Statistical Estimation of Optimal Portfolios for non-Gaussian Dependent Returns of Assets

Hiroshi Shiraishi
Waseda University
Masanobu Taniguchi*
Waseda University

Abstract
This paper discusses the asymptotic efficiency of estimators for optimal portfolios when returns are vector-valued non-Gaussian stationary processes. We give the asymptotic distribution of portfolio estimators $\hat{g}$ for non-Gaussian dependent return processes. Next we address the problem of asymptotic efficiency for the class of estimators $\hat{g}$. First, it is shown that there are some cases when the asymptotic variance of $\hat{g}$ under non-Gaussianity can be smaller than that under Gaussianity. The result shows that non-Gaussianity of the returns does not always affect the efficiency badly. Second, we give a necessary and sufficient condition for $\hat{g}$ to be asymptotically efficient when the return process is Gaussian, which shows that $\hat{g}$ is not asymptotically efficient generally. From this point of view we propose to use maximum likelihood type estimators for $g$, which are asymptotically efficient. Furthermore, we investigate the problem of predicting the one step ahead optimal portfolio return by the estimated portfolio based on $\hat{g}$ and examine the mean squares prediction error.

Key words and phrases; optimal portfolio; return process; non-Gaussian linear process; spectral density; asymptotic efficiency; prediction error

1. Introduction

In the theory of portfolio analysis, optimal portfolios are determined by the mean $\mu$ and variance $\Sigma$ of the portfolio return. 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 (e.g. Jobson and Korkie, 1980 and 1989; Lauprete, Samarov and Welsch, 2002). However, empirical studies show that financial return processes are often dependent and non-Gaussian. This symptom leads us to the assumption that financial return processes are dependent and non-Gaussian. From this point of view, Basak, Jagannathan and Sun (2002)
showed the consistency of optimal portfolio estimators when portfolio returns are stationary processes. However, in the literature there has been no study on the asymptotic efficiency of estimators for optimal portfolios. Therefore, in this paper, denoting optimal portfolios by a function $g = g(\mu, \Sigma)$ of $\mu$ and $\Sigma$, we discuss the asymptotic efficiency of estimators $\hat{g} = g(\hat{\mu}, \hat{\Sigma})$ when the return is a vector-valued non-Gaussian stationary process $\{X(t)\}$. Then it is shown that $\hat{g}$ is not asymptotically efficient generally even if $\{X(t)\}$ is Gaussian, which gives a strong warning for use of the usual estimator $\hat{g}$. We also show that there are some cases when the asymptotic variance $V_{NG}(\hat{g})$ of $\hat{g}$ under non-Gaussianity can be smaller than that under Gaussianity $V_G(\hat{g})$. Numerical studies are given to illuminate the results above. For non-Gaussian dependent return processes, we propose to use maximum likelihood type estimators for $g$, which are asymptotically efficient. Furthermore, we investigate the problem of predicting the one step ahead optimal portfolio return by the estimated portfolio based on $\hat{g}$, and evaluate the mean squares prediction error.

Numerical examples are provided. As a conclusion it seems very important to make the consideration for non-Gaussianity and dependence of return processes.

The paper is organized as follows. Section 2 describes optimal portfolios as a function $g = g(\mu, \Sigma)$ of $\mu$ and $\Sigma$. Section 3 gives the asymptotic distribution of $\hat{g}$. Also we evaluate the mean squares prediction error of the estimated portfolio. Section 4 addresses the problem of asymptotic efficiency for the class of estimators $\hat{g}$. First, it is shown that there are some cases satisfying $V_{NG}(\hat{g}) < V_G(\hat{g})$. The result shows that non-Gaussianity of $X(t)$ does not always affect the efficiency badly. Second, we give a necessary and sufficient condition for $\hat{g}$ to be asymptotically efficient when the return process is Gaussian, which shows that $\hat{g}$ is not asymptotically efficient generally. Numerical examples are provided. For non-Gaussian dependent return processes, in Section 5, we propose to use maximum likelihood type estimators for $g$, which are asymptotically efficient. We examine this approach for real financial data.

Throughout this paper, $\|A\|$ denotes the Euclidean norm of a matrix $A$ and $|A|$ denotes the sum of the absolute values of all entries of $A$. We write $X_n \xrightarrow{L} X$ if $\{X_n\}$ converges in distribution to $X$. The ‘vec’ operator transforms a matrix into a vector by stacking columns, and the ‘vech’ operator transforms a symmetric matrix into a vector by stacking elements on and below the main diagonal. For matrices $A$ and $B$, $A \otimes B$ denotes the Kronecker product of $A$ and $B$, whose $(j_1, j_2)$th block is $a_{j_1j_2}B$.

2. Optimal Portfolios

Suppose the existence of a finite number of assets indexed by $i, (i = 1, \ldots, m)$. Let $X(t) = (X_1(t), \ldots, X_m(t))'$ denote the random returns on $m$ assets at time $t$. Assuming the stationarity of $\{X(t)\}$, write $\mu = E\{X(t)\}$ and $\Sigma = \text{Cov}(X(t))$. Let $\alpha = (\alpha_1, \ldots, \alpha_m)'$ be the vector of portfolio weights. Then the return of portfolio is $X(t)\alpha'$, and the expectation and variance are, respectively, given by $\mu(\alpha) = \mu'\alpha$ and $\eta^2(\alpha) = \alpha'\Sigma\alpha$. Optimal portfolio weights have been proposed by various criteria (see Jobson and Korkie 1980 ,Gourieroux 1997 etc.).
followings are typical ones.

Consider

$$\max_{\alpha} \{ \mu(\alpha) - \beta \eta^2(\alpha) \},$$
subject to $e^T \alpha = 1$,

(2.1)

where $e = (1, \ldots, 1)'$ ($m \times 1$-vector), and $\beta$ is a given positive number. The solution for $\alpha$ is given by

$$\alpha_I = \frac{1}{2\beta} \left\{ \Sigma^{-1} \mu - \frac{e' \Sigma^{-1} \mu}{e' \Sigma^{-1} e} \Sigma^{-1} e \right\} + \frac{\Sigma^{-1} e}{e' \Sigma^{-1} e}.$$

(2.2)

Next we consider

$$\min_{\alpha} \eta^2(\alpha),$$
subject to $e^T \alpha = 1$.

(2.3)

The solution for $\alpha$ is given by

$$\alpha_{II} = \frac{\Sigma^{-1} e}{e' \Sigma^{-1} e}.$$

(2.4)

Let us now suppose that there exists a risk-free asset. We denote by $R_0$ its return, and denote by $\alpha_0$ the amount. The problem to be solved is given by

$$\max_{\alpha_0, \alpha} \{ \mu(\alpha) + R_0 \alpha_0 - \beta \eta^2(\alpha) \},$$
subject to $\sum_{j=0}^m \alpha_j = 1$.

(2.5)

Then the solution for $\alpha$ and $\alpha_0$ are

$$\alpha_{III} = \frac{1}{2\beta} \Sigma^{-1} (\mu - R_0 e),$$

$\alpha_{0III} = 1 - \frac{1}{2\beta} e' \Sigma^{-1} (\mu - R_0 e)$.

(2.6)

(2.7)

Therefore optimal portfolios can be considered as smooth functions of $\mu$ and $\Sigma$, i.e. we may put

$$g_1(\mu, \Sigma) = \frac{1}{2\beta} \left\{ \Sigma^{-1} \mu - \frac{e' \Sigma^{-1} \mu}{e' \Sigma^{-1} e} \Sigma^{-1} e \right\} + \frac{\Sigma^{-1} e}{e' \Sigma^{-1} e},$$

(2.2)'

$$g_2(\mu, \Sigma) = \frac{\Sigma^{-1} e}{e' \Sigma^{-1} e},$$

(2.4)'

$$g_3(\mu, \Sigma) = \frac{1}{2\beta} \Sigma^{-1} (\mu - R_0 e),$$

(2.6)'

$$g_4(\mu, \Sigma) = 1 - \frac{1}{2\beta} e' \Sigma^{-1} (\mu - R_0 e).$$

(2.7)'

Unifying the above we consider to estimate a general function $g(\mu, \Sigma)$ of $\mu$ and $\Sigma$. Here it should be noted that the coefficient $\alpha$ satisfies the restriction $e' \alpha = 1$. 

3
Then we have only to estimate the subvector \((\alpha_1, \ldots, \alpha_{m-1})^t\). Hence we assume that the function \(g(\cdot)\) is \((m-1)\)-dimensional, i.e.,

\[
g : (\mu, \Sigma) \rightarrow \mathbb{R}^{m-1}.
\]  
(2.8)

This paper addresses the problem of statistical estimation for \(g(\mu, \Sigma)\), which describes various optimal portfolios.

3. Asymptotic Theory for Fundamental Quantities

As we said in Introduction, empirical studies show that financial return processes are often dependent and non-Gaussian. So it is natural to suppose that the return process concerned is dependent and non-Gaussian. Henceforth we assume that the return process \(\{X(t) = (X_1(t), \ldots, X_m(t))^t; t \in \mathbb{Z}\}\) is an \(m\)-vector non-Gaussian stationary process with mean \(\mu = (\mu_1, \ldots, \mu_m)^t\) and autocovariance matrix \(R(k)\).

Initially, we assume that the return process \(\{X(t) = (X_1(t), \ldots, X_m(t))^t; t \in \mathbb{Z}\}\) is an \(m\)-vector linear process

\[
X(t) = \sum_{j=0}^{\infty} A(j) U(t-j) + \mu, \ t \in \mathbb{Z},
\]  
(3.1)

where \(\{U(t) = (u_1(t), \ldots, u_m(t))^t\}\) is a sequence of independent and identically distributed (i.i.d.) \(m\)-vector random variables with \(E U(t) = 0, Var\{U(t)\} = K\) (for short \(\{U(t)\} \sim \text{i.i.d.}(0, K)\)), and fourth order cumulants. Here

\[
K = \{K_{ab}; a, b = 1, \ldots, m\},
\]  
(3.2)

\[
\mu = \{\mu_a; a = 1, \ldots, m\},
\]  
(3.3)

\[
A(j) = \{A_{ab}(j); a, b = 1, \ldots, m\}, \ j \in \mathbb{Z}, \ A(0) = I_m.
\]  
(3.4)

We make the following assumption.

**Assumption 1.**

(i) \(\sum_{j=0}^{\infty} |j|^{1+\delta}\|A(j)\| < \infty\) for some \(\delta > 0\)

(ii) \(\operatorname{det}\left\{\sum_{j=0}^{\infty} A(j)z^j\right\} \neq 0\) on \(\{z; |z| \leq 1\}\)

The class of \(\{X(t)\}\) includes that of non-Gaussian vector-valued causal ARMA models. Hence the class is sufficiently rich. The process \(\{X(t)\}\) is a second order stationary process with spectral density matrix

\[
f(\lambda) = \{f_{ab}(\lambda); a, b = 1, \ldots, m\} = (2\pi)^{-1} A(\lambda) K A(\lambda)^*,
\]  
(3.5)

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

\[
R(k) = E \left\{ (X(t) - \mu)(X(t+k) - \mu)^t \right\},
\]  
(3.6)
we define \((m+r)\)-vector parameter \(\theta\) by
\[
\theta = (\mu', vech\{R(0)\}')'
\] (3.7)
where \(r = m(m+1)/2\). From the partial realization \(\{X(1), \cdots, X(n)\}\), we introduce
\[
\hat{\mu} = \frac{1}{n} \sum_{t=1}^{n} X(t),
\] (3.8)
\[
\hat{R}(k) = \frac{1}{n-k} \sum_{t=1}^{n-k} \left\{ (X(t) - \hat{\mu})(X(t+k) - \hat{\mu})' \right\},
\] (3.9)
\[
R^*(k) = \frac{1}{n-k} \sum_{t=1}^{n-k} \left\{ (X(t) - \mu)(X(t+k) - \mu)'ight\},
\] (3.10)
\[
\hat{\theta} = (\hat{\mu}', vech\{\hat{R}(0)\}')'.
\] (3.11)

Denote the \(\sigma\)-field generated by \(\{X(s); s \leq t\}\) by \(\mathcal{F}_t\). Also we introduce matrices;
\[
\Omega_1 = 2\pi f(0), (m \times m) - matrix
\]
\[
\Omega_2 = \left\{ 2\pi \int_{-\pi}^{\pi} \left\{ f_{a_1a_3}(\lambda)f_{a_2a_4}(\lambda) + f_{a_1a_4}(\lambda)f_{a_2a_3}(\lambda) \right\} d\lambda
\right. 
\left. + \frac{1}{(2\pi)^2} \sum_{b_1, \cdots, b_4=1}^{m} c^{U}_{b_1, \cdots, b_4} \int \int_{-\pi}^{\pi} A_{a_1b_1}(\lambda_1)A_{a_2b_2}(-\lambda_1)A_{a_3b_3}(\lambda_2)A_{a_4b_4}(-\lambda_2)d\lambda_1d\lambda_2
\right. 
\left. ; a_1, a_2, a_3, a_4 = 1, \ldots, m, a_1 \geq a_2 \text{ and } a_3 \geq a_4 \right\}, (r \times r) - matrix,
\]
\[
\Omega_3 = \left\{ \frac{1}{(2\pi)^2} \sum_{b_1, b_2, b_3=1}^{m} c^{U}_{b_1, b_2, b_3} \int \int_{-\pi}^{\pi} A_{a_1b_1}(\lambda_1 + \lambda_2)A_{a_2b_2}(-\lambda_1)A_{a_3b_3}(-\lambda_2)d\lambda_1d\lambda_2
\right. 
\left. ; a_1, a_2, a_3 = 1, \ldots, m, a_2 \geq a_3 \right\}, (m \times r) - matrix,
\]
where \(c^{U}_{b_1, \cdots, b_4}\)'s are \(j\)th order cumulants of \(U_{b_1}(t), \ldots, U_{b_4}(t) \ (j = 3, 4)\). Then we have the following result.

**Theorem 1.** Under Assumption 1,
\[
\sqrt{n}(\hat{\theta} - \theta) \overset{d}{\to} N(0, \Omega_{NG}),
\] (3.12)
where
\[
\Omega_{NG} = \begin{pmatrix} \Omega_1 & \Omega_3 \\ \Omega_3' & \Omega_2 \end{pmatrix}.
\] (3.13)

For \(g\) given by (2.8) we impose the following.

**Assumption 2.** The function \(g(\theta)\) is continuously differentiable.

As a unified estimator for optimal portfolios we introduce \(g(\hat{\theta})\). For this we have the following result.
Theorem 2. Under Assumptions 1 and 2,
\[
\sqrt{n}(g(\hat{\theta}) - g(\theta)) \overset{L}{\to} N \left( 0, \left( \frac{\partial g}{\partial \theta} \right) \Omega_N G \left( \frac{\partial g}{\partial \theta} \right)' \right),
\]
(3.14)
where \( \left( \frac{\partial g}{\partial \theta} \right) \) is the vector differentiation (see Magnus and Neudecker 1988).

Here we investigate the problem of predicting the one step ahead optimal portfolio return by the estimated portfolio based on \( \hat{g} \).

Assume that \( \{X(1), \ldots, X(n)\} \) is a realization of the \( m \)-vector linear process (3.1), and let \( \{Y(1), \ldots, Y(n)\} \) be an independent realization of the same process. If \( \hat{\theta} \) is defined by (3.11) and if we use the following
\[
\hat{P}R(n) = Y(n)'g(\hat{\theta}) \quad (3.15)
\]
as a predictor of \( PR(n+1) \equiv Y(n+1)'g(\theta) \), then the mean-square prediction error (\( PE \)) is
\[
PE = \mathbb{E}\{PR(n+1) - \hat{P}R(n)\}^2
\]
\[
= 2g(\theta)'(R(0) - R(1)) g(\theta) \quad (3.16)
\]
where
\[
B_n = \sqrt{n} \mathbb{E}\{g(\hat{\theta}) - g(\theta)\} = o(1)
\]
\[
C_n = n \mathbb{E}\{g(\hat{\theta}) - g(\theta)\} \left( g(\hat{\theta}) - g(\theta) \right)' = \left( \frac{\partial g}{\partial \theta} \right) \Omega_N G \left( \frac{\partial g}{\partial \theta} \right)' + o(1).
\]

We evaluate \( PE \) for various spectral structures, numerically.

Example 1(Prediction Error (PE)) Let \( X(1), \ldots, X(100) \) be an observed stretch from the return process \( \{X(t) = (X_1(t), X_2(t))'; t \in \mathbb{Z} \} \) generated by
\[
(1 - \alpha B)X_1(t) = (1 - \beta B)U(t) + \mu_{X_1}, \quad (3.17)
\]
\[
X_2(t) = \mu_{X_2}, \quad (3.18)
\]
where \( U(t) \sim i.i.d. T(10) \) and \( T(10) \) is \( t \)-distribution with 10 degrees of freedom.
Let the portfolio function be defined by
\[
g(\mu_{X_1}, \mu_{X_2}, R_{X_1}(0)) = \frac{\mu_{X_1} - \mu_{X_2}}{2R_{X_1}(0)}. \quad (3.19)
\]
This portfolio is one of the solution of (2.5). In this case we estimate the \( PE \) by
\[
\hat{PE} = 2g(\theta)' \left( R_{X_1}(0) - R_{X_1}(1) \right) g(\theta) \quad (\equiv PE1)
\]
\[
+ \frac{2}{\sqrt{100}} \hat{B}'_{100} \left( R_{X_1}(0) - R_{X_1}(1) \right) g(\theta) \quad (\equiv PE2)
\]
\[
+ \frac{1}{\sqrt{100}} \operatorname{tr} \left[ (\mu \mu' + R_{X_1}(0)) \hat{C}_{100} \right] \quad (\equiv PE3)
\]
where

\[
\hat{B}_{100} = \frac{1}{\sqrt{100}} \sum_{t=1}^{100} \left\{ g(\hat{\theta}_t) - g(\theta) \right\}
\]
\[
\hat{C}_{100} = \sum_{t=1}^{100} \left\{ g(\hat{\theta}_t) - g(\theta) \right\}^2
\]

Figure 1 shows the graph of $PE_2 + PE_3$ for $\alpha = -0.8(0.2)0.8$, $\beta = -0.8(0.2)0.8$, $\mu_{X_1} - \mu_{X_2} = 0.3$. We observe that if $\beta = -0.2$, and if $\alpha \nearrow 1$, then $PE$ increases.

Figure 1 is about here.

### 4. Asymptotic Efficiency of Estimators of Optimal Portfolios

In this section, we discuss the problem of asymptotic efficiency for the class of estimators $\hat{g}$. First, we compare the asymptotic variance of $\hat{g}$ under non-Gaussianity with that under Gaussianity. Second, we discuss the asymptotic efficiency of $\hat{g}$ when the return process is Gaussian.

If $\{U(t)\} \sim i.i.d.N(0, K)$, i.e., $\{X(t)\}$ is Gaussian, then it follows from Theorem 2 that

\[
\sqrt{n}(\hat{g}(\hat{\theta}) - g(\theta)) \xrightarrow{d} N \left( 0, \left( \frac{\partial g}{\partial \theta'} \right) \Omega_G \left( \frac{\partial g}{\partial \theta'} \right)' \right).
\]

(4.1)

where

\[
\Omega_G = \begin{pmatrix}
\Omega_1 & 0 \\
0 & \tilde{\Omega}_2
\end{pmatrix},
\]

(4.2)

and

\[
\tilde{\Omega}_2 = \left\{ 2\pi \int_{-\pi}^{\pi} \left\{ f_{a_1a_3}(\lambda) f_{a_2a_4}(\lambda) + f_{a_1a_4}(\lambda) f_{a_2a_3}(\lambda) \right\} d\lambda : a_1, a_2, a_3, a_4 = 1, \ldots, m, a_1 \geq a_2 and a_3 \geq a_4 \right\}, (r \times r) - matrix.
\]

First, we evaluate

\[
\mu'(V_{NG} - V_G) \mu
\]

(4.3)

where

\[
V_{NG} = \left( \frac{\partial g}{\partial \theta'} \right)' \Omega_{NG} \left( \frac{\partial g}{\partial \theta'} \right),
\]

(4.4)

\[
V_G = \left( \frac{\partial g}{\partial \theta'} \right)' \Omega_G \left( \frac{\partial g}{\partial \theta'} \right),
\]

(4.5)

for various optimal portfolios and spectral structures.
Example 2(VMA(1)model)  Let the return process be generated by
\[
X(t) = \begin{pmatrix} 1 - xB & 0 \\ 0 & 1 - yB \end{pmatrix} U(t) + \mu, \quad (|x| < 1, |y| < 1), \quad (4.6)
\]
where \(U(t) \equiv \begin{pmatrix} u_1(t) - \kappa_1 \\ u_2(t) - \kappa_2 \end{pmatrix}\). Here \(u_i(t) \sim i.i.d. E_X(\kappa_i), (i = 1, 2), E_X(\kappa_i)\) is the exponential distribution with mean \(\kappa_i\), and \(B\) is the lag operator. For \(x = 0.4, y = 0.6, \mu_1 = 0.1, \mu_2 = 0.3, \beta = 0.5, R_0 = 0.01, \kappa_1, \kappa_2 = 1.0(1.0)5.0\) in the case of \(g_3((2.6)')\) we calculated \(\mu'(V_{NG} - V_G)\mu\).

Table 1 is about here.

From Table 1 it is seen that, for some values of \(\kappa_1\) and \(\kappa_2\), \(\mu'(V_{NG} - V_G)\mu < 0\).

Example 3(VAR(1)model)  Let the return process be generated by
\[
X(t) = \begin{pmatrix} 1 - xB & 0 \\ 0 & 1 - yB \end{pmatrix} U(t) + \mu \quad (4.7)
\]
where \(\{U(t)\}\) is the same process as in Example 2. For \(x = 0.4, y = 0.6, \mu_1 = 0.1, \mu_2 = 0.3, \beta = 0.5, R_0 = 0.01, \kappa_1, \kappa_2 = 1.0(1.0)5.0\) in the cases of \(g_1((2.2)')\) and \(g_3\) we calculated \(\mu'(V_{NG} - V_G)\mu\).

Tables 2 and 3 are about here.

From Tables 2 and 3 we can see that, for some values of \(\kappa_1\) and \(\kappa_2\), \(\mu'(V_{NG} - V_G)\mu < 0\).

The above examples illuminate an interesting feature of Gaussian and non-Gaussian asymptotics of \(g(\hat{\theta})\).

Next we discuss the asymptotic Gaussian efficiency of \(g(\hat{\theta})\). Fundamental results concerning the asymptotic efficiency of sample autocovariance matrices of vector Gaussian processes were obtained by Kakizawa(1999). He compared the asymptotic variance(AV) of sample autocovariance matrices with the inverse of the corresponding Fisher information matrix(\(F^{-1}\)), and gave the condition for the asymptotic efficiency (i.e., condition for \(AV = F^{-1}\)). Based on this we will discuss the asymptotic efficiency of \(\hat{g}\).

Suppose that \(\{X(t)\}\) is a zero-mean Gaussian \(m\)-vector stationary process with spectral density matrix \(f(\lambda)\), and satisfies following assumptions.

Assumption 3.  
(i) \(f(\lambda)\) is parameterized by \(\eta = (\eta_1, \ldots, \eta_q)' \in \mathcal{H} \subset \mathbb{R}^q\) i.e., \(f_\eta = f_\eta(\lambda)\).
(ii) For \(j = 1, \ldots, q, \int_\pi^\pi \frac{\partial f_\eta(\lambda)}{\partial \eta_j} d\lambda\) can be expressed as a summable \(\sum_{l=-\infty}^{\infty} A^{(j)}(l)\) satisfying \(\sum_{l=-\infty}^{\infty} \|A^{(j)}(l)\| < \infty\).
(iii) \(q \geq m(m + 1)/2\).
**Assumption 4.** There exists a positive constant $c$ (independent of $\lambda$) such that $\mathbf{f}_\eta(\lambda) - c\mathbf{I}_m$ is positive semi-definite, where $\mathbf{I}_m$ is an $m \times m$ identity matrix.

The limit of averaged Fisher information matrix is given by

$$\mathcal{F}(\eta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \Delta(\lambda)^*\{[\mathbf{f}_\eta(\lambda)^{-1}]' \otimes \mathbf{f}_\eta(\lambda)^{-1}\}\Delta(\lambda) d\lambda$$  \hspace{1cm} (4.8)

where

$$\Delta(\lambda) = \begin{pmatrix} \text{vec}\{\partial \mathbf{f}_\eta(\lambda)/\partial \eta_1\} & \ldots & \text{vec}\{\partial \mathbf{f}_\eta(\lambda)/\partial \eta_q\} \end{pmatrix}$$ \hspace{1cm} (m^2 \times q) - \text{matrix}.$$

**Assumption 5.** The matrix $\mathcal{F}(\eta)$ is positive definite.

We introduce an $m^2 \times m^2$ matrix

$$\Phi = \begin{pmatrix} \text{vec}\{\psi_{11}\} & \ldots & \text{vec}\{\psi_{m1}\} \\
\vdots & \ddots & \vdots \\
\text{vec}\{\psi_{1m}\} & \ldots & \text{vec}\{\psi_{mm}\} \end{pmatrix},$$

where $\psi_{ab} = \frac{1}{2} \left( \text{E}_{ab} + \text{E}_{ba} \right)$ and $\text{E}_{ab}$ is a matrix of which $(a, b)$th element is 1 and others are 0. Then we have the following theorem.

**Theorem 3.** Under Assumptions 1-5, $g(\hat{\theta})$ is asymptotically efficient if and only if there exists a matrix $C$ (independent of $\lambda$) such that

$$\{\mathbf{f}_\eta(\lambda)' \otimes \mathbf{f}_\eta(\lambda)\} \Phi = \{\text{vec}\{\partial \mathbf{f}_\eta(\lambda)/\partial \eta_1\} & \ldots & \text{vec}\{\partial \mathbf{f}_\eta(\lambda)/\partial \eta_q\}\} C.$$  \hspace{1cm} (4.9)

Theorem 3 implies that if (4.9) is not satisfied, the estimator $g(\hat{\theta})$ is not asymptotically efficient. This is a strong warning to use of the ordinary portfolio estimators for even Gaussian dependent returns. The interpretation of (4.9) is difficult. But, Kakizawa(1999) showed that (4.9) is satisfied by vector AR(p) models with coefficients $\eta$.

The followings are examples, which do not satisfy (4.9).

**Example 4(VARMA($p_1, p_2$)process)**  Consider the $m \times m$ spectral density matrix of n-vector ARMA($p_1, p_2$) process,

$$\mathbf{f}(\lambda) = \frac{1}{2\pi} \Theta\{\exp(i\lambda)\}^{-1} \Psi\{\exp(i\lambda)\} \Sigma \Psi\{\exp(i\lambda)\}^* \Theta\{\exp(i\lambda)\}^{-1*}$$  \hspace{1cm} (4.10)

where $\Psi(z) = I_m - \Psi_1 z - \cdots - \Psi_{p_2} z^{p_2}$, and $\Theta(z) = I_m - \Theta_1 z - \cdots - \Theta_{p_1} z^{p_1}$ satisfy $\det \Psi(z) \neq 0$, $\det \Theta(z) \neq 0$ for all $|z| \leq 1$. From Kakizawa(1999) it follows that (4.10) does not satisfy (4.9), if $p_1 < p_2$, hence $g(\hat{\theta})$ is not asymptotically efficient if $p_1 < p_2$.

**Example 5(An exponential model)**  Consider the $m \times m$ spectral density matrix of exponential type,

$$\mathbf{f}(\lambda) = \exp \left\{ \sum_{j \neq 0} A_j \cos(j\lambda) \right\}$$  \hspace{1cm} (4.11)
where \( A_j \)'s are \( m \times m \)-matrices, and \( \exp \{ \cdot \} \) is the matrix exponential (for the definition, see Bellman[2,p.169]). Since (4.11) does not satisfy (4.9), \( g(\hat{\theta}) \) is not asymptotically efficient.

In view of the above we should be careful when we use the usual portfolio estimators \( g(\hat{\theta}) \) even if the return process is Gaussian.

Next we are interested in the degree of inefficiency of \( g(\hat{\theta}) \). Note that

\[
V_G = \text{minimum variance of } g(\hat{\theta})
\]

\[
= \left( \frac{\partial g}{\partial \theta} \right) \left( \Omega_1 \ 0 \ 0 \right) \left( \frac{\partial g}{\partial \theta} \right)^\prime - \left( \frac{\partial g}{\partial \theta} \right) \left( \Omega_1 \ 0 \ F(\eta)^{-1} \right) \left( \frac{\partial g}{\partial \theta} \right)^\prime
\]

\[
= \left( \frac{\partial g}{\partial \theta} \right) \left( 0 \ 0 \ 0 \left( \Omega_2 - F(\eta)^{-1} \right) \right) \left( \frac{\partial g}{\partial \theta} \right)^\prime.
\]

In what follows we numerically evaluate

\[
INE \equiv \det \left[ \Omega_2 - F(\eta)^{-1} \right]
\]

(4.12)

for various spectra.

**Model I** (VMA(1)model). Let the return process be generated by

\[
X(t) = \begin{pmatrix} 1 - \eta_1 B & 0 \\ 0 & 1 - \eta_1 B \end{pmatrix} U(t), \ U(t) \sim i.i.d. N \left( 0, \left( \begin{array}{cc} 0.5 & 0.1 \\ 0.1 & 0.5 \end{array} \right) \right).
\]

(4.13)

In Figure 2 we plotted the graph of \( INE = INE(VMA(1)) \) for \( \eta_1 = -0.8(0.2)0.8 \). We can see that, as \( |\eta_1| \) tends to 1, \( INE \) increases.

**Figure 2 is about here.**

**Model II** (VARMA(1,2)model). Let the return process be generated by

\[
\begin{pmatrix} 1 - \eta_1 B \\ 0 \end{pmatrix} X(t) = \begin{pmatrix} (1 - \eta_2 B)(1 - \eta_3 B) & 0 \\ 0 & (1 - \eta_2 B)(1 - \eta_3 B) \end{pmatrix} U(t), \ U(t) \sim i.i.d. N \left( 0, \left( \begin{array}{cc} 0.5 & 0.1 \\ 0.1 & 0.5 \end{array} \right) \right).
\]

(4.14)

In Figure 3 we plotted the graph of \( INE = INE(VARMA(1,2)) \) for \( \eta_1 = -0.8(0.2)0.8, \eta_2 = 0.01, \eta_3 = 0.5 \). We can see that if \( \eta_1 \searrow 1 \), \( INE \) becomes quite large.
Figure 3 is about here.

Model III (VARMA(1,2) model). Let the model be generated by (4.14) with \( \eta_1 = 0.01 \) and \( \eta_2 = 0.5 \). In Figure 4 we plotted the graph of \( \text{INE} = \text{INE(VARMA(1,2))} \) for \( \eta_3 = -0.8(0.2)0.8 \). We can see that as \( \eta_3 \nearrow 1 \), \( \text{INE} \) increases.

Figure 4 is about here.

Summarizing the above we observe that

(i) For VMA(1) model, \( \text{INE} \) increases as the absolute value of the MA coefficient \( \eta_1 \) tends to 1.

(ii) For VARMA(1,2) model with the MA coefficient \( \eta_2 \approx 0 \), \( \text{INE} \) increases as the AR coefficient \( \eta_1 \) tends to \( -1 \).

(iii) For VARMA(1,2) model with the AR coefficient \( \eta_1 \approx 0 \), \( \text{INE} \) increases as the MA coefficient \( \eta_3 \) tends to 1.

Although we just examined a few examples of dependent returns, the above studies show inefficiency of the usual portfolio estimators. Also it should be noted that the degree of inefficiency becomes quite large if some parameters tend to a boundary value.

5. Construction of Efficient Estimators

Let \( \{X(t)\} \) be the linear process defined by Sections 3 and 4 with spectral density matrix \( f_\eta(\lambda), \eta \in \mathcal{H} \). We assume \( \eta = \text{vech}(R(0)) \). Denote by \( I_x(\lambda) \), the periodogram matrix constructed from a partial realization \( \{X(1), \ldots, X(n)\} \);

\[
I_x(\lambda) = F_x(\lambda)F_x(\lambda)^*, \quad \text{with} \quad F_x(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{n} X(t)e^{it\lambda}, \quad -\pi \leq \lambda \leq \pi. \quad (5.1)
\]

To estimate \( \eta \) we introduce

\[
D(f_\eta, I_x) = \int_{-\pi}^{\pi} [\log \det f_\eta(\lambda) + \text{tr} f_\eta^{-1}(\lambda)I_x(\lambda)]d\lambda, \quad (5.2)
\]

(e.g., Hosoya and Taniguchi(1982)). A quasi-Gaussian maximum likelihood estimator \( \hat{\eta} \) of \( \eta \) is defined by

\[
\hat{\eta} = \arg \min_{\eta \in \mathcal{H}} D(f_\eta, I_x). \quad (5.3)
\]

Hosoya and Taniguchi(1982) showed that
From Hosoya and Taniguchi (1982) and Taniguchi and Kakizawa (2000) it is seen 

\[ 
\{X(t)\} \text{ is Gaussian asymptotically efficient. In (5.5), we can express } 
\]

\[ \eta \text{ respect to } g \text{ procedure is } \]

\[ \hat{\eta}(1) = \text{vech}(\hat{R}(0)) \quad (5.4) \]

\[ \hat{\eta}(k) = \hat{\eta}(k-1) - \left[ \frac{\partial^2 D(\eta, I_x)}{\partial \eta \partial \eta'} \right]^{-1} \frac{\partial D(\eta, I_x)}{\partial \eta} \bigg|_{\eta = \hat{\eta}(k-1)} \quad (k \geq 2) \quad (5.5) \]

\[ \text{From Hosoya and Taniguchi (1982) and Taniguchi and Kakizawa (2000) it is seen that } \hat{\eta}(2) \text{ is asymptotically efficient. Therefore, } g(\hat{\theta}), \hat{\theta} = (\hat{\mu}', \hat{\eta}(2)'), \text{ becomes Gaussian asymptotically efficient. In (5.5), we can express} \]

\[ \frac{\partial D(\eta, I_x)}{\partial \eta_{ij}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left[ \eta(\lambda)^{-1} E_{ij} (I_m - \eta(\lambda)^{-1} I_x(\lambda)) \right] d\lambda, \quad (5.6) \]

\[ \frac{\partial^2 D(\eta, I_x)}{\partial \eta_{ij} \partial \eta_{kl}} = \left( \frac{1}{2\pi} \right)^2 \int_{-\pi}^{\pi} \text{tr} \left[ -\eta(\lambda)^{-1} E_{kl} \eta(\lambda)^{-1} E_{ij} (I_m - \eta(\lambda)^{-1} I_x(\lambda)) + \eta(\lambda)^{-1} E_{ij} \eta(\lambda)^{-1} E_{kl} \eta(\lambda)^{-1} I_x(\lambda) \right] d\lambda, \quad (5.7) \]

where \( \eta_{ij} \) is the \((i, j)\)th element of \( \eta \) (0). To make the step (5.5) feasible, we may replace \( f_\eta \) in (5.6) and (5.7) by a nonparametric spectral estimator

\[ f_\eta(\lambda) = \int_{-\pi}^{\pi} W_n(\lambda - \mu) I_x(\mu) d\mu, \quad (5.8) \]

where \( W_n(\theta) \) satisfies appropriate regularity conditions (see Taniguchi and Kakizawa (2000)).

Next we discuss construction of efficient estimator when \( \{X(t)\} \) is non-Gaussian. Fundamental results concerning the asymptotic efficiency of estimators for unknown parameter \( \theta \) of vector linear processes were obtained by Taniguchi and Kakizawa (2000). They showed the quasi-maximum likelihood estimator \( \hat{\theta}_{QML} \) is asymptotically efficient.

Let \( \{X(t)\} \) be a non-Gaussian m-vector stationary process defined by

\[ X(t) = \sum_{j=0}^{\infty} A_\theta(j) U(t-j) + \mu, t \in \mathbb{Z}, \quad (5.9) \]

where \( U(t) \)'s are i.i.d. m-vector random variables with probability density \( p(u) > 0 \) on \( \mathbb{R}^m \), and \( A_\theta(j) = \{A_{ab}(j); a, b = 1, \ldots, m\}, j \in \mathbb{Z} \), are \( m \times m \) matrices depending on a parameter vector \( \theta = (\theta_1, \ldots, \theta_q) \in \Theta \subset \mathbb{R}^q \). Here \( A_\theta(0) = I_m \), the \( m \times m \) identity matrix. Then a quasi maximum likelihood estimator \( \hat{\theta}_{QML} \) is defined as a solution of the equation

\[ \frac{\partial}{\partial \theta} \left[ \sum_{t=1}^{n} \log p \left\{ \sum_{j=0}^{t-1} B_\theta(j) X(t-j) - \mu \right\} \right] = 0 \quad (5.10) \]
with respect to $\theta$. Under some regularity conditions, Taniguchi and Kakizawa proved that $\hat{\theta}_{QML}$ is asymptotically efficient.

As we saw in the previous paragraph the theory of linear time series is well established. However, it is not sufficient for linear models such as ARMA models to describe the real world. Using ARCH models Hafner (1998) gave an extensive analysis for finance data. The results by Hafner reveal that many relationships in real data are “nonlinear”. Thus the analysis of nonlinear time series is becoming an important component of time series analysis. H"ardle et al. (1998) introduced vector conditional heteroscedastic autoregressive nonlinear (ChARN) models, which include usual AR and ARCH models. Kato et al. (2006) established the LAN property for CHARN model.

Suppose that $\{X(t)\}$ is generated by

$$X(t) = F_\theta(X(t-1), \ldots, X(t-p_1), \mu) + H_\theta(X(t-1), \ldots, X(t-p_2))U(t), \quad (5.11)$$

where $F_\theta : \mathbb{R}^{m(p_1+1)} \to \mathbb{R}^m$ is a vector-valued measurable function, $H_\theta : \mathbb{R}^{mp_2} \to \mathbb{R}^m \times \mathbb{R}^m$ is a positive definite matrix-valued measurable function, and $\{U(t) = (u_1(t), \ldots, u_m(t))^T\}$ is a sequence of i.i.d. random variables with $E[U(t) = 0], E[U(t)] < \infty$ and $U(t)$ is independent of $\{X(s), s < t\}$. Henceforth, without loss of generality we assume $p_1 + 1 = p_2 (= p)$, and make the following assumptions.

**Assumption 6**

(A.1) $U(t)$ has a density $p(u) > 0$ a.e. on $\mathbb{R}^m$.

(A.2) There exist constants $a_{ij} \geq 0, b_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq p$ such that, as $|x| \to \infty$,

$$|F_\theta(x)| \leq \sum_{i=1}^m \sum_{j=1}^p a_{ij} |x_{ij}| + o(|x|) \quad (5.12)$$

$$|H_\theta(x)| \leq \sum_{i=1}^m \sum_{j=1}^p b_{ij} |x_{ij}| + o(|x|). \quad (5.13)$$

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

$$\lambda_{\min}\{H_\theta(x)\} \geq \lambda \quad \text{for all } x \in \mathbb{R}^{mp}, \quad (5.14)$$

where $\lambda_{\min}\{\cdot\}$ is the minimum eigenvalue of $\cdot$.

(A.4) $\max_{1 \leq i \leq m}\{\sum_{j=1}^p a_{ij} + E|U(1)| \sum_{j=1}^p b_{ij}\} < 1$.

**Assumption 7**

(B.1) $E_\theta\|F_\theta(X(t-1), \ldots, X(t-p), \mu)\|^2 < \infty$,

$$E_\theta\|H_\theta(X(t-1), \ldots, X(t-p))\| < \infty \quad \text{for all } \theta \in \Theta.$$

(B.2) There exists $c > 0$ such that

$$c \leq \|H_\theta^{1/2}(x)H_\theta(x)H_\theta^{1/2}(x)\| < \infty, \quad (5.15)$$

for all $\theta, \theta' \in \Theta$ and for all $x \in \mathbb{R}^{mp}$.
(B.3) $H_\theta$ and $F_\theta$ are continuously differentiable with respect to $\theta$, and their derivatives $\partial_j H_\theta$ and $\partial_j F_\theta (\partial_j = \partial/\partial \theta_j), j = 1, \ldots, q$, satisfy the condition that there exist square-integrable functions $A_j$ and $B_j$ such that

$$\|\partial_j H_\theta\| \leq A_j \quad \text{and} \quad \|\partial_j F_\theta\| \leq B_j, \quad (j = 1, \ldots, q), \text{for all } \theta \in \Theta. \quad (5.16)$$

(B.4) $p(\cdot)$ satisfies

$$\lim_{\|u\| \to \infty} \|u\| p(u) = 0, \quad \text{and} \quad \int uu' p(u) du = I_m, \quad (5.17)$$

where $I_m$ is the $m \times m$ identity matrix.

(B.5) The continuous derivative $Dp$ of $p(\cdot)$ exists on $\mathbb{R}^m$, and

$$\int \|p^{-1}Dp\|^4 p(u) du < \infty, \quad \text{and} \quad \int \|u\|^2 \|p^{-1}Dp\|^2 p(u) du < \infty. \quad (5.18)$$

Suppose that an observed stretch $X^{(n)} = \{X(1), \ldots, X(n)\}$ from (5.11) is available. We denote the probability distribution of $X^{(n)}$ by $P_{n, \theta}$. For two hypothetical values $\theta, \theta' \in \Theta$, the log-likelihood ratio based on $X^{(n)}$ is

$$\Lambda_n(\theta, \theta') = \log \frac{dP_{n, \theta'}}{dP_{n, \theta}}, \quad (5.19)$$

and

$$= \sum_{t=p}^n \log \frac{p[H_{\theta'}^{-1}(X(t) - F_{\theta'})]}{p[H_{\theta}^{-1}(X(t) - F_{\theta})]} \det H_{\theta}. \quad (5.20)$$

As an estimator of $\theta$, we use the following maximum likelihood estimator (MLE):

$$\hat{\theta}_{ML} \equiv \arg \max_\theta \Lambda_n(\theta_0, \theta), \quad (5.21)$$

where $\theta_0 \in \Theta$ is some fixed value. Write $\eta_t(\theta) \equiv \log p[H_{\theta}^{-1}(X(t) - F_{\theta})] \times \{\det H_{\theta}\}^{-1}$. We further impose the following assumption.

**Assumption 8.** For $i, j, k = 1, \ldots, q$, there exist functions $Q_{ij}^t = Q_{ij}^t(X(1), \ldots, X(t))$ and $T_{ijk}^t = T_{ijk}^t(X(1), \ldots, X(t))$ such that

$$|\partial_i \partial_j \eta_t(\theta)| \leq Q_{ij}^t, \quad E Q_{ij}^t < \infty, \quad (5.22)$$

and

$$|\partial_i \partial_j \partial_k \eta_t(\theta)| \leq T_{ijk}^t, \quad E T_{ijk}^t < \infty. \quad (5.23)$$

Then it is shown that under Assumptions 6-8, the MLE $\hat{\theta}_{ML}$ is asymptotically efficient (see Kato et al. (2006)).
References


Table 1: VMA(1) model (the case of $g_3$)

<table>
<thead>
<tr>
<th>$\kappa_1 \backslash \kappa_2$</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>-0.00978</td>
<td>-0.00433</td>
<td>-0.00270</td>
<td>-0.00221</td>
<td>-0.00202</td>
</tr>
<tr>
<td>2.0</td>
<td>-0.00823</td>
<td>-0.00278</td>
<td>-0.00115</td>
<td>-0.00066</td>
<td>-0.00047</td>
</tr>
<tr>
<td>3.0</td>
<td>-0.00806</td>
<td>-0.00261</td>
<td>-0.00098</td>
<td>-0.00049</td>
<td>-0.00030</td>
</tr>
<tr>
<td>4.0</td>
<td>-0.00802</td>
<td>-0.00257</td>
<td>-0.00093</td>
<td>-0.00044</td>
<td>-0.00025</td>
</tr>
<tr>
<td>5.0</td>
<td>-0.00800</td>
<td>-0.00255</td>
<td>-0.00092</td>
<td>-0.00043</td>
<td>-0.00024</td>
</tr>
</tbody>
</table>

Table 2: VAR(1) model (the case of $g_1$)

<table>
<thead>
<tr>
<th>$\kappa_1 \backslash \kappa_2$</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>-0.04410</td>
<td>-0.01584</td>
<td>-0.00653</td>
<td>-0.00320</td>
<td>-0.00178</td>
</tr>
<tr>
<td>2.0</td>
<td>-0.01675</td>
<td>-0.00736</td>
<td>-0.00309</td>
<td>-0.00233</td>
<td>-0.00188</td>
</tr>
<tr>
<td>3.0</td>
<td>-0.00717</td>
<td>-0.00231</td>
<td>0.00495</td>
<td>0.00520</td>
<td>0.00339</td>
</tr>
<tr>
<td>4.0</td>
<td>-0.00359</td>
<td>-0.00171</td>
<td>0.00703</td>
<td>0.01111</td>
<td>0.01039</td>
</tr>
<tr>
<td>5.0</td>
<td>-0.00201</td>
<td>-0.00164</td>
<td>0.00564</td>
<td>0.01244</td>
<td>0.01481</td>
</tr>
</tbody>
</table>

Table 3: VAR(1) model (the case of $g_3$)

<table>
<thead>
<tr>
<th>$\kappa_1 \backslash \kappa_2$</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>-0.11537</td>
<td>-0.02109</td>
<td>-0.01053</td>
<td>-0.00786</td>
<td>-0.00690</td>
</tr>
<tr>
<td>2.0</td>
<td>-0.11026</td>
<td>-0.01598</td>
<td>-0.00542</td>
<td>-0.00275</td>
<td>-0.00179</td>
</tr>
<tr>
<td>3.0</td>
<td>-0.10973</td>
<td>-0.01545</td>
<td>-0.00489</td>
<td>-0.00222</td>
<td>-0.00126</td>
</tr>
<tr>
<td>4.0</td>
<td>-0.10960</td>
<td>-0.01532</td>
<td>-0.00476</td>
<td>-0.00209</td>
<td>-0.00113</td>
</tr>
<tr>
<td>5.0</td>
<td>-0.10955</td>
<td>-0.01527</td>
<td>-0.00471</td>
<td>-0.00205</td>
<td>-0.00108</td>
</tr>
</tbody>
</table>
Figure 1: PE2+PE3

Figure 2: Model I: $\eta_1 = -0.8(0.2)0.8$
Figure 3: Model II: $\eta_1 = -0.8(0.2)0.8, \eta_2 = 0.01, \eta_3 = 0.5$

Figure 4: Model III: $\eta_1 = 0.01, \eta_2 = 0.5, \eta_3 = -0.8(0.2)0.8$