sigma = (0.02, 0.02, 0.05, 0.05)$ and $\sigma_{ij}$ for $i \neq j$ is $\sigma_i \sigma_j \rho$. Here, we set $\rho = 0.5$, which indicates that the asset returns have correlation 0.5. Notice that both DGP1 and DGP2 have the same mean and variance structures. The underlying true conditional densities of $u = Z - A_w b$ for DGP1 and DGP2 are a mixture of the normal $\chi^2_3$ and reversed $\chi^2_3$ distribution and the normal distributions, respectively. A simulation of the estimator, for sample size $n = 100, 500$, and $1000$ consists of 1000 replications. We choose prespecified expected return $\mu_0$ at 0.07.

For each scenario, we computed standard quantile regression estimates $\tilde{P}_i^{(n)}(\alpha)$ with corresponding portfolio weights $\tilde{\pi}_f^{(QR)} = (1 - \sum_{j=2}^d \tilde{P}_j^{(n)}(\alpha), \tilde{P}_2^{(n)}(\alpha), \ldots, \tilde{P}_d^{(n)}(\alpha))$, and our one-step estimates are defined by (3.8) for various choices of the reference density $f$ and actual density $g$ in the $\alpha$-minimizing portfolio allocation problem.

To make the problem a pure location model, we set the variance of the estimated residual to have one, that is, $\tilde{u} = \tilde{u}/\sqrt{\text{Var}(\tilde{u})}$, where $\tilde{u} = [\tilde{u}_i]_{i=1,...,n+1} = Z_i - (A_w \tilde{b}^{(n)}(\alpha))_i$. The true density $g$ can be estimated by the kernel estimator for DGP1,

$$\tilde{g}(u) = \frac{1}{(n+1)h} \sum_{i=1}^{n+1} K\left(\frac{u - \tilde{u}_i}{h}\right),$$

(4.1)

where $K$ is a kernel function, and $h$ is a bandwidth. The first derivative $g'(u)$ is estimated by

$$\tilde{g}'(u) = \frac{1}{(n+1)h^2} \sum_{i=1}^{n+1} K'(\frac{u - \tilde{u}_i}{h}).$$

(4.2)
As for the DGP2, the actual density \( g \) becomes normal because the portfolio is constructed by normally distributed returns. We use the normal distribution (\( \mathcal{N} \)), the asymmetric Laplace distribution (AL), the logistic distribution (LGT), and the asymmetric power distribution (APD) with \( \lambda = 1.5 \) for the reference density \( f \).

The density function of the asymmetric power distribution introduced by Komunjer [7] is given by

\[
\hat{f}(u) = \begin{cases} 
\frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1 + 1/\lambda)} \exp \left[ -\frac{\delta_{\alpha,\lambda}}{\alpha^\lambda} |u|^{\frac{1}{\lambda}} \right] & \text{if } u \leq 0, \\
\frac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1 + 1/\lambda)} \exp \left[ -\frac{\delta_{\alpha,\lambda}}{(1 - \alpha)^\lambda} |u|^{\frac{1}{\lambda}} \right] & \text{if } u > 0,
\end{cases}
\]

where \( 0 \leq \alpha < 1, \lambda > 0, \) and

\[
\delta_{\alpha,\lambda} = \frac{2\alpha^\lambda(1 - \alpha)^\lambda}{\alpha^\lambda + (1 - \alpha)^\lambda}.
\]

When \( \alpha = 0.5 \), the APD pdf is symmetric around zero. In this case, the APD density reduces to the standard generalized power distribution (GPD) [21, pages 194-195]. Special cases of the GPD include uniform (\( \lambda = \infty \)), Gaussian (\( \lambda = 2 \)), and Laplace (\( \lambda = 1 \)) distributions. When \( \alpha \neq 0.5 \), the APD pdf is asymmetric. Special cases include asymmetric Laplace (\( \lambda = 1 \)), the two-piece normal (\( \lambda = 2 \)) distributions.

For a given sample size, we compute simulated mean and standard deviation of \( \hat{\pi}^{(QR)} \) and \( \hat{\pi}^{(OS)} \) and the relative efficiency \( \text{Var}(\hat{\pi}^{(OS)})/\text{Var}(\hat{\pi}^{(QR)}) \) for \( i = 1, \ldots, 4 \).

Table 1 gives the results of the relative efficiencies for DGP 1. When \( \alpha = 0.1 \), we see that the efficiency gains of one-step estimators with asymmetric Laplace reference density are large compared with other reference densities with \( n = 1000 \), while these efficiency gains are less when sample size is \( n = 100 \). When \( \alpha = 0.5 \), relative efficiency of assets 3 and 4 with asymmetric Laplace reference density is minimum, while for assets 1 and 2, relative efficiency with normal reference density is minimum. This is because of the covariance structure of \( \tilde{\Sigma}_n^{(i)} \) defined by (2.10). As can be seen in Section 2, if \( \mu_i \neq \mu_p \) and \( \mu_i \neq \mu_q \), the \( (p,q) \)th element of the correlation matrix defined by (2.13) has a value close to unity. In this case, the asymptotic variance of the usual quantile regression estimator becomes large, which leads to unsatisfactorily large variances in assets 3 and 4. However, the asymptotic variance of our one-step estimator does not have such problems.

Table 2 gives the results of the relative efficiencies for DGP2. In line with efficiency at a correctly specified reference density \( f = \mathcal{N} \), we see that the relative efficiency is minimal for all assets and sample sizes with \( \alpha = 0.1 \) and 0.5. Even though we misspecify the reference density \( f \neq \mathcal{N} \), there exists some sort of efficiency gain except for assets 1 and 2 of the asymmetric Laplace reference density with \( n = 100 \) and \( \alpha = 0.1 \). Efficiency gains for the normal reference density and logistic reference density are almost the same because the underlying true density is a symmetric normal distribution, and the asymmetric power reference density with \( \lambda = 1.5 \) outperforms the asymmetric Laplace reference density.

Figure 1 plots the kernel densities for the estimated portfolio weights for DGP2 with \( \alpha = 0.5 \) and \( n = 1000 \). We see that the standard quantile regression estimators have long
Table 1: $\text{Var}(\hat{\pi}^{(\text{OS})}) / \text{Var}(\hat{\pi}^{(\text{QR})})$ for DGP 1 [3]; in this case, we estimate an unknown $g$ (which must be $\mathcal{N} - \chi^2$ mixed) using the kernel method.

| $f$ | $\pi_1$ (.1) | $\pi_2$ (.1) | $\pi_3$ (.1) | $\pi_4$ (.1) | $\pi_1$ (.5) | $\pi_2$ (.5) | $\pi_3$ (.5) | $\pi_4$ (.5) |
|-----|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| $n = 100$ |               |               |               |               |               |               |               |               |
| AL  | 1.1054        | 1.1088        | 1.0036        | 1.0082        | 1.0211        | 1.0193        | 0.9925        | 0.9570        |
| N   | 0.9773        | 0.9826        | 0.9093        | 0.9331        | 0.9412        | 0.9383        | 0.9227        | 0.9490        |
| LGT | 0.9731        | 0.9788        | 0.9017        | 0.9289        | 0.9458        | 0.9436        | 0.9228        | 0.9463        |
| APD | 0.9851        | 0.9838        | 0.9827        | 0.9893        | 0.9517        | 0.9492        | 0.9273        | 0.9471        |
| $n = 500$ |               |               |               |               |               |               |               |               |
| AL  | 1.0266        | 1.0068        | 0.9549        | 0.9871        | 0.9596        | 0.9563        | 0.3626        | 0.5478        |
| N   | 0.9762        | 0.9778        | 0.8966        | 0.9340        | 0.9186        | 0.9138        | 0.7522        | 0.8285        |
| LGT | 0.9739        | 0.9769        | 0.8893        | 0.9301        | 0.9255        | 0.9225        | 0.7277        | 0.8102        |
| APD | 0.9765        | 0.9712        | 0.9773        | 0.9896        | 0.9291        | 0.9262        | 0.6922        | 0.7842        |
| $n = 1000$ |              |               |               |               |               |               |               |               |
| AL  | 0.8702        | 0.9084        | 0.5621        | 0.8193        | 0.9150        | 0.8985        | 0.2225        | 0.3614        |
| N   | 0.9416        | 0.9485        | 0.8680        | 0.9112        | 0.7850        | 0.7714        | 0.4019        | 0.5214        |
| LGT | 0.9453        | 0.9532        | 0.8640        | 0.9085        | 0.8101        | 0.7981        | 0.4043        | 0.5199        |
| APD | 0.9290        | 0.9296        | 0.9242        | 0.9645        | 0.8206        | 0.8077        | 0.3572        | 0.4806        |

Table 2: $\text{Var}(\hat{\pi}^{(\text{OS})}) / \text{Var}(\hat{\pi}^{(\text{QR})})$ for DGP2 (multinormal). In this case, residual density is a normal distribution. Hence, we adopt $g = \mathcal{N}$.

| $f$ | $\pi_1$ (.1) | $\pi_2$ (.1) | $\pi_3$ (.1) | $\pi_4$ (.1) | $\pi_1$ (.5) | $\pi_2$ (.5) | $\pi_3$ (.5) | $\pi_4$ (.5) |
|-----|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| $n = 100$ |               |               |               |               |               |               |               |               |
| AL  | 1.0890        | 1.0996        | 0.7311        | 0.7221        | 0.7428        | 0.7584        | 0.6652        | 0.6370        |
| N   | 0.5891        | 0.6013        | 0.4105        | 0.4199        | 0.5896        | 0.6010        | 0.5451        | 0.5347        |
| LGT | 0.6153        | 0.6280        | 0.4113        | 0.4229        | 0.6026        | 0.6154        | 0.5503        | 0.5356        |
| APD | 0.8512        | 0.8603        | 0.6120        | 0.6069        | 0.6076        | 0.6204        | 0.5580        | 0.5420        |
| $n = 500$ |               |               |               |               |               |               |               |               |
| AL  | 0.8970        | 0.8814        | 0.7742        | 0.7402        | 0.6944        | 0.7030        | 0.6235        | 0.5965        |
| N   | 0.5090        | 0.4993        | 0.4883        | 0.4648        | 0.5618        | 0.5666        | 0.5169        | 0.4949        |
| LGT | 0.5184        | 0.5096        | 0.4892        | 0.4678        | 0.5703        | 0.5761        | 0.5321        | 0.5086        |
| APD | 0.7271        | 0.7183        | 0.6746        | 0.6423        | 0.5723        | 0.5784        | 0.5317        | 0.5099        |
| $n = 1000$ |              |               |               |               |               |               |               |               |
| AL  | 0.8836        | 0.8633        | 0.8015        | 0.8712        | 0.8777        | 0.8612        | 0.6726        | 0.7390        |
| N   | 0.4800        | 0.4655        | 0.4861        | 0.5095        | 0.7071        | 0.6944        | 0.5746        | 0.6261        |
| LGT | 0.4897        | 0.4750        | 0.4939        | 0.5198        | 0.7238        | 0.7094        | 0.5831        | 0.6347        |
| APD | 0.7144        | 0.7041        | 0.6617        | 0.7088        | 0.7261        | 0.7126        | 0.5845        | 0.6395        |

tails on both sides for all assets, whereas one-step estimators have a narrower interval and higher peak at the true weight. This confirms that the one-step estimators are more semiparametrically efficient than the standard ones.
Figure 1: Kernel density plots for the portfolio weights. Panels (a) to (d) correspond to the kernel density for assets 1 to 4, respectively. The density shows the standard quantile estimator (QR; solid line); the estimator with an asymmetric Laplace distribution reference density (AL; dashed line); normal distribution (N; dotted line); logistic distribution (LGT; dotted-dashed line); asymmetric power distribution (APD; long dashed line).

5. Empirical Application

We apply our methodology to weekly log returns of the 96 stocks of the TOPIX large 100 index. The samples run from January 5, 2007, to December 2, 2011, for a total of 257 observations. The stock prices are adjusted to take into account events such as stock splits on individual securities. Preliminary tests reveal that most log return series have high values of kurtosis and negative values of skewness in general, which indicates that the log returns are non-Gaussian.

We computed the optimal portfolio allocations for \( \alpha = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, \) and 0.5. We set \( \kappa = 1000 \) and \( \mu = -0.002 \), which is the third quartile of the average log-return distribution. For the first-round estimates, we used the standard quantile regression estimator, and for the one-step estimates, we chose a normal distribution as a reference
density. Since we do not have enough information about the shape of the portfolio distributions for the various choices of $\alpha$, the actual density $g$ is estimated by the kernel method.

Figure 2 plots the cumulative distribution functions of the $\alpha$-risk-minimizing portfolios obtained by the standard quantile regression estimates and one-step estimates for $\alpha = 0.1, 0.2, 0.3,$ and $0.5$. Summary statistics for the distributions of the different portfolios are reported in Table 3.

Figure 2 and Table 3 clearly show that the optimal $\alpha$-risk-minimizing portfolio manages to reduce the occurrence of events in the left tail when $\alpha$ is small for both standard QR estimates and one-step estimates. The standard deviation of the one-step estimates of an $\alpha$-minimizing portfolio is smaller than that of the standard QR estimates. We can also observe that the range of a constructed portfolio with one-step estimates is much smaller than that of standard QR estimates, due to the semiparametric efficiency properties of our one-step estimators. When $\alpha$ becomes large, the difference in the standard deviation of the constructed portfolio between standard QR estimates and one-step estimates tends to become large. Hence, efficiency gains are large for $\alpha = 0.5$, which is the mean absolute deviation portfolio.

Figure 2: Empirical cumulative distribution function of the $\alpha$-risk-minimizing portfolio based on the standard quantile regression estimator (thin line) and one-step estimator (thick line) for $\alpha = 0.1, 0.2, 0.3,$ and $0.5$, which corresponds to (a) to (d), respectively.
Figure 3: Efficient frontiers for an $\alpha$-risk-minimizing portfolio based on the standard quantile regression estimator and the one-step one. The lines with triangles and circles represent the pair of obtained standard deviation and mean for the portfolio with $\alpha = 0.5$ and 0.1, respectively. The solid and dashed lines represent risks and returns for the standard quantile regression and one-step portfolios, respectively.

Table 3: Summary statistics for the $\alpha$-minimizing-portfolio using quantile regression methods $X'\hat{\pi}_{QR}$ and one-step estimates $X'\hat{\pi}_{OS}$.

| $\alpha$ | Min   | Max   | Mean  | Std. dev. | $Q(\alpha)$ | Min   | max   | Mean  | Std. dev. | $Q(\alpha)$ |
|---------|-------|-------|-------|-----------|-------------|-------|-------|-------|-----------|-------------|
| 0.01    | −0.0210 | 0.0557 | −0.0200 | 0.0187 | −0.0210 | 0.0555 | −0.0020 | 0.0186 | −0.0209 |
| 0.05    | −0.0192 | 0.0756 | −0.00020 | 0.0196 | −0.0192 | 0.0202 | 0.0733 | −0.0020 | 0.0192 | −0.0194 |
| 0.1     | −0.0190 | 0.0649 | −0.0020 | 0.0196 | −0.0190 | 0.0227 | 0.0606 | −0.0020 | 0.0189 | −0.0191 |
| 0.2     | −0.0959 | 0.0708 | −0.0020 | 0.0189 | −0.0149 | 0.0944 | 0.0624 | −0.0020 | 0.0179 | −0.0146 |
| 0.3     | −0.1247 | 0.0512 | −0.0020 | 0.0171 | −0.0078 | 0.1200 | 0.0487 | −0.0020 | 0.0163 | −0.0077 |
| 0.4     | −0.1307 | 0.0545 | −0.0020 | 0.0173 | −0.0038 | 0.1230 | 0.0482 | −0.0020 | 0.0162 | −0.0040 |
| 0.5     | −0.1011 | 0.0605 | −0.0020 | 0.0171 | −0.0012 | 0.0958 | 0.0537 | −0.0020 | 0.0158 | −0.0013 |

Note: the corresponding summary statistics of the TOPIX log returns for minimum, maximum, mean, and standard deviation are $−0.2202$, $0.0924$, $−0.0032$, and $0.0328$, respectively. Also, quantiles for the TOPIX log returns for $\alpha = 0.05, 0.1, 0.2, 0.3, 0.4$, and $0.5$ are $−0.0981, −0.0536, −0.0158, −0.0068$, and $0.0006$, respectively.

Another interesting finding is that the standard QR-constructed portfolios have high-density peaks at the required quantiles for all values of $\alpha$, whereas the portfolio constructed by one-step estimates has a quite moderate density reduction at the required quantiles.

Consistent with economic intuition, higher risk aversion is associated with a shorter left tail. In the case where $\alpha \leq 0.1$, maximum loss is limited to less than $−0.02$. This result is particularly striking given that the sample includes the stock market crash of October 2008 due to the US subprime mortgage crisis and the bankruptcy of Lehman Brothers, which resulted in a weekly loss of more than $−0.220$ for TOPIX. The sample also includes the stock market crash of March 2011 due to the catastrophic earthquake and tsunami that hit Japan, which resulted in a weekly loss of $−0.104$.

Figure 3 presents empirical efficient frontiers corresponding to the standard quantile regression-based portfolios and one-step estimates of a portfolio with $\alpha = 0.1$ and 0.5.
Figure 3 clearly illustrates that the standard quantile regression-based portfolio is completely inefficient, far from the one-step frontier.

6. Summary and Conclusions

This paper considered a semiparametrically efficient estimation of an $\alpha$-risk-minimizing portfolio. A one-step estimator based on residual signs and ranks was proposed, and simulations were performed to compare the finite sample relative efficiencies for the standard quantile regression estimators and the one-step one. These simulations confirmed our theoretical findings. An empirical application to construct a portfolio using 96 Japanese stocks was investigated and confirms that the one-step $\alpha$-risk-minimizing portfolio has smaller variance that is obtained by the standard quantile regression estimator.

Further research topics include (1) construction of portfolios without short-sale constraints and (2) extending the results to the covariates of time series with heteroskedastic returns. For the former, we impose nonnegativity of the weights by using a penalty function containing a term that diverges to infinity as any of the weights becomes negative (see [22]). For the latter, we refer to Hallin et al. [6] and Tanaii [23].

Acknowledgments

This paper was supported by Norinchukin Bank and the Nochu Information System Endowed Chair of Financial Engineering in the Department of Management Science, Tokyo University of Science. The authors thank Professors Masanobu Taniguchi and Marc Hallin and an anonymous referee for their helpful comments.

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, “Coherent measures of risk,” Mathematical Finance, vol. 9, no. 3, pp. 203–228, 1999.
[2] H. Föllmer and A. Schied, “Convex measures of risk and trading constraints,” Finance and Stochastics, vol. 6, no. 4, pp. 429–447, 2002.
[3] G. W. J. Bassett, R. Koenker, and G. Kordas, “Pessimistic portfolio allocation and choquet expected utility,” Journal of Financial Econometrics, vol. 2, no. 4, pp. 477–492, 2004.
[4] R. Koenker and G. Bassett, Jr., “Regression quantiles,” Econometrica, vol. 46, no. 1, pp. 33–50, 1978.
[5] R. Koenker, Quantile Regression, vol. 38 of Econometric Society Monographs, Cambridge University Press, Cambridge, UK, 2005.
[6] M. Hallin, C. Vermandele, and B. J. M. Werker, “Semiparametrically efficient inference based on signs and ranks for median-restricted models,” Journal of the Royal Statistical Society B, vol. 70, no. 2, pp. 389–412, 2008.
[7] I. Komunjer, “Asymmetric power distribution: theory and applications to risk measurement,” Journal of Applied Econometrics, vol. 22, no. 5, pp. 891–921, 2007.
[8] Q. Zhao, “Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form,” Econometric Theory, vol. 17, no. 4, pp. 765–784, 2001.
[9] Y.-J. Whang, “Smoothed empirical likelihood methods for quantile regression models,” Econometric Theory, vol. 22, no. 2, pp. 173–205, 2006.
[10] I. Komunjer and Q. Vuong, “Semiparametric efficiency bound in time-series models for conditional quantiles,” Econometric Theory, vol. 26, no. 2, pp. 383–405, 2010.
[11] S. Uryasev, “Conditional value-at-risk: optimization algorithms and applications,” in Proceedings of the IEEE/IAFE/INFORMS Conference on Computational Intelligence for Financial Engineering (CIFEr ’00), pp. 49–57, 2000.
[12] G. C. Pflug, “Some remarks on the value-at-risk and the conditional value-at-risk,” in Probabilistic Constrained Optimization, vol. 49 of Nonconvex Optimization and Its Applications, pp. 272–281, Kluwer Academic Publishers, Dodrecht, The Netherlands, 2000.

[13] M. Hallin and B. J. M. Werker, “Semi-parametric efficiency, distribution-freeness and invariance,” Bernoulli, vol. 9, no. 1, pp. 137–165, 2003.

[14] A. W. van der Vaart, Asymptotic Statistics, vol. 3 of Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK, 1998.

[15] F. C. Drost, C. A. J. Klaassen, and B. J. M. Werker, “Adaptive estimation in time-series models,” The Annals of Statistics, vol. 25, no. 2, pp. 786–817, 1997.

[16] L. Schmetterer, Introduction to Mathematical Statistics, Springer, Berlin, Germany, 1974.

[17] P. J. Bickel, “On adaptive estimation,” The Annals of Statistics, vol. 10, no. 3, pp. 647–671, 1982.

[18] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, Efficient and Adaptive Estimation for Semiparametric Models, Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, Md, USA, 1993.

[19] M. Hallin, H. Oja, and D. Paindaveine, “Semi-parametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape,” The Annals of Statistics, vol. 34, no. 6, pp. 2757–2789, 2006.

[20] R. Koenker and Q. Zhao, “Conditional quantile estimation and inference for ARCH models,” Econometric Theory, vol. 12, no. 5, pp. 793–813, 1996.

[21] N. L. Johnson, S. Kotz, and N. Balakrishnan, Continuous Univariate Distributions. Vol. 1, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2nd edition, 1994.

[22] S. Leorato, F. Peracchi, and A. V. Tanase, “Asymptotically efficient estimation of the conditional expected shortfall,” Computational Statistics & Data Analysis, vol. 56, no. 4, pp. 768–784, 2012.

[23] H. Tanai, Inference for the quantiles of ARCH processes, Ph.D. thesis, Université libre de Bruxelles, Brussels, Belgium, 2009.
Research Article

Estimation for Non-Gaussian Locally Stationary Processes with Empirical Likelihood Method

Hiroaki Ogata

Waseda University, Tokyo 169-8050, Japan

Correspondence should be addressed to Hiroaki Ogata, hiroakiogata@aoni.waseda.jp

Received 28 January 2012; Revised 28 March 2012; Accepted 30 March 2012

Academic Editor: David Veredas

Copyright © 2012 Hiroaki Ogata. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An application of the empirical likelihood method to non-Gaussian locally stationary processes is presented. Based on the central limit theorem for locally stationary processes, we give the asymptotic distributions of the maximum empirical likelihood estimator and the empirical likelihood ratio statistics, respectively. It is shown that the empirical likelihood method enables us to make inferences on various important indices in a time series analysis. Furthermore, we give a numerical study and investigate a finite sample property.

1. Introduction

The empirical likelihood is one of the nonparametric methods for a statistical inference proposed by Owen [1, 2]. It is used for constructing confidence regions for a mean, for a class of M-estimates that includes quantile, and for differentiable statistical functionals. The empirical likelihood method has been applied to various problems because of its good properties: generality of the nonparametric method and effectiveness of the likelihood method. For example, we can name applications to the general estimating equations, [3] the regression models [4–6], the biased sample models [7], and so forth. Applications are also extended to dependent observations. Kitamura [8] developed the blockwise empirical likelihood for estimating equations and for smooth functions of means. Monti [9] applied the empirical likelihood method to linear processes, essentially under the circular Gaussian assumption, using a spectral method. For short- and long-range dependence, Nordman and Lahiri [10] gave the asymptotic properties of the frequency domain empirical likelihood. As we named above, some applications to time series analysis can be found but it seems that they were mainly for stationary processes. Although stationarity is the most fundamental assumption when we are engaged in a time series analysis, it is also known that real time series data are generally nonstationary (e.g., economics analysis). Therefore we need to
use nonstationary models in order to describe the real world. Recently Dahlhaus [11–13] proposed an important class of nonstationary processes, called locally stationary processes. They have so-called time-varying spectral densities whose spectral structures smoothly change in time.

In this paper we extend the empirical likelihood method to non-Gaussian locally stationary processes with time-varying spectra. First, We derive the asymptotic normality of the maximum empirical likelihood estimator based on the central limit theorem for locally stationary processes, which is stated in Dahlhaus [13, Theorem A.2]. Next, we show that the empirical likelihood ratio converges to a sum of Gamma distribution. Especially, when we consider a stationary case, that is, the time-varying spectral density is independent of a time parameter, the asymptotic distribution becomes the chi-square.

As an application of this method, we can estimate an extended autocorrelation for locally stationary processes. Besides we can consider the Whittle estimation which is stated in Dahlhaus [13].

This paper is organized as follows. Section 2 briefly reviews the stationary processes and explains about the locally stationary processes. In Section 3, we propose the empirical likelihood method for non-Gaussian locally stationary processes and give the asymptotic properties. In Section 4 we give numerical studies on confidence intervals of the autocorrelation for locally stationary processes. Proofs of theorems are given in Section 5.

2. Locally Stationary Processes

The stationary process is a fundamental setting in a time series analysis. If the process \( \{X_t\}_{t \in \mathbb{Z}} \) is stationary with mean zero, it is known to have the spectral representation:

\[
X_t = \int_{-\pi}^{\pi} \exp(i\lambda t)A(\lambda) d\xi(\lambda),
\]

where \( A(\lambda) \) is a \( 2\pi \)-periodic complex-valued function with \( A(-\lambda) = \overline{A(\lambda)} \), called transfer function, and \( \xi(\lambda) \) is a stochastic process on \([-\pi, \pi] \) with \( \xi(-\lambda) = \overline{\xi(\lambda)} \) and

\[
E[d\xi(\lambda)] = 0, \quad \text{Cov}(d\xi(\lambda_1), d\xi(\lambda_2)) = \eta(\lambda_1 - \lambda_2),
\]

where \( \eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j) \) is the \( 2\pi \)-periodic extension of the Dirac delta function. If the process is stationary, the covariance between \( X_t \) and \( X_{t+k} \) is independent of time \( t \) and a function of only the time lag \( k \). We denote it by \( \gamma(k) = \text{Cov}(X_t, X_{t+k}) \). The Fourier transform of the autocovariance function

\[
g(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) \exp(-ik\lambda)
\]

is called spectral density function. In the expression of (2.1), the spectral density function is written by \( g(\lambda) = |A(\lambda)|^2 \). It is estimated by the periodogram, defined by \( I_T(\lambda) = (2\pi)^{-1} |\sum_{t=1}^{T} X_t \exp(-i\lambda t)|^2 \). If one wants to change the weight of each data, we can insert
the function \( h(x) \) defined on \([0, 1]\) into the periodogram: 
\[
I_T (\lambda) = \left( 2\pi \sum_{t=1}^{T} h(t/T)^2 \right)^{-1} \left| \sum_{t=1}^{T} h(t/T) X_t \exp(-i\lambda t) \right|^2.
\]
The function \( h(x) \) is called data taper. Now, we give a simple example of the stationary process below.

**Example 2.1.** Consider the following AR\((p)\) process:

\[
\sum_{j=0}^{p} a_j X_{t-j} = \varepsilon_t,
\]
where \( \varepsilon_t \) are independent random variables with mean zero and variance 1. In the form of (2.1), this is obtained by letting

\[
A(\lambda) = \frac{1}{\sqrt{2\pi}} \left( \sum_{j=0}^{p} a_j \exp(-i\lambda j) \right)^{-1}.
\]

As an extension of the stationary process, Dahlhaus [13] introduced the concept of locally stationary. An example of the locally stationary processes is the following time-varying AR\((p)\) process:

\[
\sum_{j=0}^{p} a_j \left( \frac{t}{T} \right) X_{t-j,T} = \varepsilon_t,
\]
where \( a_j(u) \) is a function defined on \([0, 1]\) and \( \varepsilon_t \) are independent random variables with mean zero and variance 1. If all \( a_j(u) \) are constant, the process (2.6) reduces to stationary. To define a general class of the locally stationary processes, we can naturally consider the time-varying spectral representation

\[
X_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,T}(\frac{t}{T}, \lambda) d\xi(\lambda).
\]

However, it turns out that (2.6) has not exactly but only approximately a solution of the form of (2.7). Therefore, we only require that (2.7) holds approximately. The following is the definition of the locally stationary processes given by Dahlhaus [13].

**Definition 2.2.** A sequence of stochastic processes \( X_{t,T} \) \((t = 1, \ldots, T)\) is called locally stationary with mean zero and transfer function \( A_{t,T}^o \), if there exists a representation

\[
X_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,T}^o(\lambda) d\xi(\lambda),
\]
where the following holds.
 Consider an inference on a parameter \( \theta \) from an observed stretch \( X_{1,T}, \ldots, X_{T,T} \). We suppose that information about \( \theta \) exists through a system of general estimating equations.

For short- or long-memory processes, Nordman and Lahiri [10] supposed that \( \theta_0 \), the true value of \( \theta \), is specified from the following spectral moment condition:

\[
\int_{-\pi}^{\pi} \phi(\lambda, \theta_0) g(\lambda) d\lambda = 0,
\]

(3.1)

Moreover, we define the local periodogram \( I_N(u, \lambda) \) (for even \( N \)) as follows:

\[
d_N(u, \lambda) = \sum_{s=1}^{N} h\left(\frac{s}{N}\right) X_{[uT]-N/2:s,T} \exp(-i\lambda s),
\]

\[
H_{k,N} = \sum_{s=1}^{N} h\left(\frac{s}{N}\right)^k, \quad (2.11)
\]

\[
I_N(u, \lambda) = \frac{1}{2\pi H_{2,N}} |d_N(u, \lambda)|^2.
\]

Here, \( h : \mathbb{R} \to \mathbb{R} \) is a data taper with \( h(x) = 0 \) for \( x \notin [0, 1] \). Thus, \( I_N(u, \lambda) \) is nothing but the periodogram over a segment of length \( N \) with midpoint \( [uT] \). The shift from segment to segment is denoted by \( S \), which means we calculate \( I_N \) with midpoints \( t_j = S(j-1) + N/2 \) (for \( j = 1, \ldots, M \)), where \( T = S(M - 1) + N \), or, written in rescaled time, at time points \( u_j := t_j/T \). Hereafter we set \( S = 1 \) rather than \( S = N \). That means the segments overlap each other.

### 3. Empirical Likelihood Approach for Non-Gaussian Locally Stationary Processes

Consider an inference on a parameter \( \theta \in \Theta \subset \mathbb{R}^d \) based on an observed stretch \( X_{1,T}, \ldots, X_{T,T} \). We suppose that information about \( \theta \) exists through a system of general estimating equations. For short- or long-memory processes, Nordman and Lahiri [10] supposed that \( \theta_0 \), the true value of \( \theta \), is specified from the following spectral moment condition:
where \( \phi(\lambda, \theta) \) is an appropriate function depending on \( \theta \). Following this manner, we naturally suppose that \( \theta_0 \) satisfies the following time-varying spectral moment condition:

\[
\int_0^1 \int_{-\pi}^{\pi} \phi(u, \lambda, \theta_0) g(u, \lambda) d\lambda du = 0
\]

(3.2)

in a locally stationary setting. Here \( \phi : [0, 1] \times [-\pi, \pi] \times \mathbb{R}^q \to \mathbb{C}^q \) is a function depending on \( \theta \) and satisfies Assumption 3.4(i). We give brief examples of \( \phi \) and corresponding \( \theta_0 \).

**Example 3.1** (autocorrelation). Let us set

\[
\phi(u, \lambda, \theta) = \theta - \exp(ik).
\]

Then (3.2) leads to

\[
\theta_0 = \frac{\int_0^1 \int_{-\pi}^{\pi} \exp(ik) g(u, \lambda) d\lambda du}{\int_0^1 \int_{-\pi}^{\pi} g(u, \lambda) d\lambda du}.
\]

(3.4)

When we consider the stationary case, that is, \( g(u, \lambda) \) is independent of the time parameter \( u \), (3.4) becomes

\[
\theta_0 = \frac{\int_{-\pi}^{\pi} \exp(ik) g(\lambda) d\lambda}{\int_{-\pi}^{\pi} g(\lambda) d\lambda} = \frac{y(k)}{y(0)} = \rho(k),
\]

(3.5)

which corresponds to the autocorrelation with lag \( k \). So, (3.4) can be interpreted as a kind of autocorrelation with lag \( k \) for the locally stationary processes.

**Example 3.2** (Whittle estimation). Consider the problem of fitting a parametric spectral model to the true spectral density by minimizing the disparity between them. For the stationary process, this problem is considered in Hosoya and Taniguchi [14] and Kakizawa [15]. For the locally stationary process, the disparity between the parametric model \( g_0(u, \lambda) \) and the true spectral density \( g(u, \lambda) \) is measured by

\[
\mathcal{L}(\theta) = \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \left\{ \log g_0(u, \lambda) + \frac{g(u, \lambda)}{g_0(u, \lambda)} \right\} d\lambda du
\]

(3.6)

and we seek the minimizer

\[
\theta_0 = \arg \min_{\theta \in \Theta} \mathcal{L}(\theta).
\]

(3.7)

Under appropriate conditions, \( \theta_0 \) in (3.7) is obtained by solving the equation \( \partial \mathcal{L}(\theta)/\partial \theta = 0 \). Suppose that the fitting model is described as \( g_0(u, \lambda) = \sigma^2(u) f_0(u, \lambda) \), which means \( \theta \) is free from innovation part \( \sigma^2(u) \). Then, by Kolmogorov’s formula (Dahlhaus [11, Theorem 3.2])
we can see that \( \int_{-\pi}^{\pi} \log g_\theta(u, \lambda) d\lambda \) is independent of \( \theta \). So the differential condition on \( \theta_0 \) becomes

\[
\int_0^1 \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} g_\theta(u, \lambda)^{-1} \bigg|_{\theta=\theta_0} g(u, \lambda) d\lambda du = 0. \tag{3.8}
\]

This is the case when we set

\[
\phi(u, \lambda, \theta) = \frac{\partial}{\partial \theta} g_\theta(u, \lambda)^{-1}. \tag{3.9}
\]

Now, we set

\[
m_j(\theta) = \int_{-\pi}^{\pi} \phi(u_j, \lambda, \theta) I_N(u_j, \lambda) d\lambda \quad (j = 1, \ldots, M) \tag{3.10}
\]

as an estimating function and use the following empirical likelihood ratio function \( R(\theta) \) defined by

\[
R(\theta) = \max_w \left\{ \prod_{j=1}^M M w_j \mid \sum_{j=1}^M w_j m_j(\theta) = 0, \ w_j \geq 0, \ \sum_{j=1}^M w_j = 1 \right\}. \tag{3.11}
\]

Denote the maximum empirical likelihood estimator by \( \tilde{\theta} \), which maximizes the empirical likelihood ratio function \( R(\theta) \).

**Remark 3.3.** We can also use the following alternative estimating function:

\[
m_j^{(T)}(\theta) = \frac{2\pi}{T} \sum_{t=1}^T \phi(u_j, \Delta_t, \theta) I_N(u_j, \Delta_t) \quad \left( \Delta_t = \frac{2\pi t}{T} \right) \tag{3.12}
\]

instead of \( m_j(\theta) \) in (3.10). The asymptotic equivalence of \( m_j(\theta) \) and \( m_j^{(T)}(\theta) \) can be proven if

\[
E \left| m_j^{(T)}(\theta) - m_j(\theta) \right| = o(1) \tag{3.13}
\]

is satisfied for any \( j \), and this is shown by straightforward calculation.

To show the asymptotic properties of \( \tilde{\theta} \) and \( R(\theta_0) \), we impose the following assumption.

**Assumption 3.4.** (i) The functions \( A(u, \lambda) \) and \( \phi(u, \lambda, \theta) \) are \( 2\pi \)-periodic in \( \lambda \), and the periodic extensions are differentiable in \( u \) and \( \lambda \) with uniformly bounded derivative \( \partial / \partial \lambda \) (resp. \( \partial / \partial u \)) \( A \) (resp. \( \phi \)).

(ii) The parameters \( N \) and \( T \) fulfill the relations \( T^{1/4} \ll N \ll T^{1/2} / \log T \).
Theorem 3.6. Suppose that Assumption 3.4 holds and the data taper \( h : \mathbb{R} \to \mathbb{R} \) with \( h(x) = 0 \) for all \( x \not\in (0,1) \) is continuous on \( \mathbb{R} \) and twice differentiable at all \( x \not\in p \) where \( p \) is a finite set and \( \sup_{x \not\in p} |h''(x)| < \infty \).

(iv) For \( k = 1, \ldots, 8 \),

\[
q_k(\lambda_1, \ldots, \lambda_{k-1}) = c_k \quad \text{(constant).}
\]

Remark 3.5. Assumption 3.4(ii) seems to be restrictive. However, this is required to use the central limit theorem for locally stationary processes (cf. Assumption A.1 and Theorem A.2 of Dahlhaus [13]) (Most of the restrictions on \( N \) result from the \( \sqrt{T} \)-unbiasedness in the central limit theorem). See also A.3. Remarks of Dahlhaus [13] for the detail.

Now we give the following theorem.

Theorem 3.6. Suppose that Assumption 3.4 holds and \( X_{1,T}, \ldots, X_{T,T} \) is realization of the locally stationary process which has the representation (2.8). Then,

\[
\sqrt{T}(\hat{\theta} - \theta_0) \overset{d}{\to} N(0, \Sigma)
\]

as \( T \to \infty \), where

\[
\Sigma = 4\pi \left( \Sigma_3 \Sigma_2^{-1} \Sigma_3 \right)^{-1} \left( \Sigma_1 \Sigma_2^{-1} \Sigma_3 \right)^{-1}
\]

(3.16)

Here \( \Sigma_1 \) and \( \Sigma_2 \) are the \( q \) by \( q \) matrices whose \( (i,j) \) elements are

\[
(\Sigma_1)_{ij} = \frac{1}{2\pi} \int_{0}^{1} \int_{-\pi}^{\pi} \phi_i(u, \lambda, \theta_0) \left( \phi_j(u, \lambda, \theta_0) + \phi_j(u, -\lambda, \theta_0) \right) g(u, \lambda) d\lambda du,
\]

\[
+ c_4 \int_{-\pi}^{\pi} \phi_i(u, \lambda, \theta_0) g(u, \lambda) d\lambda \int_{-\pi}^{\pi} \phi_j(u, \mu, \theta_0) g(u, \mu) d\mu du,
\]

(3.17)

\[
(\Sigma_2)_{ij} = \frac{1}{2\pi} \int_{0}^{1} \int_{-\pi}^{\pi} \phi_i(u, \lambda, \theta_0) \left( \phi_j(u, \lambda, \theta_0) + \phi_j(u, -\lambda, \theta_0) \right) g(u, \lambda) d\lambda du,
\]

\[
+ \int_{-\pi}^{\pi} \phi_i(u, \lambda, \theta_0) g(u, \lambda) d\lambda \int_{-\pi}^{\pi} \phi_j(u, \mu, \theta_0) g(u, \mu) d\mu du,
\]

(3.18)

respectively, and \( \Sigma_3 \) is the \( q \) by \( q \) matrix which is defined as

\[
\Sigma_3 = \int_{-\pi}^{\pi} \frac{\partial \phi(u, \lambda, \theta)}{\partial \theta} g(u, \lambda) d\lambda du.
\]

(3.19)

In addition, we give the following theorem on the asymptotic property of the empirical likelihood ratio \( \mathcal{R}(\theta_0) \).
Theorem 3.7. Suppose that Assumption 3.4 holds and \( X_{1,T}, \ldots, X_{T,T} \) is realization of a locally stationary process which has the representation (2.8). Then,

\[
-\frac{1}{n} \log R(\theta_0) \xrightarrow{d} (\text{FN})'(\text{FN})
\]  

(3.20)
as \( T \to \infty \), where \( \mathbf{N} \) is a \( q \)-dimensional normal random vector with zero mean vector and covariance matrix \( \mathbf{I}_q \) (identity matrix) and \( \mathbf{F} = \mathbf{\Sigma}_2^{-1/2} \mathbf{\Sigma}_1^{1/2} \). Here \( \mathbf{\Sigma}_1 \) and \( \mathbf{\Sigma}_2 \) are same matrices in Theorem 3.6.

Remark 3.8. Denote the eigenvalues of \( \mathbf{FF} \) by \( a_1, \ldots, a_q \), then we can write

\[
(\text{FN})'(\text{FN}) = \sum_{i=1}^{q} Z_i
\]  

(3.21)
where \( Z_i \) is distributed as \( \text{Gamma}(1/2, 1/(2a_i)) \), independently.

Remark 3.9. If the process is stationary, that is, the time-varying spectral density is independent of the time parameter \( u \), we can easily see that \( \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 \) and the asymptotic distribution becomes the chi-square with degree of freedom \( q \).

Remark 3.10. In our setting, the number of the estimating equations and that of the parameters are equal. In that case, the empirical likelihood ratio at the maximum empirical likelihood estimator, \( R(\hat{\theta}) \), becomes one (cf. [3, page 305]). That means the test statistic in Theorem 3.7 becomes zero when we evaluate it at the maximum empirical likelihood estimator.

4. Numerical Example

In this section, we present simulation results of the estimation of the autocorrelation in locally stationary processes which is stated in Example 3.1. Consider the following time-varying AR(1) process:

\[
X_{t,T} - a \left( \frac{t}{T} \right) X_{t-1,T} = \varepsilon_t \quad \text{for} \ t \in \mathbb{Z},
\]  

(4.1)
where \( \varepsilon_t \overset{i.i.d.}{\sim} \text{Gamma}(3/\pi, (3/\pi)^{1/2}) - (3/\pi)^{1/2} \) and \( a(u) = (u - b)^2 \), \( b = 0.1, 0.5, 0.9 \). The observations \( X_{1,T}, \ldots, X_{T,T} \) are generated from the process (4.1), and we make the confidence intervals of the autocorrelation with lag \( k \) = 1, which is expressed as

\[
\hat{\theta}_0 = \frac{\int_{0}^{1/\pi} \int_{-\pi}^{\pi} x e^{i\lambda} g(u, \lambda) d\lambda du}{\int_{0}^{1/\pi} \int_{-\pi}^{\pi} g(u, \lambda) d\lambda du},
\]  

(4.2)

based on the result of Theorem 3.7. The several combinations of the sample size \( T \) and the window length \( N \) are chosen: \( (T, N) = (100, 10), (500, 10), (500, 50), (1000, 10), (1000, 100) \), and the data taper is set as \( h(x) = (1/2) \{1 - \cos(2\pi x)\} \). Then we calculate the values of the test statistic \( -\pi^{-1} \log R(\hat{\theta}) \) at numerous points \( \theta \) and obtain confidence intervals by
Table 1: 90% confidence intervals of the autocorrelation with lag \( k = 1 \).

| \( (T, N) \)         | Lower bound | Upper bound | Interval length | Successful rate |
|----------------------|-------------|-------------|-----------------|-----------------|
| \( b = 0.1, \theta_0 = 0.308 \)                     |             |             |                 |                 |
| (100, 10)            | 0.057       | 0.439       | 0.382           | 0.854           |
| (500, 10)            | 0.172       | 0.382       | 0.210           | 0.866           |
| (500, 50)            | 0.203       | 0.332       | 0.129           | 0.578           |
| (1000, 10)           | 0.203       | 0.356       | 0.154           | 0.826           |
| (1000, 100)          | 0.225       | 0.308       | 0.084           | 0.444           |
| \( b = 0.5, \theta_0 = 0.085 \)                     |             |             |                 |                 |
| (100, 10)            | -0.087      | 0.225       | 0.312           | 0.890           |
| (500, 10)            | 0.001       | 0.169       | 0.168           | 0.910           |
| (500, 50)            | 0.028       | 0.104       | 0.076           | 0.515           |
| (1000, 10)           | 0.023       | 0.139       | 0.116           | 0.922           |
| (1000, 100)          | 0.047       | 0.087       | 0.040           | 0.384           |
| \( b = 0.9, \theta_0 = 0.308 \)                     |             |             |                 |                 |
| (100, 10)            | 0.060       | 0.449       | 0.388           | 0.841           |
| (500, 10)            | 0.176       | 0.393       | 0.216           | 0.871           |
| (500, 50)            | 0.201       | 0.332       | 0.131           | 0.586           |
| (1000, 10)           | 0.203       | 0.359       | 0.156           | 0.827           |
| (1000, 100)          | 0.226       | 0.310       | 0.083           | 0.467           |

Collecting the points \( \theta \) which satisfy \(-\pi^{-1} \log R(\theta) < z_\alpha\), where \( z_\alpha \) is \( \alpha \)-percentile of the asymptotic distribution in Theorem 3.7. We admit that Assumption 3.4. (ii) is hard to hold in a finite sample experiment, but this Monte Carlo simulation is purely illustrative and just for investigating how the sample size and the window length affect the results of confidence intervals.

We set a confidence level as \( \alpha = 0.90 \) and carry out the above procedure 1000 times for each case. Table 1 shows the averages of lower and upper bounds, lengths of the intervals, and the successful rates. Looking at the results, we find out that the larger sample size gives the shorter length of the interval, as expected. Furthermore, the results indicate that the larger window length leads to the worse successful rate. We can predict that the best rate \( N/T \) lies around 0.02 because the combination \( (T, N) = (500, 10) \) seems to give the best result among all.

5. Proofs

5.1. Some Lemmas

In this subsection we give the three lemmas to prove Theorems 3.6 and 3.7. First of all, we introduce the following function \( L_N : \mathbb{R} \to \mathbb{R} \), which is defined by the \( 2\pi \)-periodic extension of

\[
L_N(\alpha) := \begin{cases}
N, & |\alpha| \leq \frac{1}{N} \\
1, & \frac{1}{N} \leq |\alpha| \leq \pi.
\end{cases}
\]  

(5.1)

The properties of the function \( L_N \) are described in Lemma A.4 of Dahlhaus [13].
Lemma 5.1. Suppose (3.2) and Assumption 3.4 hold. Then for $1 \leq k \leq 8$,

$$
\text{cum}[d_N(u_1, \lambda_1), \ldots, d_N(u_k, \lambda_k)]
= (2\pi)^{k-1} c_k \left\{ \prod_{j=1}^{k} A(u_j, \lambda_j) \right\} \exp \left\{ -i \sum_{j=1}^{k} \lambda_j ([u_j T] - [u_j T]) \right\}
\times \sum_{s=1}^{N} \left\{ \prod_{j=1}^{k} h \left( s + [u_k T] - [u_j T] \right) \right\} \exp \left\{ -i \left( \sum_{j=1}^{k} \lambda_j \right) s \right\}
+ O \left( \frac{N^2}{T} \right) + O \left( (\log N)^{k-1} \right) = O \left( L_N \left( \sum_{j=1}^{k} \lambda_j \right) \right) + O \left( \frac{N^2}{T} \right) + O \left( (\log N)^{k-1} \right).
$$

(5.2)

Proof. Let $\Pi = (-\pi, \pi]$ and let $\omega = (\omega_1, \ldots, \omega_k)$. Since

$$
\text{cum}(X_{t_1, T}, \ldots, X_{t_k, T}) = c_k \int_{\Pi^k} \exp \left\{ i \sum_{j=1}^{k} \omega_j T \right\} \left( \prod_{j=1}^{k} A_{t_j, T}^{\omega_j}(\omega_j) \right) \eta \left( \sum_{j=1}^{k} \omega_j \right) d\omega,
$$

(5.3)

the $k$th cumulant of $d_N$ is equal to

$$
c_k \int_{\Pi^k} \exp \left\{ i \sum_{j=1}^{k} \omega_j \left( [u_j T] - \frac{N}{2} \right) \right\} \eta \left( \sum_{j=1}^{k} \omega_j \right)
\times \prod_{j=1}^{k} \sum_{s=1}^{N} h \left( \frac{S}{N} \right) A_{[u_j T] - N/2 + s, T}^{\omega_j}(\omega_j) \exp \{-i(\lambda_j - \omega_j)s\} d\omega.
$$

(5.4)

As in the proof of Theorem 2.2 of Dahlhaus [12] we replace $A_{[u_1 T] - N/2 + s_1, T}^{\omega_1}(\omega_1)$ by $A(u_1 + (-N/2 + s_1)/T, \lambda_1)$ and we obtain

$$
\left| \sum_{s=1}^{N} h \left( \frac{S}{N} \right) \left\{ A_{[u_1 T] - N/2 + s, T}^{\omega_1}(\omega_1) - A \left( u_1 + \frac{-N/2 + s}{T}, \lambda_1 \right) \right\} \exp \{-i(\lambda_1 - \omega_1)s\} \right| \leq K
$$

(5.5)

with some constant $K$ while

$$
\left| \sum_{s=1}^{N} h \left( \frac{S}{N} \right) A_{[u_1 T] - N/2 + s, T}^{\omega_j}(\omega_j) \exp \{-i(\lambda_j - \omega_j)s\} \right| \leq KL_N(\lambda_j - \omega_j)
$$

(5.6)
for \( j = 2, \ldots, k \). The replacement error is smaller than

\[
K \int_{\Omega^k} \prod_{j=2}^{k} L_N(\lambda_j - \omega_j) d\omega \leq K (\log N)^{k-1}.
\]  

(5.7)

In the same way we replace \( A_{[u_0 T] - N/2 + s_j T}(\omega_j) \) by \( A(u_j + (-N/2 + s_j)/T, \lambda_j) \) for \( j = 2, \ldots, k \), and then we obtain

\[
c_k \sum_{s_1, \ldots, s_k = 1}^{N} \left\{ \prod_{j=1}^{k} h\left( \frac{s_j}{N} \right) A\left( u_j + \frac{-N/2 + s_j}{T}, \lambda_j \right) \right\} \exp\left( -i \sum_{j=1}^{k} \lambda_j s_j \right) \\
\times \int_{\Omega^k} \eta\left( \sum_{j=1}^{k} \omega_j \right) \exp\left\{ i \sum_{j=1}^{k} \omega_j \left( [u_j T] - \frac{N}{2} + s_j \right) \right\} d\omega + O\left( (\log N)^{k-1} \right).
\]

(5.8)

The integral part is equal to

\[
\prod_{j=1}^{k-1} \int_{\Omega} \exp\left\{ i \omega_j \left( [u_j T] - [u_k T] + s_j - s_k \right) \right\} d\omega_j.
\]

(5.9)

So we get

\[
(2\pi)^{k-1} c_k \sum_{s=1}^{N} \left\{ \prod_{j=1}^{k} h\left( \frac{s + [u_k T] - [u_j T]}{N} \right) A\left( u_j + \frac{-N/2 + s + [u_k T] - [u_j T]}{T}, \lambda_j \right) \right\} \\
\times \exp\left\{ -i \sum_{j=1}^{k} \lambda_j (s + [u_k T] - [u_j T]) \right\} + O\left( (\log N)^{k-1} \right).
\]

(5.10)

Since \( h(x) = 0 \) for \( x \notin (0, 1) \), we only have to consider the range of \( s \) which satisfies \( 1 \leq s + [u_k T] - [u_j T] \leq N - 1 \). Therefore we can regard \((-N/2 + s + [u_k T] - [u_j T])/T \) as \( O(N/T) \), and Taylor expansion of \( A \) around \( u_j \) gives the first equation of the desired result. Moreover, as in the same manner of the proof of Lemma A.5 of Dahlhaus [13] we can see that

\[
\sum_{s=1}^{N} \left\{ \prod_{j=1}^{k} h\left( \frac{s + [u_k T] - [u_j T]}{N} \right) \right\} \exp\left\{ -i \left( \sum_{j=1}^{k} \lambda_j \right) s \right\} = O\left( L_N\left( \sum_{j=1}^{k} \lambda_j \right) \right),
\]

(5.11)

which leads to the second equation.
Lemma 5.2. Suppose (3.2) and Assumption 3.4 hold. Then,

$$P_M := \frac{1}{2\pi \sqrt{M}} \sum_{j=1}^{M} m_j(\theta_0) \xrightarrow{d} N(0, \Sigma).$$  \hfill (5.12)

Proof. We set

$$J_T(\phi) := \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \phi(u_j, \lambda, \theta_0) I_N(u_j, \lambda) d\lambda,$$

$$J(\phi) := \int_{0}^{1} \int_{-\pi}^{\pi} \phi(u, \lambda) g(u) d\lambda du.$$

Henceforth we denote $\phi(u, \lambda, \theta_0)$ by $\phi(u, \lambda)$ for simplicity. This lemma is proved by proving the convergence of the cumulants of all orders. Due to Lemma A.8 of Dahlhaus [13] the expectation of $P_M$ is equal to

$$\frac{\sqrt{M}}{2\pi} \left\{ J(\phi) + o(T^{-1/2}) \right\}. \hfill (5.14)$$

By (3.2) and $O(M) = O(T)$, this converges to zero.

Next, we calculate the covariance of $P_M$. From the relation $T = M + N - 1$ we can rewrite

$$P_M = \sqrt{\frac{M \sqrt{T}}{T}} J_T(\phi) = \sqrt{1 - \frac{N + 1}{T}} \frac{\sqrt{T}}{2\pi} J_T(\phi). \hfill (5.15)$$

Then the $(\alpha, \beta)$-element of the covariance matrix of $P_M$ is equal to

$$\frac{1}{(2\pi)^2} \left( 1 - \frac{N + 1}{T} \right) T \operatorname{cov} \{ J_T(\phi_\alpha), J_T(\phi_\beta) \}. \hfill (5.16)$$

Due to Lemma A.9 of Dahlhaus [13], this converges to

$$\frac{1}{2\pi} \int_{0}^{1} \left[ \int_{-\pi}^{\pi} \phi_\alpha(u, \lambda) \{ \phi_\beta(u, \lambda) + \phi_\beta(u, -\lambda) \} g(u, \lambda)^2 d\lambda \ight. \left. + \int_{-\pi}^{\pi} \phi_\alpha(u, \lambda) \phi_\beta(u, \mu) g(u, \lambda) g(u, \mu) q_4(\lambda, -\lambda, \mu) d\lambda d\mu \right] du. \hfill (5.17)$$

By Assumption 3.4(iv) the covariance tends to $\Sigma_1$.

The $k$th cumulant for $k \geq 3$ tends to zero due to Lemma A.10 of Dahlhaus [13]. Then we obtain the desired result. □
Lemma 5.3. Suppose (3.2) and Assumption 3.4 hold. Then,

\[ S_M := \frac{1}{2\pi M} \sum_{j=1}^{M} m_j(\theta_0)m_j(\theta_0)' \rightarrow^P \Sigma_2. \]  

(5.18)

Proof. First we calculate the mean of \((\alpha, \beta)\)-element of \(S_M\):

\[
E \left[ \frac{1}{2\pi M} \sum_{j=1}^{M} m_j(\theta_0)m_j(\theta_0)' \right]_{\alpha\beta} = \frac{1}{2\pi M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \phi_\alpha(u_j, \lambda)\phi_\beta(u_j, \mu)E[I_N(u_j, \lambda)I_N(u_j, \mu)] d\lambda d\mu
\]  

(5.19)

Due to Dahlhaus [12, Theorem 2.2 (i)] the second term of (5.19) becomes

\[
\frac{1}{2\pi M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \phi_\alpha(u_j, \lambda) \left\{ g(u_j, \lambda) + O\left(\frac{N^2}{T^2}\right) + O\left(\frac{\log N}{N}\right) \right\} d\lambda \\
\times \int_{-\pi}^{\pi} \phi_\beta(u_j, \mu) \left\{ g(u_j, \mu) + O\left(\frac{N^2}{T^2}\right) + O\left(\frac{\log N}{N}\right) \right\} d\mu
\]  

(5.20)

\[
= \frac{1}{2\pi} \int_{0}^{1} \left\{ \int_{-\pi}^{\pi} \phi_\alpha(u, \lambda) g(u, \lambda) d\lambda \int_{-\pi}^{\pi} \phi_\beta(u, \mu) g(u, \mu) d\mu \right\}
\times \left( O\left(\frac{1}{M}\right) + O\left(\frac{N^2}{T^2}\right) + O\left(\frac{\log N}{N}\right) \right). 
\]

Next we consider

\[
\text{cov}\{I_N(u_j, \lambda), I_N(u_j, \mu)\}
\]

\[
= \frac{1}{(2\pi H_{2,n})^2} \left[ \text{cum}\{d_N(u_j, \lambda), d_N(u_j, \mu)\} \text{cum}\{d_N(u_j, -\lambda), d_N(u_j, -\mu)\} \\
+ \text{cum}\{d_N(u_j, \lambda), d_N(u_j, -\mu)\} \text{cum}\{d_N(u_j, -\lambda), d_N(u_j, \mu)\} \\
+ \text{cum}\{d_N(u_j, \lambda), d_N(u_j, -\lambda), d_N(u_j, \mu), d_N(u_j, -\mu)\} \\
+ \text{cum}\{d_N(u_j, \lambda), d_N(u_j, -\lambda), d_N(u_j, -\mu), d_N(u_j, \mu)\} \\
+ \text{cum}\{d_N(u_j, \lambda), d_N(u_j, \mu), d_N(u_j, -\lambda), d_N(u_j, -\mu)\} \\
+ \text{cum}\{d_N(u_j, \lambda), d_N(u_j, -\lambda), d_N(u_j, \mu), d_N(u_j, -\mu)\} \right]. 
\]  

(5.21)
We calculate the three terms separately. From Lemma 5.1 the first term of (5.21) is equal to

\[
\frac{1}{(2\pi H_{2,N}^2)^2} \left\{ \sum_{s=1}^{N} \mathcal{h} \left( \frac{S}{N} \right)^2 \exp \left\{ -i(\lambda + \mu)s \right\} + O \left( \frac{N^2}{T} \right) + O(\log N) \right\}
\]

\[
\times \left\{ \sum_{s=1}^{N} \mathcal{h} \left( \frac{S}{N} \right)^2 \exp \left\{ -i(-\lambda - \mu)s \right\} + O \left( \frac{N^2}{T} \right) + O(\log N) \right\}.
\]

(5.22)

It converges to zero when \( \lambda \neq -\mu \) and is equal to

\[
g(u_j, \lambda)^2 + O \left( \frac{N}{T} \right) + O \left( \frac{\log N}{N} \right)
\]

(5.23)

when \( \lambda = -\mu \). Similarly the second term of (5.21) converges to zero when \( \lambda \neq \mu \) and is equal to (5.23) when \( \lambda = \mu \). We can also apply Lemma 5.1 to the third term of (5.21), and analogous calculation shows that it converges to zero. After all we can see that (5.19) converges to \((\Sigma_2)_{\alpha\beta}\), the \((\alpha, \beta)\)-element of \(\Sigma_2\).

Next we calculate the second-order cumulant:

\[
\text{cum} \left\{ \left[ \frac{1}{2\pi M} \sum_{j=1}^{M} m_j(\theta_0) \right] \phi_\alpha(u_j, \lambda) \phi_\beta(u_j, \mu) \right\}, \quad \left( \frac{1}{2\pi M} \sum_{j=1}^{M} m_j(\theta_0) \right)_{\alpha\beta_1} \left( \frac{1}{2\pi M} \sum_{j=1}^{M} m_j(\theta_0) \right)_{\alpha\beta_2}.
\]

(5.24)

This is equal to

\[
(2\pi M)^{-2}(2\pi H_{2,N})^{-4} \sum_{j_1=1}^{M} \sum_{j_2=1}^{M} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi_{\alpha_1}(u_{j_1}, \lambda_1) \phi_{\beta_1}(u_{j_1}, \mu_1) \phi_{\alpha_2}(u_{j_2}, \lambda_2) \phi_{\beta_2}(u_{j_2}, \mu_2)
\]

\[
\times \text{cum} \left\{ d_N(u_{j_1}, \lambda_1) d_N(u_{j_1}, -\lambda_1) d_N(u_{j_1}, \mu_1) d_N(u_{j_1}, -\mu_1) \right\}
\]

\[
\times d_N(u_{j_2}, \lambda_2) d_N(u_{j_2}, -\lambda_2) d_N(u_{j_2}, \mu_2) d_N(u_{j_2}, -\mu_2) \right\} \text{d}\lambda_1 \text{d}\mu_1 \text{d}\lambda_2 \text{d}\mu_2.
\]

(5.25)

Using the product theorem for cumulants (cf. [16, Theorem 2.3.2]) we have to sum over all indecomposable partitions \(\{P_1, \ldots, P_m\} \) with \(|P_i| \geq 2\) of the scheme

\[
d_N(u_{j_1}, \lambda_1) \quad d_N(u_{j_1}, -\lambda_1) \quad d_N(u_{j_1}, \mu_1) \quad d_N(u_{j_1}, -\mu_1)
\]

\[
d_N(u_{j_2}, \lambda_2) \quad d_N(u_{j_2}, -\lambda_2) \quad d_N(u_{j_2}, \mu_2) \quad d_N(u_{j_2}, -\mu_2).
\]

(5.26)

We can apply Lemma 5.1 to all cumulants which is seen in (5.25), and the dominant term of the cumulants is \(o(N^4)\) so (5.25) tends to zero. Then we obtain the desired result.
5.2. Proof of Theorem 3.6

Using the lemmas in Section 5.1, we prove Theorem 3.6. To find the maximizing weights $w_j$ of (3.11), we proceed by the Lagrange multiplier method. Write

$$
G = \sum_{j=1}^{M} \log (Mw_j) - Ma' \sum_{j=1}^{M} w_j m_j(\theta) + \gamma \left( \sum_{j=1}^{M} w_j - 1 \right),
$$

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

$$
\frac{\partial G}{\partial w_j} = \frac{1}{w_j} - Ma'm_j(\theta) + \gamma = 0.
$$

So the equation $\sum_{j=1}^{M} w_j (\partial G / \partial w_j) = 0$ gives $\gamma = -M$. Then, we may write

$$
w_j = \frac{1}{M + \alpha'm_j(\theta)},
$$

where the vector $\alpha = \alpha(\theta_0)$ satisfies $q$ equations given by

$$
\frac{1}{M} \sum_{j=1}^{M} \frac{m_j(\theta)}{1 + \alpha'm_j(\theta)} = 0.
$$

Therefore, $\tilde{\theta}$ is a minimizer of the following (minus) empirical log likelihood ratio function

$$
l(\theta) := \sum_{j=1}^{M} \log \left( 1 + \alpha'm_j(\theta) \right)
$$

and satisfies

$$
0 = \frac{\partial l(\theta)}{\partial \theta}\bigg|_{\theta=\tilde{\theta}} = \sum_{j=1}^{M} \left( \frac{\partial \alpha'(\theta)}{\partial \theta} \right) m_j(\theta) + \left( \frac{\partial m'_j(\theta)}{\partial \theta} \right) \alpha(\theta) \bigg|_{\theta=\tilde{\theta}}
$$

$$
= \sum_{j=1}^{M} \left( \frac{\partial m'_j(\theta)}{\partial \theta} \right) \alpha(\theta) \bigg|_{\theta=\tilde{\theta}}.
$$
Denote

\[
Q_{1M}(\theta, \alpha) := \frac{1}{M} \sum_{j=1}^{M} \frac{m_j(\theta)}{1 + \alpha'(\theta)m_j(\theta)},
\]

\[
Q_{2M}(\theta, \alpha) := \frac{1}{M} \sum_{j=1}^{M} \frac{1}{1 + \alpha'(\theta)m_j(\theta)} \frac{\partial m_j(\theta)}{\partial \theta} \alpha(\theta).
\]  

(5.33)

Then, from (5.30) and (5.32), we have

\[
0 = Q_{1M}(\tilde{\theta}, \tilde{\alpha})
\]

\[
= Q_{1M}(\theta_0, 0) + \frac{\partial Q_{1M}(\theta_0, 0)}{\partial \theta'} (\tilde{\theta} - \theta_0) + \frac{\partial Q_{1M}(\theta_0, 0)}{\partial \alpha'} (\tilde{\alpha} - 0) + o_p(\delta_M),
\]

\[
0 = Q_{2M}(\tilde{\theta}, \tilde{\alpha})
\]

\[
= Q_{2M}(\theta_0, 0) + \frac{\partial Q_{2M}(\theta_0, 0)}{\partial \theta'} (\tilde{\theta} - \theta_0) + \frac{\partial Q_{2M}(\theta_0, 0)}{\partial \alpha'} (\tilde{\alpha} - 0) + o_p(\delta_M),
\]

(5.34)

(5.35)

where \( \tilde{\alpha} = \alpha(\tilde{\theta}) \) and \( \delta_M = ||\tilde{\theta} - \theta_0|| + ||\tilde{\alpha}||. \) Let us see the asymptotic properties of the above four derivatives. First,

\[
\frac{\partial Q_{1M}(\theta_0, 0)}{\partial \theta'} = \frac{1}{M} \sum_{j=1}^{M} \frac{\partial m_j(\theta_0)}{\partial \theta'} = \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \frac{\partial \phi(u_j, \lambda, \theta)}{\partial \theta'} I_N(u_j, \lambda) d\lambda.
\]  

(5.36)

From Lemmas A.8 and A.9 of Dahlhaus [13], we have

\[
E \left[ \frac{\partial Q_{1M}(\theta_0, 0)}{\partial \theta'} \right] = \int_{0}^{1} \int_{-\pi}^{\pi} \frac{\partial \phi(u, \lambda, \theta)}{\partial \theta'} g(u, \lambda) d\lambda du + o(M^{-1/2}),
\]

\[
\text{cov} \left[ \left[ \frac{\partial Q_{1M}(\theta_0, 0)}{\partial \theta'} \right]_{u, \lambda}, \left[ \frac{\partial Q_{1M}(\theta_0, 0)}{\partial \theta'} \right]_{u', \lambda} \right] = O(M^{-1}),
\]

(5.37)

which leads to

\[
\frac{\partial Q_{1M}(\theta_0, 0)}{\partial \theta'} \rightarrow_{p} \int_{0}^{1} \int_{-\pi}^{\pi} \frac{\partial \phi(u, \lambda, \theta)}{\partial \theta'} g(u, \lambda) d\lambda du = \Sigma_3.
\]  

(5.38)

Similarly, we have

\[
\frac{\partial Q_{2M}(\theta_0, 0)}{\partial \alpha'} = \frac{1}{M} \sum_{j=1}^{M} \frac{\partial m_j(\theta_0)}{\partial \theta'} \rightarrow_{p} \Sigma_3'.
\]  

(5.39)
Next, from Lemma 5.3, we obtain
\[
\frac{\partial Q_1}{\partial \alpha'}(\theta_0, 0) = - \frac{1}{M} \sum_{j=1}^{M} m_j(\theta_0)m_j(\theta_0)' \overset{p}{\rightarrow} -2\pi \Sigma_2.
\] (5.40)

Finally, we have
\[
\frac{\partial Q_2}{\partial \theta'}(\theta_0, 0) = 0.
\] (5.41)

Now, (5.34), (5.35) and (5.38)–(5.41) give
\[
\left( \begin{array}{c} \tilde{\alpha} \\ \tilde{\theta} - \theta_0 \end{array} \right) = \left( \begin{array}{cc} \frac{\partial Q_1}{\partial \alpha'} & \frac{\partial Q_1}{\partial \theta'} \\ \frac{\partial Q_2}{\partial \alpha'} & \frac{\partial Q_2}{\partial \theta'} \end{array} \right)_{(\theta_0, 0)}^{-1} \left( \begin{array}{c} -Q_1(\theta_0, 0) + o_p(\delta_M) \\ o_p(\delta_M) \end{array} \right),
\] (5.42)

where
\[
\left( \begin{array}{cc} \frac{\partial Q_1}{\partial \alpha'} & \frac{\partial Q_1}{\partial \theta'} \\ \frac{\partial Q_2}{\partial \alpha'} & \frac{\partial Q_2}{\partial \theta'} \end{array} \right)_{(\theta_0, 0)} \overset{p}{\rightarrow} \left( \begin{array}{cc} -2\pi \Sigma_2 & \Sigma_3 \\ \Sigma_3' & 0 \end{array} \right).
\] (5.43)

Because of Lemma 5.2, we have
\[
Q_1(\theta_0, 0) = \frac{1}{M} \sum_{j=1}^{M} m_j(\theta_0) = O_p\left(M^{-1/2}\right).
\] (5.44)

From this and the relation (5.42), (5.43), we can see that \(\delta_M = O_p\left(M^{-1/2}\right)\). Again, from (5.42), (5.43), and Lemma 5.2, direct calculation gives that
\[
\sqrt{M}\left( \tilde{\theta} - \theta_0 \right) \overset{d}{\rightarrow} N(0, \Sigma).
\] (5.45)

5.3. Proof of Theorem 3.7

Using the lemmas in Section 5.1, we prove Theorem 3.7. The proof is the same as that of Theorem 3.6 up to (5.30). Let \(\alpha = \|\alpha\|e\) where \(\|e\| = 1\), and we introduce
\[
Y_j := \alpha'm_j(\theta_0), \quad Z_{M}^* := \max_{1 \leq j \leq M} \|m_j(\theta_0)\|.
\] (5.46)
Note $1/(1 + Y_j) = 1 - Y_j/(1 + Y_j)$ and from (5.30) we find that

$$e' \left\{ \frac{1}{M} \sum_{j=1}^{M} \left( 1 - \frac{Y_j}{1 + Y_j} \right) m_j(\theta_0) \right\} = 0,$$

$$e' \left( \frac{1}{M} \sum_{j=1}^{M} \alpha' m_j(\theta_0) m_j(\theta_0)' \right) = e' \left( \frac{1}{M} \sum_{j=1}^{M} m_j(\theta_0) \right), \quad (5.47)$$

$$\| \alpha \| e' \left( \frac{1}{M} \sum_{j=1}^{M} m_j(\theta_0) m_j(\theta_0)' \right) = e' \left( \frac{1}{M} \sum_{j=1}^{M} m_j(\theta_0) \right).$$

Every $w_j > 0$, so $1 + Y_j > 0$, and therefore by (5.47), we get

$$\| \alpha \| e' S_M e \leq \| \alpha \| e' \left( \frac{1}{2\pi M} \sum_{j=1}^{M} \frac{m_j(\theta_0) m_j(\theta_0)'}{1 + Y_j} \right) e \cdot \left( 1 + \max_j Y_j \right)$$

$$\leq \| \alpha \| e' \left( \frac{1}{2\pi} \sum_{j=1}^{M} \frac{m_j(\theta_0) m_j(\theta_0)'}{1 + Y_j} \right) e \cdot (1 + \| \alpha \| Z_M^*),$$

$$= e' M^{-1/2} P_M (1 + \| \alpha \| Z_M^*),$$

where $S_M$ and $P_M$ are defined in Lemmas 5.2 and 5.3. Then by (5.48), we get

$$\| \alpha \| \left\{ e' S_M e - Z_M^* e' \left( M^{-1/2} P_M \right) \right\} \leq e' \left( M^{-1/2} P_M \right).$$

(5.49)

From Lemmas 5.2 and 5.3 we can see that

$$M^{-1/2} P_M = O_P \left( M^{-1/2} \right), \quad S_M = O_P(1).$$

(5.50)

We evaluate the order of $Z_M^*$. We can write

$$Z_M^* \leq \max_{1 \leq j \leq M} \int_{-\pi}^{\pi} \left\| \phi_{\theta_0}(u_j, \lambda) \right\| I_N(u_j, \lambda) d\lambda =: \max_{1 \leq j \leq M} m_j^* (\theta_0) \quad \text{(say).}$$

(5.51)
Then, for any $\varepsilon > 0$,

$$
P\left( \max_{1 \leq j \leq M} m_j^*(\theta_0) > \varepsilon \sqrt{M} \right) \leq \sum_{j=1}^{M} P\left( m_j^*(\theta_0) > \varepsilon \sqrt{M} \right)
$$

$$
= \sum_{j=1}^{M} P\left( m_j^*(\theta_0)^3 > \left( \varepsilon \sqrt{M} \right)^3 \right) \leq \sum_{j=1}^{M} \frac{1}{\varepsilon^3 M^{3/2}} E \left| m_j^*(\theta_0) \right|^3
$$

$$
= \frac{1}{\varepsilon^3 M^{3/2}} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \| \phi_0(u_j, \lambda_1) \phi_0(u_j, \lambda_2) \phi_0(u_j, \lambda_3) \|
$$

$$
\times E \left[ I_N(u_j, \lambda_1) I_N(u_j, \lambda_2) I_N(u_j, \lambda_3) \right] d\lambda_1 d\lambda_2 d\lambda_3.
$$

The above expectation is written as

$$
E \left[ I_N(u_j, \lambda_1) I_N(u_j, \lambda_2) I_N(u_j, \lambda_3) \right]
$$

$$
= \frac{1}{(2\pi H_{2,N})^3} \text{cum} \left[ d_N(u_j, \lambda_1) d_N(u_j, -\lambda_1) d_N(u_j, \lambda_2) \right.
$$

$$
\left. \times d_N(u_j, -\lambda_2) d_N(u_j, \lambda_3) d_N(u_j, -\lambda_3) \right].
$$

From Lemma 5.1 this is of order $O_p(1)$, so we can see that (5.52) tends to zero, which leads

$$
Z_M^* = o_p\left( M^{1/2} \right).
$$

(5.54)

From (5.49), (5.50), and (5.54), it is seen that

$$
\| \alpha \| \left[ O_p(1) - o_p\left( M^{-1/2} \right) O_p\left( M^{-1/2} \right) \right] \leq O_p\left( M^{-1/2} \right).
$$

(5.55)

Therefore,

$$
\| \alpha \| = O_p\left( M^{-1/2} \right).
$$

(5.56)

Now we have from (5.54) that

$$
\max_{1 \leq j \leq T} |Y_j| = O_p\left( M^{-1/2} \right) o_p\left( M^{1/2} \right) = o_p(1)
$$

(5.57)
and from (5.30) that

\[
0 = \frac{1}{M} \sum_{j=1}^{M} m_j(\theta_0) - \frac{1}{1 + Y_j}
\]
\[
= \frac{1}{M} \sum_{j=1}^{M} m_j(\theta_0) \left( 1 - Y_j + \frac{Y_j^2}{1 + Y_j} \right) \tag{5.58}
\]
\[
= 2\pi M^{-1/2} P_M - 2\pi S_M \alpha + \frac{1}{M} \sum_{j=1}^{M} m_j(\theta_0) Y_j^2
\]

Noting that

\[
\frac{1}{M} \sum_{j=1}^{M} \| m_j(\theta_0) \|^3 \leq \frac{1}{M} \sum_{j=1}^{M} Z_{M}^* \| m_j(\theta_0) \|^2 = o_p\left(M^{1/2}\right) \tag{5.59}
\]

we can see that the final term in (5.58) has a norm bounded by

\[
\frac{1}{M} \sum_{j=1}^{M} \| m_j(\theta_0) \|^3 \| \alpha \|^2 |1 + Y_j|^{-1} = o_p\left(M^{1/2}\right)O_p\left(M^{-1}\right)O_p(1) = o_p\left(M^{-1/2}\right). \tag{5.60}
\]

Hence, we can write

\[
\alpha = M^{-1/2} S_M^{-1} P_M + \epsilon, \tag{5.61}
\]

where \(\epsilon = o_p(M^{-1/2})\). By (5.57), we may write

\[
\log(1 + Y_j) = Y_j - \frac{1}{2} Y_j^2 + \eta_j, \tag{5.62}
\]

where for some finite \(K\)

\[
Pr\left( |\eta_j| \leq K |Y_j|^3, 1 \leq j \leq M \right) \longrightarrow 1 \quad (T \rightarrow \infty). \tag{5.63}
\]
We may write
\[
-\frac{1}{\pi} \log R(\theta_0) = -\frac{1}{\pi} \sum_{j=1}^{M} \log(Tw_j) = \frac{1}{\pi} \sum_{j=1}^{M} \log(1 + Y_j)
\]
\[
= \frac{1}{\pi} \sum_{j=1}^{M} Y_j - \frac{1}{2\pi} \sum_{j=1}^{M} Y_j^2 + \frac{1}{\pi} \sum_{j=1}^{M} \eta_j
\]
\[
= \mathbf{p}_M' \mathbf{S}_M^{-1} \mathbf{P}_M - M \mathbf{e}' \mathbf{S}_M \mathbf{e} + \frac{1}{\pi} \sum_{j=1}^{M} \eta_j
\]
\[
= (A) - (B) + (C) \quad \text{(say)}.
\]
Here it is seen that
\[
(B) = M \mathbf{O}_p \left( M^{-1/2} \right) \mathbf{O}_p (1) \mathbf{O}_p \left( M^{-1/2} \right) = \mathbf{o}_p (1),
\]
\[
(C) \leq K \| \mathbf{a} \|^3 \sum_{j=1}^{M} \| m_j(\theta_0) \|^3 = \mathbf{O}_p \left( M^{-3/2} \right) \mathbf{O}_p \left( M^{3/2} \right) = \mathbf{o}_p (1).
\]

And finally from Lemmas 5.2 and 5.3, we can show that
\[
(A) \xrightarrow{d} \left( \mathbf{S}_2^{-1/2} \mathbf{S}_1^{-1/2} \mathbf{S}_1^{-1/2} \mathbf{P}_M \right)' \left( \mathbf{S}_2^{-1/2} \mathbf{S}_1^{-1/2} \mathbf{S}_1^{-1/2} \mathbf{P}_M \right) \xrightarrow{d} (\mathbf{F} \mathbf{N})'(\mathbf{F} \mathbf{N}).
\]

Then we can obtain the desired result.

**Acknowledgments**

The author is grateful to Professor M. Taniguchi, J. Hirukawa, and H. Shiraishi for their instructive advice and helpful comments. Thanks are also extended to the two referees whose comments are useful. This work was supported by Grant-in-Aid for Young Scientists (B) (22700291).

**References**

[1] A. B. Owen, “Empirical likelihood ratio confidence intervals for a single functional,” *Biometrika*, vol. 75, no. 2, pp. 237–249, 1988.

[2] A. B. Owen, “Empirical likelihood ratio confidence regions,” *The Annals of Statistics*, vol. 18, no. 1, pp. 90–120, 1990.

[3] J. Qin and J. Lawless, “Empirical likelihood and general estimating equations,” *The Annals of Statistics*, vol. 22, no. 1, pp. 300–325, 1994.

[4] A. B. Owen, “Empirical likelihood for linear models,” *The Annals of Statistics*, vol. 19, no. 4, pp. 1725–1747, 1991.
[5] S. X. Chen, “On the accuracy of empirical likelihood confidence regions for linear regression model,” *Annals of the Institute of Statistical Mathematics*, vol. 45, no. 4, pp. 621–637, 1993.

[6] S. X. Chen, “Empirical likelihood confidence intervals for linear regression coefficients,” *Journal of Multivariate Analysis*, vol. 49, no. 1, pp. 24–40, 1994.

[7] J. Qin, “Empirical likelihood in biased sample problems,” *The Annals of Statistics*, vol. 21, no. 3, pp. 1182–1196, 1993.

[8] Y. Kitamura, “Empirical likelihood methods with weakly dependent processes,” *The Annals of Statistics*, vol. 25, no. 5, pp. 2084–2102, 1997.

[9] A. C. Monti, “Empirical likelihood confidence regions in time series models,” *Biometrika*, vol. 84, no. 2, pp. 395–405, 1997.

[10] D. J. Nordman and S. N. Lahiri, “A frequency domain empirical likelihood for short- and long-range dependence,” *The Annals of Statistics*, vol. 34, no. 6, pp. 3019–3050, 2006.

[11] R. Dahlhaus, “On the Kullback-Leibler information divergence of locally stationary processes,” *Stochastic Processes and Their Applications*, vol. 62, no. 1, pp. 139–168, 1996.

[12] R. Dahlhaus, “Asymptotic statistical inference for nonstationary processes with evolutionary spectra,” in *Proceedings of the Athens Conference on Applied Probability and Time Series Analysis*, vol. 115 of *Lecture Notes in Statistics*, pp. 145–159, Springer, 1996.

[13] R. Dahlhaus, “Fitting time series models to nonstationary processes,” *The Annals of Statistics*, vol. 25, no. 1, pp. 1–37, 1997.

[14] Y. Hosoya and M. Taniguchi, “A central limit theorem for stationary processes and the parameter estimation of linear processes,” *The Annals of Statistics*, vol. 10, no. 1, pp. 132–153, 1982, Correction: vol 21, pp. 1115–1117, 1993.

[15] Y. Kakizawa, “Parameter estimation and hypothesis testing in stationary vector time series,” *Statistics & Probability Letters*, vol. 33, no. 3, pp. 225–234, 1997.

[16] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Holden-Day, San Francisco, Calif, USA, 2001.
Research Article

A Simulation Approach to Statistical Estimation of Multiperiod Optimal Portfolios

Hiroshi Shiraishi

The Jikei University School of Medicine, Tokyo 1828570, Japan

Correspondence should be addressed to Hiroshi Shiraishi, hiroshi.shiraishi@jikei.ac.jp

Received 24 February 2012; Accepted 9 April 2012

Academic Editor: Kenichiro Tamaki

Copyright © 2012 Hiroshi Shiraishi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper discusses a simulation-based method for solving discrete-time multiperiod portfolio choice problems under AR(1) process. The method is applicable even if the distributions of return processes are unknown. We first generate simulation sample paths of the random returns by using AR bootstrap. Then, for each sample path and each investment time, we obtain an optimal portfolio estimator, which optimizes a constant relative risk aversion (CRRA) utility function. When an investor considers an optimal investment strategy with portfolio rebalancing, it is convenient to introduce a value function. The most important difference between single-period portfolio choice problems and multiperiod ones is that the value function is time dependent. Our method takes care of the time dependency by using bootstrapped sample paths. Numerical studies are provided to examine the validity of our method. The result shows the necessity to take care of the time dependency of the value function.

1. Introduction

Portfolio optimization is said to be “myopic” when the investor does not know what will happen beyond the immediate next period. In this framework, basic results about single period portfolio optimization (such as mean-variance analysis) are justified for short-term investments without portfolio rebalancing. Multiperiod problems are much more realistic than single-period ones. In this framework, we assume that an investor makes a sequence of decisions to maximize a utility function at each time. The fundamental method to solve this problem is the dynamic programming. In this method, a value function which expresses the expected terminal wealth is introduced. The recursive equation with respect to the value function is so-called Bellman equation. The first order conditions (FOCs) to satisfy the Bellman equation are key tool in order to solve the dynamic problem.

The original literature on dynamic portfolio choice, pioneered by Merton [1] in continuous time and by Samuelson [2] and Fama [3] in discrete time, produced many important
insights into the properties of optimal portfolio policies. Unfortunately, since it is known that
the closed-form solutions are obtained only for a few special cases, the recent literature uses a
variety of numerical and approximate solution methods to incorporate realistic features into
the dynamic portfolio problem such as Ait-Sahalia and Brandt [4] and Brandt et al. [5].

We introduce an procedure to construct the dynamic portfolio weights based on AR bootstrap.

The simulation algorithm is as follows; first, we generate simulation sample paths
of the vector random returns by using AR bootstrap. Based on the bootstrapping samples,
an optimal portfolio estimator, which is applied from time $T - 1$ to the end of trading
time $T$, is obtained under a constant relative risk aversion (CRRA) utility function. Note
that this optimal portfolio corresponds “myopic” (single period) optimal portfolio. Next
we approximate the value function by linear functions of the past observation. This idea is
similar to that of [4, 5]. Then, optimal portfolio weight estimators at each trading time are
obtained based on the value function. Finally, we construct an optimal investment strategy
as a sequence of the optimal portfolio weight estimators.

This paper is organized as follows. We describe the basic idea to solve multiperiod
optimal portfolio weights under a CRRA utility function in Section 2. In Section 3, we
discuss an algorithm to construct the estimator involving the method of AR bootstrap. The
applications of our method are in Section 4.

2. Multiperiod Optimal Portfolio

Suppose the existence of a finite number of risky assets indexed by $i$, $(i = 1, \ldots, m)$. Let $X_t = (X_1(t), \ldots, X_m(t))'$ denote the random excess returns on $m$ assets from time $t$ to $t + 1$ (suppose
that $S_i(t)$ is a value of asset $i$ at time $t$. Then, the return is described as $1 + X_i(t) = S_i(t)/S_i(t-1)$).
Suppose too that there exists a risk-free asset with the excess return $X_f$. Suppose that $B(t)$ is
a value of risk-free asset at time $t$. Then, the return is described as $1 + X_f = B(t)/B(t - 1)$.

Based on the process $[X_t]_{t=1}^T$ and $X_f$, we consider an investment strategy from time 0 to time $T$
where $T \in \mathbb{N}$ denotes the end of the investment time. Let $w_t = (w_1(t), \ldots, w_m(t))'$ be vectors
of portfolio weight for the risky assets at the beginning of time $t + 1$. Here we assume that
the portfolio weights $w_t$ can be rebalanced at the beginning of time $t + 1$ and measurable
(predictable) with respect to the past information $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$. Here we make the
following assumption.

Assumption 2.1. There exists an optimal portfolio weight $\tilde{w}_t \in \mathbb{R}^m$ satisfied with $|\tilde{w}_t X_{t+1} + (1 - \tilde{w}_t) e X_f| \ll 1$ (we assume that the risky assets exclude ultra high-risk and high-return ones,
for instance, the asset value $S_i(t + 1)$ may be larger than $2S_i(t)$), almost surely for each time
t = 0, 1, \ldots, T − 1 where $e = (1, \ldots, 1)'$.

Then the return of the portfolio from time $t$ to $t + 1$ is written as $1 + X_f + w'_t(X_{t+1} - X_f e)$
(assuming that $S_t := (S_1(t), \ldots, S_m(t))' = B(t)e$, the portfolio return is written as $(w'_t S_{t+1} + (1 - w_t e) B(t + 1))/(w'_t S_t + (1 - w_t e) B(t)) = 1 + X_f + w'_t(X_{t+1} - X_f e)$) and the return from time
0 to time $T$ (called terminal wealth) is written as

$$W_T := \prod_{t=0}^{T-1} (1 + X_f + w'_t(X_{t+1} - X_f e)). \quad (2.1)$$
Suppose that a utility function $U : x \mapsto U(x)$ is differentiable, concave, and strictly increasing for each $x \in \mathbb{R}$. Consider an investor’s problem

$$\max_{\{w_t\}_{t=0}^{T-1}} E[U(W_T)].$$ \hfill (2.2)

Following a formulation by the dynamic programming (e.g., Bellman [6]), it is convenient to express the expected terminal wealth in terms of a value function $V_t$:

$$V_t \equiv \max_{\{w_t\}_{t=0}^{T-1}} E[U(W_T) \mid \mathcal{F}_t]$$

$$= \max_{w_t} E \left[ \max_{\{w_s\}_{s=t}^{T-1}} E[U(W_T) \mid \mathcal{F}_{t+1}] \mid \mathcal{F}_t \right]$$ \hfill (2.3)

$$= \max_{w_t} E[V_{t+1} \mid \mathcal{F}_t],$$

subject to the terminal condition $V_T = U(W_T)$. The recursive equation (2.3) is the so-called Bellman equation and is the basis for any recursive solution of the dynamic portfolio choice problem. The first-order conditions (FOCs) (here $(\partial/\partial w_t) E[V_{t+1} \mid \mathcal{F}_t] = E[(\partial/\partial w_t)V_{t+1} \mid \mathcal{F}_t]$ is assumed). in order to obtain an optimal solution at each time $t$ are

$$\frac{\partial V_t}{\partial w_t} = E[\partial_t U(W_T)(X_{t+1} - X_t e) \mid \mathcal{F}_t] = 0,$$ \hfill (2.4)

where $\partial_t U(x_0) = (\partial/\partial x)U(x)|_{x=x_0}$. These FOCs make up a system of nonlinear equations involving integrals that can in general be solved for $w_t$ only numerically.

According to the literature (e.g., [5]), we can simplify this problem in case of a constant relative risk aversion (CRRA) utility function, that is,

$$U(W) = \frac{W^{1-\gamma}}{1-\gamma}, \quad \gamma \neq 1,$$ \hfill (2.5)
where $\gamma$ denotes the coefficient of relative risk aversion. In this case, the Bellman equation simplifies to

$$V_t = \max_{w_t} E \left[ \max_{\{w_s\}_{t+1}^{T-1}} \mathbb{E} \left[ \frac{1}{1-\gamma} (W_{t+1})^{1-\gamma} | \mathcal{F}_{t+1} \right] \middle| \mathcal{F}_t \right]$$

$$= \max_{w_t} E \left[ \max_{\{w_s\}_{t+1}^{T-1}} \mathbb{E} \left[ \frac{1}{1-\gamma} \left( \prod_{s=0}^{T-1} (1 + X_f + w'_s(X_{s+1} - X_f e)) \right)^{1-\gamma} \right] \middle| \mathcal{F}_{t+1} \right]$$

$$= \max_{w_t} E \left[ \frac{1}{1-\gamma} \left( \prod_{s=0}^{T-1} (1 + X_f + w'_s(X_{s+1} - X_f e)) \right)^{1-\gamma} \right]$$

$$= \max_{w_t} E \left[ U(W_{t+1}) \Psi_{t+1} \middle| \mathcal{F}_t \right],$$

(2.6)

where $W_{t+1}^T = \prod_{s=0}^{T-1} (1 + X_f + w'_s(X_{s+1} - X_f e))$ and $\Psi_{t+1} = \max_{\{w_s\}_{t+1}^{T-1}} E [(W_{t+1}^T)^{1-\gamma} | \mathcal{F}_{t+1}]$. From this, the value function $V_t$ can be expressed as

$$V_t = U(W_t) \Psi_t,$$

(2.7)

and $\Psi_t$ also satisfies a Bellman equation

$$\Psi_t = \max_{w_t} E \left[ (1 + X_f + w'_t(X_{t+1} - X_f e))^{1-\gamma} \Psi_{t+1} \middle| \mathcal{F}_t \right],$$

(2.8)

subject to the terminal condition $\Psi_T = 1$.

The corresponding FOCs (in terms of $\Psi_t$) are

$$E \left[ (1 + X_f + w'_t(X_{t+1} - X_f e))^{1-\gamma} \Psi_{t+1} (X_{t+1} - X_f e) \middle| \mathcal{F}_t \right] = 0.$$  

(2.9)

### 3. Estimation

Suppose that $\{X_t = (X_1(t), \ldots, X_m(t))'; \ t \in \mathbb{Z} \}$ is an $m$-vector AR(1) process defined by

$$X_t = \mu + A(X_{t-1} - \mu) + \epsilon_t,$$

(3.1)

where $\mu = (\mu_1, \ldots, \mu_m)'$ is a constant $m$-dimensional vector, $\epsilon_t = (\epsilon_1(t), \ldots, \epsilon_m(t))'$ are independent and identically distributed (i.i.d.) random $m$-dimensional vectors with $E[\epsilon_t] = 0$.
and \( E[\epsilon_i \epsilon_j] = \Gamma \) (\( \Gamma \) is a nonsingular \( m \) by \( m \) matrix), and \( A \) is a nonsingular \( m \) by \( m \) matrix. We make the following assumption.

**Assumption 3.1.** \( \det \{ I_m - Az \} \neq 0 \) on \( \{ z \in \mathbb{C} ; |z| \leq 1 \} \).

Given \( \{ X_{n+1}, \ldots, X_0, X_1, \ldots, X_t \} \), the least-squares estimator \( \hat{A}^{(t)} \) of \( A \) is obtained by solving

\[
\hat{\Gamma}^{(t)} \hat{A}^{(t)} = \sum_{s=-n+1}^{t} \tilde{Y}_s^{(t)} (\tilde{Y}_s^{(t)})^\prime,
\]

where \( \tilde{Y}_s^{(t)} = X_s - \hat{\mu}^{(t)} \), \( \hat{\Gamma}^{(t)} = \sum_{s=-n+1}^{t} \tilde{Y}_s^{(t)} (\tilde{Y}_s^{(t)})^\prime \) and \( \hat{\mu}^{(t)} = (1/(n+t)) \sum_{s=-n+1}^{t} X_s \). Then, the error \( \hat{e}_s^{(t)} = (\hat{e}_1^{(t)}(s), \ldots, \hat{e}_m^{(t)}(s))^\prime \) is “recovered” by

\[
\hat{e}_s^{(t)} := \tilde{Y}_s^{(t)} - \hat{A}^{(t)} \tilde{Y}_s^{(t)}, \quad s = -n + 2, \ldots, t.
\]

Let \( F_n^{(t)}(\cdot) \) denote the distribution which puts mass \( 1/(n+t) \) at \( \hat{e}_s^{(t)} \). Let \( \{ e_s^{(b,t)^*}\}_{s=t+1}^{\gamma} \) (for \( b = 1, \ldots, B(\in \mathbb{N}) \)) be i.i.d. bootstrapped observations from \( F_n^{(t)} \).

Given \( \{ e_s^{(b,t)^*}\} \), define \( Y_s^{(b,t)^*} \) and \( X_s^{(b_1,b_2,t)^*} \) by

\[
Y_s^{(b,t)^*} = (\hat{A}^{(t)})^{-1} (X_s - \hat{\mu}^{(t)}) + \sum_{k=t}^{\gamma} (\hat{A}^{(t)})^{-k} e_k^{(b,t)^*},
\]

\[
X_s^{(b_1,b_2,t)^*} = \hat{\mu}^{(t)} + \hat{A}^{(t)} Y_s^{(b,t)^*} + e_s^{(b_2,t)^*},
\]

for \( s = t + 1, \ldots, T \).

Based on the above \( \{ X_s^{(b_1,b_2,t)^*}\}_{b_1,b_2=1,\ldots,B; s=t+1,\ldots,T} \) for each \( t = 0, \ldots, T - 1 \), we construct an estimator of the optimal portfolio weight \( \tilde{w}_t \) as follows.

**Step 1.** First, we fix the current time \( t \) which implies that the observed stretch \( n + t \) is fixed. Then, we can generate \( \{ X_s^{(b_1,b_2,t)^*}\} \) by (3.4).

**Step 2.** Next, for each \( b_0 = 1, \ldots, B \), we obtain \( \tilde{w}_{T-1}^{(b_0,t)} \) as the maximizer of

\[
E_{T-1}^* \left[ \left( 1 + X_f + w' \left( X_T^{(b_0,b,t)^*} - X_f e \right) \right)^{1-\gamma} \right] = \frac{1}{B} \sum_{b=1}^{B} \left( 1 + X_f + w' \left( X_T^{(b_0,b,t)^*} - X_f e \right) \right)^{1-\gamma},
\]

or the solution of

\[
E_{T-1}^* \left[ \left( 1 + X_f + w' \left( X_T^{(b_0,b,t)^*} - X_f e \right) \right)^{1-\gamma} (X_T^{(b_0,b,t)^*} - X_f e) \right]
\]

\[
= \frac{1}{B} \sum_{b=1}^{B} \left( 1 + X_f + w' \left( X_T^{(b_0,b,t)^*} - X_f e \right) \right)^{1-\gamma} (X_T^{(b_0,b,t)^*} - X_f e)
\]

\[
= 0,
\]
with respect to \( w \). Here we introduce a notation “\( E^*_1 \)” as an estimator of conditional expectation \( E[\cdot | \mathcal{F}_s] \), which is defined by \( E^*_1[h(X^{(b_j,b_j^\star)})] = (1/B) \sum_{b=1}^B h(X^{(b_j,b_j^\star)}) \) for any function \( h \) of \( X^{(b_j,b_j^\star)} \). This \( \hat{w}^{(b_j)}_{T-1} \) corresponds to the estimator of myopic (single period) optimal portfolio weight.

**Step 3.** Next, we construct estimators of \( \Psi_{T-1} \). Since it is difficult to express the explicit form of \( \Psi_{T-1} \), we parameterize it as linear functions of \( X_{T-1} \) as follows;

\[
\Psi^{(1)}(X_{T-1}, \theta_{T-1}) : = \left[ 1, X'_{T-1} \right] \theta_{T-1},
\]

\[
\Psi^{(2)}(X_{T-1}, \theta_{T-1}) : = \left[ 1, X'_{T-1}, \text{vech} (X_{T-1} X_{T-1}') \right] \theta_{T-1}.
\]

Note that the dimensions of \( \theta_{T-1} \) in \( \Psi^{(1)} \) and \( \Psi^{(2)} \) are \( m+1 \) and \( m(m+1)/2+m+1 \), respectively. The idea of \( \Psi^{(1)} \) and \( \Psi^{(2)} \) is inspired by the parameterization of the conditional expectations in [5].

In order to construct the estimators of \( \Psi^{(i)} \) \( (i = 1, 2) \), we introduce the conditional least squares estimators of the parameter \( \theta^{(i)}_{T-1} \), that is,

\[
\hat{\theta}^{(i)}_{T-1} = \arg \min_{\theta} Q^{(i)}_{T-1}(\theta),
\]

where

\[
Q^{(i)}_{T-1}(\theta) = \frac{1}{B} \sum_{b=1}^B E^*_1 \left[ \left( \Psi_{T-1} - \Psi^{(i)} \right)^2 \right]
\]

\[
= \frac{1}{B} \sum_{b=1}^B \left[ \frac{1}{B} \sum_{b=1}^B \left( \Psi_{T-1} \left( X^{(b_j,b_j^\star)}_{T-1} \right) - \Psi^{(i)} \left( X^{(b_j,b_j^\star)}_{T-1} \theta \right) \right)^2 \right],
\]

\[
\Psi_{T-1} \left( X^{(b_j,b_j^\star)}_{T-1} \right) = \left( 1 + X_f + \hat{w}^{(b_j)}_{T-1} \right) \left( X^{(b_j,b_j^\star)}_{T-1} - X_f e \right)^{1-T}.
\]

Then, by using \( \hat{\theta}^{(i)}_{T-1} \), we can compute \( \Psi^{(i)} \left( X^{(b_j,b_j^\star)}_{T-1}, \hat{\theta}^{(i)}_{T-1} \right) \).

**Step 4.** Based on the above \( \Psi^{(i)} \), we obtain \( \hat{w}^{(b_j)}_{T-2} \) as the maximizer of

\[
E^*_2 \left[ \left( 1 + X_f + w^* \left( X^{(b_j,b_j^\star)}_{T-1} - X_f e \right) \right)^{1-T} \Psi^{(i)} \left( X^{(b_j,b_j^\star)}_{T-1}, \theta^{(i)}_{T-1} \right) \right]
\]

\[
= \frac{1}{B} \sum_{b=1}^B \left( 1 + X_f + w^* \left( X^{(b_j,b_j^\star)}_{T-1} - X_f e \right) \right)^{1-T} \Psi^{(i)} \left( X^{(b_j,b_j^\star)}_{T-1}, \hat{\theta}^{(i)}_{T-1} \right),
\]
or the solution of

\[
E_T^* \left[ \left( 1 + X_f + w^T (X_{T-1}^{(b,t)})^* - X_f \right) \right]^{1-T} \left( X_{T-1}^{(b,t)} - X_f \right) \Psi i \left( \sum_{i=1}^{B} \left( 1 + X_f + w^T (X_{T-1}^{(b,t)})^* - X_f \right) \Psi i \left( X_{T-1}^{(b,t)} - X_f \right) \Psi i \left( X_{T-1}^{(b,t)} - X_f \right) \Psi i \left( X_{T-1}^{(b,t)} - X_f \right) \right] = 0. 
\]

(3.12)

with respect to \( w \). This \( \hat{w}_{T-2}^{(b,t)} \) does not correspond to the estimator of myopic (single period) optimal portfolio weight due to the effect of \( \Psi i \).

**Step 5.** In the same manner of Steps 3–4, we can obtain \( \hat{\theta}_s \) and \( \hat{w}_s^{(b,t)} \), recursively, for \( s = T-2, T-1, \ldots, t+1 \).

**Step 6.** Then, we define an optimal portfolio weight estimator at time \( t \) as \( \hat{w}_t^{(b,t)} := \hat{w}_t^{(b,t)} \) by Step 4. Note that \( \hat{w}_t^{(b,t)} \) is obtained as only one solution because \( X_{t+1}^{(b,t)}(\hat{\mu}_t + A(X_t - \hat{\mu}_t) + \epsilon_t^{(b,t)}) \) is independent of \( b_0 \).

**Step 7.** For each time \( t = 0, 1, \ldots, T-1 \), we obtain \( \hat{w}_t^{(b,t)} \) by Steps 1–6. Finally, we can construct an optimal investment strategy as \( \{ \hat{w}_t^{(b,t)} \}_{t=0}^{T-1} \).

### 4. Examples

In this section we examine our approach numerically. Suppose that there exists a risky asset with the excess return \( X_t \) at time \( t \) and a risk-free asset with the excess return \( X_f = 0.01 \). We assume that \( X_t \) is defined by the following univariate AR(1) model:

\[
X_t = \mu + A(X_{t-1} - \mu) + e_t, \quad e_t \sim N(0, \Gamma).
\]

(4.1)

Let \( w_t \) be a portfolio weight for the risky asset at the beginning of time \( t+1 \). Suppose that an investor is interested in the investment strategy from time 0 to time \( T \). Then the terminal wealth is written as (2.1). Applying our method, the estimator \( \hat{W}_T \) can be obtained by

\[
\hat{W}_T = \prod_{t=0}^{T-1} (1 + X_f + \hat{w}_t(X_{t+1} - X_f)),
\]

(4.2)

where \( \hat{w}_t \) is the estimator of optimal portfolio under the CRRA utility function defined by (2.5). In what follows, we examine the effect of \( \hat{W}_T \) for a variety of \( n \) (initial sample size), \( B \) (resampling size), \( A \) (AR parameter), \( \Gamma \) (variance of \( e_t \)), \( \gamma \) (relative risk aversion parameter), and \( \Psi \) (defined by (3.7) or (3.8)).

**Example 4.1** (myopic (single period) versus dynamic (multiPeriod)). Let \( \mu = 0.02, A = 0.1, \Gamma = 0.05, n = 100, T = 10, \) and \( B = 100 \). We generate the excess return process \( \{X_t\}_{t=-n+1,\ldots,T} \)
Figure 1: Resampled excess return.

Figure 2: Myopic and dynamic portfolio return.
by (4.1). First, for each $t = 0, \ldots, T - 1$ we generate $\{X_{t}^{(b_{1}, b_{2}, t)}\}_{b_{1}, b_{2}=1, \ldots, B_{s}=1, \ldots, B_{s}}$ by (3.4) based on $\{X_{s}\}_{s=-n+1}$ (as Step 1). We plot $\{X_{t}\}_{t=1, \ldots, T}$ and $\{X_{t}^{(b_{1}, b_{2}, t)}\}_{b_{1}, b_{2}=1, \ldots, B_{s}=1, \ldots, B_{s}}$ in Figure 1.

It can be seen that $X_{t}^{(b_{1}, b_{2}, t)}$ show similar behavior with $X_{t}$.
### Table 1: Dynamic portfolio returns for $\gamma = 5$.  

| $T$ | Mean | Myopic ($q_{0.25}, q_{0.5}, q_{0.75}$) | Mean | Dynamic ($\Psi^{(1)}$) | Mean | Dynamic ($\Psi^{(2)}$) |
|-----|------|--------------------------------------|------|--------------------------|------|--------------------------|
|     | Mean | (0.9924, 1.0091, 1.0192)             | Mean | (0.9920, 1.0096, 1.0176) | Mean | (0.9920, 1.0096, 1.0176) |
| 1   | 1.013564 | (0.9924, 1.0091, 1.0192) | 1.013667 | (0.9920, 1.0096, 1.0176) | 1.013814 | (0.9920, 1.0096, 1.0176) |
| 2   | 1.024329 | (0.9917, 1.0192, 1.0445) | 1.024396 | (0.9923, 1.0177, 1.0436) | 1.024667 | (0.9924, 1.0183, 1.0437) |
| 5   | 1.065986 | (1.0021, 1.0504, 1.1125) | 1.065988 | (1.0000, 1.0509, 1.1115) | 1.066355 | (0.9999, 1.0505, 1.1106) |
| 10  | 1.137727 | (1.0273, 1.1062, 1.2024) | 1.137707 | (1.0264, 1.1041, 1.2005) | 1.138207 | (1.0265, 1.1043, 1.2002) |

#### A: Terminal wealth

| $T$ | Mean | Myopic ($q_{0.25}, q_{0.5}, q_{0.75}$) | Mean | Dynamic ($\Psi^{(1)}$) | Mean | Dynamic ($\Psi^{(2)}$) |
|-----|------|--------------------------------------|------|--------------------------|------|--------------------------|
| 1   | -0.24158 | (-0.257, -0.241, -0.231) | -0.24139 | (-0.258, -0.240, -0.233) | -0.24130 | (-0.258, -0.240, -0.233) |
| 2   | -0.23609 | (-0.258, -0.231, -0.210) | -0.23595 | (-0.257, -0.233, -0.210) | -0.23578 | (-0.257, -0.232, -0.210) |
| 5   | -0.21761 | (-0.247, -0.205, -0.163) | -0.21761 | (-0.249, -0.204, -0.163) | -0.21703 | (-0.250, -0.205, -0.164) |
| 10  | -0.18349 | (-0.224, -0.166, -0.119) | -0.18339 | (-0.225, -0.168, -0.120) | -0.18287 | (-0.225, -0.168, -0.120) |

### Table 2: Dynamic portfolio returns for $\gamma = 10$.  

| $T$ | Mean | Myopic ($q_{0.25}, q_{0.5}, q_{0.75}$) | Mean | Dynamic ($\Psi^{(1)}$) | Mean | Dynamic ($\Psi^{(2)}$) |
|-----|------|--------------------------------------|------|--------------------------|------|--------------------------|
|     | Mean | (1.0011, 1.0095, 1.0146)             | Mean | (1.0010, 1.0098, 1.0138) | Mean | (1.0010, 1.0098, 1.0138) |
| 1   | 1.011802 | (1.0011, 1.0095, 1.0146) | 1.011859 | (1.0010, 1.0098, 1.0138) | 1.011944 | (1.0010, 1.0098, 1.0138) |
| 2   | 1.022249 | (1.0059, 1.0196, 1.0323) | 1.022286 | (1.0065, 1.0190, 1.0319) | 1.022439 | (1.0065, 1.0192, 1.0319) |
| 5   | 1.058344 | (1.0276, 1.0512, 1.0825) | 1.058373 | (1.0254, 1.0509, 1.0823) | 1.058584 | (1.0253, 1.0507, 1.0818) |
| 10  | 1.120369 | (1.0658, 1.1070, 1.1544) | 1.120323 | (1.0687, 1.1068, 1.1533) | 1.120595 | (1.0666, 1.1060, 1.1532) |

#### A: Terminal wealth

| $T$ | Mean | Myopic ($q_{0.25}, q_{0.5}, q_{0.75}$) | Mean | Dynamic ($\Psi^{(1)}$) | Mean | Dynamic ($\Psi^{(2)}$) |
|-----|------|--------------------------------------|------|--------------------------|------|--------------------------|
| 1   | -0.10224 | (-0.109, -0.101, -0.097) | -0.10215 | (-0.110, -0.101, -0.098) | -0.10210 | (-0.110, -0.101, -0.098) |
| 2   | -0.09530 | (-0.105, -0.093, -0.083) | -0.09523 | (-0.104, -0.093, -0.083) | -0.09515 | (-0.104, -0.093, -0.083) |
| 5   | -0.07581 | (-0.086, -0.070, -0.054) | -0.07582 | (-0.088, -0.071, -0.054) | -0.07557 | (-0.088, -0.071, -0.054) |
| 10  | -0.05007 | (-0.062, -0.044, -0.030) | -0.05003 | (-0.061, -0.044, -0.030) | -0.04986 | (-0.062, -0.044, -0.030) |
Next, we construct the optimal portfolio estimator $\tilde{\omega}_s^{(b,t)}$ along the lines with Steps 2–7. Here we apply the approximated solution for (3.5) or (3.11) following [5], that is,

$$
\tilde{\omega}_s^{(b,t)} = \frac{1}{2E_s[D_{3,s+1}^{(b,t)}]} \left\{ E_s [D_{2,s+1}^{(b,t)*}] + 3 \left( \tilde{\omega}_s^{(b,t)} \right)^2 E_s [D_{4,s+1}^{(b,t)}] + 4 \left( \tilde{\omega}_s^{(b,t)} \right)^3 E_s [D_{5,s+1}^{(b,t)*}] \right\},
$$

(4.3)

where

$$
\begin{align*}
D_{2,s+1}^{(b,t)*} &= (1 + X_f)^{-\gamma} (X_{s+1}^{(b,t)*} - X_f) \psi(i) \left( X_{s+1}^{(b,t)*} - \theta_{s+1}^{(i)} \right), \\
D_{3,s+1}^{(b,t)} &= -\frac{\gamma}{2} (1 + X_f)^{-1-\gamma} \left( X_{s+1}^{(b,t)*} - X_f \right)^2 \psi(i) \left( X_{s+1}^{(b,t)*} - \theta_{s+1}^{(i)} \right), \\
D_{4,s+1}^{(b,t)*} &= \frac{(-\gamma)(-1-\gamma)}{6} (1 + X_f)^{-2-\gamma} \left( X_{s+1}^{(b,t)*} - X_f \right)^3 \psi(i) \left( X_{s+1}^{(b,t)*} - \theta_{s+1}^{(i)} \right), \\
D_{5,s+1}^{(b,t)*} &= \frac{(-\gamma)(-1-\gamma)(-2-\gamma)}{24} (1 + X_f)^{-3-\gamma} \left( X_{s+1}^{(b,t)*} - X_f \right)^4 \psi(i) \left( X_{s+1}^{(b,t)*} - \theta_{s+1}^{(i)} \right),
\end{align*}
$$

(4.4)
This approximate solution describes a fourth-order expansion of the value function around $1 + X_f$ ($\hat{\omega}_s$ describes a second-order expansion). According to [5], a second-order expansion of the value function is sometimes not sufficiently accurate, but a fourth-order expansion includes adjustments for the skewness and kurtosis of returns and their effects on the utility of the investor.

Figure 2 shows time series plots for single portfolio return ($=1+X_f+\hat{\omega}_t(X_{i+1}-X_f)$, Line 1), cumulative portfolio return ($=\hat{W}_T$, Line 2), and value of utility function ($=1/(1-\hat{g})\hat{W}_T^{1-\hat{g}}$, Line 3) for $\gamma = 5, 10$ and 20. The solid line shows the investment only for risk-free asset (i.e., $\hat{\omega}_t = 0$), the dotted line with $\triangle$ shows myopic (single period) portfolio (i.e., $\Psi^{(0)} = 1$) and the dotted line with $+$ shows dynamic (multiperiod) portfolio by using $\Psi^{(1)}$.

Regarding the single-portfolio return, we can not argue the best investment strategy among the risk-free, the myopic portfolio and the dynamic portfolio investment. However, to look at the cumulative portfolio return or the value of utility function, it is obviously that the dynamic portfolio investment is the best one. The difference between the myopic and dynamic portfolio is due to $\Psi$ and is called “hedging demands” because by deviating from the single period portfolio choice, the investor tries to hedge against changes in the investment opportunities. In view of the effect of $\gamma$, we can see that the magnitude of the hedging demands decreases with increased amount of $\gamma$.

Next, we repeat the above algorithm 100 times using the different generated data. Tables 1, 2, and 3 show means, 25 percentiles ($q_{0.25}$), medians ($q_{0.5}$), and 75 percentiles ($q_{0.75}$) of terminal wealth ($\hat{W}_T$) and the values of utility function ($1/(1-\hat{g})\hat{W}_T^{1-\hat{g}}$) for $T = 1, 2, 5, 10, \gamma = 5, 10, 20$.

We can see that for all $T$, the means of terminal wealth $\hat{W}_T$ are larger than that of risk-free investment (i.e., $(1+X_f)^T$). In view of the distribution of $\hat{W}_T$, the means are larger than the medians ($q_{0.5}$) which shows the asymmetry of the distribution. Among the myopic, dynamic portfolio using $\Psi^{(1)}$ and $\Psi^{(2)}$, dynamic portfolio using $\Psi^{(2)}$ is the best investment strategy in view of the means of $\hat{W}_T$ or $1/(1-\hat{g})\hat{W}_T^{1-\hat{g}}$. There are some cases that the means of $\hat{W}_T$ for dynamic portfolio using $\Psi^{(1)}$ are smaller than those for myopic portfolio. This phenomenon would show the inaccuracy of the approximation of $\Psi$. In addition, in view of the dispersion of $\hat{W}_T$, the dynamic portfolio’s one is relatively smaller than the myopic portfolio’s one.

**Example 4.2** (sample size ($n$) and resampling size ($B$)). In this example, we examine effect of the initial sample size ($n$) and the resample size ($B$). Let $\mu = 0.02$, $A = 0.1$, $\Gamma = 0.05$, $T = 10$, and $\gamma = 5$. In the same manner as Example 4.1, we consider the effect of $\hat{W}_T$ for $n = 10, 100, 1000$ and $B = 5, 20, 100$. Figure 3 shows the box plots of the terminal wealth $\hat{W}_T$ for each $n$ and $B$.

It can be seen that the medians tend to increase with increased amount of $n$ and $B$. In addition, the wideness of the box plots decreases with increased amount of $n$ and $B$. This phenomenon shows the accuracy of the approximation of $X^*_t$.

**Example 4.3** (AR Parameter ($A$) and variance of $\varepsilon_t$ ($\Gamma$)). In this example, we examine effect of the AR parameter ($A$) and the variance of $\varepsilon_t$ ($\Gamma$). Let $\mu = 0.02$, $n = 100$, $B = 100$, $T = 10$, and $\gamma = 5$. In the same manner as Example 4.1, we consider the effect of $\hat{W}_T$ for $A = 0.01, 0.1, 0.2$, and $\Gamma = 0.01, 0.05, 0.10$. Figure 4 shows the box plots of the terminal wealth $\hat{W}_T$ for each $A$ and $\Gamma$.

Obviously, the medians increase with decreased amount of $\Gamma$ which shows that the investment result is preferred when the amount of $\varepsilon_t$ is small. On the other hand, the wideness
of the box plots increases with increased amount of $A$ which shows that the difference of the investment result is wide when the amount of $A$ is large.

References

[1] R. C. Merton, “Lifetime portfolio selection under uncertainty: the continuous-time case,” *The Review of Economics and Statistics*, vol. 51, no. 3, pp. 247–257, 1969.
[2] P. A. Samuelson, “Lifetime portfolio selection by dynamic stochastic programming,” *Review of Economics and Statistics*, vol. 51, no. 3, pp. 239–246, 1969.
[3] E. F. Fama, “Multiperiod consumption-investment decisions,” *American Economic Review*, vol. 60, no. 1, pp. 163–174, 1970.
[4] Y. Aït-Sahalia and M. W. Brandt, “Variable selection for portfolio choice,” *Journal of Finance*, vol. 56, no. 4, pp. 1297–1351, 2001.
[5] M. W. Brandt, A. Goyal, P. Santa-Clara, and J. R. Stroud, “A simulation approach to dynamic portfolio choice with an application to learning about return predictability,” *Review of Financial Studies*, vol. 18, no. 3, pp. 831–873, 2005.
[6] R. Bellman, *Dynamic Programming*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, USA, 2010.
Research Article

On the Causality between Multiple Locally Stationary Processes

Junichi Hirukawa
Faculty of Science, Niigata University, 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan
Correspondence should be addressed to Junichi Hirukawa, hirukawa@math.sc.niigata-u.ac.jp
Received 14 January 2012; Accepted 25 March 2012
Academic Editor: Kenichiro Tamaki

Copyright © 2012 Junichi Hirukawa. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When one would like to describe the relations between multivariate time series, the concepts of dependence and causality are of importance. These concepts also appear to be useful when one is describing the properties of an engineering or econometric model. Although the measures of dependence and causality under stationary assumption are well established, empirical studies show that these measures are not constant in time. Recently one of the most important classes of nonstationary processes has been formulated in a rigorous asymptotic framework by Dahlhaus in [1996, 1997, 2000], called locally stationary processes. Locally stationary processes have time-varying spectral densities whose spectral structures smoothly change in time. Here, we generalize measures of linear dependence and causality to multiple locally stationary processes. We give the measures of linear dependence, linear causality from one series to the other, and instantaneous linear feedback, at time $t$ and frequency $\lambda$.

1. Introduction

In discussion of the relations between time series, concepts of dependence and causality are frequently invoked. Geweke [1] and Hosoya [2] have proposed measures of dependence and causality for multiple stationary processes (see also Taniguchi et al. [3]). They have also showed that these measures can be additively decomposed into frequency-wise. However, it seems to be restrictive that these measures are constants all the time. Priestley [4] has developed the extensions of prediction and filtering theory to nonstationary processes which have evolutionary spectra. Alternatively, in this paper we generalize measures of dependence and causality to multiple locally stationary processes.

When we deal with nonstationary processes, one of the difficult problems to solve is how to set up an adequate asymptotic theory. To meet this Dahlhaus [5–7] introduced an important class of nonstationary processes and developed the statistical inference. We give the precise definition of multivariate locally stationary processes which is due to Dahlhaus [8].
Definition 1.1. A sequence of multivariate stochastic processes $Z_{t,T} = (Z_{t,T}^{(1)}, \ldots, Z_{t,T}^{(d_Z)})^t$, $(t = 2-N/2, \ldots, 1, \ldots, T, \ldots, T+N/2; T, N \geq 1)$ is called locally stationary with mean vector $\theta$ and transfer function matrix $A^\circ$ if there exists a representation

$$Z_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,T}^\circ(\lambda) d\xi(\lambda),$$

(1.1)

where

(i) $\xi(\lambda) = (\xi^{(1)}(\lambda), \ldots, \xi^{(d_Z)}(\lambda))^t$ is a complex valued stochastic vector process on $[-\pi, \pi]$ with $\xi^{(a)}(\lambda) = \xi^{(a)}(-\lambda)$ and

$$\text{cum}\left\{d\xi^{(a_1)}(\lambda_1), \ldots, d\xi^{(a_k)}(\lambda_k)\right\} = \eta \left(\sum_{j=1}^{k} \lambda_j\right) \frac{\kappa_{a_1,\ldots,a_k}}{(2\pi)^{k-1}} d\lambda_1 \cdots d\lambda_{k-1},$$

(1.2)

for $k \geq 2$, $a_1, \ldots, a_k = 1, \ldots, d_Z$, where cum{⋅⋅⋅} denotes the cumulant of kth order, and $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ is the period $2\pi$ extension of the Dirac delta function.

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

$$\sup_{t,\lambda} \left| A_{t,T}^\circ(\lambda)_{a,b} - A \left(\frac{t}{T}, \lambda\right)_{a,b}\right| \leq KT^{-1}$$

(1.3)

for all $a,b = 1, \ldots, d_Z$ and $T \in \mathbb{N}$. $A(u,\lambda)$ is assumed to be continuous in $u$.

We call $f(u,\lambda) := A(u,\lambda)\Omega A(u,\lambda)^*$ the time-varying spectral density matrix of the process, where $\Omega = (\kappa_{a,b})_{a,b=1,\ldots,d_Z}$. Write

$$\epsilon_t := \int_{-\pi}^{\pi} \exp(i\lambda t) d\xi(\lambda),$$

(1.4)

then $\{\epsilon_t\}$ becomes a white noise process with $E(\epsilon_t) = 0$ and $\text{Var}(\epsilon_t) = \Omega$.

Our objective is the generalization of dependence and causality measures to locally stationary processes and construction of test statistics which can examine the nonstationary effect of actual time series data. The paper, organized as follows. Section 2 explains the generalization of causality measures to multiple locally stationary processes. Since this extension is natural, we do is not explain the original idea of the causality measures in stationary case and recommend to refer Geweke [1] and Hosoya [2] for it. In Section 3 we introduce the nonparametric spectral estimator of multivariate locally stationary processes and explain their asymptotic properties. Finally, we propose the test statistics for linear dependence and show their performance in terms of empirical numerical example in Section 4.
2. Measurements of Linear Dependence and Causality for Nonstationary Processes

Here, we generalize measures of dependence and causality to multiple locally stationary processes. The assumptions and results of this section are straightforward extension of the original idea in stationary case. To avoid repetition, Geweke [1] and Hosoya [2] should be referred to for the original idea of causality.

For the $d$-dimensional locally stationary process $\{Z_t,T\}$, we introduce $H$, the Hilbert space spanned by $Z_{j,t,T}$, $j = 1,\ldots,d$, $t = 0,\pm 1,\ldots$, and call $\mathcal{H}(Z_{t,T})$ the closed subspace spanned by $Z_{j,t,T}$, $j = 1,\ldots,d$, $s \leq t$. We obtain the best one-step linear predictor of $Z_{t+1,T}$ by projecting the components of the vector onto $\mathcal{H}(Z_{t,T})$, so here projection implies component-wise projection. We denote the error of prediction by $\xi_{t,T}$.

Then, for locally stationary process we have

$$E(\xi_s,T\xi_t') = \delta_{s,t}G_{t,T}, \quad (2.1)$$

where $\delta_{s,t}$ is the Kronecker delta function. Note that $\xi_{t,T}$’s are uncorrelated but do not have identical covariance matrices; namely, $G_{t,T}$ are time-dependent. Now, we impose the following assumption on $G_{t,T}$.

**Assumption 2.1.** The covariance matrices of errors $G_{t,T}$ are nonsingular for all $t$ and $T$.

Define

$$u_{t,T} = \sum_{j=0}^{\infty} H_{t,T}(j)\xi_{t-j,T}, \quad \text{tr} \left\{ \sum_{j=0}^{\infty} H_{t,T}(j)G_{t-j,T}H_{t,T}(j)' \right\} < \infty \quad (2.2)$$

as a one-sided linear process and

$$v_{t,T} = Z_{t,T} - u_{t,T}, \quad (2.3)$$

where coefficient matrices are

$$H_{t,T}(j) = E(Z_{t,T}\xi_{t-j,T}')G_{t,T}^{-1}, \quad j \geq 1, \quad H_{t,T}(0) = I_d. \quad (2.4)$$

Note that each $H_{t,T}(j)\xi_{t-j,T}$, $j = 0,1,\ldots$ is projection of $Z_{t,T}$ onto the closed subspace spanned by $\xi_{t-j,T}$. Now, we have the following Wold decomposition for locally stationary processes.

**Lemma 2.2** (Wold decomposition). If $\{Z_{t,T}\}$ is a locally stationary vector process of $d(Z)$ components, then $Z_{t,T} = u_{t,T} + v_{t,T}$, where $u_{t,T}$ is given by (2.1), (2.2), and (2.4), $v_{t,T}$ is deterministic, and $E(v_{s,T}\xi_{t,T}') \equiv 0$.

If only $u_{t,T}$ occurs, we say that $Z_{t,T}$ is purely nondeterministic.

**Assumption 2.3.** $Z_{t,T}$ is purely nondeterministic.
In view of Lemma 2.2, we can see that under Assumptions 2.1 and 2.3, \( Z_{t,T} \) becomes a one-side linear process given by (2.2). For locally stationary process, if we choose an orthonormal basis \( \epsilon_t^{(j)}, j = 1, \ldots, d(Z) \), in the closed subspace spanned by \( \xi_{t,T} \), then \( \{ \epsilon_t \} \) will be an uncorrelated stationary process. We call \( \{ \epsilon_t \} \) a fundamental process of \( \{ Z_{t,T} \} \) and \( C_{t,T}(j); j = 0, 1, \ldots \) denote the corresponding coefficients, that is,

\[
Z_{t,T} = \sum_{j=0}^{\infty} C_{t,T}(j) \epsilon_{t-j}.
\]

Let \( f_{t,T}(\lambda) \) be the time-varying spectral density matrix of \( Z_{t,T} \). A process is said to have the maximal rank if it has nondegenerate spectral density matrix a.e.

**Assumption 2.4.** The locally stationary process \( \{ Z_{t,T} \} \) has the maximal rank for all \( t \) and \( T \). In particular

\[
\int_{-\pi}^{\pi} \log |f_{t,T}(\lambda)| d\lambda > -\infty, \quad \forall t, T,
\]

where \( |D| \) denotes the determinant of the matrix \( D \).

We will say that a function \( \phi(z) \), analytic in the unit disc, belongs to the class \( \mathcal{H}_2 \) if

\[
H_2(\phi) = \sup_{0 \leq r < 1} \int_{-\pi}^{\pi} \left| \phi \left( r e^{i\lambda} \right) \right|^2 d\lambda < \infty.
\]

Under Assumptions 2.1–2.4, it follows that \( \{ Z_{t,T} \} \) has a time-varying spectral density \( f_{t,T}(\lambda) \) which has rank \( d(Z) \) for almost all \( \lambda \), and is representable in the form

\[
f_{t,T}(\lambda) = \frac{1}{2\pi} \Phi_{t,T}(e^{i\lambda}) \Phi_{t,T}^*(e^{i\lambda}),
\]

where \( D^* \) denotes the complex conjugate of matrix \( D \) and \( \Phi_{t,T}(e^{i\lambda}) \) is the boundary value of a \( d(Z) \times d(Z) \) analytic function

\[
\Phi_{t,T}(z) = \sum_{j=0}^{\infty} C_{t,T}(j) z^j,
\]

in the unit disc, and it holds that \( \Phi_{t,T}(0) \Phi_{t,T}(0)^* = G_{t,T} \).

Now, we introduce measures of linear dependence, linear causality, and instantaneous linear feedback at time \( t \). Let \( Z_{t,T} = (X'_{t,T}, Y'_{t,T})' \) be \( d(Z) = (d(X) + d(Y)) \)-dimensional locally stationary process, which has time-varying spectral density matrix:

\[
f_{t,T}(\lambda) = \begin{pmatrix}
    f_{t,T}^{(xx)}(\lambda) & f_{t,T}^{(xy)}(\lambda) \\
    f_{t,T}^{(yx)}(\lambda) & f_{t,T}^{(yy)}(\lambda)
\end{pmatrix}.
\]

\[
f_{t,T}(\lambda) = \begin{pmatrix}
    f_{t,T}^{(xx)}(\lambda) & f_{t,T}^{(xy)}(\lambda) \\
    f_{t,T}^{(yx)}(\lambda) & f_{t,T}^{(yy)}(\lambda)
\end{pmatrix}.
\]
We will find the partitions \( \xi_{t,T}^{(1)} \) and \( \xi_{t,T}^{(2)} \),

\[
\begin{pmatrix}
\xi_{t,T}^{(1)} \\
\xi_{t,T}^{(2)}
\end{pmatrix}
\]

\( d(X) \times 1 \)

and

\[
\text{Cov}(\xi_{t,T}, \xi_{t,T}) = G_{t,T} = \begin{pmatrix}
G_{t,T}^{(1,1)} & G_{t,T}^{(1,2)} \\
G_{t,T}^{(2,1)} & G_{t,T}^{(2,2)}
\end{pmatrix}
\]

useful. Meanwhile \( G_{t,T}^{(X)} \) and \( G_{t,T}^{(Y)} \) denote the covariance matrices of the one-step-ahead errors \( s_{t,T}^{(X)} \) and \( s_{t,T}^{(Y)} \) when \( X_{t,T} \) and \( Y_{t,T} \) are forecasts from their own pasts alone; namely, \( s_{t,T}^{(X)} \) and \( s_{t,T}^{(Y)} \) are the residuals of the projections of \( X_{t,T} \) and \( Y_{t,T} \) onto \( \mathcal{L}(X_{t-1,T}) \) and \( \mathcal{L}(Y_{t-1,T}) \), respectively.

We define the measures of linear dependence, linear causality from \{\( Y_{t,T} \)\} to \{\( X_{t,T} \)\}, from \{\( X_{t,T} \)\} to \{\( Y_{t,T} \)\} and instantaneous linear feedback, at time \( t \) as

\[
M_{t,T}^{(XY)} = \log \frac{G_{t,T}^{(X)}}{|G_{t,T}|},
\]

\[
M_{t,T}^{(Y \rightarrow X)} = \log \frac{|G_{t,T}|}{G_{t,T}^{(1,1)}},
\]

\[
M_{t,T}^{(X \rightarrow Y)} = \log \frac{G_{t,T}^{(2,2)}}{|G_{t,T}|},
\]

\[
M_{t,T}^{(XY)} = \log \frac{|G_{t,T}^{(1,1)}|}{|G_{t,T}^{(2,2)}|},
\]

respectively; then we have

\[
M_{t,T}^{(XY)} = M_{t,T}^{(Y \rightarrow X)} + M_{t,T}^{(X \rightarrow Y)} + M_{t,T}^{(XY)}.
\]

Next, we decompose measures of linear causality into frequency-wise. To define frequency-wise measures of causality, we introduce the following analytic facts.

**Lemma 2.5.** The analytic matrix \( \Phi_{t,T}(z) \) corresponding to a fundamental process \( \{e_t\} \) (for \( \{Z_{t,T}\} \)) is maximal among analytic matrices \( \Psi_{t,T}(z) \) with components from the class \( \mathcal{L}_2 \), and satisfying the boundary condition (2.8); that is,

\[
\Phi_{t,T}(0)\Phi_{t,T}(0)^* \geq \Psi_{t,T}(0)\Psi_{t,T}(0)^*.
\]

Although the following assumption is natural extension of Kolmogorov’s formula in stationary case (see, e.g., [9]), it is not straightforward and unfortunately, so far, we cannot prove it from more simple assumption. We guess it requires another completely technical paper.
Assumption 2.6 (Kolmogorov’s formula). Under Assumptions 2.1–2.4, an analytic matrix \( \Phi_{t,T}(z) \) satisfying the boundary condition (2.8) will be maximal if and only if

\[
|\Phi_{t,T}(0)|^2 = |G_{t,T}| = \exp \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|2\pi f_{t,T}(\lambda)| d\lambda. \tag{2.16}
\]

Now we define the process \( \{\eta_{t,T}\} \) as

\[
\begin{pmatrix}
\eta_{t,T}^{(1)} \\
\eta_{t,T}^{(2)}
\end{pmatrix} = \begin{pmatrix}
I_{d(X)} & -G_{t,T}^{(1,2)} G_{t,T}^{(2,1)}^{-1} \\
-G_{t,T}^{(2,1)} G_{t,T}^{(1,1)}^{-1} & I_{d(Y)}
\end{pmatrix} \begin{pmatrix}
\hat{\eta}_{t,T}^{(1)} \\
\hat{\eta}_{t,T}^{(2)}
\end{pmatrix}, \tag{2.17}
\]

then \( \eta_{t,T}^{(1)} \) is the residuals of the projection of \( X_{t,T} \) onto \( \mathcal{E}(X_{t-1,T}, Y_{t,T}) \), whereas \( \eta_{t,T}^{(2)} \) is the residuals of the projection of \( Y_{t,T} \) onto \( \mathcal{E}(X_{t,T}, Y_{t-1,T}) \).

Furthermore, we have

\[
\text{Cov} \left\{ \begin{pmatrix}
\hat{\eta}_{t,T}^{(1)} \\
\hat{\eta}_{t,T}^{(2)}
\end{pmatrix}, \begin{pmatrix}
\hat{\eta}_{t,T}^{(1)} \\
\hat{\eta}_{t,T}^{(2)}
\end{pmatrix} \right\} = \begin{pmatrix}
G_{t,T}^{(1,1)} & 0 \\
0 & G_{t,T}^{(2,2)} - G_{t,T}^{(2,1)} G_{t,T}^{(1,1)}^{-1} G_{t,T}^{(1,2)}
\end{pmatrix}, \tag{2.18}
\]

so we can see that \( \eta_{t,T}^{(2)} \) is orthogonal to \( \hat{\eta}_{t,T}^{(1)} \). For a \( d(Z) \times d(Z) \) matrix

\[
F_{t,T} = \begin{bmatrix}
I_{d(X)} & 0 \\
-G_{t,T}^{(2,1)} G_{t,T}^{(1,1)}^{-1} & I_{d(Y)}
\end{bmatrix}, \tag{2.19}
\]

we have \( \begin{pmatrix}
\hat{\eta}_{t,T}^{(1)} \\
\hat{\eta}_{t,T}^{(2)}
\end{pmatrix} = F_{t,T} \begin{pmatrix}
\hat{\eta}_{t,T}^{(1)} \\
\hat{\eta}_{t,T}^{(2)}
\end{pmatrix} \).

If we set

\[
\tilde{\Phi}_{t,T}(z) = \Phi_{t,T}(z) \Phi_{t,T}(0)^{-1} F_{t,T}^{-1} G_{t,T}^{-1/2}
\]

\[
= F_{t,T}(z) G_{t,T}^{1/2}, \tag{2.20}
\]

we have the following lemma.

**Lemma 2.7.** \( \tilde{\Phi}_{t,T}(z) \) is an analytic function in the unit disc with \( \tilde{\Phi}_{t,T}(0) \tilde{\Phi}_{t,T}(0)^* = G_{t,T} \) and thus maximal, such that the time-varying spectral density \( f_{t,T}(\lambda) \) has a factorization

\[
f_{t,T}(\lambda) = \frac{1}{2\pi} \tilde{\Phi}_{t,T}(e^{i\lambda}) \tilde{\Phi}_{t,T}(e^{i\lambda})^*. \tag{2.21}
\]
Advances in Decision Sciences

From this lemma, it is seen that time-varying spectral density is decomposed into two parts:

\[
\tilde{f}_{t,T}^{(xy)}(\lambda) = \frac{1}{2\pi} \left\{ \Gamma_{t,T}^{(1,1)} (\epsilon^{(1)}) \Gamma_{t,T}^{(1,1)} (\epsilon^{(1)})^* + \Gamma_{t,T}^{(1,2)} (\epsilon^{(1)}) \Gamma_{t,T}^{(1,2)} (\epsilon^{(1)})^* \right\},
\]

(2.22)

where \( \Gamma_{t,T}^{(1,1)}(z) \) is a \( d^{(X)} \times d^{(X)} \) left-upper submatrix of \( \Gamma_{t,T}(z) \). The former part is related to the process \{\mathbf{x}_{t,T}\} whereas the latter part is related to the process \{\mathbf{y}_{t,T}\}, which is orthogonal to \{\mathbf{x}_{t,T}\}. This relation suggests frequency-wise measure of causality, from \{\mathbf{y}_{t,T}\} to \{\mathbf{x}_{t,T}\} at time \( t \):

\[
M_{t,T}^{(Y \rightarrow X)} (\lambda) = \log \left| \frac{\tilde{f}_{t,T}^{(xy)}(\lambda)}{(1/2\pi) \left\{ \Gamma_{t,T}^{(1,1)} (\epsilon^{(1)}) \Gamma_{t,T}^{(1,1)} (\epsilon^{(1)})^* \right\}} \right|.
\]

(2.23)

Similarly, we propose

\[
M_{t,T}^{(X \rightarrow Y)} (\lambda) = \log \left| \frac{\tilde{f}_{t,T}^{(yx)}(\lambda)}{(1/2\pi) \left\{ \Delta_{t,T}^{(2,2)} (\epsilon^{(1)}) \Delta_{t,T}^{(2,2)} (\epsilon^{(1)})^* \right\}} \right|,
\]

\[
M_{t,T}^{(XY)} (\lambda) = -\log \left| I_{(i)} - \tilde{f}_{t,T}^{(xy)}(\lambda) \tilde{f}_{t,T}^{(yx)}(\lambda) - \tilde{f}_{t,T}^{(xx)}(\lambda) \tilde{f}_{t,T}^{(yy)}(\lambda) \right|,
\]

\[
M_{t,T}^{(XY)} (\lambda) = \log \left| \frac{(1/2\pi) \left\{ \Gamma_{t,T}^{(1,1)} (\epsilon^{(1)}) \Gamma_{t,T}^{(1,1)} (\epsilon^{(1)})^* \right\}}{(1/2\pi) \left\{ \Delta_{t,T}^{(2,2)} (\epsilon^{(1)}) \Delta_{t,T}^{(2,2)} (\epsilon^{(1)})^* \right\}} \right|,
\]

(2.24)

where \( \Delta_{t,T}^{(2,2)}(z) \) is in the same manner of \( \Gamma_{t,T}^{(1,1)}(z) \).

Now, we introduce the following assumption.

**Assumption 2.8.** The roots of \( |\Gamma_{t,T}^{(1,1)}(z)| \) and \( |\Delta_{t,T}^{(2,2)}(z)| \) all lie outside the unit circle.

The relation of frequency-wise measure to overall measure is addressed in the following result.

**Theorem 2.9.** Under Assumptions 2.1–2.8, we have

\[
M_{t,T}^{(c)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} M_{t,T}^{(c)}(\lambda) d\lambda.
\]

(2.25)

If Assumptions 2.1–2.6 hold, but Assumption 2.8 does not hold, then

\[
M_{t,T}^{(Y \rightarrow X)} > \frac{1}{2\pi} \int_{-\pi}^{\pi} M_{t,T}^{(Y \rightarrow X)}(\lambda) d\lambda, \quad M_{t,T}^{(X \rightarrow Y)} > \frac{1}{2\pi} \int_{-\pi}^{\pi} M_{t,T}^{(X \rightarrow Y)}(\lambda) d\lambda.
\]

(2.26)
Assumption 3.1. A stationary processes. First, we make the following assumption on the transfer function matrix.

In this section we introduce the nonparametric spectral estimator of multivariate locally stationary processes. Since \( \mathcal{A}(Z_{t,T}) = \mathcal{A}(X_{t,T}, \mathbf{\eta}_{t,T}) \) and \( \mathcal{A}(Z_{t,T}) = \mathcal{A}(X_{t,T}, \mathbf{\eta}_{t,T}) \), we can see that \( \mathcal{A}(Z_{t,T}) = \mathcal{A}(X_{t,T}, \mathbf{\eta}_{t,T}) \). Therefore, the best one-step prediction error of the process \( \{X_{t,T}, \eta_{t,T}\} \) is given by \( \{\hat{\mathbf{e}}_{t,T}(\lambda)\} \). Let \( \tilde{\mathbf{f}}_{t,T}(\lambda) \) be a time-varying spectral density matrix of the process \( \{X_{t,T}, \eta_{t,T}\} \) and denote the partition by

\[
\tilde{\mathbf{f}}_{t,T}(\lambda) = \begin{pmatrix}
\tilde{f}_{t,T}^{(xx)}(\lambda) & \tilde{f}_{t,T}^{(12)}(\lambda) \\
\tilde{f}_{t,T}^{(21)}(\lambda) & \frac{1}{2\pi} \tilde{G}_{t,T}^{(22)}
\end{pmatrix}.
\]

Then, we obtain another representation of frequency-wise measure of causality, from \( \{Y_{t,T}\} \) to \( \{X_{t,T}\} \) at time \( t \):

\[
M_{t,T}^{(Y \rightarrow X)}(\lambda) = \log \left| \frac{\tilde{f}_{t,T}^{(xx)}(\lambda)}{\tilde{f}_{t,T}^{(xx)}(\lambda) - 2\pi \tilde{f}_{t,T}^{(12)}(\lambda) \tilde{G}_{t,T}^{(22)} \tilde{f}_{t,T}^{(21)}(\lambda)} \right|.
\]

This relation suggests that we apply the nonparametric time-varying spectral density estimator of the residual process \( \{X_{t,T}, \eta_{t,T}\} \). However, this problem requires another paper. We will make it as a further work.

3. Nonparametric Spectral Estimator of Multivariate Locally Stationary Processes

In this section we introduce the nonparametric spectral estimator of multivariate locally stationary processes. First, we make the following assumption on the transfer function matrix \( \mathbf{A}(u, \lambda) \).

Assumption 3.1. (i) The transfer function matrix \( \mathbf{A}(u, \lambda) \) is \( 2\pi \)-periodic in \( \lambda \), and the periodic extension is twice differentiable in \( u \) and \( \lambda \) with uniformly bounded continuous derivatives \( \frac{\partial^2}{\partial u^2} \mathbf{A}, \frac{\partial^2}{\partial \lambda^2} \mathbf{A} \) and \( (\partial^2/\partial u \partial \lambda)(\partial \mathbf{A}/\partial \lambda) \). Furthermore, the uniformly bounded continuous derivative \( (\partial^2/\partial u \partial \lambda)(\partial \mathbf{A}/\partial \lambda) \) also exists.

(ii) All the eigenvalues of \( \mathbf{f}(u, \lambda) \) are bounded from below and above by some constants \( \delta_1, \delta_2 > 0 \) uniformly in \( u \) and \( \lambda \).

As an estimator of \( \mathbf{f}(u, \lambda) \), we use the nonparametric estimator of kernel type defined by

\[
\hat{\mathbf{f}}(u, \lambda) = \int_{-\pi}^{\pi} W_T(\lambda - \mu) I_N(u, \mu) d\mu,
\]

where \( W_T \) is a window function and \( I_N \) is a kernel function.
where \( W_T(\omega) = M \sum_{t=-\infty}^{\infty} W(M(\omega + 2\pi t)) \) is the weight function and \( M > 0 \) depends on \( T \), and \( I_N(u, \lambda) \) is the localized periodogram matrix over the segment \( \{[uT] - N/2 + 1, [uT] + N/2\} \) defined as

\[
I_N(u, \lambda) = \frac{1}{2\pi H_{2,N}} \left\{ \sum_{s=1}^{N} h\left(\frac{s}{N}\right) Z_{[uT]-N/2+s,T} \exp\{i\lambda s\} \right\} \times \left\{ \sum_{r=1}^{N} h\left(\frac{r}{N}\right) Z_{[uT]-N/2+r,T} \exp\{i\lambda r\} \right\}^*.
\]

Here \( h : [0, 1] \to \mathbb{R} \) is a data taper and \( H_{2,N} = \sum_{s=1}^{N} h(s/N)^2 \). It should be noted that \( I_N(u, \lambda) \) is not a consistent estimator of the time-varying spectral density. To make a consistent estimator of \( f(u, \lambda) \) we have to smooth it over neighbouring frequencies.

Now we impose the following assumptions on \( W(\cdot) \) and \( h(\cdot) \).

**Assumption 3.2.** The weighted function \( W : \mathbb{R} \to [0, \infty] \) satisfies \( W(x) = 0 \) for \( x \notin [-1/2, 1/2] \) and is continuous and even function satisfying \( \int_{-1/2}^{1/2} W(x) dx = 1 \) and \( \int_{-1/2}^{1/2} x^2 W(x) dx < \infty \).

**Assumption 3.3.** The data taper \( h : \mathbb{R} \to \mathbb{R} \) satisfies (i) \( h(x) = 0 \) for all \( x \notin [0,1] \) and \( h(x) = h(1-x) \); (ii) \( h(x) \) is continuous on \( \mathbb{R} \), twice differentiable at all \( x \notin U \) where \( U \) is a finite set of \( \mathbb{R} \), and \( \sup_{x \in U} |h''(x)| < \infty \). Write

\[
K_t(x) := \left\{ \int_{0}^{t} h(x)^2 dx \right\}^{-1} h\left( x + \frac{1}{2} \right)^2, \quad x \in \left[ -\frac{1}{2}, \frac{1}{2} \right],
\]

which plays a role of kernel in the time domain.

Furthermore, we assume the following.

**Assumption 3.4.** \( M = M(T) \) and \( N = N(T) \), \( M \ll N \ll T \) satisfy

\[
\frac{\sqrt{T}}{M^2} = o(1), \quad \frac{N^2}{T^{3/2}} = o(1), \quad \frac{\sqrt{T} \log N}{N} = o(1).
\]

The following lemmas are multivariate version of Theorem 2.2 of Dahlhaus [10] and Theorem A.2 of Dahlhaus [7] (see also [11]).

**Lemma 3.5.** Assume that Assumptions 3.1–3.4 hold. Then

(i)

\[
E(I_N(u, \lambda)) = f(u, \lambda) + \frac{N^2}{2T^2} \int_{-1/2}^{1/2} x^2 K_t(x)^2 dx \frac{\partial^2}{\partial u^2} f(u, \lambda)
\]

\[
+ o\left( \frac{N^2}{T^2} \right) + o\left( \frac{\log N}{N} \right).
\]
(ii)

\[
E\left( \hat{f}(u, \lambda) \right) = f(u, \lambda) + \frac{N^2}{2T^2} \int_{-1/2}^{1/2} x^2 K_t(x) dx \frac{\partial^2}{\partial u^2} f(u, \lambda) + \frac{1}{2M^2} \int_{-1/2}^{1/2} x^2 W(x)^2 dx \frac{\partial^2}{\partial \lambda^2} f(u, \lambda) + o\left( \frac{N^2}{T^2} + M^{-2} \right) + O\left( \frac{\log N}{N} \right),
\]

(3.6)

(iii)

\[
\sum_{i,j=1}^{m} \text{Var}\left( \hat{f}_{ij}(u, \lambda) \right) = \frac{M}{N} \sum_{i,j=1}^{m} f_{ij}(u, \lambda)^2 \int_{-1/2}^{1/2} K_t(x)^2 dx \times \int_{-1/2}^{1/2} W(x)^2 dx (2\pi + 2\pi \{ \lambda \equiv 0 \text{ mod } \pi \}) + o\left( \frac{M}{N} \right).
\]

(3.7)

Hence, we have

\[
E \left\| \hat{f}(u, \lambda) - f(u, \lambda) \right\|^2 = O\left( \frac{M}{N} \right) + O\left( M^{-2} + N^2T^{-2} \right)^2 = O\left( \frac{M}{N} \right),
\]

(3.8)

where \( \|D\| \) is the Euclidean norm of the matrix \( D \) and \( \|D\| = \{ \text{tr} \{DD^*\} \}^{1/2} \).

**Lemma 3.6.** Assume that Assumptions 3.1–3.4 hold. Let \( \phi_j(u, \lambda), \ j = 1, \ldots, k \) be \( d^{(Z)} \times d^{(Z)} \) matrix-valued continuous function on \( [0,1] \times [-\pi, \pi] \) which satisfies the same conditions as the transfer function matrix \( A(u, \lambda) \) in Assumption 3.1 and \( \hat{\phi}_j(u, \lambda)^* = \phi_j(u, \lambda), \phi_j(u, -\lambda) = \phi_j(u, \lambda)' \). Then

\[
L_T(\phi_j) = \sqrt{T} \left\{ \frac{1}{T} \sum_{t=1}^{T} \int_{-\pi}^{\pi} \text{tr}\left\{ \phi_j \left( \frac{t}{T}, \lambda \right) I_N \left( \frac{t}{T}, \lambda \right) \right\} d\lambda - \int_{0}^{1} \int_{-\pi}^{\pi} \text{tr}\{\phi_j(u, \lambda)f(u, \lambda)\} d\lambda d\mu \right\}, \ j = 1, \ldots, k
\]

(3.9)
have, asymptotically, a normal distribution with zero mean vector and covariance matrix \( V \) whose \((i, j)\)-th element is

\[
4\pi \int_{0}^{\pi} \left[ \int_{-\pi}^{\pi} \text{tr}\{ \phi_i(u, \lambda) f(u, \lambda) \phi_j(u, \lambda) f(u, \lambda) \} d\lambda \right. \\
+ \left. \frac{1}{4\pi^2} \sum_{a_1, a_2, a_3, a_4} \sum_{b_1, b_2, b_3, b_4} \kappa_{b_1, b_2, b_3, b_4} \right] \\
\times \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi_i(u, \lambda) a_{1, a_2} \phi_j(u, \mu) a_{4, a_3} \cdot A(u, \lambda)_{a_2, b_1} A(u, -\lambda)_{a_1, b_2} \\
\times A(u, -\mu)_{a_1, b_3} A(u, \mu)_{a_3, b_4} d\lambda d\mu \] \\
\left[ \right] du.

(3.10)

Assumption 3.4 does not coincide with Assumption A.1(ii) of Dahlhaus [7]. As mentioned in A.3 Remarks of Dahlhaus [7, page 27], Assumption A.1(ii) of Dahlhaus [7] is required because of the \( \sqrt{T} \)-unbiasedness at the boundary 0 and 1. If we assume that \( \{Z_{2-N/2T, \ldots, Z_{0,T}}\} \) and \( \{Z_{T+1,T, \ldots, Z_{T+2-N/2T}}\} \) are available with Assumption 3.4, then from Lemma 3.5 (i)

\[
E(L_T(\phi_j)) = \sqrt{T} E \left\{ \frac{1}{T} \sum_{t=1}^{T} \int \text{tr} \left\{ \phi_j(t/T, \lambda) I_N \left(t/T, \lambda \right) \right\} d\lambda \right. \\
\left. - \int_{0}^{1} \int_{-\pi}^{\pi} \text{tr} \{ \phi_j(u, \lambda) f(u, \lambda) \} d\lambda du \right\} \\
= O \left( \sqrt{T} \left( \frac{N^2}{T^2} + \frac{\log N}{N} + \frac{1}{T} \right) \right) = o(1).

(4.11)

4. Testing Problem for Linear Dependence

In this section we discuss the testing problem for linear dependence. The average measure of linear dependence is given by the following integral functional of time varying spectral density:

\[
\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} M_{\phi_T}^{(X,Y)} = \int_{0}^{1} \int_{-\pi}^{\pi} - \frac{1}{2\pi} \log \left| L_{d^{(\gamma)}} - f_{xx}(u, \lambda) f_{xy}(u, \lambda)^{-1} f_{yx}(u, \lambda) f_{yy}(u, \lambda)^{-1} \right| d\lambda du \\
= \int_{0}^{1} \int_{-\pi}^{\pi} K_{(X,Y)}(f(u, \lambda)) d\lambda du,

(4.1)
where

\[ K_{(X,Y)}(f(u,\lambda)) \equiv -\frac{1}{2\pi} \log \left| I_{d(u)} - f_{yx}(u,\lambda)f_{xy}(u,\lambda)^{-1}f_{yy}(u,\lambda)^{-1} \right|. \]  

(4.2)

We consider the testing problem for existence of linear dependence:

\[ H : \int_{0}^{1} \int_{-\pi}^{\pi} K_{(X,Y)}(f(u,\lambda)) \, d\lambda \, du = 0 \]  

(4.3)

against

\[ A : \int_{0}^{1} \int_{-\pi}^{\pi} K_{(X,Y)}(f(u,\lambda)) \, d\lambda \, du \neq 0. \]  

(4.4)

For this testing problem, we define the test statistics \( S_T \) as

\[ S_T = \sqrt{T} \int_{0}^{1} \int_{-\pi}^{\pi} K_{(X,Y)}(\hat{f}(u,\lambda)) \, d\lambda \, du, \]  

(4.5)

then, we have the following result.

**Theorem 4.1.** Under \( H \),

\[ S_T \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, V_{K_{(X,Y)}}^2 \right), \]  

(4.6)

where the asymptotic variance of \( S_T \) is given by

\[ V_{K_{(X,Y)}}^2 = 4\pi \int_{0}^{1} \int_{-\pi}^{\pi} \text{tr} \left[ f(u,\lambda) K_{(X,Y)}^{(1)}(f(u,\lambda))' \right]^2 \, d\lambda \]

\[ + \frac{1}{4\pi^2} \sum_{a,b,c,d} \kappa_{a,b,c,d} \left( \gamma_{a,b,c,d}(u) \right) du, \]  

(4.7)

with

\[ \Gamma(u) = \{ \gamma(u) \}_{a,b=1,\ldots,d(z)} = \int_{-\pi}^{\pi} A(u,\lambda)^* K_{(X,Y)}^{(1)}(f(u,\lambda)) A(u,\lambda) d\lambda, \]  

(4.8)

and \( K_{(X,Y)}^{(1)}(\cdot) \) is the first derivative of \( K_{(X,Y)}(\cdot) \).
To simplify, \( \{Z_{t,T}\} \) is assumed to be Gaussian locally stationary process. Then, the asymptotic variance of \( S_T \) becomes the integral functional of the time-varying spectral density:

\[
V_{K_{X,Y}}^2 = 4\pi \int_0^1 \int_{-\pi}^\pi \text{tr} \left[ f(u,\lambda)K_{X,Y}^{(1)}(f(u,\lambda)) \right]^2 d\lambda
\]

(4.9)

If we take \( \tilde{V}_{K_{X,Y}}^2 = V_{K_{X,Y}}^2 \{f(u,\lambda)\} \), then \( \tilde{V}_{K_{X,Y}}^2 \) is consistent estimator of asymptotic variance, so, we have

\[
L_T = \frac{S_T}{\sqrt{V_{K_{X,Y}}^2}} \overset{\mathcal{D}}{\rightarrow} \mathcal{N}(0,1).
\]

(4.10)

Next, we introduce a measure of goodness of our test. Consider a sequence of alternative spectral density matrices:

\[
g_T(u,\lambda) = f(u,\lambda) + \frac{1}{\sqrt{T}} b(u,\lambda),
\]

(4.11)

where \( b(u,\lambda) \) is a \( d^{(Z)} \times d^{(Z)} \) matrix whose entries \( b_{ab}(u,\lambda) \) are square-integrable functions on \( [0,1] \times [-\pi,\pi] \).

Let \( E_{g_T}(\cdot) \) and \( V_0(\cdot) \) denote the expectation under \( g_T(u,\lambda) \) and the variance under \( f(u,\lambda) \), respectively. It is natural to define an efficacy of \( L_T \) by

\[
\text{eff}(L_T) = \lim_{T \to \infty} \frac{E_{g_T}(S_T)}{\sqrt{V_0(S_T)}}
\]

(4.12)

in line with the usual definition for a sequence of “parametric alternatives.” Then we see that

\[
\text{eff}(L_T) = \lim_{T \to \infty} \frac{\sqrt{T} \int_{-\pi}^\pi \left[ K_{X,Y}(g_T(u,\lambda)) - K_{X,Y}(f(u,\lambda)) \right] d\lambda}{V_{K_{X,Y}}^2}
\]

\[
= \int_{-\pi}^\pi \text{tr} \left[ K_{X,Y}^{(1)}(f(u,\lambda)) b(u,\lambda) \right] d\lambda
\]

(4.13)

For another test \( L_T^* \) we can define an asymptotic relative efficiency (ARE) of \( L_T \) relative to \( L_T^* \) by

\[
\text{ARE}(L_T, L_T^*) = \left( \frac{\text{eff}(L_T)}{\text{eff}(L_T^*)} \right)^2.
\]

(4.14)
Table 1: $L_T$ in (4.10) for each two companies.

|       | 1:Hi | 2:Ma | 3:Sh | 4:So | 5:Ho | 6:Ni | 7:To |
|-------|------|------|------|------|------|------|------|
| 1:Hi  | —    | —    | —    | —    | —    | —    | —    |
| 2:Ma  | 18.79| —    | —    | —    | —    | —    | —    |
| 3:Sh  | 19.86| 18.93| —    | —    | —    | —    | —    |
| 4:So  | 19.22| 19.18| 18.27| —    | —    | —    | —    |
| 5:Ho  | 15.35| 14.46| 15.17| 15.42| —    | —    | —    |
| 6:Ni  | 15.18| 15.03| 15.84| 16.58| 19.24| —    | —    |
| 7:To  | 15.86| 16.06| 16.00| 16.61| 20.57| 19.12| —    |

Figure 1: The daily linear dependence between HONDA and TOYOTA.

If we take the test statistic based on stationary assumption as another test $L_T^*$, we can measure the effect of nonstationarity when the process concerned is locally stationary process.

Finally, we discuss a testing problem of linear dependence for stock prices of Tokyo Stock Exchange. The data are daily log returns of 7 companies; 1: HITACHI 2: MATSUSHITA 3: SHARP 4: SONY 5: HONDA 6: NISSAN 7: TOYOTA. The individual time series are 1174 data points since December 28, 1999 until October 1, 2004. We compute $L_T$ in (4.10) for each two companies. The selected parameters are $T = 1000$, $N = 175$, and $M = 8$, where $N$ is the length of segment which the localized periodogram is taken over and $M$ is the bandwidth of the weight function.

The results are listed in Table 1. It shows that all values for each two companies are large. Since under null hypothesis the limit distribution of $L_T$ is standard normal, we can conclude hypothesis is rejected. Namely, the linear dependencies exist at each two companies. In particular, the values both among electric appliance companies and among automobile companies are significantly large. Therefore, we can see that the companies in the same business have strong dependence.

In Figures 1 and 2, the daily linear dependence between HONDA and TOYOTA and between HITACHI and SHARP is plotted. They show that the daily dependencies are not
constant and change in time. So, it seems to be reasonable that we use the test statistic based on nonstationary assumption.

References

[1] J. Geweke, “Measurement of linear dependence and feedback between multiple time series,” *Journal of the American Statistical Association*, vol. 77, no. 378, pp. 304–313, 1982.

[2] Y. Hosoya, “The decomposition and measurement of the interdependency between second-order stationary processes,” *Probability Theory and Related Fields*, vol. 88, no. 4, pp. 429–444, 1991.

[3] M. Taniguchi, M. L. Puri, and M. Kondo, “Nonparametric approach for non-Gaussian vector stationary processes,” *Journal of Multivariate Analysis*, vol. 56, no. 2, pp. 259–283, 1996.

[4] M. B. Priestley, *Spectral Analysis and Time Series*, Academic Press, London, UK, 1981.

[5] R. Dahlhaus, “On the Kullback-Leibler information divergence of locally stationary processes,” *Stochastic Processes and Their Applications*, vol. 62, no. 1, pp. 139–168, 1996.

[6] R. Dahlhaus, “Maximum likelihood estimation and model selection for locally stationary processes,” *Journal of Nonparametric Statistics*, vol. 6, no. 2-3, pp. 171–191, 1996.

[7] R. Dahlhaus, “Fitting time series models to nonstationary processes,” *The Annals of Statistics*, vol. 25, no. 1, pp. 1–37, 1997.

[8] R. Dahlhaus, “A likelihood approximation for locally stationary processes,” *The Annals of Statistics*, vol. 28, no. 6, pp. 1762–1794, 2000.

[9] Yu. A. Rozanov, *Stationary Random Processes*, Holden-Day, San Francisco, Calif, USA, 1967.

[10] R. Dahlhaus, “Asymptotic statistical inference for nonstationary processes with evolutionary spectra,” in *Athens Conference on Applied Probability and Time Series 2*, vol. 115 of Lecture Notes in Statist, pp. 145–159, Springer, New York, NY, USA, 1996.

[11] K. Sakiyama and M. Taniguchi, “Discriminant analysis for locally stationary processes,” *Journal of Multivariate Analysis*, vol. 90, no. 2, pp. 282–300, 2004.
Research Article

Optimal Portfolios with End-of-Period Target

Hiroshi Shiraishi, Hiroaki Ogata, Tomoyuki Amano, Valentin Patilea, David Veredas, and Masanobu Taniguchi

1 The Jikei University School of Medicine, Tokyo 1828570, Japan
2 School of International Liberal Studies, Waseda University, Tokyo 1698050, Japan
3 Faculty of Economics, Wakayama University, Wakayama 6408510, Japan
4 CREST, Ecole Nationale de la Statistique et de l’Analyse de l’Information, France
5 ECARES, Solvay Brussels School of Economics and Management, Université libre de Bruxelles, CP114/04, Avenue F.D. Roosevelt 50, 1050, Brussels, Belgium
6 Department of Applied Mathematics, School of Fundamental Science and Engineering, Waseda University, Tokyo 1698555, Japan

Correspondence should be addressed to Hiroshi Shiraishi, hiroshi.shiraishi@jikei.ac.jp

Received 7 November 2011; Accepted 22 December 2011

Academic Editor: Junichi Hirukawa

Copyright © 2012 Hiroshi Shiraishi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study the estimation of optimal portfolios for a Reserve Fund with an end-of-period target and when the returns of the assets that constitute the Reserve Fund portfolio follow two specifications. In the first one, assets are split into short memory (bonds) and long memory (equity), and the optimality of the portfolio is based on maximizing the Sharpe ratio. In the second, returns follow a conditional heteroskedasticity autoregressive nonlinear model, and we study when the distribution of the innovation vector is heavy-tailed stable. For this specification, we consider appropriate estimation methods, which include bootstrap and empirical likelihood.

1. Introduction

The Government Pension Investment Fund (GPIF) of Japan was established in April 1st 2006 as an independent administrative institution with the mission of managing and investing the Reserve Fund of the employees’ pension insurance and the national pension (http://www.gpif.go.jp/ for more information) [1]. It is the world’s largest pension fund ($1.4 trillions in assets under management as of December 2009), and it has a mission of managing and investing the Reserve Funds in safe and efficient investment with a long-term perspective. Business management targets to be achieved by GPIF are set by the Minister of Health, Labour, and Welfare based on the law on the general rules of independent administrative agencies. In the actuarial science, “required Reserve Fund” for pension insurance has been investigated for a long time. The traditional approach focuses on the expected value of future obligations and
interest rate. Then, the investment strategy is determined for exceeding the expected value of interest rate. Recently, solvency for the insurer is defined in terms of random values of future obligations (e.g., Olivieri and Pitacco [2]). In this paper, we assume that the Reserve Fund is defined in terms of the random interest rate and the expected future obligations. Then, we propose optimal portfolios by optimizing the randomized Reserve Fund.

The GPIF invests in a portfolio of domestic and international stocks and bonds. In this paper, we consider the optimal portfolio problem of the Reserve Fund under two econometric specifications for the asset’s returns.

First, we select the optimal portfolio weights based on the maximization of the Sharpe ratio under three different functional forms for the portfolio mean and variance, two of them depending on the Reserve Fund at the end-of-period target (about 100 years). Following the asset structure of the GPIF, we split the assets into cash and domestic and foreign bonds on one hand and domestic and foreign equity on the other. The first type of assets are assumed to be short memory, while the second type are long memory. To obtain the optimal portfolio weights, we rely on bootstrap. For the short memory returns, we use wild bootstrap (WB). Early work focuses on providing first- and second-order theoretical justification for the wild bootstrap in the classical linear regression model (see, e.g., [3]). Gonçalves and Kilian [4] show that WB is applicable for the linear regression model with conditional heteroscedastic such as stationary ARCH, GARCH, and stochastic volatility effects. For the long memory returns, we apply sieve bootstrap (SB). Bühlmann [5] establishes consistency of the autoregressive sieve bootstrap. Assuming that the long memory process can be written as AR($\infty$) and MA($\infty$) processes, we estimate the long memory parameter by means of the Whittle’s approximate likelihood [6]. Given this estimator, the residuals are computed and resampled for the construction of the bootstrap samples, from which the optimal portfolio estimated weights are obtained. We study the usefulness of these procedures with an application to the GPIF assets.

Second, we consider the case when the returns are time dependent and follow a heavy-tailed. It is known that one of the stylized facts of financial returns are heavy tails. It is, therefore, reasonable to use the stable distribution, instead of the Gaussian, since it allows for skewness and fat tails. We couple this distribution with the conditional heteroskedasticity autoregressive nonlinear (CHARN) model that nests many well-known time series models, such as ARMA and ARCH. We estimate the parameters and the optimal portfolio by means of empirical likelihood.

The paper is organized as follows. Section 2 sets the Reserve Fund portfolio problem. Section 3 focuses on the first part, that is, estimation in terms of the Sharpe ratio and discusses the bootstrap procedure. Section 4 covers the CHARN model under stable innovations and the estimation by means of empirical likelihood. Section 5 concludes.

2. Reserve Funds Portfolio with End-of-Period Target

Let $S_{i,t}$ be the price of the $i$th asset at time $t$ ($i = 1, \ldots, k$), and let $X_{i,t}$ be its log-return. Time runs from 0 to $T$. The paper, we consider that today is $T_0$ and $T$ is the end-of-period target. Hence the past and present observations run for $t = 0, \ldots, T_0$, and the future until the end-of-period target for $t = T_0 + 1, \ldots, T$. The price $S_{i,t}$ can be written as

$$S_{i,t} = S_{i,t-1} \exp\{X_{i,t}\} = S_{i,0} \exp\left(\sum_{s=1}^{t}X_{i,s}\right),$$

(2.1)
where $S_{i,0}$ is the initial price. Let $F_{i,t}$ denote the Reserve Fund on asset $i$ at time $t$ and be defined by

$$F_{i,t} = F_{i,t-1} \exp \{X_{i,t}\} - c_{i,t},$$

(2.2)

where $c_{i,t}$ denotes the maintenance cost at time $t$. By recursion, $F_{i,t}$ can be written as

$$F_{i,t} = F_{i,t-1} \exp \{X_{i,t}\} - c_{i,t}
= F_{i,t-2} \exp \left( \sum_{s=1}^{t} X_{i,s} \right) - \sum_{s=1}^{t} c_{i,s} \exp \left( \sum_{s' = s+1}^{t} X_{i,s'} \right)
= F_{i,0} \exp \left( \sum_{s=1}^{t} X_{i,s} \right) - \sum_{s=1}^{t} c_{i,s} \exp \left( \sum_{s' = s+1}^{t} X_{i,s'} \right),$$

(2.3)

where $F_{i,0} = S_{i,0}$.

We gather the Reserve Funds in the vector $F_t = (F_{1,t}, \ldots, F_{k,t})$. Let $F_t(\alpha) = \alpha' F_t$ be a portfolio form by the $k$ Reserve Funds, which depend on the vector of weights $\alpha = (\alpha_1, \ldots, \alpha_k)$. The portfolio Reserve Fund can be expressed as a function of all past returns

$$F_t(\alpha) \equiv \sum_{i=1}^{k} \alpha_i F_{i,t}
= \sum_{i=1}^{k} \alpha_i \left( F_{i,0} \exp \left( \sum_{s=1}^{t} X_{i,s} \right) - \sum_{s=1}^{t} c_{i,s} \exp \left( \sum_{s' = s+1}^{t} X_{i,s'} \right) \right).$$

(2.4)

We are interested in maximizing $F_t(\alpha)$ at the end-of-period target $F_T(\alpha)$

$$F_T(\alpha) = \sum_{i=1}^{k} \alpha_i \left( F_{i,T_0} \exp \left( \sum_{s=T_0+1}^{T} X_{i,s} \right) - \sum_{s=T_0+1}^{T} c_{i,s} \exp \left( \sum_{s' = s+1}^{T} X_{i,s'} \right) \right).$$

(2.5)

It depends on the future returns, the maintenance cost, and the portfolio weights. While the first two are assumed to be constant from $T_0$ to $T$ (the constant return can be seen as the average return over the $T - T_0$ periods), we focus on the optimality of the weights that we denote by $\alpha^{opt}$.

### 3. Sharpe-Ratio-Based Optimal Portfolios

In the first specification, the estimation of the optimal portfolio weights is based on the maximization of the Sharpe ratio:

$$\alpha^{opt} = \arg \max_{\alpha} \frac{\mu(\alpha)}{\sigma(\alpha)},$$

(3.1)
under different functional forms of the expectation $\mu(\alpha)$ and the risk $\sigma(\alpha)$ of the portfolio. We propose three functional forms, two of them depending on the Reserve Fund. The first one is the traditional based on the returns:

$$
\mu(\alpha) = \alpha' E(X_T), \quad \sigma(\alpha) = \sqrt{\alpha' V(X_T)\alpha},
$$

where $E(X_T)$ and $V(X_T)$ are the expectation and the covariance matrix of the returns at the end-of-period target. The second form for the portfolio expectation and risk is based on the vector of Reserve Funds:

$$
\mu(\alpha) = \alpha' E(F_T), \quad \sigma(\alpha) = \sqrt{\alpha' V(F_T)\alpha},
$$

where $E(F_T)$ and $V(F_T)$ indicate the mean and covariance of the Reserve Funds at time $T$. Last, we consider the case where the portfolio risk depends on the lower partial moments of the Reserve Funds at the end-of-period target:

$$
\mu(\alpha) = \alpha' E(F_T), \quad \sigma(\alpha) = E\left\{ \left( \bar{F} - F_T(\alpha) \right) \mathbb{1}\left( F_T(\alpha) < \bar{F} \right) \right\},
$$

where $\bar{F}$ is a given value.

Standard portfolio management rules are based on a mean-variance approach, for which risk is measured by the standard deviation of the future portfolio value. However, the variance often does not provide a correct assessment of risk under dependency and non-Gaussianity. To overcome this problem, various optimization models have been proposed such as mean-semivariance model, mean-absolute deviation model, mean-variance-skewness model, mean-(C)VaR model, and mean-lower partial moment model. The mean-lower partial moment model is an appropriate model for reducing the influence of heavy tails.

The $k$ returns are split into $p$- and $q$-dimensional vectors $\{X^S_t; t \in \mathbb{Z}\}$ and $\{X^L_t; t \in \mathbb{Z}\}$, where $S$ and $L$ stand for short and long memory, respectively. The short memory returns correspond to cash and domestic and foreign bonds, which we generically denote by bonds. The long memory returns correspond to domestic and foreign equity, which we denote as equity.

Cash and bonds follow the nonlinear model

$$
X^S_t = \mu^S + \mathbf{H}\left(X^S_{t-1}, \ldots, X^S_{t-m}\right)\epsilon^S_t,
$$

where $\mu^S$ is a vector of constants, $\mathbf{H}: \mathbb{R}^{np} \rightarrow \mathbb{R}^p \times \mathbb{R}^p$ is a positive definite matrix-valued measurable function, and $\epsilon^S_t = (\epsilon^S_{1,t}, \ldots, \epsilon^S_{p,t})$ are i.i.d. random vectors with mean $0$ and covariance matrix $\Sigma^S$. By contrast, equity returns follow a long memory nonlinear model

$$
X^L_t = \sum_{\nu=0}^{\infty} \phi^L_{t-\nu} + \epsilon^L_t, \quad \epsilon^L_t = \sum_{\nu=0}^{\infty} q^L_{t-\nu} X^L_{t-\nu},
$$
where

\[
\phi_\nu = \frac{\Gamma(v + d)}{\Gamma(v + 1)\Gamma(d)}, \quad q_\nu = \frac{\Gamma(v - d)}{\Gamma(v + 1)\Gamma(-d)}
\] (3.7)

with \(-1/2 < d < 1/2\), and \(e^L_t = (e^L_{1,t}, \ldots, e^L_{p,t})\) are i.i.d. random vectors with mean 0 and covariance matrix \(\Sigma^L\).

We estimate the optimal portfolio weights by means of bootstrap. Let the superindexes \((S, b)\) and \((L, b)\) denote the bootstrapped samples for the bonds and equity, respectively, and \(B\) the total number of bootstrapped samples. In the sequel, we show the bootstrap procedure for both types of assets.

**Bootstrap Procedure for \(X^{(S,b)}_t\)**

**Step 1.** Generate the i.i.d. sequences \(\{\epsilon^{(S,b)}_t\}\) for \(t = T_0 + 1, \ldots, T\) and \(b = 1, \ldots, B\) from \(N(0, I_p)\).

**Step 2.** Let \(Y^S_t = X^S_t - \hat{\mu}^S\), where \(\hat{\mu}^S = (1/T_0) \sum_{s=1}^{T_0} X^S_s\). Generate the i.i.d. sequences \(\{Y^{(S,b)}_t\}\) for \(t = T_0 + 1, \ldots, T\) and \(b = 1, \ldots, B\) from the empirical distribution of \(\{Y^S_t\}\).

**Step 3.** Compute \(\{X^{(S,b)}_t\}\) for \(t = T_0 + 1, \ldots, T\) and \(b = 1, \ldots, B\) as

\[
X^{(S,b)}_t = \hat{\mu}^S + Y^{(S,b)}_t \odot \epsilon^{(S,b)}_t,
\] (3.8)

where \(\odot\) denotes the cellwise product.

**Bootstrap Procedure for \(X^{(L,b)}_t\)**

**Step 1.** Estimate \(\hat{d}\) from the observed returns by means of Whittle’s approximate likelihood:

\[
\hat{d} = \arg \min_{d \in (0,1/2)} L(d, \Sigma),
\] (3.9)

where

\[
L(d, \Sigma) = \frac{2}{T_0} \sum_{j=1}^{(T_0-1)/2} \left\{ \log \det f(\lambda_j, T_0, d, \Sigma) + tr \left( f(\lambda_j, T_0, d, \Sigma)^{-1} I(\lambda_j, T_0) \right) \right\},
\]

\[
f(\lambda, d, \Sigma) = \frac{1 - \exp(i\lambda)}{2\pi} \Sigma^d,
\]

\[
I(\lambda) = \frac{1}{\sqrt{2\pi T_0}} \left| \sum_{t=1}^{T_0} X^L_t e^{i\lambda t} \right|^2,
\]

\[
\lambda_j, T_0 = \frac{2\pi j}{T_0}.
\] (3.10)
Step 2. Compute \( \{ \tilde{e}_t^L \} \) for \( t = 1, \ldots, T_0 \),

\[
\tilde{e}_t^L = \sum_{k=0}^{T_0-1} \tau_k X_{t-k}^L,
\]

where \( \tau_k = \begin{cases} 
\frac{\Gamma(k - \hat{d})}{\Gamma(k + 1)\Gamma(-\hat{d})}, & k \leq 100, \\
\frac{\Gamma(k - \hat{d})}{k^{-d-1} \Gamma(-\hat{d})}, & k > 100.
\end{cases}
\] (3.11)

Step 3. Generate \( \{ e_t^{(L,b)} \} \) for \( t = T_0 + 1, \ldots, T \) and \( b = 1, \ldots, B \) from the empirical distribution of \( \{ \tilde{e}_t^L \} \).

Step 4. Generate \( \{ X_t^{(L,b)} \} \) for \( t = T_0 + 1, \ldots, T \) and \( b = 1, \ldots, B \) as

\[
X_t^{(L,b)} = \sum_{k=0}^{T_0-1} \tau_k e_{t-k}^{(L,b)} + \sum_{k=T_0}^{T-1} \tau_k \tilde{e}_{t-k}.
\] (3.12)

We gather \( X_t^{(S,b)} \) and \( X_t^{*(L,b)} \) into \( X_t^{(b)} = (X_t^{(S,b)}, X_t^{(L,b)}) = (X_{1,t}^{(b)}, \ldots, X_{T+q,T}^{(b)}) \) for \( t = T_0 + 1, \ldots, T \) and \( b = 1, \ldots, B \). The bootstrapped Reserve Funds \( F_{i,T}^{(b)} = (F_{1,T}^{(b)}, \ldots, F_{p+q,T}^{(b)}) \)

\[
F_{i,T}^{(b)} = F_{i,T_0} \exp \left( \sum_{s=T_0+1}^{T} X_{i,s}^{(b)} \right) - \sum_{s=T_0+1}^{T} \epsilon_i \exp \left( \sum_{s=s+1}^{T} X_{i,s}^{(b)} \right).
\] (3.13)

And the bootstrapped Reserve Fund portfolio is

\[
F_T^{(b)}(\alpha) = \alpha' F_T^{(b)} = \sum_{i=1}^{p+q} \alpha_i F_{i,T}^{(b)}.
\] (3.14)

Finally, the estimated portfolio weights that give the optimal portfolio are

\[
\tilde{\alpha}^{opt} = \arg \max_{\alpha} \frac{\mu^{(b)}(\alpha)}{\sigma^{(b)}(\alpha)},
\] (3.15)

where \( \mu^{(b)}(\alpha) \) and \( \sigma^{(b)}(\alpha) \) may take any of the three forms introduced earlier but be evaluated in the bootstrapped returns or Reserve Funds.

### 3.1. An Illustration

We consider monthly log-returns from January 31 1971 to October 31 2009 (466 observations) of the five types of assets considered earlier: domestic bond (DB), domestic equity (DE), foreign bond (FB), foreign equity (FE), and cash (cash). Cash and bonds are gathered in the short-memory panel \( X_i^S = (X_i^{(DB)}, X_i^{(FB)}, X_i^{(cash)}) \) and follow (3.5). Equities are gathered into...
We estimate the optimal portfolio weights, $\hat{\alpha}_{\text{opt1}}$, $\hat{\alpha}_{\text{opt2}}$, and $\hat{\alpha}_{\text{opt3}}$, corresponding to the three forms for the expectation and risk of the Sharpe ratio, and we compute the trajectory of the optimal Reserve Fund for $t = T_0 + 1, \ldots, T$. Because of liquidity reasons, the portfolio weight for cash is kept constant to 5%. The target period is fixed to 100 years, and the maintenance cost is based on the 2004 Pension Reform.

Table 1 shows the estimated optimal portfolio weights for the three different choices of the portfolio expectation and risk. The weight of domestic bonds is very high and clearly dominates over the other assets. Domestic bonds are low risk and medium return, which is in contrast with equity that shows higher return but also higher risk, and with foreign bonds that show low return and risk. Therefore, in a sense, domestic bonds are a compromise between the characteristic of the four equities and bonds.

Figure 2 shows the trajectory of the future Reserve Fund for different values of the yearly return (assumed to be constant from $T_0 + 1$ to $T$) ranging from 2.7% to 3.7%. Since the investment term is extremely long, 100 years, the Reserve Fund is quite sensitive to the choice of the yearly return. In the 2004 Pension Reform, authorities assumed a yearly return of the portfolio of 3.2%, which corresponds to the middle thick line of the figure.

### 4. Optimal Portfolio with Time-Dependent Returns and Heavy Tails

In this section, we consider the second scenario where returns follow a dependent model with stable innovations. The theory of portfolio choice is mostly based on the assumption that investors maximize their expected utility. The most well-known utility is the Markowitz’s mean-variance function that is optimal under Gaussianity. However, it is widely
acknowledged that financial returns show fat tails and, frequently, skewness. Moreover, the variance may not always be the best risk measure. Since the purpose of GPIF is to avoid making a big loss at a certain point in future, risk measures that summarize the probability that the Reserve Fund is below the prescribed level at a certain future point, such as value at risk (VaR), are more appropriate [7]. In addition, the traditional mean-variance approach considers that returns are i.i.d., which is not realistic as past information may help to explain today’s distribution of returns.

We need a specification that allows for heavy tails and skewness and time depend-encies. This calls for a general model with location and scale that are a function of past observations and with innovations that are stable distributed. The location-scale model for the returns is the conditional heteroscedastic autoregressive nonlinear (CHARN), which is fairly general and it nests important models such as ARMA and ARCH.

Estimation of the parameters in a stable framework is not straightforward since the density does not have a closed form (Maximum likelihood is feasible for the i.i.d. univariate case thanks to the STABLE packages developed by John Nolan—see Nolan [8] and the website http://academic2.american.edu/~jpnolan/stable/stable.html. For more complicated cases, including dynamics, maximum likelihood is a quite complex task.). We rely on empirical likelihood, which is one of the nonparametric methods, as it has been already studied in this context [9]. Once the parameters are estimated, we simulate samples of the returns, which are used to compute the Reserve Fund at the end-of-period target, and estimate the optimal portfolio weights by means of minimizing the empirical VaR of the Reserve Fund at time $T$.

Suppose that the vector of returns $X_t \in \mathbb{R}^k$ follows the following CHARN model:

$$X_t = F_\mu(X_{t-1}, \ldots, X_{t-p}) + H_\sigma(X_{t-1}, \ldots, X_{t-p}) \varepsilon_t,$$  \hspace{1cm} (4.1)

where $F_\mu : \mathbb{R}^{kp} \rightarrow \mathbb{R}^k$ is a vector-valued measurable function with a parameter $\mu \in \mathbb{R}^p$ and $H_\sigma : \mathbb{R}^{kq} \rightarrow \mathbb{R}^k \times \mathbb{R}^k$ is a positive definite matrix-valued measurable function with a
parameter $\sigma \in \mathbb{R}^n$. Each element of the vector of innovations $\epsilon_i \in \mathbb{R}^k$ is standardized stable distributed: $\epsilon_{ij} \sim S(\alpha_i, \beta_i, 1, 0)$ and $\epsilon_{ij}$’s are independent with respect to both $i$ and $t$. We set $\theta = (\mu, \sigma, \alpha, \beta)$, where $\alpha = (\alpha_1, \ldots, \alpha_k)$ and $\beta = (\beta_1, \ldots, \beta_k)$.

The stable distribution is often represented by its characteristic function:

$$\phi(v) = E[\exp(iv\epsilon_{i,t})] = \exp\left(\gamma|\nu|^{\alpha}\left(1 + i\beta \text{sgn}(\nu) \tan\left(\frac{\pi \alpha}{2}\left(|\nu|^{1-\alpha} - 1\right)\right)\right) + iv\delta\right),$$  \hspace{1cm} (4.2)

where $\delta \in \mathbb{R}$ is a location parameter, $\gamma > 0$ is a scale parameter, $\beta \in [-1, 1]$ is a skewness parameter, and $\alpha \in (0, 2]$ is a characteristic exponent that captures the tail thickness of the distribution: the smaller the $\alpha$ the heavier the tail. The distributions with $\alpha = 2$ correspond to the Gaussian. The existence of moments is given by $\alpha$: moments of order higher than $\alpha$ do not exist, with the case of $\alpha = 2$ being an exception, for which all moments exist.

The lack of important moments may, in principle, render estimation by the method of moments difficult. However, instead of matching moments, it is fairly simple to match the theoretical and empirical characteristic function evaluated at a grid of frequencies [9]. Let

$$\epsilon_i = H^{-1}_\sigma(X_i - F_\mu)$$  \hspace{1cm} (4.3)

be the residual of the CHARN model. If the parameters $\mu$ and $\sigma$ are the true ones, the residuals $\epsilon_{ij}$ should be independently and identically distributed to $S(\alpha_i, \beta_i, 1, 0)$. So the aim is to find the estimated parameters such that the residuals are i.i.d. and stable distributed, meaning that their probability law is expressed by the above characteristic function. Or, in other words, estimate the parameters by matching the empirical and theoretical characteristic functions and minimizing their distance. Let $J$ be the number of frequencies at which we evaluate the characteristic function: $\nu_1, \ldots, \nu_J$. That makes, in principle, a system of $J$ matching equations. But since the characteristic function can be split into the real and imaginary parts, $\phi(v) = E[\cos(v\epsilon_{i,t})] + iE[\sin(v\epsilon_{i,t})]$, we double the dimension of the system by matching these parts. Let $\Re(\phi(v))$ and $\Im(\phi(v))$ be the real and imaginary parts of the theoretical characteristic function, and $\cos(v\epsilon_{i,t})$ and $\sin(v\epsilon_{i,t})$ the empirical counterparts. The estimating functions are

$$\psi_{\theta}(\epsilon_{i,t}) = \begin{pmatrix} 
\cos(\nu_1 \epsilon_{i,t}) - \Re(\phi(\nu_1)) \\
\vdots \\
\cos(\nu_j \epsilon_{i,t}) - \Re(\phi(\nu_j)) \\
\sin(\nu_1 \epsilon_{i,t}) - \Im(\phi(\nu_1)) \\
\vdots \\
\sin(\nu_j \epsilon_{i,t}) - \Im(\phi(\nu_j))
\end{pmatrix},$$  \hspace{1cm} (4.4)

for each $i = 1, \ldots, k$, and gather them into the vector

$$\psi_{\theta}(\epsilon_i) = (\psi_{\theta}(\epsilon_{i,1}), \ldots, \psi_{\theta}(\epsilon_{i,k})).$$  \hspace{1cm} (4.5)
The number of frequencies \( J \) and the frequencies themselves are chosen arbitrary. Feuerverger and McDunnough [10] show that the asymptotic variance can be made arbitrarily close to the Cramér-Rao lower bound if the number of frequencies is sufficiently large and the grid is sufficiently fine and extended. Similarly, Yu [11, Section 2.1] argues that, from the viewpoint of the minimum asymptotic variance, many and fine frequencies are the appropriate. However, Carrasco and Florens [12] show that too fine frequencies lead to a singular asymptotic variance matrix and we cannot calculate its inverse.

Given the estimating functions (4.5), the natural estimator is constructed by GMM:

\[
\hat{\theta} = \arg \min_{\theta} E[\psi_\theta(e_t)] \cdot W E[\psi_\theta(e_t)],
\]

where \( W \) is a weighting matrix defining the metric (its optimal choice is typically the inverse of the covariance matrix of \( \psi_\theta(e_t) \)) and the expectations are replaced by sample moments. GMM estimator can be generalized to the empirical likelihood estimator, which was originally proposed by Owen [13] as nonparametric methods of inference based on a data-driven likelihood ratio function (see also [14], for a review and applications). It produces a better variance estimate in one step, while, in general, the optimal GMM requires a preliminary step and a preliminary estimation of an optimal \( W \) matrix. We define the empirical likelihood ratio function for \( \theta \) as

\[
\mathcal{R}(\theta) = \max_{p} \left\{ T_0 \prod_{t=1}^{T_0} p_t \left| \sum_{t=1}^{T_0} p_t \psi_\theta(e_t) = 0, \sum_{t=1}^{T_0} p_t = 1, p_t \geq 0 \right. \right\},
\]

where \( p = (p_1, \ldots, p_{T_0}) \) and the maximum empirical likelihood estimator is

\[
\tilde{\theta} = \arg \max_{\theta} \mathcal{R}(\theta).
\]

Qin and Lawless [15] show that this estimator is consistent, asymptotically Gaussian, and with covariance matrix \((B^*_{\theta_0}, A^{-1}_{\theta_0}B_{\theta_0})^{-1}\), where

\[
B_{\theta} = E\left[\frac{\partial \psi_\theta}{\partial \theta}\right], \quad A_{\theta} = E[\psi_\theta(e_t)\psi_\theta(e_t)'].
\]

Once the parameters are estimated, we compute the optimal portfolio weights and the portfolio Reserve Fund at the end-of-period target. Because of a notational conflict, the weights are now denoted by \( a = (a_1, \ldots, a_k) \). And, for simplicity, we assume that there is no maintenance cost, so (2.5) simplifies to

\[
F_T(a) = \sum_{i=1}^{k} a_i F_{i,T_0} \exp \left( \sum_{t=T_0+1}^{T} X_{i,t} \right).
\]

The procedure to estimate the optimal portfolio weights is as follows.
Advances in Decision Sciences

Step 1. For each asset $i = 1, \ldots, k$, we simulate the innovation process

$$
\tilde{\epsilon}_{it} \sim S(\tilde{\alpha}_i, \tilde{\beta}_i, 1, 0), \quad t = T_0 + 1, \ldots, T
$$

based on the maximum empirical likelihood estimator $(\tilde{\alpha}_i, \tilde{\beta}_i)$.

Step 2. We calculate the predicted log-returns

$$
\tilde{X}_t = F_\tilde{\mu} (\tilde{X}_{t-1}, \ldots, \tilde{X}_{t-p}) + H_\tilde{\sigma} (\tilde{X}_{t-1}, \ldots, \tilde{X}_{t-p}) \tilde{\epsilon}_t
$$

for $t = T_0 + 1, \ldots, T$ and based on the estimators $(\tilde{\mu}, \tilde{\sigma})$ and the simulated $\tilde{\epsilon}_t$ obtained in Step 1.

Step 3. For a given portfolio weight $a$, we calculate the predicted values of fund at time $T$, $F_T(a)$, with (4.10).

Step 4. We repeat Step 1–Step 3 $M$ times and save $F_T^{(1)}(a), \ldots, F_T^{(M)}(a)$. Then we calculate the proportion that the predicted values fail below the prescribed level $F$, that is,

$$
g(a) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{I} \{F_T^{(m)}(a) < F\}
$$

Step 5. Minimize $g(a)$ with respect to $a$: $a^* = \arg \min_a g(a)$.

4.1. An Illustration

In this section, we apply the above procedure to real financial data. We consider the same monthly log-returns data in Section 3.1. domestic bond (DB), domestic equity (DE), foreign bond (FB), and foreign equity (FE) are assumed to follow the following ARCH(1) model:

$$
X_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = b X_{t-1}^2
$$

respectively. Here $b > 0$ and $\epsilon_t \sim S(\alpha, \beta, 1, 0)$. Cash is virtually constant so we assume the log-return of cash as 0, permanently. Set the present Reserve Fund $F_{T_0} = 1$ and the target period is fixed to half years.

Table 2 shows the estimated optimal portfolio weights for the different prescribed level $F$. The weights of domestic and foreign bonds tend to be high when $F$ is small. Small $F$ implies that we want to avoid the loss. On the contrary, the weights of equities become higher when $F$ is larger. Large $F$ implies that we do not want to miss the chance of big gain. This result seems to be natural because bonds are lower risk (less volatile) than equities.

5. Conclusions

In this paper, we study the estimation of optimal portfolios for a Reserve Fund with an end-of-period target in two different settings. In the first setting, one assets are split into short
memory (bonds) and long memory (equity), and the optimality of the portfolio is based on maximizing the Sharpe ratio. The simulation result shows that the portfolio weight of domestic bonds is quite high. The reason is that the investment term is extremely long (100 years). Because the investment risk for the Reserve Fund is exponentially amplified year by year, the portfolio selection problem for the Reserve Fund is quite sensitive to the year-based portfolio risk. In the second setting, returns follow a conditional heteroskedasticity autoregressive nonlinear model, and we study when the distribution of the innovation vector is heavy-tailed stable. Simulation studies show that we prefer the bonds when we want to avoid the big loss in the future. The result seems to be natural because the bonds are less volatile than the equities.

**Acknowledgments**

This paper was supported by Government Pension Investment Fund (GPIF). The authors thank all the related people of GPIF, especially Dr. Takashi Yamashita.

**References**

[1] GPIF Home Page, http://www.gpif.go.jp/eng/index.html.
[2] A. Olivieri and E. Pitacco, “Solvency requirements for life annuities,” in Proceedings of the Proceedings of the Actuarial Approach for Financial Risks (AFIR ’00), Colloquium, 2000.
[3] C.-F. J. Wu, “Jackknife, bootstrap and other resampling methods in regression analysis,” The Annals of Statistics, vol. 14, no. 4, pp. 1261–1350, 1986.
[4] S. Gonçalves and L. Kilian, “Bootstrapping autoregressions with conditional heteroskedasticity of unknown form,” Journal of Econometrics, vol. 123, no. 1, pp. 89–120, 2004.
[5] P. Bühlmann, “Sieve bootstrap for time series,” Bernoulli, vol. 3, no. 2, pp. 123–148, 1997.
[6] J. Beran, Statistics for Long-Memory Processes, vol. 61 of Monographs on Statistics and Applied Probability, Chapman and Hall, New York, NY, USA, 1994.
[7] T. Shiohama, M Hallin, D. Veredas, and M. Taniguchi, “Dynamic portfolio optimization using generalized conditional heteroskedasticity factor models,” ECARES WP 2010-30, 2010.
[8] J. P. Nolan, “Numerical calculation of stable densities and distribution functions,” Communications in Statistics—Stochastic Models, vol. 13, no. 4, pp. 759–774, 1997.
[9] H. Ogata, “Empirical likelihood estimation for a class of stable processes,” Mimeo, 2009.
[10] A. Feuerverger and P. McDunnough, “On the efficiency of empirical characteristic function procedures,” Journal of the Royal Statistical Society B, vol. 43, no. 1, pp. 20–27, 1981.
[11] I. Yu, “Empirical characteristic function estimation and its applications,” Econometric Reviews, vol. 23, no. 2, pp. 93–123, 2004.
[12] M. Carrasco and J. Florens, “Efficient GMM estimation using the empirical characteristic function,” IDEI Working Papers 140, Institut d’Economie Industrielle (IDEI), Toulouse, France, 2002.
[13] A. B. Owen, “Empirical likelihood ratio confidence intervals for a single functional,” Biometrika, vol. 75, no. 2, pp. 237–249, 1988.

**Table 2:** Estimated optimal portfolio weights (Section 4).

| $F$   | DB | DE | FB | FE | Cash |
|-------|----|----|----|----|------|
| 0.90  | 0.79| 0.03| 0.07| 0.05| 0.05 |
| 0.95  | 0.49| 0.03| 0.38| 0.05| 0.05 |
| 1.00  | 0.39| 0.01| 0.54| 0.01| 0.05 |
| 1.05  | 0.17| 0.36| 0.36| 0.06| 0.05 |
| 1.10  | 0.37| 0.27| 0.26| 0.05| 0.05 |
[14] A. B. Owen, *Empirical Likelihood*, Chapman-Hall/CRC, Boca Raton, Fla, USA, 2001.
[15] J. Qin and J. Lawless, “Empirical likelihood and general estimating equations,” *The Annals of Statistics*, vol. 22, no. 1, pp. 300–325, 1994.
Research Article

Least Squares Estimators for Unit Root Processes with Locally Stationary Disturbance

Junichi Hirukawa and Mako Sadakata

Faculty of Science, Niigata University, 8050, Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan

Correspondence should be addressed to Junichi Hirukawa, hirukawa@math.sc.niigata-u.ac.jp

Received 4 November 2011; Accepted 26 December 2011

Academic Editor: Hiroshi Shiraishi

Copyright © 2012 J. Hirukawa and M. Sadakata. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The random walk is used as a model expressing equitableness and the effectiveness of various finance phenomena. Random walk is included in unit root process which is a class of nonstationary processes. Due to its nonstationarity, the least squares estimator (LSE) of random walk does not satisfy asymptotic normality. However, it is well known that the sequence of partial sum processes of random walk weakly converges to standard Brownian motion. This result is so-called functional central limit theorem (FCLT). We can derive the limiting distribution of LSE of unit root process from the FCLT result. The FCLT result has been extended to unit root process with locally stationary process (LSP) innovation. This model includes different two types of nonstationarity. Since the LSP innovation has time-varying spectral structure, it is suitable for describing the empirical financial time series data. Here we will derive the limiting distributions of LSE of unit root, near unit root and general integrated processes with LSP innovation. Testing problem between unit root and near unit root will be also discussed. Furthermore, we will suggest two kind of extensions for LSE, which include various famous estimators as special cases.

1. Introduction

Since the random walk is a martingale sequence, the best predictor of the next term becomes the value of this term. In this sense, the random walk is used as a model expressing equitableness and the effectiveness of various finance phenomena in economics. Furthermore, because the random walk is a unit root process, taking the difference of the random walk, we can recover the independent sequence. However, the information of the original sequence will be lost by taking the difference when it does not include a unit root. Therefore, the testing of the existence of unit root in the original sequence becomes important.
In this section, we review the fundamental asymptotic results for unit root processes. Let \( \{ \varepsilon_j \} \) be i.i.d. \((0, \sigma^2)\) random variables, where \( \sigma^2 > 0 \), and define the partial sum

\[
    r_j = r_{j-1} + \varepsilon_j \quad (r_0 = 0)
    \]

\[
    = \sum_{i=1}^{j} \varepsilon_i, \quad (j = 1, \ldots, T),
    \]

which is the so-called random walk process. Random walk corresponds to the first-order autoregressive (AR(1)) model with unit coefficient. Therefore, random walk is included in unit root \((I(1))\) processes which is a class of nonstationary processes. Let \( C = C[0,1] \) be the space of all real-valued continuous functions defined on \([0,1]\). For random walk process, we construct the sequence of the processes of the partial sum \( \{ \tilde{R}_T \} \) in \( C \) as

\[
    \tilde{R}_T(t) = \frac{1}{\sigma \sqrt{T}} r_j + T \left( t - \frac{j}{T} \right) \frac{1}{\sigma \sqrt{T}} \varepsilon_j, \quad \left( \frac{j-1}{T} \leq t \leq \frac{j}{T} \right).
    \]

It is well known that the partial sum process \( \{ \tilde{R}_T \} \) converge weakly to a standard Brownian motion on \([0,1]\), namely,

\[
    \mathcal{L}(\tilde{R}_T) \rightarrow \mathcal{L}(W) \quad \text{as } T \rightarrow \infty,
    \]

where \( \mathcal{L}(\cdot) \) denotes the distribution law of the corresponding random elements. This result is the so-called functional central limit theorem (FCLT) (see Billingsley [1]).

The FCLT result can be extended to the unit root process where the innovation is general linear process. We consider a sequence \( \{ \tilde{R}_T \} \) of a stochastic process in \( C \) defined by

\[
    \tilde{R}_T(t) = \frac{1}{\sqrt{T}} \tilde{r}_j + T \left( t - \frac{j}{T} \right) \frac{1}{\sqrt{T}} u_j, \quad \left( \frac{j-1}{T} \leq t \leq \frac{j}{T} \right),
    \]

where \( \tilde{r}_j = \sum_{i=1}^{j} u_i \) and \( \{ u_j \} \) is assumed to be generated by

\[
    u_j = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{j-i}, \quad \alpha_0 = 1.
    \]

Here, \( \{ \varepsilon_j \} \) is a sequence of i.i.d. \((0, \sigma^2)\) random variables, and \( \{ \alpha_i \} \) is a sequence of constants which satisfies \( \sum_{i=0}^{\infty} \alpha_i < \infty \); therefore, \( \{ u_j \} \) becomes stationary process. Using the Beveridge and Nelson [2] decomposition, it holds (see, e.g., Tanaka [3])

\[
    \mathcal{L}(\tilde{R}_T) \rightarrow \mathcal{L}(\alpha W), \quad \alpha = \sum_{i=0}^{\infty} \alpha_i.
    \]

The asymptotic property of LSE for stationary autoregressive models has been well established (see, e.g., Hannan [4]). On the other hand, due to its nonstationarity, the
LSE of random walk does not satisfy asymptotic normality. However, we can derive the limiting distribution of LSE of unit root process from the FCLT result. For more detailed understanding about unit root process with i.i.d. or stationary innovation, refer to, for example, Billingsley [1] and Tanaka [3].

In the above case, the \{u_j\}'s are stationary and hence, have constant variance, while covariances depend on only time differences. This is referred to as the homogeneous case, which is too restrictive to interpret empirical data, for example, empirical financial data. Recently, an important class of nonstationary processes have been proposed by Dahlhaus (see, e.g., Dahlhaus [5, 6]), called locally stationary processes. In this paper, we alternatively adopt locally stationary innovation process, which has smoothly changing variance. Since the LSP innovation has time-varying spectral structure, it is suitable for describing the empirical financial time series data.

This paper is organized as follows. In the appendix, we review the extension of the FCLT results to the cases that the innovations are locally stationary process. Namely, we explain the FCLT for unit root, near unit root, and general integrated processes with LSP innovations. In Section 2, we obtain the asymptotic distribution of the least squares estimator for each case of the appendix. In Section 3, we also consider the testing problem for unit root with LSP innovation. Finally, in Section 4, we discuss the extensions of LSE, which include various famous estimators as special cases.

2. The Property of Least Squares Estimator

In this section, we investigate the asymptotic properties of least squares estimators for unit root, near unit root, and \(I(d)\) processes with locally stationary process innovations. Testing problem for unit root is also discussed. For the notations which are not defined in this section, refer to the appendix.

2.1. Least Squares Estimator for Unit Root Process

Here, we consider the following statistics:

\[
\hat{\rho} = \frac{\sum_{j=2}^{T} x_{j-1,T} x_{j,T}}{\sum_{j=2}^{T} x_{j-1,T}^2},
\]

obtained from model (A.3), which can be regarded as the least squares estimator (LSE) of autoregressive coefficient in the first-order autoregressive (AR(1)) model \(x_{j,T} = \rho x_{j-1,T} + u_{j,T}\).

Define

\[
U_{1,T} = \frac{1}{T \sigma^2} \sum_{j=2}^{T} x_{j-1,T} (x_{j,T} - x_{j-1,T})
\]

\[
= \frac{1}{2} X_T(1)^2 - \frac{1}{2} X(0)^2 - \frac{1}{2T \sigma^2} \sum_{j=1}^{T} u_{j,T}^2 - \frac{X(0) u_{1,T}}{\sqrt{T} \sigma},
\]

\[
V_{1,T} = \frac{1}{T \sigma^2} \sum_{j=2}^{T} x_{j-1,T}^2 = \frac{1}{T} \sum_{j=1}^{T} X_T \left( \frac{j}{T} \right)^2 - \frac{1}{T} X_T(1)^2,
\]

obtained from model (A.3), which can be regarded as the least squares estimator (LSE) of autoregressive coefficient in the first-order autoregressive (AR(1)) model \(x_{j,T} = \rho x_{j-1,T} + u_{j,T}\).
then we have

\[ S_{1,T} = T(\hat{\rho} - 1) = \frac{U_{1,T}}{V_{1,T}}. \]  \quad (2.3)

Let us define a continuous function \( H_1(x) = (H_{11}(x), H_{12}(x)) \) for \( x \in C \), where

\[
H_{11}(x) = \frac{1}{2} \left\{ x(1)^2 - x(0)^2 - \int_0^1 \sum_{l=0}^{\infty} a_l(v)^2 dv \right\}, \quad H_{12}(x) = \int_0^1 x(v)^2 dv. \quad (2.4)
\]

It is easy to check

\[
U_{1,T} = H_{11}(X_T) + o_p(1), \quad V_{1,T} = H_{12}(X_T) + o_p(1). \quad (2.5)
\]

Therefore, the continuous mapping theorem (CMT) leads to \( \mathcal{L}(U_{1,T}, V_{1,T}) \to \mathcal{L}(H_1(X)) \) and

\[
\mathcal{L}(S_{1,T}) = \mathcal{L}(T(\hat{\rho} - 1))
\]

\[ \to \mathcal{L} \left( \frac{H_{11}(X)}{H_{12}(X)} \right) = \mathcal{L} \left( \frac{(1/2) \left\{ X(1)^2 - X(0)^2 - \int_0^1 \sum_{l=0}^{\infty} a_l(v)^2 dv \right\}}{\int_0^1 X(v)^2 dv} \right) \quad (2.6) \]

\[ = \mathcal{L} \left( \frac{\int_0^1 X(v) dX(v) + (1/2) \int_0^1 \left[ \sum_{l=0}^{\infty} a_l(v) \right]^2 - \sum_{l=0}^{\infty} a_l(v)^2 \right]}{\int_0^1 X(v)^2 dv} \right). \]

### 2.2. Least Squares Estimator for Near Unit Root Process

We next consider the least squares estimator \( \hat{\rho}_T \) for model (A.11) in the case that \( \beta(t) \equiv \beta \) is a constant on \([0, 1]\), namely,

\[ y_{j,T} = \rho_T y_{j-1,T} + u_{j,T}, \quad (j = 1, \ldots, T), \]  \quad (2.7)

with \( \rho_T = 1 - \beta/T \). Then, we have

\[
\hat{\rho}_T = 1 - \frac{\hat{\beta}}{T} = \frac{\sum_{j=2}^{T} y_{j-1,T} y_{j,T}}{\sum_{j=2}^{T} y_{j-1,T}^2}, \quad S_{2,T} = T(\hat{\rho}_T - 1) = -\hat{\beta} = \frac{U_{2,T}}{V_{2,T}} \quad (2.8)
\]
where

\[ U_{2,T} = \frac{1}{T\sigma^2} \sum_{j=2}^{T} (y_{j-1,T} - y_{j-1,T})^2 \]

\[ = \frac{1}{2} Y_T^2(1)^2 - \frac{1}{2} Y(0)^2 - \frac{1}{2T\sigma^2} \sum_{j=1}^{T} \left( u_{j,T} - \frac{\beta}{T} y_{j-1,T} \right)^2 - \frac{1}{\sqrt{T}\sigma} Y(0) \left( u_{1,T} - \frac{\beta}{T} y_{0,T} \right), \quad (2.9) \]

\[ V_{2,T} = \frac{1}{T^2\sigma^2} \sum_{j=2}^{T} y_{j-1,T}^2 = \frac{1}{T} \sum_{j=1}^{T} \left( \frac{j}{T} \right)^2 - \frac{1}{T} Y_T(1)^2. \]

Let us define a continuous function \( H_2(x) = (H_{21}(x), H_{22}(x)) \) for \( x \in C \), where

\[ H_{21}(x) = \frac{1}{2} \left\{ x(1)^2 - x(0)^2 - \int_0^1 \sum_{l=0}^{\infty} a_l(v)^2 dv \right\}, \quad H_{22}(x) = \int_0^1 x(v)^2 dv. \quad (2.10) \]

It is easy to check

\[ U_{2,T} = H_{21}(Y_T) + o_P(1), \quad V_{2,T} = H_{22}(Y_T) + o_P(1). \quad (2.11) \]

Therefore, the CMT leads to \( \mathcal{L}(U_{2,T}, V_{2,T}) \to \mathcal{L}(H_2(Y)) \) and

\[ \mathcal{L}(S_{2,T}) = \mathcal{L}(T(\hat{\rho} - 1)) = \mathcal{L}(-\hat{\beta}) \]

\[ \to \mathcal{L} \left( \frac{H_{21}(Y)}{H_{22}(Y)} \right) = \mathcal{L} \left( \frac{(1/2) \left\{ Y(1)^2 - Y(0)^2 - \int_0^1 \sum_{l=0}^{\infty} a_l(v)^2 dv \right\}}{\int_0^1 Y(v)^2 dv} \right) \]

\[ = \mathcal{L} \left( \int_0^1 Y(v) dY(v) + (1/2) \int_0^1 \left\{ \left[ \sum_{l=0}^{\infty} a_l(v) \right]^2 - \sum_{l=0}^{\infty} a_l(v)^2 \right\} dv \right). \quad (2.12) \]

### 2.3. Least Squares Estimator for \( I(d) \) Process

Furthermore, we consider the least squares estimator

\[ \hat{\rho}^{(d)} = \frac{\sum_{j=2}^{T} x_j^{(d)} x_j^{(-d)}}{\sum_{j=2}^{T} x_j^{(d)}}, \quad S_{3,T} = T \left( \hat{\rho}^{(d)} - 1 \right) = \frac{U_{3,T}}{V_{3,T}}, \quad (2.13) \]
obtained from model \( x_{j,T}^{(d)} = \rho x_{j-1,T}^{(d)} + x_{j,T}^{(d-1)} \), where

\[
U_{3,T} = \frac{1}{T^{2d-1}} \sigma^2 \sum_{j=2}^{T} \left( x_{j-1,T}^{(d)} - x_{j,T}^{(d-1)} \right)
\]

\[
= \frac{1}{2} X_T^{(d)} (1)^2 - \frac{1}{2T^2} \sum_{j=1}^{T} \left( X_T^{(d-1)} \left( \frac{j}{T} \right) \right)^2 - \frac{1}{T^2} X_T^{(d)} (0) X_T^{(d-1)} \left( \frac{1}{T} \right)
\]

\[
V_{3,T} = \frac{1}{T^{2d-1}} \sigma^2 \sum_{j=2}^{T} \left( x_{j-1,T}^{(d)} \right)^2 = \frac{1}{T} \sum_{j=1}^{T} \left\{ X_T^{(d)} \left( \frac{j}{T} \right) \right\}^2 - \frac{1}{T} \left\{ X_T^{(d)} (1) \right\}^2.
\]

(2.14)

Let us define a continuous function \( H_3(x) = (H_{31}(x), H_{32}(x)) \) for \( x \in \mathcal{C} \), where

\[
H_{31}(x) = \frac{1}{2} x(1)^2, \quad H_{32}(x) = \int_0^1 x(\nu)^2 d\nu.
\]

(2.15)

It is easy to check

\[
U_{3,T} = H_{31} \left( X_T^{(d)} \right) + o_P(1), \quad V_{3,T} = H_{32} \left( X_T^{(d)} \right) + o_P(1).
\]

(2.16)

Therefore, the CMT leads to \( \mathcal{L}(U_{3,T}, V_{3,T}) \to \mathcal{L}(H_3(X^{(d-1)})) \) and

\[
\mathcal{L}(S_{3,T}) = \mathcal{L} \left( \hat{\rho}^{(d)} - 1 \right)
\]

\[
\to \mathcal{L} \left( \frac{H_{31}(X^{(d-1)})}{H_{32}(X^{(d-1)})} \right)
\]

\[
= \mathcal{L} \left( \frac{(1/2) \left\{ X^{(d-1)} (1) \right\}^2}{\int_0^1 \left\{ X^{(d-1)} (\nu) \right\}^2 d\nu} \right)
\]

\[
= \mathcal{L} \left( \int_0^1 X^{(d-1)} (\nu) dX^{(d-1)} (\nu) \right) \left( \int_0^1 \left\{ X^{(d-1)} (\nu) \right\}^2 d\nu \right).
\]

(2.17)

The equality above is due to \( (d - 1) \)-times differentiability of \( X^{(d-1)} \).

### 3. Testing for Unit Root

In the analysis of empirical financial data, the existence of the unit root is an important problem. However, as we see in the previous section, the asymptotic results between unit root and near unit root processes are quite different (the drift term appeared in the limiting
process of near unit root. Therefore, we consider the following testing problem against the local alternative hypothesis:

\[ H_0 : \rho = 1 \quad H_1 : \rho = 1 - \frac{\beta}{T}. \] (3.1)

We should assume that \( \sigma^2 \) is a unit to identify the models. Let the statistics \( S_{1,T} \) be constructed in (2.3). Recall that, as \( T \to \infty \), under \( H_0 \),

\[
\mathcal{L}(S_{1,T}) \to \mathcal{L}
\left( 
\frac{\int_0^1 X(v)dX(v) + (1/2) \int_0^1 \left[ \left( \sum_{l=0}^\infty a_l(v) \right)^2 - \sum_{l=0}^\infty a_l(v)^2 \right] dv}{\int_0^1 X(v)^2dv}
\right)
\]

\[ = \mathcal{L}
\left( 
\frac{U}{V} + \frac{1}{2} \int_0^1 \left[ \left( \sum_{l=0}^\infty a_l(v) \right)^2 - \sum_{l=0}^\infty a_l(v)^2 \right] dv
\right),
\]

where

\[ U = \int_0^1 X(v)dX(v), \quad V = \int_0^1 X(v)^2dv. \] (3.3)

Since \( \left( \sum_{l=0}^\infty a_l(v) \right)^2 \), \( \sum_{l=0}^\infty a_l(v)^2 \) are unknown, we construct a test statistic

\[ Z_\rho = T(\hat{\rho} - 1) + \frac{(1/T) \sum_{j=1}^T \tilde{u}_{j,T}^2 - (1/T) \sum_{j=1}^T \tilde{f}(t/T,0)}{2(1/T^2) \sum_{j=2}^T \tilde{x}_j^2}, \] (3.4)

where \( \tilde{u}_{j,T} = x_{j,T} - x_{j-1,T} \). A nonparametric time-varying spectral density estimator \( \tilde{f}(u, \lambda) \) is given by

\[
\tilde{f}(u, \lambda_l) = M \int K(M(\lambda_l - \mu_k))I_N(u, \mu_k)d\mu_k
\]

\[ \approx \frac{2\pi M}{T} \sum_{k=-T/4\pi M}^{T/4\pi M} K(M(\lambda_l - \mu_k))I_N(u, \mu_k), \] (3.5)

where \( \lambda_l = (2\pi/T)l - \pi, l = 1, \ldots, T - 1 \) and \( \mu_k = (2\pi/T)k - \pi, k = 1, \ldots, T - 1 \). Here, \( I_N(u, \lambda) \) is the local periodogram around time \( u \) given by

\[ I_N(u, \lambda) = \left| \frac{1}{2\pi N} \sum_{j=1}^N h\left( \frac{s}{N} \right) \tilde{u}_{[uT]-N/2+s,T} e^{-i\lambda s} \right|^2, \] (3.6)
where \([\cdot]\) denotes Gauss symbol and, for real number \(a\), \([a]\) is the greatest integer that is less than or equal to \(a\). Furthermore, we employ the following kernel functions and the orders of bandwidth for smoothing in time and frequency domain, respectively,

\[
K(x) = 6\left(\frac{1}{4} - x^2\right), \quad x \in \left[-\frac{1}{2}, \frac{1}{2}\right], \quad h(x) = \{6x(1-x)\}^{1/2}, \quad x \in [0, 1],
\]

\[
M = T^{1/6}, \quad N = T^{5/6},
\]

which are optimal in the sense that they minimize the mean squared error of nonparametric estimator (see Dahlhaus [6]); however, we simply multiply the orders of bandwidth by the constants equal to one. Then, it can be established that, under \(H_0\),

\[
\mathcal{L}(Z_{\rho}) \rightarrow \mathcal{L}\left(\frac{U}{V}\right).
\]

We now have to deal with statistics for which numerical integration must be elaborated. Let \(R\) be such a statistic, which takes the form \(R = U/V\). Using Imhof’s [7] formula gives us distribution function of \(R\),

\[
F_R(x) = P(R \leq x) = P(xV - U \geq 0) = \frac{1}{2} + \frac{1}{\pi} \int_0^1 \frac{1}{s} \text{Im}\{\phi(s;x)\} ds,
\]

where \(\phi(s;x)\) is the characteristic function of \(xV - U\), namely,

\[
\phi(-is;x) = E[\exp\{s(xV - U)\}] = E\left[\exp\left\{s \left( x \int_0^1 X(\nu)^2 d\nu - \int_0^1 X(\nu)dX(\nu) \right) \right\}\right].
\]

However, so far we do not have the explicit form of the distribution function of the estimator. Therefore, we cannot perform a numerical experiment except for the clear simple cases. It includes the complicated problem in the differential equation and requires one further paper for solution.

4. Extensions of LSE

In this section, we consider the extensions of LSE \(\hat{\rho}_T\) for near random walk model \(y_{j,T} = \rho_T y_{j-1,T} + u_{j,T}, \rho_T = 1 - \beta/T\).
4.1. Ochi Estimator

Ochi [8] proposed the class of estimators of the following form, which are the extensions of LSE for autoregressive coefficient:

\[
\tilde{\rho}_T^{(\theta_1, \theta_2)} = 1 - \frac{\tilde{\beta}^{(\theta_1, \theta_2)}}{T} = \frac{\sum_{j=2}^{T} y_{j-1,T} y_{j,T}}{\sum_{j=2}^{T-1} y_{j,T}^2 + \theta_1 y_{1,T}^2 + \theta_2 y_{T,T}^2}, \quad \theta_1, \theta_2 \geq 0, \tag{4.1}
\]

\[
S_{4,T} = T \left( \tilde{\rho}_T^{(\theta_1, \theta_2)} - 1 \right) = -\tilde{\beta}^{(\theta_1, \theta_2)} = \frac{U_{4,T}}{V_{4,T}}.
\]

where

\[
U_{4,T} = \frac{1}{T \sigma^2} \left\{ \sum_{j=2}^{T} y_{j-1,T} y_{j,T} - \frac{T-1}{T} \sum_{j=2}^{T-1} y_{j,T}^2 - \theta_1 y_{1,T}^2 - \theta_2 y_{T,T}^2 \right\}
\]

\[
= \left\{ \frac{1}{2} (1-2\theta_1) + \frac{\beta}{T} (2\theta_1 - 1) + \frac{\beta^2}{T^2} (1 - \theta_1) \right\} Y(0)^2 + \frac{1}{2} (1 - \theta_2) Y_T(1)^2 - \frac{1}{2 \sigma^2} \sum_{j=1}^{T} \left( u_{j,T} - \frac{\beta}{T} y_{j-1,T} \right)^2
\]

\[
+ \frac{1}{\sqrt{T}} \left\{ 1 - 2\theta_1 + 2 \frac{\beta}{T} (\theta_1 - 1) \right\} u_{1,T} Y(0) + \frac{1}{T \sigma^2} (1 - \theta_1) u_{1,T}^2,
\]

\[
V_{4,T} = \frac{1}{T^2 \sigma^2} \left\{ \sum_{j=2}^{T} y_{j,T}^2 + \theta_1 y_{1,T}^2 + \theta_2 y_{T,T}^2 \right\}
\]

\[
= \frac{1}{T} \sum_{j=1}^{T} Y_T \left( \frac{j}{T} \right)^2 + \left( \theta_1 - 1 \right) \frac{1}{T} Y_T \left( \frac{1}{T} \right)^2 + \left( \theta_2 - 1 \right) \frac{1}{T} Y_T(1)^2.
\]

This class of estimators includes LSE \( \hat{\rho}_T^{(1,0)} \), Daniels’s estimator \( \hat{\rho}_T^{(1/2,1/2)} \), and Yule-Walker estimator \( \hat{\rho}_T^{(1,1)} \) as the special cases.

Define for \( x \in C \), \( H_4(x) = (H_{41}(x), H_{42}(x)) \),

\[
H_{41}(x) = \frac{1}{2} \left\{ (1 - 2\theta_1)x(0)^2 + (1 - 2\theta_2)x(1)^2 - \int_0^1 \sum_{i=0}^{\infty} \alpha_i(v)^2 dv \right\}, \tag{4.3}
\]

\[
H_{42}(x) = \int_0^1 x(v)^2 dv,
\]

then we see that \( H_4(x) \) is continuous and

\[
U_{4,T} = H_{41}(Y_T) + o_p(1), \quad V_{4,T} = H_{42}(Y_T) + o_p(1). \tag{4.4}
\]
From the CMT, we obtain \( \mathcal{L}(U_{4,T}, V_{4,T}) \rightarrow \mathcal{L}(H_4(Y)) \), and therefore,

\[
\mathcal{L}(S_{4,T}) = \mathcal{L}\left( T\left( \hat{\rho}_T^{(\theta_0, \theta_1)} - 1 \right) \right) = \mathcal{L}\left( -\hat{\beta}_{(\theta_1, \theta_2)} \right) \rightarrow \mathcal{L}\left( \frac{H_{41}(Y)}{H_{42}(Y)} \right),
\]

(4.5)

where

\[
H_{41}(Y) = \frac{1}{2} \left\{ (1 - 2\theta_1)Y(0)^2 + (1 - 2\theta_2)Y(1)^2 - \int_0^1 \sum_{l=0}^\infty \alpha_l(v)^2 dv \right\}
\]

\[
= (1 - 2\theta_2) \int_0^1 Y(v)dY(v) + (1 - \theta_1 - \theta_2)Y(0)^2
\]

\[
+ \frac{1}{2} \int_0^1 \left[ (1 - 2\theta_2) \left\{ \sum_{l=0}^\infty \alpha_l(v) \right\}^2 - \sum_{l=0}^\infty \alpha_l(v)^2 \right] dv,
\]

(4.6)

\[
H_{42}(Y) = \int_0^1 Y(v)^2 dv.
\]

### 4.2. Another Extension of LSE

Next, we suggest another class of estimators which are also the extensions of LSE. Define for \( \theta(u) \in \mathcal{C} \) with continuous derivative \( \theta'(u) = (\partial/\partial u)\theta(u) \),

\[
\hat{\rho}_T^\theta = 1 - \frac{\tilde{\beta}_0}{T} = \frac{\sum_{j=2}^T \theta((j-1)/T)y_{j-1,T}y_{j,T}}{\sum_{j=2}^T \theta((j-1)/T)y_{j-1,T}^2}, \quad S_{5,T} = T\left( \hat{\rho}_T^\theta - 1 \right) = -\hat{\beta}_0 = \frac{U_{5,T}}{V_{5,T}},
\]

(4.7)

where

\[
U_{5,T} = \frac{1}{T\sigma^2} \sum_{j=2}^T \theta\left( \frac{j-1}{T} \right) y_{j-1,T}y_{j,T} (y_{j,T} - y_{j-1,T})
\]

\[
= -\frac{1}{2} \left\{ \theta\left( \frac{j}{T} \right) - \theta\left( \frac{j-1}{T} \right) \right\} Y_T\left( \frac{j}{T} \right)^2 + \frac{1}{2} \theta(1)Y_T(1)^2 - \frac{1}{2}\theta(0)Y(0)^2
\]

\[
- \frac{1}{2} \frac{1}{2T\sigma^2} \sum_{j=1}^T \theta\left( \frac{j}{T} \right) \left( u_{j,T} - \frac{\beta}{T} y_{j-1,T} \right)^2 + \frac{1}{2T\sigma^2} \theta\left( \frac{1}{T} \right) \left( u_{1,T} - \frac{\beta}{T} y_{0,T} \right)^2
\]

\[
+ \frac{1}{2T\sigma^2} \theta(0) \left\{ u_{1,T}(u_{1,T} + 2y_{0,T}) - \frac{2\beta}{T^2} y_{0,T}(y_{0,T} + u_{1,T}) + \frac{\beta^2}{T^2} y_{0,T}^2 \right\},
\]

(4.8)

\[
V_{5,T} = \frac{1}{T^2\sigma^2} \sum_{j=2}^T \theta\left( \frac{j-1}{T} \right) y_{j-1,T}^2 - \frac{1}{T} \sum_{j=1}^T \theta\left( \frac{j}{T} \right) Y_T\left( \frac{j}{T} \right)^2 - \frac{1}{T}\theta(1)Y_T(1)^2.
\]

If we take the taper function as \( \theta(u) \), this estimator corresponds to the local LSE.
Define for \( x \in C, H_5(x) = (H_{51}(x), H_{52}(x)) \),

\[
H_{51}(x) = -\frac{1}{2} \left\{ \int_0^1 \theta'(v)X(v)^2dv - \theta(1)X(1)^2 + \theta(0)X(0)^2 \right\}
- \frac{1}{2} \left\{ \int_0^1 \theta(v) \sum_{i=0}^\infty a_i(v)^2dv \right\},
\]

\[
H_{52}(x) = \int_0^1 \theta(v)X(v)^2dv,
\]

where \( \theta'(u) = (\partial/\partial u)\theta(u) \), then we see that \( H_5(x) \) is continuous and

\[
U_{5,T} = H_{51}(Y_T) + \sigma_P(1), \quad V_{5,T} = H_{52}(Y_T) + \sigma_P(1).
\]  

From the CMT, we obtain \( \mathcal{L}(U_{5,T}, V_{5,T}) \to \mathcal{L}(H_5(X)) \), and therefore,

\[
\mathcal{L}(S_{5,T}) = \mathcal{L}(T(\tilde{\beta}^\theta_t - 1)) = \mathcal{L}(-\tilde{\beta}^\theta) \to \mathcal{L}(\frac{H_{51}(Y)}{H_{52}(Y)}) \equiv \mathcal{L}(\gamma^\theta),
\]

where

\[
H_{51}(Y) = -\frac{1}{2} \left\{ \int_0^1 \theta'(v)Y(v)^2dv - \theta(1)Y(1)^2 + \theta(0)Y(0)^2 \right\}
- \frac{1}{2} \left\{ \int_0^1 \theta(v) \sum_{i=0}^\infty a_i(v)^2dv \right\},
\]

\[
H_{52}(Y) = \int_0^1 \theta(v)Y(v)^2dv.
\]

The integration by part leads to

\[
\gamma^\theta = \frac{(1/2) \left\{ \int_0^1 \theta(v)dY^{(1)}(v) - \int_0^1 \theta(v) \sum_{i=0}^\infty a_i(v)^2dv \right\}}{\int_0^1 \theta(v)Y(v)^2dv},
\]

with \( Y^{(1)}(t) = Y(t)^2 \). Hence, using Ito's formula,

\[
dY^{(1)}(t) = d\left\{ Y(t)^2 \right\} = 2Y(t)dY(t) + \left\{ \sum_{i=0}^\infty a_i(t) \right\}^2dt,
\]
we have
\[ Y^0 = \left( \int_0^1 \theta(v)Y(v)dY(v) + \frac{1}{2} \int_0^1 \theta(v) \left[ \left( \sum_{l=0}^{\infty} \alpha_l(v) \right)^2 - \sum_{l=0}^{\infty} \alpha_l(v)^2 \right] dv \right) \int_0^1 \theta(v)Y(v)^2 dv. \] (4.15)

**Appendices**

In this appendix, we review the extensions of functional central limit theorem to the cases that innovations are locally stationary processes, which are used for the main results of this paper.

**A. FCLT for Locally Stationary Processes**

Hirukawa and Sadakata [9] extended the FCLT results to the unit root processes which have locally stationary process innovations. Namely, they derived the FCLT for unit root, near unit root, and general integrated processes with LSP innovations. In this section, we briefly review these results which are applied in previous sections.

**A.1. Unit Root Process with Locally Stationary Disturbance**

First, we introduce locally stationary process innovation. Let \( \{ u_{j,T} \} \) be generated by the following time-varying MA (\( \infty \)) model:

\[ u_{j,T} = \sum_{l=0}^{\infty} \alpha_l \left( \frac{j}{T} \right) \varepsilon_{j-1} := \sum_{l=0}^{\infty} \alpha_l \left( \frac{j}{T} \right) L^l \varepsilon_j = a \left( \frac{j}{T}, L \right) \varepsilon_j, \] (A.1)

where \( L \) is the lag-operator which acts as \( L \varepsilon_j = \varepsilon_{j-1} \) and \( a(u, L) = \sum_{l=0}^{\infty} \alpha_l(u) L^l \), and time-varying MA coefficients satisfy

\[ \sum_{l=0}^{\infty} \sup_{0 \leq u \leq 1} | \alpha_l(u) | < \infty, \quad \sum_{l=0}^{\infty} \sup_{0 \leq u \leq 1} \left| \frac{\partial}{\partial u} \alpha_l(u) \right| < \infty. \] (A.2)

Then, these \( \{ u_{j,T} \} \)'s become locally stationary processes (see Dahlhaus [5], Hirukawa and Taniguchi [10]). Using this innovation process, define the partial sum \( \{ x_{j,T} \} \) as

\[ x_{j,T} = x_{j-1,T} + u_{j,T} = x_{0,T} + \sum_{i=1}^{j} u_{i,T}, \] (A.3)

where \( x_{0,T} = \sigma \sqrt{T} X(0), X(0) \sim N(\gamma X, \delta X^2) \) and is independent of \( \{ \varepsilon_j \} \).

We consider a sequence \( \{ X_T \} \) of partial sum stochastic processes in \( C \) defined by

\[ X_T(t) = \frac{1}{\sigma \sqrt{T}} x_{j,T} + T \left( t - \frac{j}{T} \right) \frac{1}{\sigma \sqrt{T}} u_{j,T}, \quad \left( \frac{j-1}{T} \leq t \leq \frac{j}{T} \right). \] (A.4)
Now, we define on $\mathbb{R} \times \mathbb{C}$

$$h^{(1)}_t(x, y) = x + \alpha(t, 1)y(t) - \int_0^t \alpha'(v, 1)y(v)dv,$$

$$\alpha(t, 1) = \sum_{i=0}^{\infty} \alpha_i(t), \quad \alpha'(t, 1) = \frac{\partial}{\partial t} \alpha(t, 1) = \sum_{i=0}^{\infty} \frac{\partial}{\partial t} \alpha_i(t).$$

(A.5)

Then, we can obtain

$$\mathcal{L}(X_T) \rightarrow \mathcal{L}\left\{h^{(1)}(X(0), W) \right\} \equiv \mathcal{L}(X).$$

(A.6)

The integration by parts leads to

$$X(t) = X(0) + \alpha(t, 1)W(t) - \int_0^t \alpha'(v, 1)W(v) dv
\begin{align*}
&= X(0) + \int_0^t \alpha(v, 1) dW(v),
&dX(t) = \alpha(t, 1) dW(t).
\end{align*}$$

(A.7)

Note that the time-varying MA ($\infty$) process $u_{j,T}$ in (A.1) has the spectral representation

$$u_{j,T} = \int_{-\pi}^{\pi} A\left(\frac{j}{T}, \lambda \right)e^{ij\lambda}d\xi(\lambda),$$

(A.8)

where $\xi(\lambda)$ is the spectral measure of i.i.d. process $\{\varepsilon_j\}$ which satisfies $\varepsilon_j = \int_{-\pi}^{\pi} e^{ij\lambda}d\xi(\lambda)$, and the transfer function $A(t, \lambda)$ is given by

$$A(t, \lambda) = \sum_{i=0}^{\infty} \alpha_i(t)e^{-i\lambda}, \quad A(t, 0) = \sum_{i=0}^{\infty} \alpha_i(t) = \alpha(t, 1).$$

(A.9)

Therefore, stochastic differential in (A.7) can be written as

$$dX(t) = A(t, 0)dW(t).$$

(A.10)

A.2. Near Unit Root Process with Locally Stationary Disturbance

In this section, we consider the following near unit root process $\{y_{j,T}\}$ with locally stationary disturbance:

$$y_{j,T} = \rho_{j,T}y_{j-1,T} + u_{j,T}, \quad (j = 1, \ldots, T)$$

$$= \prod_{i=1}^{j} \rho_{i,T}y_{0,T} + \sum_{i=1}^{j} \left( \prod_{k=i+1}^{j} \rho_{k,T} \right)u_{i,T},$$

(A.11)
where \( \{ u_{j,T} \} \) is generated from the time-varying MA (\( \infty \)) model in (A.1), \( \rho_{j,T} = 1 - (1/T) \beta(j/T) \), \( \beta(t) \in \mathbb{C}[0,1] \), \( y_{0,T} = \sqrt{T} \sigma Y(0) \), and \( Y(0) \sim \mathcal{N}(\gamma_Y, \delta_Y) \) is independent of \( \{ \varepsilon_j \} \) and \( X(0) \). Then, we define a sequence \( \{ Y_T \} \) of partial sum processes in \( \mathcal{C} \) as

\[
Y_T(t) = \frac{1}{\sigma \sqrt{T}} y_{j,T} + T \left( t - \frac{j}{T} \right) \frac{y_{j,T} - y_{j-1,T}}{\sigma \sqrt{T}}, \quad \left( \frac{j-1}{T} \leq t \leq \frac{j}{T} \right). \tag{A.12}
\]

Define on \( \mathbb{R}^2 \times \mathcal{C} \)

\[
h^{(2)}_t(x, y, z) = e^{-\int_0^t \beta(v) dv} (y - x) - \int_0^t \beta(v) e^{-\int_v^t \beta(s) ds} z(v) dv + z(t). \tag{A.13}
\]

Then, we can obtain

\[
\mathcal{L}(Y_T) \longrightarrow \mathcal{L}\left\{ h^{(2)}(X(0), Y(0), X) \right\} \equiv \mathcal{L}(Y). \tag{A.14}
\]

The integration by parts and Ito’s formula lead to

\[
Y(t) = e^{-\int_0^t \beta(s) ds} \left( Y(0) - X(0) - \int_0^t \beta(v) e^{-\int_0^v \beta(s) ds} X(v) dv \right) + X(t)
\]

\[
= e^{-\int_0^t \beta(s) ds} \left( Y(0) + \int_0^t e^{-\int_0^v \beta(s) ds} \mu_1 dX(v) \right)
\]

\[
= e^{-\int_0^t \beta(s) ds} \left( Y(0) + \int_0^t e^{-\int_0^v \beta(s) ds} \mu_1 dW(v) \right), \tag{A.15}
\]

\[
dY(t) = -\beta(t) Y(t) + \alpha(t, 1) dW(t)
\]

\[
= -\beta(t) Y(t) + A(t, 0) dW(t)
\]

\[
= -\beta(t) Y(t) + dX(t).
\]

### A.3. \( I(d) \) Process with Locally Stationary Disturbance

Let \( I(d) \) process \( \{ x_{j,T}^{(d)} \} \) be generated by

\[
(1 - L)^d x_{j,T}^{(d)} = u_{j,T}, \quad (j = 1, \ldots, T), \tag{A.16}
\]

with \( x_{-d+1,T}^{(d)} = \cdots = x_{0,T}^{(d)} = 0 \), and \( \{ u_{j,T} \} \) being the time-varying MA (\( \infty \)) process in (A.1). Note that the relation (A.16) can be rewritten as

\[
(1 - L) x_{j,T}^{(d)} = x_{j,T}^{(d-1)}. \tag{A.17}
\]
Then, we construct the partial sum process \( \{ X_T^{(d)} \} \) as

\[
X_T^{(d)}(t) = \frac{1}{T^{d-1}} \left\{ \frac{1}{\sigma \sqrt{T}} x_{j,T}^{i(d)} + T \left( t - \frac{j}{T} \right) \frac{1}{\sigma \sqrt{T}} x_{j,T}^{i-1(d)} \right\},
\]

(A.18)

for \((j - 1)/T \leq t \leq j/T\), \(d \geq 2\), and \(X_T^{(1)}(t) \equiv X_T(t)\), where the partial sum process \( \{ X_T \} \) is defined in (A.4). Let us first discuss weak convergence to the onefold integrated process \( \{ X^{(1)} \} \) defined by

\[
X^{(1)}(t) = \int_{0}^{t} X(\nu)d\nu = \int_{0}^{t} \left\{ X(0) + \int_{0}^{\nu} \sigma(\mu, 1)dW(\mu) \right\} d\nu.
\]

(A.19)

For \(d = 2\), the partial sum process in (A.18) becomes

\[
X_T^{(2)}(t) = \frac{1}{T} \left\{ \sum_{i=1}^{j} X_T \left( \frac{i}{T} \right) + T \left( t - \frac{j}{T} \right) X_T \left( \frac{j}{T} \right) \right\}, \quad \left( \frac{j-1}{T} \leq t \leq \frac{j}{T} \right).
\]

(A.20)

Define on \( C \)

\[
h^{(3)}_{i}(x) = \int_{0}^{x} x(\nu)d\nu.
\]

(A.21)

Then, we can see that

\[
\mathcal{L} \left\{ X^{(2)}_T \right\} \rightarrow \mathcal{L} \left\{ h^{(3)}(X) \right\} = \mathcal{L} \left\{ X^{(1)} \right\}.
\]

(A.22)

For the general integer \(d\), define the \(d\)-fold integrated process \( \{ X^{(d)} \} \) by

\[
X^{(d)}(t) = \int_{0}^{t} X^{(d-1)}(\nu)d\nu, \quad X^{(0)}(t) = X(t).
\]

(A.23)

From the similar argument in the case of \(d = 2\), we can see that the partial sum process \( \{ X_T^{(d)} \} \) satisfies

\[
\mathcal{L} \left\{ X^{(d)}_T \right\} \rightarrow \mathcal{L} \left\{ h^{(3)}(X^{(d-1)}) \right\} = \mathcal{L} \left\{ X^{(d-1)} \right\}.
\]

(A.24)

**Acknowledgments**

The authors would like to thank the referees for their many insightful comments, which improved the original version of this paper. The authors would also like to thank Professor Masanobu Taniguchi who is the lead guest editor of this special issue for his efforts and celebrate his sixtieth birthday.
References

[1] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, NY, USA, 1968.
[2] S. Beveridge and C. R. Nelson, “A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’,” *Journal of Monetary Economics*, vol. 7, no. 2, pp. 151–174, 1981.
[3] K. Tanaka, *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1996.
[4] E. J. Hannan, *Multiple Time Series*, John Wiley and Sons, London, UK, 1970.
[5] R. Dahlhaus, “Maximum likelihood estimation and model selection for locally stationary processes,” *Journal of Nonparametric Statistics*, vol. 6, no. 2-3, pp. 171–191, 1996.
[6] R. Dahlhaus, “Asymptotic statistical inference for nonstationary processes with evolutionary spectra,” in *Athens Conference on Applied Probability and Time Series Analysis, Vol. II*, vol. 115 of *Lecture Notes in Statist.*, pp. 145–159, Springer, New York, NY, USA, 1996.
[7] J. P. Imhof, “Computing the distribution of quadratic forms in normal variables,” *Biometrika*, vol. 48, pp. 419–426, 1961.
[8] Y. Ochi, “Asymptotic expansions for the distribution of an estimator in the first-order autoregressive process,” *Journal of Time Series Analysis*, vol. 4, no. 1, pp. 57–67, 1983.
[9] J. Hirukawa and M. Sadakata, “Asymptotic properties of unit root processes with locally stationary disturbance,” Preprint.
[10] J. Hirukawa and M. Taniguchi, “LAN theorem for non-Gaussian locally stationary processes and its applications,” *Journal of Statistical Planning and Inference*, vol. 136, no. 3, pp. 640–688, 2006.
Research Article

Statistical Portfolio Estimation under the Utility Function Depending on Exogenous Variables

Kenta Hamada, Dong Wei Ye, and Masanobu Taniguchi

Department of Applied Mathematics, School of Fundamental Science and Engineering, Waseda University, Tokyo 169-8050, Japan

Correspondence should be addressed to Masanobu Taniguchi, taniguchi@waseda.jp

Received 8 September 2011; Accepted 15 November 2011

Copyright © 2012 Kenta Hamada et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the estimation of portfolios, it is natural to assume that the utility function depends on exogenous variable. From this point of view, in this paper, we develop the estimation under the utility function depending on exogenous variable. To estimate the optimal portfolio, we introduce a function of moments of the return process and cumulant between the return processes and exogenous variable, where the function means a generalized version of portfolio weight function.

First, assuming that exogenous variable is a random process, we derive the asymptotic distribution of the sample version of portfolio weight function. Then, an influence of exogenous variable on the return process is illuminated when exogenous variable has a shot noise in the frequency domain.

Second, assuming that exogenous variable is nonstochastic, we derive the asymptotic distribution of the sample version of portfolio weight function. Then, an influence of exogenous variable on the return process is illuminated when exogenous variable has a harmonic trend. We also evaluate the influence of exogenous variable on the return process numerically.

1. Introduction

In the usual theory of portfolio analysis, optimal portfolios are determined by the mean $\mu$ and the variance $\Sigma$ of the portfolio return $X = \{X(t)\}$. Several authors proposed estimators of optimal portfolios as functions of the sample mean $\hat{\mu}$ and the sample variance $\hat{\Sigma}$ for independent returns of assets. However, empirical studies show that financial return processes are often dependent and non-Gaussian. Shiraishi and Taniguchi [1] showed that the above estimators are not asymptotically efficient generally if the returns are dependent. Under the non-Gaussianity, if we consider a general utility function $U(\cdot)$, the expected utility should depend on higher-order moments of the return. From this point of view, Shiraishi and Taniguchi [1] proposed the portfolios including higher-order moments of the return.

However, empirical studies show that the utility function often depends on exogenous variable $Z = \{Z(t)\}$. From this point of view, in this paper, we develop the estimation under
the utility function depending on exogenous variable. Denote the optimal portfolio estimator by a function \( \hat{g} = g(\hat{\theta}) = g(\hat{E}(X), \hat{\text{cov}}(X, X), \hat{\text{cov}}(X, Z), \hat{\text{cum}}(X, X, Z)) \) where hat (\( \hat{\cdot} \)) means the sample version of (\( \cdot \)). Although Shiraishi and Taniguchi’s [1] setting does not include the exogenous variable \( Z(t) \) in \( \hat{g} \), we can develop the asymptotic theory in the light of their work.

First, assuming that \( Z = \{ Z(t) \} \) is a random process, we derive the asymptotic distribution of \( \hat{g} \). Then, an influence of \( Z \) on the return process is illuminated when \( Z \) has a shot noise in the frequency domain. Second, assuming that \( Z \) is a nonrandom sequence of variables which satisfy Grenander’s conditions, we also derive the asymptotic distribution of \( \hat{g} \). Then an influence of \( Z \) on \( X \) is evaluated when \( Z \) is a sequence of harmonic functions. Numerical studies will be given, and they show some interesting features.

The paper is organized as follows. Section 2 introduces the optimal portfolio of the form \( \hat{g} \) and provides the asymptotic distribution of \( \hat{g} \). Assuming that \( Z \) is a stochastic process, we derive the asymptotics of \( \hat{g} \) when \( Z \) has a shot noise in the frequency domain. The influence of \( Z \) on \( X \) is numerically evaluated in Section 2.2. Assuming that \( Z \) is a nonrandom sequence satisfying Grenander’s conditions, we derive the asymptotic distribution of \( \hat{g} \). Section 3 provides numerical studies for the influence of \( Z \) on \( X \) when \( Z \) is a sequence of harmonic functions. The appendix gives the proofs of all the theorems.

2. Optimal Portfolio with the Exogenous Variables

Suppose the existence of a finite number of assets indexed by \( i, (i = 1, \ldots, p) \). Let \( X(t) = (X_1(t), \ldots, X_p(t))' \) denote the random returns on \( p \) assets at time \( t \), and let \( Z(t) = (Z_1(t), \ldots, Z_q(t))' \) denote the exogenous variables influencing on the utility function at time \( t \). We write \( Y(t) = (X(t)', Z(t)')' = (X_1(t), \ldots, X_p(t), Z_1(t), \ldots, Z_q(t))' \).

Since it is empirically observed that \( \{ X(t) \} \) is non-Gaussian and dependent, we will assume that it is a non-Gaussian stationary process with the 3rd-order cumulants. Also, suppose that there exists a risk-free asset whose return is denoted by \( Y_0(t) \). Let \( a_0 \) and \( a = (\overbrace{a_1, \ldots, a_p}^{q}, 0, \ldots, 0)' \) be the portfolio weights at time \( t \), and the portfolio is \( M(t) = Y(t)'a + Y_0(t)a_0 \) whose higher-order cumulants are written as

\[
\begin{align*}
c_1^M(t) &= \text{cum}\{M(t)\} = c^{a_1}a_1 + Y_0(t)a_0, \\
c_2^M(t) &= \text{cum}\{M(t), M(t)\} = c^{a_1a_2}a_1a_2, \\
c_3^M(t) &= \text{cum}\{M(t), M(t), M(t)\} = c^{a_1a_2a_3}a_1a_2a_3.
\end{align*}
\] (2.1)

We use Einstein’s summation convention here and throughout the paper. For a utility function \( U(\cdot) \), the expected utility can be approximated as

\[
\mathbb{E}[U(M(t))] \approx U\left(c_1^M(t)\right) + \frac{1}{2!}D^2U\left(c_1^M(t)\right)c_2^M(t) + \frac{1}{3!}D^3U\left(c_1^M(t)\right)c_3^M(t),
\] (2.2)
by Taylor expansion of order 3. The approximate optimal portfolio may be described as

\[
\max_{x_0, \mathbf{a}} \quad \{ \text{the right hand side of (2.2)} \},
\]

subject to \( a_0 + \sum_{i=1}^{p} a_i = 1. \) \( \tag{2.3} \)

Solving (2.4), Shiraishi and Taniguchi [1] introduced the optimal portfolio depending on the mean, variance, and the third-order cumulants, and then derived the asymptotic distribution of a sample version estimator. Although our problem is different from that of Shiraishi and Taniguchi [1], we develop the discussion with the methods inspired by them.

Introduce a portfolio estimator function based on observed higher-order cumulants,

\[
\tilde{g} \equiv g(\hat{\theta}) = g \left( \tilde{E}(X), \ c\text{ov}(X, X), \ c\text{ov}(X, Z), \ c\text{um}(X, X, Z) \right),
\]

and assume that the function \( g(\cdot) \) is \( p \)-dimensional and measurable, that is,

\[
g(\theta) : g(E(X), \ c\text{ov}(X, X), \ c\text{ov}(X, Z), \ c\text{um}(X, X, Z)) \rightarrow \mathbb{R}^p. \tag{2.5} \]

Let the random process \( \{Y(t) = (Y_1(t), \ldots, Y_{p+q}(t))'\} \) be a \( (p + q) \)-vector linear process generated by

\[
Y(t) = \sum_{j=0}^{\infty} G(j) \varepsilon(t - j) + \mu, \tag{2.6}
\]

where \( \{\varepsilon(t)\} \) is a \( (p + q) \)-dimensional stationary process such that \( E\{\varepsilon(t)\} = 0 \) and \( E\{\varepsilon(s)\varepsilon(t)\}' = \delta(s, t)K \), with \( K \) a nonsingular \( (p + q) \times (p + q) \)-matrix, \( G(j)'s \) are \( (p + q) \times (p + q) \)-matrices, and \( \mu = (\mu_1, \ldots, \mu_{p+q}) \) is the mean vector of \( \{Y(t)\} \). All the components of \( Y, \varepsilon, G, \mu \) are real. Assuming that \( \{\varepsilon(t)\} \) has all order cumulants, let \( Q^{(\varepsilon)}_{a_1, a_2}(t_1, \ldots, t_{j-1}) \) be the joint \( j \)-order cumulant of \( \varepsilon_{a_1}(t), \varepsilon_{a_2}(t + t_1), \ldots, \varepsilon_{a_j}(t + t_{j-1}) \). In what follows we assume that, for each \( j = 1, 2, 3, \ldots, \)

\[
\sum_{t_1, \ldots, t_{j-1} = -\infty}^{\infty} \sum_{a_1, \ldots, a_j = 1}^{p+q} \left| Q^{(\varepsilon)}_{a_1, a_2}(t_1, \ldots, t_{j-1}) \right| < \infty, \tag{2.7}
\]

\[
\sum_{t = 0}^{\infty} \|G(t)\| < \infty.
\]

Letting \( Q^{Y}_{a_1, a_2}(t_1, \ldots, t_{j-1}) \) be the joint \( j \)-th order cumulant of \( Y_{a_1}(t), Y_{a_2}(t + t_1), \ldots, Y_{a_j}(t + t_{j-1}) \), we define the \( j \)-th order cumulant spectral density by

\[
f_{a_1, \ldots, a_j}(\lambda_1, \ldots, \lambda_{j-1}) = \left( \frac{1}{2\pi} \right)^{\frac{j-1}{2}} \sum_{t_1, \ldots, t_{j-1} = -\infty}^{\infty} \exp\{-i(\lambda_1 t_1 + \cdots + \lambda_{j-1} t_{j-1})\} \times Q^{Y}_{a_1, a_2}(t_1, \ldots, t_{j-1}), \tag{2.8}
\]
which is expressed as

\[ f_{\lambda_1, \ldots, \lambda_{J-1}}(\lambda_1, \ldots, \lambda_{J-1}) = \sum_{b_1, \ldots, b_{J-1} = 1}^{n+q} k_{a_1b_1}(\lambda_1 + \cdots + \lambda_{J-1}) \cdots k_{a_{J-1}b_{J-1}}(\lambda_J) \tilde{Q}_{a_1, \ldots, a_{J-1}}(\lambda_1, \ldots, \lambda_{J-1}), \]  

where \( \tilde{Q}_{a_1, \ldots, a_{J-1}} \) is the \( J \)-th order cumulant spectral density of \( \varepsilon_{a_1}(t), \ldots, \varepsilon_{a_{J-1}}(t) \), \( k_{ab}(\lambda) = \sum_{l=0}^{\infty} G_{ab}(l)e^{i\lambda l} \) and \( G_{ab}(l) \) is the \((a,b)\)th element of \( G(l) \). We introduce the following quantities:

\[ \tilde{c}^{a_1} = \frac{1}{n} \sum_{s=1}^{n} Y_{a_1}(s), \]

\[ \tilde{c}^{a_1a_2} = \frac{1}{n} \sum_{s=1}^{n} (Y_{a_2}(s) - \tilde{c}^{a_2}))(Y_{a_3}(s) - \tilde{c}^{a_3}), \]

\[ \tilde{c}^{a_1a_3} = \frac{1}{n} \sum_{s=1}^{n} (Y_{a_3}(s) - \tilde{c}^{a_3}))(Y_{a_5}(s) - \tilde{c}^{a_5}), \]

\[ \tilde{c}^{a_1a_5a_5} = \frac{1}{n} \sum_{s=1}^{n} (Y_{a_6}(s) - \tilde{c}^{a_6}))(Y_{a_7}(s) - \tilde{c}^{a_7}))(Y_{a_8}(s) - \tilde{c}^{a_8}), \]

where \( 1 \leq a_1, a_2, a_3, a_4, a_6, a_7 \leq p \) and \( p+1 \leq a_5, a_8 \leq p+q \). Write the quantities that appeared in (2.4) by

\[ \hat{\theta} = (\tilde{c}^{a_1}, \tilde{c}^{a_1a_2}, \tilde{c}^{a_1a_3}, \tilde{c}^{a_1a_5a_5}), \]

\[ \theta = (c^{a_1}, c^{a_1a_2}, c^{a_1a_3}, c^{a_1a_5a_5}), \]

where \( c^{a_1, \ldots, a_{J-1}} = Q_{a_1, \ldots, a_{J-1}}(0, \ldots, 0) \). Then \( \operatorname{dim} \theta = \operatorname{dim} \hat{\theta} = a+b+c+d \), where \( a = p, b = p(p+1)/2, c = pq, d = p(p+1)q/2 \).

First, we derive the asymptotics of the fundamental quantity \( \hat{\theta} \).

**Theorem 2.1.** Under the assumptions,

\[ \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{p} N(0, \Omega), \quad (n \to \infty), \]

where

\[ \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} & \Omega_{14} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & \Omega_{24} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & \Omega_{34} \\ \Omega_{41} & \Omega_{42} & \Omega_{43} & \Omega_{44} \end{pmatrix}, \]
and the typical element of $\Omega_{ij}$ corresponding to the covariance between $\hat{c}^\Delta$ and $\hat{c}^\nabla$ is denoted by $V \{(\Delta) (\nabla)\}$, and

\[
V\{(a_1)(a'_1)\} = \Omega_{11} = 2\pi f_{a_1a'_1}(0),
\]
\[
V\{(a_2a_3)(a'_1)\} = \Omega_{12} = 2\pi \int_{-\pi}^{\pi} f_{a_2a_3a'_1}(\lambda, 0) d\lambda (= \Omega_{21}),
\]
\[
V\{(a_4a_5)(a'_1)\} = \Omega_{13} = 2\pi \int_{-\pi}^{\pi} f_{a_4a_5a'_1}(\lambda, 0) d\lambda (= \Omega_{31}),
\]
\[
V\{(a_6a_7a_8)(a'_1)\} = \Omega_{14} = 2\pi \int_{-\pi}^{\pi} f_{a_6a_7a_8a'_1}(\lambda_1, \lambda_2, 0) d\lambda_1 d\lambda_2 (= \Omega_{41}),
\]
\[
V\{(a_2a_3)(a'_2a'_3)\} = \Omega_{22} = 2\pi \int_{-\pi}^{\pi} f_{a_2a_3a'_2a'_3}(\lambda_1, \lambda_2, \lambda_3, \lambda_4) d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4 (= \Omega_{42}),
\]
\[
V\{(a_4a_5)(a'_4a'_5)\} = \Omega_{33} = 2\pi \int_{-\pi}^{\pi} f_{a_4a_5a'_4a'_5}(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4 d\lambda_5 (= \Omega_{43}),
\]
\[
V\{(a_6a_7a_8)(a'_6a'_7a'_8)\} = \Omega_{44} = 2\pi \int_{-\pi}^{\pi} f_{a_6a_7a_8a'_6a'_7a'_8}(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) d\lambda_1 \cdots d\lambda_6
\]
\[
+ 2\pi \int_{-\pi}^{\pi} \sum_{\nu_1} f_{a_1a_2a_3a'_1}(\lambda_1, \lambda_2, \lambda_3) f_{a_4a_5a'_6}
\]
\[
\times (-\lambda_2 - \lambda_3) d\lambda_1 d\lambda_2 d\lambda_3
\]
\[
+ 2\pi \int_{-\pi}^{\pi} \sum_{\nu_2} f_{a_1a_2a_3a'_1}(\lambda_1, \lambda_2) f_{a_4a_5a'_6a'_7}
\]
\[
\times (\lambda_3 - \lambda_2 - \lambda_3) d\lambda_1 d\lambda_2 d\lambda_3
\]
\[
+ 2\pi \int_{-\pi}^{\pi} \sum_{\nu_3} f_{a_1a_2a_3a'_1}(\lambda_1) f_{a_4a_5a'_6a'_7a'_8}
\]
\[
\times (-\lambda_1 - \lambda_2) d\lambda_1 d\lambda_2.
\]

(2.14)
Table 1: Standardized 3rd-order cumulants of the returns of five stocks from 2005/11/08 to 2011/11/08.

|       | IBM    | Ford   | Merck  | HP     | EXXON  |
|-------|--------|--------|--------|--------|--------|
| $\hat{s}$ | 0.05608 | 2.50879 | 0.79521 | 9.84777 | 0.34485 |

In what follows we place all the proofs of theorems in the appendix.

Next we discuss the estimation of portfolio $g(\theta)$. For this we assume that the portfolio function $g(\theta)$ is continuously differentiable. Henceforth, we use a unified estimator $\hat{g}(\hat{\theta})$ for $g(\theta)$. The $\delta$-method and Slutsky’s lemma imply the following.

**Theorem 2.2.** Under the assumptions

$$\sqrt{n}(\hat{g}(\hat{\theta}) - g(\theta)) \overset{\mathcal{D}}{\longrightarrow} N\left(0, (Dg)\Omega(Dg)'\right), \quad (n \to \infty),$$

where $Dg = \{\partial_i g^j; \ i = 1, \ldots, \dim \theta, \ j = 1, \ldots, p + q\}$.

The quantities $\hat{c}_a^3, a^3$'s are the 3rd-order cumulants of the process, which show the non-Gaussianity. For the returns of five financial stocks IBM, Ford, Merck, HP, and EXXON, we calculated the standardized 3rd-order cumulants $\hat{s} = \hat{c}_a^3 / \hat{v}^2$ where $\hat{v}^2$ is the sample variance of the stock. Table 1 below shows their values.

From Table 1 we observe that the five returns are non-Gaussian. In view of Theorem 2.1, it is possible to construct the $(1 - \alpha)$ confidence interval for $c = c^a_c a^3$ in the following form:

$$\left[\hat{c} - \frac{z_{\alpha}}{\sqrt{n}} \hat{\Omega}^{1/2}_{44}, \hat{c} + \frac{z_{\alpha}}{\sqrt{n}} \hat{\Omega}^{1/2}_{44}\right],$$

where $z_{\alpha}$ is the upper level-$\alpha$ point of $N(0,1)$ and $\hat{\Omega}_{44}$ is a consistent estimator of $\Omega_{44}$ calculated by the method of Keenan [2] and Taniguchi [3].

### 2.1. Influence of Exogenous Variable

In this subsection we investigate an influence of the exogenous variables $Z(t)$ on the asymptotics of the portfolio estimator $g(\hat{\theta})$.

Assume that the exogenous variables have “shot noise” in the frequency domain, that is,

$$Z_{a_i}(\lambda) = \delta(\lambda_{a_i} - \lambda),$$

where $\delta(\cdot)$ is the Dirac delta function with period $2\pi$, and $\lambda_{a_i} \neq 0$, hence $Z_{a_i}(\lambda)$ has one peak at $\lambda + \lambda_{a_i} \equiv 0 \pmod{2\pi}$. 
Theorem 2.3. For (2.17), denote \( \Omega_{ij} \) and \( V((\Delta)(\nabla)) \) in Theorem 2.1 by \( \Omega'_{ij} \) and \( V'((\Delta)(\nabla)) \), respectively. That is, \( \Omega'_{ij} \) and \( V'((\Delta)(\nabla)) \) represent the asymptotic variance when the exogenous variables are shot noise. Then,

\[
V'(a_4a_5(a_4')) = \Omega'_{13} = 0 (= \Omega'_{31}),
\]

\[
V'(a_6a_7a_8(a_4')) = \Omega'_{14} = 2\pi f_{a_6a_7a_8}(\lambda_{a_7}, 0) (= \Omega'_{41}),
\]

\[
V'(a_2a_3(a_4'a_5')) = \Omega'_{23} = 2\pi f_{a_2a_3a_4}(\lambda_{a_2}, 0) + 2\pi f_{a_2a_4}(\lambda_{a_3}) f_{a_3a_5}(-\lambda_{a_5}) + 2\pi f_{a_3a_5}(\lambda_{a_3}) f_{a_3a_5}(-\lambda_{a_5}) = \Omega'_{13} = 0 (= \Omega'_{31}),
\]

\[
V'(a_2a_3(a_4'a_5')) = \Omega'_{24} = 2\pi \int_{-\pi}^{\pi} f_{a_2a_3a_4a_5}(\lambda_{a_2}, \lambda_{a_3}, -\lambda_{a_5}, \lambda_{a_5}) d\lambda_{a_2} d\lambda_{a_3} (= \Omega'_{12}),
\]

\[
V'(a_4a_5(a_4'a_5')) = \Omega'_{33} = 2\pi f_{a_4a_5a_4a_5}(\lambda_{a_4}, \lambda_{a_5}, -\lambda_{a_5}) + 2\pi f_{a_4a_5}(\lambda_{a_4}) f_{a_4a_5}(-\lambda_{a_5}) + 2\pi f_{a_4a_5}(\lambda_{a_4}) f_{a_4a_5}(-\lambda_{a_5}) = \Omega'_{13} = 0 (= \Omega'_{31}).
\]

\[
V'(a_4a_5(a_4'a_5')) = \Omega'_{34} = 2\pi \int_{-\pi}^{\pi} f_{a_4a_5a_4a_5}(\lambda_{a_4}, \lambda_{a_5}, -\lambda_{a_5}, -\lambda_{a_5}) d\lambda (= \Omega'_{33}).
\]

2.2. Numerical Studies for Stochastic Exogenous Variables

This subsection provides some numerical examples which show the influence of \( Z(t) \) on \( \Omega_{ij} \).

Example 2.4. For a risk-free asset \( X_0(t) \) and risky asset \( X(t) \), we consider construction of optimal portfolios \( aX(t) + a_0X_0(t) \). Here \( \{X(t)\} \) is the return process of the risky asset, which is generated by

\[
X(t) = \theta X(t - 1) + \varepsilon(t) + \mu_1, \tag{2.19}
\]

where \( \mathbb{E}[\varepsilon(t)] = 0, \text{Var}[\varepsilon(t)] = \sigma^2 \). We assume that \( X_0(t) = \mu \), and that the exogenous variable in the frequency domain is given by \( Z(\lambda) = \delta(\lambda) \). Write,

\[
Y(t) = (X(t)X_0(t)Z(t')), \tag{2.20}
\]

then

\[
\Omega'_{13} = \Omega'_{31} = V'(a_4a_5(a_4')) = 0,
\]

\[
\Omega'_{23} = \Omega'_{32} = V'(a_2a_3(a_4'a_5')) = 0,
\]

\[
\Omega'_{33} = V'(a_4a_5(a_4'a_5')) = \text{cum}(a_1, a'_1) \text{cum}(a_3, a'_3)
\]

\[
= \sigma^2 \frac{1}{\left(1 - \theta e^{i\lambda_{a_5}}\right)^2}.
\]
which are covariances between $Z(t)$ and $X(t)$, and show an influence $Z(t)$ on $X(t)$. From Figure 1, it is seen that as $\theta$ tends to 1, and $\lambda_{a_3}$ tends to 0, then $\Omega'_{33}$ increases. If $\theta$ tends to $-1$ and $\lambda_{a_3}$ tends to $-\pi$, $\pi$, then $\Omega'_{33}$ also increases, which entails that the exogenous variables have big influence on the asymptotics of estimators when $\theta$ is close to the unit root of AR(2.2).

**Remark 2.5.** $\Omega'_{13}$ is robust for the shot noise in $Z(t)$ at $\lambda = \lambda_{a_3}$.

### 3. Portfolio Estimation for Nonstochastic Exogenous Variables

So far we assumed that the sequence of exogenous variables $\{Z(t)\}$ is a random stochastic process. In this section, assuming that $\{Z(t)\}$ is a nonrandom sequence, we will propose a portfolio estimator, and elucidate the asymptotics. We introduce the following quantities,

$$
\hat{A}_{j,k} = \frac{\sum_{t=1}^{n} X_j(t) Z_k(t)}{\sqrt{n\sum_{t=1}^{n} Z_k^2(t)}},$

$$
\hat{B}_{j,m,k} = \frac{\sum_{t=1}^{n} X_j(t) X_m(t) Z_k(t)}{\sqrt{n\sum_{t=1}^{n} Z_k^2(t)}}. 
$$

We assume that $Z(t)$'s satisfy Grenander's conditions (G1)–(G4) with

$$
a^{(n)}_{j,k}(h) = \sum_{t=1}^{n-h} Z_j(t+h)Z_k(t). 
$$

(G1) $\lim_{n \to \infty} a^{(n)}_{j,j}(0) = \infty$, $(j = 1, \ldots, q)$.

(G2) $\lim_{n \to \infty} Z_j(n+1)^2 / a^{(n)}_{j,j}(0) = 0$, $(j = 1, \ldots, q)$. 

![Figure 1: Values of $\Omega'_{33}$ for $\theta = -0.9(0.018)0.9$, $\lambda_{a_3} = -\pi(0.06)\pi$.](image)
Grenander’s conditions and the assumption

\[
\Phi(h) = \int_{-\pi}^{\pi} e^{ih\lambda} d\Phi(\lambda). \tag{3.3}
\]

\(\Phi(\lambda)\) is the regression spectral measure of \(\{Z(t)\}\). Next we discuss the asymptotics for sample versions of \(\text{cov}(X,Z)\) and \(\text{cov}(XX,Z)\). For this we need the following assumption. There exists constant \(b > 0\) such that

\[
\det\{f_X(\lambda)\} \geq b, \tag{3.4}
\]

where \(f_X(\lambda)\) is the spectral density matrix of \(\{X(t)\}\).

**Theorem 3.1.** Under Grenander’s conditions and the assumption

\[
\sqrt{n}\left\{ \hat{A}_{j,k} - A_{j,k} \right\} \xrightarrow{\mathcal{D}} N(0, \Omega_{j,k}), \tag{3.5}
\]

where the \((j', k')\)-th element of \(\Omega_{j,k}\) is given by

\[
V(j, k : j', k') = 2\pi \int_{-\pi}^{\pi} f_{jj'}(\lambda) dM_{k,k}(\lambda). \tag{3.6}
\]

**Theorem 3.2.** Under Grenander’s conditions and the assumption

\[
\sqrt{n}\left\{ \hat{B}_{j,m,k} - B_{j,m,k} \right\} \xrightarrow{\mathcal{D}} N(0, \Omega_{j,m,k}), \tag{3.7}
\]

where \(\Omega_{j,m,k} = \{V(j, m, k : j', m', k')\}\) with

\[
V(j, m, k : j', m', k') = 2\pi \int_{-\pi}^{\pi} \left[ \int_{-\pi}^{\pi} \left\{ f_{jm'}(\lambda - \lambda_1) f_{m'j}(\lambda_1) + f_{jj'}(\lambda - \lambda_1) f_{mm'}(\lambda_1) \right\} d\lambda_1 
+ \int_{-\pi}^{\pi} f_{jj'm'}(\lambda_1, \lambda_2 - \lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \right] dM_{k,k}(\lambda). \tag{3.8}
\]

### 3.1. Numerical Studies for Nonstochastic Exogenous Variables

Letting \(\{X(t)\}\) and \(\{Z(t)\}\) be scalar processes, we investigate an influence of non-stochastic process \(\{Z(t)\}\) on \(\{X(t)\}\). The figures below show influence of harmonic trends \(\{Z(t)\}\) on \(V(j, m, k : j', m', k') \in \Omega_{j,m,k}\). In these cases \(V(j, m, k : j', m', k')\) measures the amount of covariance between \(XX\) and \(Z\).
Example 3.3. Let the return process \( \{X(t)\} \) and the exogenous process \( \{Z(t)\} \) be generated by
\[
X(t) = \varepsilon(t) - \eta \varepsilon(t-1),
\]
\[
Z(t) = \cos(\mu t) + \cos(0.25 \mu t),
\]
where \( \varepsilon(t) \)'s are i. i. d. \( N(0,1) \) variables. Next, suppose that \( \{Z(t)\} \) consists of harmonic trends with period \( \mu \) and the quarter period. We plotted the graph of \( V(j, m, k : j', m', k') \) in Figure 2.

Example 3.4. Assume that \( \{X(t)\} \) and \( \{Z(t)\} \) are generated by
\[
X(t) - \eta X(t-1) = \varepsilon(t),
\]
\[
Z(t) = \cos(\mu t) + \cos(0.25 \mu t).
\]
We observe that there exist two peaks in Figure 3. If \( \mu \approx 0 \) and \( | \eta | \approx 1 \), \( V(j, m, k : j', m', k') \) increases rapidly. For further study it may be noted that Cheng et al. [4] discussed statistical estimation of generalized multiparameter likelihood models. Although these models are for independent samples, there is some possibility to apply them to our portfolio problem.

**Appendix**

This section provides the proofs of theorems.

**Proof of Theorem 2.1.** Our setting includes the exogenous variables. Although Shiraishi and Taniguchi’s [1] setting does not include them, we can prove the theorem in line with Shiraishi and Taniguchi [1].

Let

\[
\tilde{c}_{a_1 a_2 a_3} = \frac{1}{n} \sum_{s=1}^{n} (Y_{a_1} - \mu_{a_1})(Y_{a_2} - \mu_{a_2}),
\]

\[
\tilde{c}_{a_6 a_7 a_8} = \frac{1}{n} \sum_{s=1}^{n} (Y_{a_6} - \mu_{a_6})(Y_{a_7} - \mu_{a_7})(Y_{a_8} - \mu_{a_8}).
\]

From Fuller [5], it is easy to see that

\[
\left( \tilde{c}_{a_6} - \mu_{a_6} \right) \left( \tilde{c}_{a_7} - \mu_{a_7} \right) \left( \tilde{c}_{a_8} - \mu_{a_8} \right) = o_p \left( \frac{1}{\sqrt{n}} \right),
\]

\[
\tilde{c}_{a_6 a_7 a_8} - \mu_{a_6} a_7 a_8 = c_{a_6 a_7 a_8} + o_p \left( \frac{1}{\sqrt{n}} \right),
\]

where

\[
1 \leq a_6, a_7, a_8 \leq p, \quad p + 1 \leq a_6 + a_7 + a_8.
\]

Then we can see that

\[
\tilde{c}_{a_6 a_7 a_8} = \tilde{c}_{a_6 a_7 a_8} - \sum_{k=6}^{8} c_{a_k a_k a_k} (\tilde{c}_{a_k} - \mu_{a_k}) + o_p \left( \frac{1}{\sqrt{n}} \right),
\]

where \( \sum_{k=6}^{8} \) is the sum over \( k = 6, 7, 8 \) with \( i_k \) and \( j_k \) \( \in \{6, 7, 8\} \) satisfying \( i_k < j_k; k \neq i_k, j_k \).

Hence it follows that

\[
\begin{align*}
\frac{n}{\text{Cov}} & \left( \tilde{c}_{a_6 a_7 a_8} - c_{a_6 a_7 a_8} \right) \left( \tilde{c}_{a_6 a_7 a_8} - c_{a_6 a_7 a_8} \right) \\
& = n \sum_{k=6}^{8} \sum_{k=6}^{8} c_{a_k a_k a_k} (\tilde{c}_{a_k} - \mu_{a_k}) - \sum_{k=6}^{8} \sum_{k=6}^{8} c_{a_k a_k a_k} (\tilde{c}_{a_k} - \mu_{a_k}) - \sum_{k=6}^{8} \sum_{k=6}^{8} c_{a_k a_k a_k} (\tilde{c}_{a_k} - \mu_{a_k}) \left( \tilde{c}_{a_k} - \mu_{a_k} \right)
\end{align*}
\]

\[
+ n \sum_{k=6}^{8} \sum_{k=6}^{8} \sum_{k=6}^{8} c_{a_k a_k a_k} (\tilde{c}_{a_k} - \mu_{a_k}) c_{a_k a_k a_k} (\tilde{c}_{a_k} - \mu_{a_k}) + o(1).
\]
In what follows we assume that \((a_i, \ldots, a_n)\) is an arbitrary permutation of \((a_1, a_2, a_3, a_4)\),

\[
1 \leq a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12} \leq p, \quad p + 1 \leq a_{13}, a_{14} \leq p + q.
\]  

(A.6)

Then we get

\[
\sum_{t=-\infty}^{\infty} \left\{ Q_{a_6 a_7 a_8 a_9 a_{10}}^{Y}(0,0,t,t,t) + \sum_{v_1} Q_{a_4 a_5 a_6 a_7}^{Y}(0,t,t)Q_{a_8 a_9 a_{10}}^{Y}(t) 
+ \sum_{v_2} Q_{a_1 a_2 a_3}^{Y}(t)Q_{a_4 a_5 a_6}^{Y}(t) + \sum_{v_3} Q_{a_1 a_2 a_3}^{Y}(t)Q_{a_4 a_5 a_6}^{Y}(t) + Q_{a_8 a_9 a_{10}}^{Y}(t) \right\}.
\]  

(A.7)

By use of Fourier transform, we see that

\[
\begin{align*}
(A1) &= 2\pi \int \int \int \int_{-\pi}^{\pi} f_{a_1 a_2 a_3 a_4 a_5 a_6} \left( \lambda_1, \lambda_2, \lambda_3, \lambda_4, -\lambda_3, -\lambda_4 \right) d\lambda_1 \cdots d\lambda_4 \\
&+ 2\pi \int \int \int_{-\pi}^{\pi} \sum_{v_1} f_{a_1 a_2 a_3 a_4} \left( \lambda_1, \lambda_2, \lambda_3 \right) f_{a_5 a_6} \left( -\lambda_2 - \lambda_3 \right) d\lambda_1 d\lambda_2 d\lambda_3 \\
&+ 2\pi \int \int \int_{-\pi}^{\pi} \sum_{v_2} f_{a_1 a_2 a_3 a_4} \left( \lambda_1, \lambda_2 \right) f_{a_5 a_6} \left( -\lambda_1 - \lambda_2 \right) d\lambda_1 d\lambda_2 d\lambda_3 \\
&+ 2\pi \int \int \int_{-\pi}^{\pi} \sum_{v_3} f_{a_1 a_2 a_3 a_4} \left( \lambda_1 \right) f_{a_5 a_6} \left( -\lambda_1 \right) d\lambda_1 d\lambda_2.
\end{align*}
\]  

(A.8)

The other asymptotic covariances are similarly evaluated. Finally, it suffices to prove the asymptotic normality of \(\sqrt{n}(\hat{\theta} - \theta)\). For this we prove

\[
\text{cum}\{ \sqrt{n}(\hat{\theta}^{a_1} - \theta^{a_1}), \ldots, \sqrt{n}(\hat{\theta}^{a_j} - \theta^{a_j}) \} \rightarrow 0, \quad j \geq 3,
\]  

(A.9)

where \(\hat{\theta}^{a_i}\) and \(\theta^{a_i}\) are the \(i\)th component of \(\hat{\theta}\) and \(\theta\), respectively. Let

\[
\hat{\theta}^{a_i} - \theta^{a_i} = \begin{cases} 
\hat{c}^{b_1} - c^{b_1} & \text{if } i = 1, \ldots, j_1 \\
\hat{c}^{b_2 b_3} - c^{b_2 b_3} & \text{if } i = j_1 + 1, \ldots, j_1 + j_2 \\
\hat{c}^{b_1 b_3} - c^{b_1 b_3} & \text{if } i = j_1 + j_2 + 1, \ldots, j_1 + j_2 + j_3 \\
\hat{c}^{b_1 b_2 b_3} - c^{b_1 b_2 b_3} & \text{if } i = j_1 + j_2 + j_3, \ldots, j_1 + j_2 + j_3 + j_4 (= j) . 
\end{cases}
\]  

(A.10)
Next we evaluate the covariance:

\[
\text{Cov}\left(\sqrt{n}\left(\hat{A}_{j,k} - A_{j,k}\right), \sqrt{n}\left(\hat{A}_{j',k'} - A_{j',k'}\right)\right)
\]

\[
= n \mathbb{E}\left[\frac{\left(\hat{A}_{j,k} - A_{j,k}\right) - \mathbb{E}\left(\hat{A}_{j,k} - A_{j,k}\right)}{\sqrt{n}} \right] \mathbb{E}\left[\frac{\left(\hat{A}_{j',k'} - A_{j',k'}\right) - \mathbb{E}\left(\hat{A}_{j',k'} - A_{j',k'}\right)}{\sqrt{n}}\right]
\]

\[
= n \mathbb{E}\left[\left(\hat{A}_{j,k} - A_{j,k}\right) + o\left(\frac{1}{\sqrt{n}}\right)\right] \left(\hat{A}_{j',k'} - A_{j',k'}\right) + o(1)
\]

\[
= \sum_{t=1}^{n} \sum_{s=1}^{n} R_{j,f}(s-t) \frac{Z_k(t)Z_k(s)}{d_{t,k}(n)d_{s,k}(n)}
\]

\[
= \sum_{l=-n+1}^{n-1} R_{j,f}(l) \sum_{s=1}^{n} \frac{Z_k(s-l)Z_k(s)}{d_{s-k}(n)d_{s-k}(n)}
\]

\[
\rightarrow 2\pi \int_{-\pi}^{\pi} f_{jj}(\lambda) dM_{k,k}(\lambda) = V(j,k:j',k'),
\]

where \(R_{j,f}(s-t)\) is the covariance function of \(X_j(t)\) and \(X_f(s)\).

The asymptotic normality of \(\sqrt{n}(\hat{A}_{j,k} - A_{j,k})\) can be shown if we prove

\[
\text{Cov}\left(\sqrt{n}\left(\hat{A}_{j,k} - A_{j,k}\right), \sqrt{n}\left(\hat{A}_{j,k} - A_{j,k}\right)\right) \rightarrow 0, \quad l \geq 3.
\]
Similarly as in Theorem 5.11.1 of Brillinger [6], we can see that
\[
\text{cum}\left\{ \sqrt{n}\left( \hat{A}_{j,k_1} - A_{j,k_1} \right), \ldots, \sqrt{n}\left( \hat{A}_{j,k_l} - A_{j,k_l} \right) \right\} \\
= n^{l/2} \text{cum}\left\{ \hat{A}_{j,k_1}, \ldots, \hat{A}_{j,k_l} \right\} \\
= O\left(n^{1-l/2}\right), \quad \text{for } l \geq 3. \tag{A.16}
\]

Proof of Theorem 3.2. First, it is seen that
\[
\lim_{n \to \infty} E\left[ \tilde{B}_{j,m,k} \right] = \lim_{n \to \infty} R_{j,m}(0) \frac{\sum_{k=1}^{n} Z_k(t)}{\sqrt{n}d_{l,k}(n)} \\
\to R_{j,m}(0) \int_{-\pi}^{\pi} dM_{0,k}(\lambda) \overset{\text{say}}{=} B_{j,m,k} \quad \text{(by (G3)).}
\tag{A.17}
\]

We can evaluate the covariance as follows:
\[
\text{Cov}\left( \sqrt{n}\left( \tilde{B}_{j,m,m} - B_{j,k,m} \right), \sqrt{n}\left( \tilde{B}_{j',k',m'} - B_{j',k',m'} \right) \right) \\
= n E\left[ \left( \tilde{B}_{j,m,m} - B_{j,k,m} \right) \left( \tilde{B}_{j',k',m'} - B_{j',k',m'} \right) \right] \\
= n E\left[ \tilde{B}_{j,m,m} \tilde{B}_{j',k',m'} - B_{j,k,m} B_{j',k',m'} \right] \\
= \sum_{j=1}^{n} \sum_{s=1}^{n} \text{Cov}(X_j(t)X_m(t), X_{j'}(s)X_{m'}(s)) \int_{-\pi}^{\pi} dM_{k,k'}(\lambda) + o(1) \\
= \int_{-\pi}^{\pi} \sum_{l=1}^{\infty} \left\{ \text{cum}_{j,m,j',m'}(0, l, l) + R_{j,m}(l) R_{j',m'}(l) + R_{j,j'}(l) R_{m,m'}(l) \right\} e^{-il\lambda} \text{d}M_{k,k'}(\lambda) \\
\to 2\pi \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left\{ f_{jm}(\lambda - \lambda_1) f_{mm}(\lambda_1) + f_{jm'}(\lambda - \lambda_1) f_{m'm}(\lambda_1) \right\} d\lambda_1 \\
\quad + 2\pi \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f_{jm'm'}(\lambda_1, \lambda_2 + \lambda, -\lambda_2) d\lambda_1 d\lambda_2 \text{d}M_{k,k'}(\lambda) \\
= V(j, m, k: j', m', k'). \tag{A.18}
\]

Next we derive the asymptotic normality of $\sqrt{n}(\tilde{B}_{j,m,k} - B_{j,m,k})$. For this we prove
\[
\text{cum}\left\{ \sqrt{n}\left( \tilde{B}_{j_1,k_1,m_1} - B_{j_1,k_1,m_1} \right), \ldots, \sqrt{n}\left( \tilde{B}_{j_l,k_l,m_l} - B_{j_l,k_l,m_l} \right) \right\} \to 0, \quad l \geq 3. \tag{A.19}
\]
Similarly as in Theorem 5.11.1 of Brillinger [6], it is shown that

\[
\begin{aligned}
\text{cum}\left\{ \sqrt{n} \left( \hat{B}_{j_1,m_1,k_1} - B_{j_1,m_1,k_1} \right), \ldots, \sqrt{n} \left( \hat{B}_{j_l,m_l,k_l} - B_{j_l,m_l,k_l} \right) \right\} \\
= n^{1/2} \text{cum}\left\{ \hat{B}_{j_1,m_1,k_1}, \ldots, \hat{B}_{j_l,m_l,k_l} \right\} \\
= O\left( n^{1-1/2} \right),
\end{aligned}
\]

which proves the asymptotic normality.

\[\square\]

**Acknowledgments**

The authors thank Professor Cathy Chen and two referees for their kind comments.

**References**

[1] H. Shiraishi and M. Taniguchi, “Statistical estimation of optimal portfolios depending on higher order cumulants,” *Annales de l’ISUP*, vol. 53, no. 1, pp. 3–18, 2009.
[2] D. M. Keenan, “Limiting behavior of functionals of higher-order sample cumulant spectra,” *The Annals of Statistics*, vol. 15, no. 1, pp. 134–151, 1987.
[3] M. Taniguchi, “On estimation of the integrals of the fourth order cumulant spectral density,” *Biometrika*, vol. 69, no. 1, pp. 117–122, 1982.
[4] M.-Y. Cheng, W. Zhang, and L.-H. Chen, “Statistical estimation in generalized multiparameter likelihood models,” *Journal of the American Statistical Association*, vol. 104, no. 487, pp. 1179–1191, 2009.
[5] W. A. Fuller, *Introduction to Statistical Time Series*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1996.
[6] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Holden-Day, San Francisco, Calif, USA, 2nd edition, 1981.
Research Article

Statistical Estimation for CAPM with Long-Memory Dependence

Tomoyuki Amano,1 Tsuyoshi Kato,2 and Masanobu Taniguchi2

1 Faculty of Economics, Wakayama University, Wakayama 6408510, Japan
2 Department of Applied Mathematics, School of Fundamental Science and Engineering, Waseda University, Tokyo 1698555, Japan

Correspondence should be addressed to Tomoyuki Amano, tomchami@center.wakayama-u.ac.jp

Received 24 June 2011; Revised 27 August 2011; Accepted 10 September 2011

Academic Editor: Junichi Hirukawa

Copyright © 2012 Tomoyuki Amano et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We investigate the Capital Asset Pricing Model (CAPM) with time dimension. By using time series analysis, we discuss the estimation of CAPM when market portfolio and the error process are long-memory process and correlated with each other. We give a sufficient condition for the return of assets in the CAPM to be short memory. In this setting, we propose a two-stage least squares estimator for the regression coefficient and derive the asymptotic distribution. Some numerical studies are given. They show an interesting feature of this model.

1. Introduction

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 [1] did the groundwork of this model. In his research, he cast the investor’s portfolio selection problem in terms of expected return and variance. Sharpe [2] and Lintner [3] developed Markowitz’s idea for economical implication. Black [4] 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. [5] discussed the estimation of CAPM, but in their work they did not discuss the time dimension. However, in the econometric analysis, it is necessary to investigate this model with the time dimension; that is, the model is represented as $R_{it} = \alpha_{im} + \beta_{im}R_{mt} + \epsilon_{it}$. Recently from the empirical analysis, it is known that the return of
asset follows a short-memory process. But Granger \[6\] showed that the aggregation of short-memory processes yields long-memory dependence, and it is known that the return of the market portfolio follows a long-memory process. From this point of view, first, we show that the return of the market portfolio and the error process \(e_t\) are long-memory dependent and correlated with each other.

For the 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 variable. This method was first used by Wright \[7\], and many researchers developed this method (see Reiersøl \[8\], Geary \[9\], etc.). Comprehensive reviews are seen in White \[10\]. However, the instrumental variable method has been discussed in the case where the error process does not follow long-memory process, and this makes the estimation difficult.

For the analysis of long-memory process, Robinson and Hidalgo \[11\] considered a stochastic regression model defined by
\[
y_t = \alpha + \beta x_t + u_t,
\]
where \(\alpha, \beta = (\beta_1, \ldots, \beta_K)'\) are unknown parameters and the \(K\)-vector processes \(\{x_t\}\) and \(\{u_t\}\) are long-memory dependent with \(E(x_t) = 0, E(u_t) = 0\). Furthermore, in Choy and Taniguchi \[12\], 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 \[12\] introduced a ratio estimator, the least squares estimator, and the best linear unbiased estimator for \(\beta\). However, Robinson and Hidalgo \[11\] and Choy and Taniguchi \[12\] assume that the explanatory process \(\{x_t\}\) and the error process \(\{u_t\}\) are independent.

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

This paper is organized as follows. Section 2 gives our definition of the CAPM, and we give a sufficient condition that return of assets as short dependence is generated by the returns of market portfolio and error process which are long-memory dependent and mutually correlated each other. In Section 3 we propose 2SLS estimator for this model and show its consistency and asymptotic normality. Section 4 provides some numerical studies which show interesting features of our estimator. The proof of theorem is relegated to Section 5.

2. CAPM (Capital Asset Pricing Model)

For Sharpe and Lintner version of the CAPM (see Sharpe \[2\] and Lintner \[3\]), the expected return of asset \(i\) is given by
\[
E[R_i] = R_f + \beta_{im}(E[R_m - R_f]),
\]
where
\[
\beta_{im} = \frac{\text{Cov}[R_i, R_m]}{\text{V}[R_m]},
\]
and
\[
E[R_i] = \alpha + \beta x_t + u_t,
\]
where \(\alpha, \beta = (\beta_1, \ldots, \beta_K)'\) are unknown parameters and the \(K\)-vector processes \(\{x_t\}\) and \(\{u_t\}\) are long-memory dependent with \(E(x_t) = 0, E(u_t) = 0\). Furthermore, in Choy and Taniguchi \[12\], 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 \[12\] introduced a ratio estimator, the least squares estimator, and the best linear unbiased estimator for \(\beta\). However, Robinson and Hidalgo \[11\] and Choy and Taniguchi \[12\] assume that the explanatory process \(\{x_t\}\) and the error process \(\{u_t\}\) are independent.

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

This paper is organized as follows. Section 2 gives our definition of the CAPM, and we give a sufficient condition that return of assets as short dependence is generated by the returns of market portfolio and error process which are long-memory dependent and mutually correlated each other. In Section 3 we propose 2SLS estimator for this model and show its consistency and asymptotic normality. Section 4 provides some numerical studies which show interesting features of our estimator. The proof of theorem is relegated to Section 5.
$R_m$ is the return of the market portfolio, and $R_f$ is the return of the risk-free asset. Another Sharpe-Lintner’s CAPM (see Sharpe [2] and Lintner [3]) is defined for $Z_t \equiv R_t - R_f$,

$$E[Z_t] = \beta_{im} E[Z_m], \quad (2.3)$$

where

$$\beta_{im} = \frac{\text{Cov}[Z_t, Z_m]}{\text{V}[Z_m]} \quad (2.4)$$

and $Z_m = R_m - R_f$.

Black [4] derived a more general version of CAPM, which is written as

$$E[R_i] = \alpha_{im} + \beta_{im} E[R_m], \quad (2.5)$$

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

Since CAPM is a single-period model, (2.1) and (2.5) 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:

$$Y_{i,t} = \alpha_i + \beta_i Z_t + \epsilon_{i,t}, \quad (2.6)$$

where $i$ denotes the asset, $t$ denotes the period, and $Y_{i,t}$ and $Z_t$, $i = 1, \ldots, n$ and $t = 1, \ldots, T$, are, respectively, the returns of the asset $i$ and the market portfolio at $t$.

Empirical features of the realized returns for assets and market portfolios are well known.

We plot the autocorrelation function ($ACF(l)$ ($l$ : time lag)) of returns of IBM stock and S&P500 (squared transformed) in Figures 1 and 2, respectively.

From Figures 1 and 2, we observe that the return of stock (i.e., IBM) shows the short-memory dependence and that a market index (i.e., S&P500) shows the long-memory dependence.

Suppose that an $n$-dimensional process $\{Y_t = (Y_{1,t}, \ldots, Y_{n,t})\}$ is generated by

$$Y_t = \alpha + B'Z_t + \epsilon_t \quad (t = 1, 2, \ldots, T), \quad (2.7)$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)'$ and $B = \{\beta_{ij}; \ i = 1, \ldots, p, \ j = 1, \ldots, n\}$ are unknown vector and matrix; respectively, $\{Z_t = (Z_{1,t}, \ldots, Z_{p,t})'\}$ is an explanatory stochastic regressor process, and $\{\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{n,t})'\}$ is a sequence of disturbance process. The $i$th component is written as

$$Y_{i,t} = \alpha_i + \beta_{i}'Z_t + \epsilon_{i,t}, \quad (2.8)$$

where $\beta_{i}' = (\beta_{i1}, \ldots, \beta_{ip})$.

In the CAPM, $Y_t$ is the return of assets and $Z_t$ is the return of the market portfolios. As we saw, empirical studies suggest that $\{Y_t\}$ is short-memory dependent and that $\{Z_t\}$ is long-memory dependent. On this ground, we investigate the conditions that the CAPM (2.7)
is well defined. It is seen that, if the model (2.7) is valid, we have to assume that \( \{ \epsilon_t \} \) is also long-memory dependent and is correlated with \( \{ Z_t \} \).

Hence, we suppose that \( \{ Z_t \} \) and \( \{ \epsilon_t \} \) are defined by

\[
Z_t = \sum_{j=0}^{\infty} \gamma_j a_{t-j} + \sum_{j=0}^{\infty} \rho_j b_{t-j},
\]

\[
\epsilon_t = \sum_{j=0}^{\infty} \eta_j e_{t-j} + \sum_{j=0}^{\infty} \xi_j b_{t-j},
\]  

(2.9)

where \( \{ a_t \} \), \( \{ b_t \} \), and \( \{ \epsilon_t \} \) are \( p \)-dimensional zero-mean uncorrelated processes, and they are mutually independent. Here the coefficients \( \{ \gamma_j \} \) and \( \{ \rho_j \} \) are \( p \times p \)-matrices, and all the components of \( \gamma_j \) are \( \ell^1 \)-summable, (for short, \( \gamma_j \in \ell^1 \)) and those of \( \rho_j \) are \( \ell^2 \)-summable (for
short, \( \rho_j \in \ell^2 \). The coefficients \( \{ \eta_j \} \) and \( \{ \xi_j \} \) are \( n \times p \)-matrices, \( \eta_j \in \ell^1 \), and \( \xi_j \in \ell^2 \). From (2.9) it follows that

\[
Y_t = \alpha + \sum_{j=0}^{\infty} \left( B' \eta_j a_{t-j} + \eta_j \epsilon_{t-j} + \sum_{j=0}^{\infty} (B' \rho_j + \xi_j) b_{t-j} \right).
\]

(2.10)

Although \( (B' \rho_j + \xi_j) \in \ell^2 \) generally, if \( B' \rho_j + \xi_j = O(1/j^a) \), \( \alpha > 1 \), then \( (B' \rho_j + \xi_j) \in \ell^1 \), which leads to the following.

**Proposition 2.1.** If \( B' \rho_j + \xi_j = O(j^{-a}) \), \( \alpha > 1 \), then the process \( \{ Y_t \} \) is short-memory dependent.

Proposition 2.1 provides an important view for the CAPM; that is, if we assume natural conditions on (2.7) based on the empirical studies, then they impose a sort of “curved structure”: \( B' \rho_j + \xi_j = O(j^{-a}) \) on the regressor and disturbance. More important view is the statement implying that the process \( \{ \beta_j' Z_t + \epsilon_{1,t} \} \) is fractionally cointegrated. Here \( \beta_j \) and \( \epsilon_{1,t} \) are called the cointegrating vector and error, respectively, (see Robinson and Yajima [13]).

### 3. Two-Stage Least Squares Estimation

This section discusses estimation of (2.7) satisfying Proposition 2.1. Sinc \( E(Z_t' \epsilon_t') \neq 0 \), the least squares estimator for \( B \), is known to be inconsistent. In what follows we assume that \( \alpha = 0 \) in (2.7), because it can be estimated consistently by the sample mean. However, by use of the econometric theory, it is often possible to find other variables that are uncorrelated with the errors \( \epsilon_t \), which we call instrumental variables, and to overcome this difficulty. Without instrumental variables, correlations between the observables \( \{ Z_t \} \) and unobservables \( \{ \epsilon_{1,t} \} \) persistently contaminate our estimator for \( B \). Hence, instrumental variables are useful in allowing us to estimate \( B \).

Let \( \{ X_t \} \) be \( r \times 1 \)-dimensional vector \( (p \leq r) \) instrumental variables with \( E[X_t] = 0 \), \( \text{Cov}(X_t, Z_t) \neq 0 \), and \( \text{Cov}(X_t, \epsilon_{1,t}) = 0 \). Consider the OLS regression of \( Z_t \) on \( X_t \). If \( Z_t \) can be represented as

\[
Z_t = \delta' X_t + u_t,
\]

(3.1)

where \( \delta \) is a \( r \times p \) matrix and \( \{ u_t \} \) is a \( p \)-dimensional vector process which is independent of \( \{ X_t \} \), \( \delta \) can be estimated by the OLS estimator

\[
\hat{\delta} = \left[ \sum_{t=1}^{T} X_t X_t' \right]^{-1} \sum_{t=1}^{T} X_t Z_t'.
\]

(3.2)

From (2.7) with \( \alpha = 0 \) and (3.1), \( Y_t \) has the form:

\[
Y_t = B' \delta' X_t + B' u_t + \epsilon_{1,t},
\]

(3.3)
\[ B_{OLS} = \left[ \sum_{t=1}^{T} (\delta'X_t)(\delta'X_t)' \right]^{-1} \left[ \sum_{t=1}^{T} (\delta'X_t)Y_t \right]. \]  

(3.4)

Using (3.2) and (3.4), we can propose the 2SLS estimator:

\[ B_{2SLS} = \left[ \sum_{t=1}^{T} (\tilde{\delta}'X_t)(\tilde{\delta}'X_t)' \right]^{-1} \left[ \sum_{t=1}^{T} (\tilde{\delta}'X_t)Y_t \right]. \]  

(3.5)

Now, we aim at proving the consistency and asymptotic normality of the 2SLS estimator \( B_{2SLS} \). For this we assume that \( \{e_t\} \) and \( \{X_t\} \) jointly constitute the following linear process:

\[
\begin{pmatrix}
  e_t \\
  X_t
\end{pmatrix} = \sum_{j=0}^{\infty} G(j) \Gamma(t-j) = A_t \text{ (say)},
\]

(3.6)

where \( \{\Gamma(t)\} \) is uncorrelated \((n + r)\)-dimensional vector process with

\[
E[\Gamma(t)] = 0,
\]

\[
E[\Gamma(t)\Gamma(s)'] = \delta(t, s)K,
\]

\[
\delta(t, s) = \begin{cases} 1, & t = s, \\ 0, & t \neq s, \end{cases}
\]

(3.7)

and \( G(j) \)'s are \((n + r) \times (n + r)\) matrices which satisfy \( \sum_{j=0}^{\infty} tr\{G(j)K\} < \infty \). Then \( \{A_t\} \) has the spectral density matrix:

\[
f(\omega) = \frac{1}{2\pi} k(\omega)Kk(\omega)' = \{f_{ab}(\omega); \; 1 \leq a, b \leq (n + r)\} \quad (-\pi < \omega \leq \pi),
\]

(3.8)

where

\[
k(\omega) = \sum_{j=0}^{\infty} G(j)e^{ij\omega} = \{k_{ab}(\omega); \; 1 \leq a, b \leq (n + r)\} \quad (-\pi < \omega \leq \pi).
\]

(3.9)

Further, we assume that \( \int_{-\pi}^{\pi} \log \det f(\omega)d\omega > -\infty \), so that the process \( \{A_t\} \) is nondeterministic. For the asymptotics of \( B_{2SLS} \), from page 108, line 1–page 109, line 7 of Hosoya [14], we impose the following assumption.

Assumption 3.1. (i) There exists \( \epsilon > 0 \) such that, for any \( t < t_1 \leq t_2 \leq t_3 \leq t_4 \) and for each \( \beta_1, \beta_2 \),

\[
\text{var}[E\{\Gamma_{\beta_1}(t_1)\Gamma_{\beta_2}(t_2) \mid B(t) \} - \delta(t_1 - t_2, 0)K_{\beta_1\beta_2}] = O\{ (t_1 - t)^{-2-\epsilon} \},
\]

(3.10)
and also

\[ E\left[ \left( \Gamma_{\beta_1}(t_1)\Gamma_{\beta_2}(t_2)\Gamma_{\beta_3}(t_3)\Gamma_{\beta_4}(t_4) \right) \mid B(t) \right] = \mathcal{O}\left( \left( t_1 - t \right)^{-1 - \epsilon} \right), \]

(3.11)

uniformly in \( t \), where \( B(t) \) is the \( \sigma \)-field generated by \( \{ \Gamma(s); s \leq t \} \).

(ii) For any \( \epsilon > 0 \) and for any integer \( M \geq 0 \), there exists \( B_\epsilon > 0 \) such that

\[ E\left[ T(n, s)^2 \mid T(n, s) > B_\epsilon \right] < \epsilon, \]

(3.12)

uniformly in \( n, s \), where

\[ T(n, s) = \left[ \sum_{a, b=1}^{p} \sum_{r=0}^{M} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\Gamma_a(t + s)\Gamma_b(t + s + r) - K_{a,b}\delta(0, r)) \right\} \right]^{1/2}, \]

(3.13)

and \( \{ T(n, s) > B_\epsilon \} \) is the indicator, which is equal to 1 if \( T(n, s) > B_\epsilon \) and equal to 0 otherwise.

(iii) Each \( f_{ab}(\omega) \) is square-integrable.

Under the above assumptions, we can establish the following theorem.

**Theorem 3.2.** Under Assumption 3.1, it holds that

(i)

\[ \tilde{B}_{2SLS} \xrightarrow{p} B, \]

(3.14)

(ii)

\[ \sqrt{T}(\tilde{B}_{2SLS} - B) \xrightarrow{d} Q^{-1} E[X_i] E[Z_i]^{-1} U, \]

(3.15)

where

\[ Q = [E(Z_i X_i')] [E(X_i X_i')]^{-1} [E(X_i Z_i')], \]

(3.16)

and \( U = \{ U_{i,j}; 1 \leq i \leq r, 1 \leq j \leq n \} \) is a random matrix whose elements follow normal distributions with mean 0 and

\[ \text{Cov}[U_{i,j}, U_{k,l}] = 2\pi \int_{-\pi}^{\pi} f_{n+i,n+k}(\omega) f_{j,j}(\omega) + f_{n+i,l}(\omega) f_{j,n+k}(\omega) d\omega \]

\[ + 2\pi \sum_{\beta_1, ..., \beta_4=1}^{p} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \kappa_{n+i,\beta_1}(\omega_1) \kappa_{j,j}(\omega_1) \kappa_{n+k,\beta_4}(\omega_2) \kappa_{j,j}(\omega_2) Q^{r}_{\beta_1, ..., \beta_4} \]

\[ \times (\omega_1, -\omega_2, \omega_2) d\omega_1, \]

(3.17)
The next example prepares the asymptotic variance formula of $\hat{B}_{2SLS}$ to investigate its features in simulation study.

**Example 3.3.** Let $\{Z_t\}$ and $\{X_t\}$ be scalar long-memory processes, with spectral densities $\left\{2\pi|1 - e^{i\lambda}|^{2d_z}\right\}^{-1}$ and $\left\{2\pi|1 - e^{i\lambda}|^{2d_x}\right\}^{-1}$, respectively, and cross spectral density $(1/(2\pi))(1 - e^{i\lambda})^{-d_x}(1 - e^{-i\lambda})^{-d_z}$, where $0 < d_Z < 1/2$ and $0 < d_X < 1/2$. Then

$$E(X_tZ_t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{(1 - e^{i\lambda})^{d_x}} \frac{1}{(1 - e^{-i\lambda})^{d_z}} d\lambda. \tag{3.18}$$

Suppose that $\{\epsilon_t\}$ is a scalar uncorrelated process with $\sigma^2 = E(\epsilon_t^2)$. Assuming Gaussianity of $\{A_t\}$, it is seen that the right hand of (3.17) is

$$2\pi \int_{-\pi}^{\pi} \frac{1}{2\pi|1 - e^{i\lambda}|^{2d_x}2\pi} \sigma^2_{\epsilon} d\lambda, \tag{3.19}$$

which entails

$$\lim_{T \to \infty} \text{var} \left[ \sqrt{T}(\hat{B}_{2SLS} - B) \right] = \frac{2\pi \int_{-\pi}^{\pi} \left(1/(2\pi)|1 - e^{i\lambda}|^{2d_x}\right) \sigma^2_{\epsilon} (1/(2\pi)|1 - e^{i\lambda}|^{2d_x}) \sigma^2_{\epsilon} (1/(2\pi)|1 - e^{-i\lambda}|^{2d_x}) d\lambda}{\left(1/(2\pi) \int_{-\pi}^{\pi} (1/(1 - e^{i\lambda})^{d_x})(1/(1 - e^{-i\lambda})^{d_z}) d\lambda\right)^2} \tag{3.20}$$

$$= \sigma^2_{\epsilon} \left( \int_{-\pi}^{\pi} (1/(1 - e^{i\lambda})^{d_x})(1/(1 - e^{-i\lambda})^{d_z}) d\lambda \right)^2$$

$$= \sigma^2_{\epsilon} \times V_*(d_X, d_Z).$$

**4. Numerical Studies**

In this section, we evaluate the behaviour of $\hat{B}_{2SLS}$ in the case $p = 1$ in (2.7) numerically.

**Example 4.1.** Under the condition of Example 3.3, we investigate the asymptotic variance behaviour of $\hat{B}_{2SLS}$ by simulation. Figure 3 plots $V_*(d_X, d_Z)$ for $0 < d_X < 1/2$ and $0 < d_Z < 1/2$.

From Figure 3, we observe that, if $d_Z \downarrow 0$ and if $d_X \uparrow 1/2$, 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 behaviour of $Z_t$ makes the asymptotic variance of the 2SLS estimator small, but one of $X_t$ makes it large.
Figure 3: $V_\ast (d_x, d_z)$ in Section 4.

Table 1: 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   |

Example 4.2. In this example, we consider the following model:

\begin{align*}
Y_t &= Z_t + \epsilon_t, \\
Z_t &= X_t + u_t, \\
\epsilon_t &= w_t + u_t, \tag{4.1}
\end{align*}

where $X_t$, $w_t$, and $u_t$ are the 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

\begin{equation}
\tilde{B}_{OLS} = \left[ \sum_{t=1}^{T} Z_t^2 \right]^{-1} \left[ \sum_{t=1}^{T} Z_t Y_t \right]. \tag{4.2}
\end{equation}

The lengths of $X_t$, $Y_t$, and $Z_t$ are set by 100, and based on 5000 times simulation we report the mean square errors (MSE) of $\hat{B}_{2SLS}$ and $\tilde{B}_{OLS}$. We set $d_1, d_2 = 0.1, 0.2, 0.3$ in Table 1.

In most cases of $d_1$ and $d_2$ in Table 1, MSE of $\hat{B}_{2SLS}$ is smaller than that of $\tilde{B}_{OLS}$. Hence, from this Example we can see that our estimator $\hat{B}_{2SLS}$ is better than $\tilde{B}_{OLS}$ in the sense of MSE. Furthermore, from Table 1, 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 4.3. In this example, we calculate $\hat{B}_{2SLS}$ based on the actual financial data. We choose S&P500 (square transformed) as $Z_t$ and the Nikkei stock average as an instrumental variable $X_t$. Assuming that $Y_t(5 \times 1)$ consists of the return of IBM, Nike, Amazon, American Expresses and Ford; the 2SLS estimates for $B_i$, $i = 1, \ldots, 5$ are recorded in Table 2. We chose the Nikkei stock average as the instrumental variable, because we got the following correlation analysis between the residual processes of returns and Nikkei.

| Stock   | IBM | Nike | Amazon | American Express | Ford |
|---------|-----|------|--------|------------------|------|
| $B_{2SLS}$ | 0.75 | 1.39 | 1.71   | 2.61             | −1.89|

Correlation of IBM’s residual and Nikkei’s return: −0.000311
Correlation of Nike’s residual and Nikkei’s return: −0.00015
Correlation of Amazon’s residual and Nikkei’s return: −0.000622
Correlation of American Express’s residual and Nikkei’s return: 0.000147
Correlation of Ford’s residual and Nikkei’s return: −0.000536,
which supports the assumption $\text{Cov}(X_t, \epsilon_t) = 0$.

From Table 2, we observe that the return of the finance stock (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.

5. Proof of Theorem

This section provides the proof of Theorem 3.2. First for convenience we define $\tilde{Z}_t = (\tilde{Z}_{1,t}, \ldots, \tilde{Z}_{p,t})' \equiv \delta'X_t$. Let $\tilde{u}_t = (\tilde{u}_{1,t}, \ldots, \tilde{u}_{p,t})'$ be the residual from the OLS estimation of (3.1); that is,

$$\tilde{u}_{i,t} = Z_{i,t} - \tilde{Z}_{i,t}. \quad (5.1)$$

The OLS makes this residual orthogonal to $X_t$:

$$\sum_{i=1}^{T} X_i \tilde{u}_{i,t} = 0, \quad (5.2)$$

which implies the residual is orthogonal to $\tilde{Z}_{j,t}$,

$$\sum_{i=1}^{T} \tilde{Z}_{j,t} \tilde{u}_{i,t} = \left( \sum_{i=1}^{T} X_i \tilde{u}_{i,t} \right) \delta_j = 0, \quad (5.3)$$
where $\hat{\delta}_j$ is $j$th column vector of $\hat{\delta}$. Hence, we can obtain

$$
\sum_{t=1}^{T} \hat{z}_{j,t} z_{i,t} = \sum_{t=1}^{T} \hat{z}_{j,t} \left( \hat{Z}_{i,t} + \hat{u}_{i,t} \right) = \sum_{t=1}^{T} \hat{z}_{j,t} \hat{Z}_{i,t},
$$

(5.4)

for all $i$ and $j$. This means

$$
\sum_{t=1}^{T} \hat{z}_i z_i' = \sum_{t=1}^{T} \hat{Z}_i \hat{Z}_i'.
$$

(5.5)

So, the $i$th column vector of the 2SLS estimator (3.5) $\hat{\beta}_{2SLS,i}$ (say) can be represented as

$$
\hat{\beta}_{2SLS,i} = \left[ \sum_{t=1}^{T} \hat{Z}_i \hat{Z}_i' \right]^{-1} \left[ \sum_{t=1}^{T} \hat{Z}_i y_{i,t} \right],
$$

(5.6)

which leads to

$$
\hat{\beta}_{2SLS,i} - \beta_i = \left[ \sum_{t=1}^{T} \hat{Z}_i \hat{Z}_i' \right]^{-1} \left[ \sum_{t=1}^{T} \hat{Z}_i \epsilon_{i,t} \right].
$$

(5.7)

Hence, we can see that

$$
\sqrt{T} \left( \hat{B}_{2SLS} - B \right) = \left[ \frac{1}{T} \sum_{t=1}^{T} \hat{Z}_i \hat{Z}_i' \right]^{-1} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{Z}_i \epsilon_i' \right].
$$

(5.8)

Note that, by the ergodic theorem (e.g., Stout [15] p179–181),

$$
\frac{1}{T} \sum_{t=1}^{T} \hat{z}_i z_i' = \frac{1}{T} \hat{\delta} \sum_{t=1}^{T} x_i z_i'
$$

$$
= \left[ \frac{1}{T} \sum_{t=1}^{T} z_i x_i' \right] \left[ \frac{1}{T} \sum_{t=1}^{T} x_i x_i' \right]^{-1} \left[ \frac{1}{T} \sum_{t=1}^{T} x_i z_i' \right]
$$

$$
\overset{P}{\longrightarrow} Q.
$$

(5.9)

Furthermore, the second term of the right side of (5.8) can be represented as

$$
\left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{z}_i \epsilon_i' \right] = \hat{\delta} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_i \epsilon_i'.
$$

(5.10)
and by the ergodic theorem (e.g., Stout [15] pp. 179–181), we can see
\[
\delta_i = \left[ \sum_{t=1}^{T} Z_t X_i^t \right] \left[ \sum_{t=1}^{T} X_i X_i^t \right]^{-1} \rightarrow P \left[ E(Z_t X_i^t) \right] \left[ E(X_i X_i^t) \right]^{-1}.
\]

**Proof of (i).** From the above,
\[
\mathbb{B}^{2SLS} - \mathbb{B} = O_P \left[ \frac{1}{T} \sum_{t=1}^{T} X_i \epsilon_i^t \right].
\]

In view of Theorem 1.2 (i) of Hosoya [14], the right-hand side of (5.12) converges to 0 in probability. \qed

**Proof of (ii).** From Theorem 3.2 of Hosoya [14], if Assumption 3.1 holds, it follows that
\[
(1/\sqrt{T}) \sum_{t=1}^{T} X_i \epsilon_i^t \overset{d}{\rightarrow} \mathbb{U}.
\]
Hence, Theorem 3.2 is proved. \qed

### Acknowledgments

The author would like to thank the Editor and the referees for their comments, which improved the original version of this paper.

### References

[1] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons, New York, NY, USA, 1991.

[2] W. Sharpe, “Capital asset prices: a theory of market equilibrium under conditions of risk,” *Journal of Finance*, vol. 19, pp. 425–442, 1964.

[3] J. Lintner, “The valuation of risk assets and the selection of risky in vestments in stock portfolios and capital budgets,” *The Review of Economics and Statistics*, vol. 47, pp. 13–37, 1965.

[4] F. Black, “Capital market equilibrium with restricted borrowing,” *The Journal of Business*, vol. 45, pp. 444–455, 1972.

[5] J. Y. Campbell, A. W. Lo, and A. C. Mackinlay, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ, USA, 1997.

[6] C. W. J. Granger, “Long memory relationships and the aggregation of dynamic models,” *Journal of Econometrics*, vol. 14, no. 2, pp. 227–238, 1980.

[7] P. G. Wright, *The Tarring on Animal and Vegetable Oils*, Macmillan, New York, NY, USA, 1928.

[8] O. Reiersol, “Confluence analysis by means of instrumental sets of variables,” *Arkiv för Matematik, Astronomi Och Fysik*, vol. 32A, no. 4, pp. 1–119, 1945.

[9] R. C. Geary, “Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown,” *Econometrica*, vol. 17, pp. 30–58, 1949.

[10] H. White, *Asymptotic Theory for Econometricians*, Academic Press, New York, NY, USA, 2001.

[11] P. M. Robinson and F. J. Hidalgo, “Time series regression with long-range dependence,” *The Annals of Statistics*, vol. 25, no. 1, pp. 77–104, 1997.

[12] K. Choy and M. Taniguchi, “Stochastic regression model with dependent disturbances,” *Journal of Time Series Analysis*, vol. 22, no. 2, pp. 175–196, 2001.

[13] P. M. Robinson and Y. Yajima, “Determination of cointegrating rank in fractional systems,” *Journal of Econometrics*, vol. 106, no. 2, pp. 217–241, 2002.

[14] Y. Hosoya, “A limit theory for long-range dependence and statistical inference on related models,” *The Annals of Statistics*, vol. 25, no. 1, pp. 105–137, 1997.

[15] W. F. Stout, *Almost Sure Convergence*, Academic Press, New York, NY, USA, 1974.