RESAMPLING PROCEDURE IN ESTIMATION OF OPTIMAL PORTFOLIOS FOR TIME-VARYING ARCH PROCESSES

HIROSHI SHIRAISHI

Abstract. This paper discusses a resampling procedure in estimation of optimal portfolios when the returns are the class of nonstationary ARCH models with time-varying parameters. The asymptotic properties of weighted Gaussian quasi maximum likelihood estimators $\hat{\boldsymbol{\theta}}_{GQML}$ of time-varying ARCH(p) processes are studied, including asymptotic normality. In particular, the extra bias due to nonstationarity of the process is investigated. We consider bias adjusted estimators $\boldsymbol{\theta}^*_{GQML}$ by use of resampling. In this paper we assume that the optimal portfolio weight g depends on the ARCH parameter $\boldsymbol{\theta}$, i.e., $g = g(\boldsymbol{\theta})$. Then the asymptotic distribution of the optimal portfolio estimator $g(\boldsymbol{\theta}^*_{GQML})$ is derived. We numerically evaluate the magnitude of $g(\hat{\boldsymbol{\theta}}_{GQML})$ and $g(\boldsymbol{\theta}^*_{GQML})$ for actual financial data, which shows eventually the effect of bias adjustment.

1. Introduction

In the theory of portfolio analysis, optimal portfolio weights are determined by the mean $\boldsymbol{\mu}$ and variance Σ of the portfolio return. Several authors proposed estimators of the optimal portfolio weights as functions of the sample mean $\hat{\boldsymbol{\mu}}$ and the sample variance $\hat{\Sigma}$ for independent returns of assets. However, empirical studies show that financial return processes are often dependent. ¿From this point of view, Shiraishi (2005), Shiraishi and Taniguchi (2008) considered the asymptotic efficiency of optimal portfolio weight estimators when the returns are Gaussian or non-Gaussian linear stationary processes. Furthermore, Shiraishi and Taniguchi (2006) considered that when the returns are non-Gaussian linear locally stationary processes. Although the above papers took care of the dependency of the return processes, they are not enough because it is known that financial return processes often have non-linearity of the past observations. To describe this phenomenon, the ARCH model is introduced by Engle (1982) and its related models such as the GARCH model are also introduced by many researchers. Furthermore, Dahlhaus and Rao (2006) introduced a time-varying ARCH (tvARCH) model which is a class of ARCH models with time-varying parameters. They studied the parameter estimation for tvARCH(p) models by weighted Gaussian quasi maximum likelihood methods. Furthermore, they showed that the estimator $\hat{\boldsymbol{\theta}}_{GQML}$ has asymptotic normality and the bias can be explained in terms of the derivatives of the tvARCH process. In this paper, denoting the optimal portfolio weights by a function $g = g(\boldsymbol{\theta})$ of ARCH parameter $\boldsymbol{\theta}$, we discuss the asymptotic property of optimal portfolio weight estimators (i.e. $g(\hat{\boldsymbol{\theta}}_{GQML})$) when the returns are vector-valued non-Gaussian tvARCH(p) processes with time dependent mean.

Since the nonstationarity of the process causes the estimator to be biased, we also consider bias adjusted estimators by use of resampling. In general, it is difficult to apply resampling to dependent data, because the idea is to simulate sampling from the population by sampling from the sample under the i.i.d. assumption. Hall and Yao (2003)

MSC : 62F40, 62M10
Key words and phrases : optimal portfolio, bootstrap, locally stationary, quasi-likelihood estimates, time-varying ARCH process

and Shiraishi (2010) suggested a resampling procedure when the process is GARCH model. In a similar way to the procedure, we show that stationary ARCH processes (which locally approximate tvARCH processes) can be generated. Then, based on the approximated ARCH processes, unbiased estimators $\boldsymbol{\theta}^*_{GQML}$ for ARCH parameter $\boldsymbol{\theta}$ can be constructed, which implies that we can construct unbiased optimal portfolio weight estimators $g(\boldsymbol{\theta}^*_{GQML})$.

This paper is organized as follows. First, following Dahlhaus and Rao (2006), we show that a weighted Gaussian quasi maximum likelihood estimator has a good property when the tvARCH process with time dependent mean is non Gaussian in Section 2. According to Dahlhaus and Rao (2006), the tvARCH process can be locally approximated by stationary ARCH processes. Therefore, the tvARCH processes can be called locally stationary. We elucidate the asymptotics of the estimator. Furthermore, we generate approximated stationary ARCH processes by use of resampling. Then, we construct an unbiased estimator for ARCH parameter and prove asymptotic normality of the estimator. In Section 3, we propose an optimal portfolio weight depending on the ARCH parameter. Moreover, we examine our approach numerically. The result shows eventually the effect of bias adjustment. We place the proofs of the theorems and lemmas in Section 4.

Throughout this paper, $|\boldsymbol{a}|$ and $|A|$ denote the Euclidean norm of a vector $\boldsymbol{a}$ and a matrix A defined by $\sqrt{\boldsymbol{a}'\boldsymbol{a}}$ and $\sqrt{tr(A'A)}$, respectively. We write $X_n \xrightarrow{d} X$ (or $\xrightarrow{p}$ or $\xrightarrow{a.s.}$) if $\{X_n\}$ converges in distribution (or in probability or almost surely) 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.

2. Asymptotic Theory for Fundamental Quantities

We suppose that the return process $\{\boldsymbol{X}_{t,N} = (X_{1,t,N}, \ldots, X_{m,t,N})'; t = 1, \ldots, N , N \in \mathbb{N}\}$ is an m-vector ARCH process with time-varying parameter $\{\boldsymbol{\theta}_{t/N} = (\theta_{1,t/N}, \ldots, \theta_{q,t/N})'; t = 1, \ldots, N\}$, defined by

$$\boldsymbol{X}_{t,N} = \boldsymbol{\mu}(\boldsymbol{\theta}_{t/N}) + D_{t,N}(\boldsymbol{\theta}_{t/N})\boldsymbol{\epsilon}_t \tag{1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ is an mean vector function, $D_{t,N} = diag(h^{\frac{1}{2}}_{1,t,N}, \ldots, h^{\frac{1}{2}}_{m,t,N})$ and $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \ldots, \epsilon_{m,t})'$ are i.i.d. random vectors with $E(\boldsymbol{\epsilon}_t) = \boldsymbol{0}, E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') = I_m$ and $E(|\boldsymbol{\epsilon}_t|^{12+\delta}) < \infty$ for some $\delta > 0$. Here $\boldsymbol{H}_{t,N} = (h_{1,t,N}, \ldots, h_{m,t,N})'$ are m vectors defined by

$$\boldsymbol{H}_{t,N}(\boldsymbol{\theta}_{t/N}) = \boldsymbol{U}(\boldsymbol{\theta}_{t/N}) + \sum_{j=1}^{p} A_j(\boldsymbol{\theta}_{t/N})\boldsymbol{Y}^2_{t-j,N}(\boldsymbol{\theta}_{(t-j)/N}) \tag{2}$$

with $\boldsymbol{Y}^2_{t-j,N} = (Y^2_{1,t-j,N}, \ldots, Y^2_{m,t-j,N})', Y_{i,t-j,N} = X_{i,t-j,N} - \mu_i, \boldsymbol{U} = (U_1, \ldots, U_m)'$ and $A_j = (A_{ab,j})_{a,b=1,\ldots,m}$.

This model extends Dahlhaus and Rao (2006)'s model to multidimensional case with a mean function. As you may see below, the asymptotic property of weighted Gaussian quasi maximum likelihood estimators are essentially same as Dahlhaus and Rao (2006)'s result.

We also assume that the time-varying parameter $\boldsymbol{\theta}_{t/N}$ are unknown and included a compact subset Θ of $\mathbb{R}^q$, i.e. $\boldsymbol{\theta}_{t/N} \in \Theta \subset \mathbb{R}^q$ for $\forall t = 1, \ldots, N, N \in \mathbb{N}$. We introduce the notation $\nabla_i = \frac{\partial}{\partial\theta_i}, \nabla_{ij} = \frac{\partial^2}{\partial\theta_i\partial\theta_j}, \nabla_{ijk} = \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}, \nabla = (\nabla_1, \ldots, \nabla_q)'$ and $\nabla^2 = (\nabla_{ij})_{i,j=1,\ldots,q}$ for $\forall\boldsymbol{\theta} \in \Theta$.

We call the sequence of stochastic processes $\{\boldsymbol{X}_{t,N} : t = 1, \ldots, N\}$ satisfied with (1) and (2), a time-varying ARCH process with order p (tvARCH(p) process). As shown below, the tvARCH(p) process can be locally approximated by stationary ARCH processes.

Therefore, we also call tvARCH processes locally stationary. We make the following assumptions.

Assumption 1. *There exist* $0 < \rho, Q, M < \infty, 0 < \nu < 1$ *and a positive sequence* $\{l(j)\}$.

(i) *For* $\forall \boldsymbol{\theta} \in \Theta, i = 1, \dots, m$ *and* $j = 1, \dots, p$

$$\rho < U_i(\boldsymbol{\theta}), \quad |\boldsymbol{\mu}(\boldsymbol{\theta})| \le Q, \quad |\boldsymbol{U}(\boldsymbol{\theta})| \le Q \quad \textit{and} \quad |A_j(\boldsymbol{\theta})| \le \frac{Q}{l(j)}$$

where $\{l(j)\}$ *satisfies* $mQ \sum_{j=1}^{\infty} \frac{1}{l(j)} \le 1 - \nu$ *and* $\sum_{j=1}^{\infty} \frac{j}{l(j)} < \infty$.

(ii) *For each* $u \in (0, 1]$, *we assume* $\boldsymbol{\theta}_u \in Int(\boldsymbol{\Theta})$ *and*

$$|\boldsymbol{\mu}(\boldsymbol{\theta}_u) - \boldsymbol{\mu}(\boldsymbol{\theta}_{u'})| \le M|u - u'|, \quad |\boldsymbol{U}(\boldsymbol{\theta}_u) - \boldsymbol{U}(\boldsymbol{\theta}_{u'})| \le M|u - u'|$$

and

$$|A_j(\boldsymbol{\theta}_u) - A_j(\boldsymbol{\theta}_{u'})| \le \frac{M}{l(j)}|u - u'|,$$

for $u, u' \in (0, 1]$.

(iii) *The third derivatives of* $\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta})$ *and* $A_j(\boldsymbol{\theta})$ *exist with*

$$|\nabla_{i_1 \dots i_k} \boldsymbol{\mu}_a(\boldsymbol{\theta})| \le C, \quad |\nabla_{i_1 \dots i_k} \boldsymbol{U}_a(\boldsymbol{\theta})| \le C \quad \textit{and} \quad |\nabla_{i_1 \dots i_k} A_{ab,j}(\boldsymbol{\theta})| \le C$$

for $k = 1, 2, 3, i_1, \dots, i_k = 1, \dots, q, a, b = 1, \dots, m$ *and* $\forall \boldsymbol{\theta} \in \Theta$, *where* C *is a finite constant independent of* i, a, b *and* $\boldsymbol{\theta}$.

(iv) *The third derivatives of* $\boldsymbol{\theta}_u$ *exist with*

$$\left| \frac{\partial^j \boldsymbol{\theta}_u}{\partial u^j} \right| \le C < \infty$$

for $j = 1, 2, 3$ *and* $\forall u \in (0, 1]$.

(v) *The random vector* $\boldsymbol{\epsilon}_t$ *has a positive density on an interval containing zero.*

For each given $u_0 \in (0, 1]$, we assume that there exist a stochastic process $\{\tilde{\boldsymbol{X}}_t(u_0) = (\tilde{X}_{1,t}(u_0), \dots, \tilde{X}_{m,t}(u_0))'; t \in \mathbb{N}\}$, that is, a stationary ARCH process associated with the tvARCH(p) process at time point u_0 which satisfies

$$\tilde{\boldsymbol{X}}_t(u_0) = \boldsymbol{\mu}(\boldsymbol{\theta}_{u_0}) + \tilde{D}_t(u_0, \boldsymbol{\theta}_{u_0})\boldsymbol{\epsilon}_t \tag{3}$$

where $\tilde{D}_t(u_0) = diag(\tilde{h}_{1,t}(u_0)^{\frac{1}{2}}, \dots, \tilde{h}_{m,t}(u_0)^{\frac{1}{2}})$. Here $\tilde{\boldsymbol{H}}_t(u_0) = (\tilde{h}_{1,t}(u_0), \dots, \tilde{h}_{m,t}(u_0))'$ are m vectors defined by

$$\tilde{\boldsymbol{H}}_t(u_0, \boldsymbol{\theta}_{u_0}) = \boldsymbol{U}(\boldsymbol{\theta}_{u_0}) + \sum_{j=1}^{p} A_j(\boldsymbol{\theta}_{u_0}) \tilde{\boldsymbol{Y}}^2_{t-j}(u_0, \boldsymbol{\theta}_{u_0}) \tag{4}$$

with $\tilde{\boldsymbol{Y}}^2_{t-j}(u_0) = (\tilde{Y}_{1,t-j}(u_0)^2, \dots, \tilde{Y}_{m,t-j}(u_0)^2)', \tilde{Y}_{i,t-j}(u_0) = \tilde{X}_{i,t-j}(u_0) - \mu_i$.

Comparing (1) and (2) with (3) and (4), it seems clear that if t/N is close to u_0. Then, $\boldsymbol{Y}^2_{t,N}(\boldsymbol{\theta}_{t/N})$ and $\tilde{\boldsymbol{Y}}_t^2(u_0, \boldsymbol{\theta}_{u_0})$ should be close and the degree of the approximation should depend both on the rescaling factor N and the deviation $|t/N - u_0|$. The following lemma corresponds to Theorem 1 of Dahlhaus and Rao (2006).

Lemma 1. *Suppose* $\{\boldsymbol{X}_{t,N}\}$ *and* $\{\tilde{\boldsymbol{X}}_t(u_0)\}$ *are tvARCH(p) and ARCH processes defined by (1),(2) and (3),(4), respectively. Then, under Assumption 1, we have*

$$\boldsymbol{Y}^2_{t,N}(\boldsymbol{\theta}_{t/N}) = \tilde{\boldsymbol{Y}}_t^2(u_0, \boldsymbol{\theta}_{u_0}) + O_p\left(\left| \frac{t}{N} - u_0 \right| + \frac{1}{N} \right).$$

The proofs of Theorem and Lemma will be given in Section 4

In what follows, we consider a kernel type estimator of the parameter of a tvARCH(p) model given the sample $\{\boldsymbol{X}_{t,N}; t = 1, \dots, N\}$. We now define the segment (kernel) estimator of $\boldsymbol{\theta}_{u_0}$ for $u_0 \in (0,1)$. Let $t_0 \in \mathbb{N}$ such that $|u_0 - t_0/N| < 1/N$. The estimator is the minimizer of the weighted conditional likelihood

$$\mathcal{L}_{t_0,N}(\boldsymbol{\theta}) := \sum_{k=p+1}^{N} \frac{1}{bN} W\left(\frac{t_0 - k}{bN}\right) l_{k,N}(\boldsymbol{\theta}) \tag{5}$$

where

$$l_{k,N}(\boldsymbol{\theta}) = \frac{1}{2}\left\{\log\det\left(D_{k,N}(\boldsymbol{\theta})^2\right) + (\boldsymbol{X}_{k,N} - \boldsymbol{\mu}(\boldsymbol{\theta}))' D_{k,N}(\boldsymbol{\theta})^{-2} (\boldsymbol{X}_{k,N} - \boldsymbol{\mu}(\boldsymbol{\theta}))\right\}$$

and $W : [-1/2, 1/2] \to \mathbb{R}$ is a kernel function of bounded variation with $\int_{-1/2}^{1/2} W(x)dx = 1$ and $\int_{-1/2}^{1/2} xW(x)dx = 0$. That is, we consider

$$\hat{\boldsymbol{\theta}}_{t_0/N} = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathcal{L}_{t_0,N}(\boldsymbol{\theta}). \tag{6}$$

In the derivation of the asymptotic properties of this estimator, we make use of the local approximation of $\boldsymbol{Y}_{t,N}^2$ by the stationary process $\tilde{\boldsymbol{Y}}_t^2(u_0)$ defined in (4). Similarly to the above, we therefore define the weighted likelihood

$$\tilde{\mathcal{L}}_N(u_0, \boldsymbol{\theta}) := \sum_{k=p+1}^{N} \frac{1}{bN} W\left(\frac{t_0 - k}{bN}\right) \tilde{l}_k(u_0, \boldsymbol{\theta}) \tag{7}$$

where $|u_0 - t_0/N| < 1/N$ and

$$\begin{aligned}
\tilde{l}_k(u_0, \boldsymbol{\theta}) = \frac{1}{2}\Big\{&\log\det\left(\tilde{D}_k(u_0,\boldsymbol{\theta})^2\right) \\
&+ \left(\tilde{\boldsymbol{X}}_k(u_0) - \boldsymbol{\mu}(\boldsymbol{\theta})\right)' \tilde{D}_k(u_0,\boldsymbol{\theta})^{-2} \left(\tilde{\boldsymbol{X}}_k(u_0) - \boldsymbol{\mu}(\boldsymbol{\theta})\right)\Big\}.
\end{aligned}$$

Similarly to Dahlhaus and Rao (2006), it is shown that both $\mathcal{L}_{t_0,N}(\boldsymbol{\theta})$ and $\tilde{\mathcal{L}}_N(\boldsymbol{\theta})$ converge to

$$\mathcal{L}(u_0, \boldsymbol{\theta}) := E\left(\tilde{l}_0(u_0, \boldsymbol{\theta})\right)$$

as $N \to \infty, b \to 0, bN \to \infty$ and $|u_0 - t_0/N| < 1/N$. It is easy to show that $\mathcal{L}(u_0, \boldsymbol{\theta})$ is minimized by $\boldsymbol{\theta} = \boldsymbol{\theta}_{u_0}$. For later reference, we introduce the followings:

$$\begin{aligned}
\Sigma(u_0) &:= -E\left(\nabla^2 \tilde{l}_0(u_0, \boldsymbol{\theta}_{u_0})\right) = -\left\{\Sigma_{ij}(u_0)\right\}_{i,j=1,\ldots,q} \\
K(u_0) &:= \left(\frac{1}{2}\int_{-1/2}^{1/2} W(x)^2 dx\right) E\left(\nabla\tilde{l}_0(u_0,\boldsymbol{\theta}_{u_0})\nabla\tilde{l}_0(u_0,\boldsymbol{\theta}_{u_0})'\right) \\
&= \left\{\left(\frac{1}{2}\int_{-1/2}^{1/2} W(x)^2 dx\right) K_{ij}(u_0)\right\}_{i,j=1,\ldots,q} \\
\boldsymbol{B}(u_0) &:= \left(\frac{1}{2}\int_{-1/2}^{1/2} x^2 W(x) dx\right) \Sigma(u_0)^{-1} \left(\frac{\partial^2}{\partial u^2} \nabla\mathcal{L}(u, \boldsymbol{\theta}_{u_0})\Big|_{u=u_0}\right)
\end{aligned}$$

where

$$\begin{aligned}
\Sigma_{ij}(u_0) &= \sum_{l=1}^{m} E\left(\frac{\nabla_i \mu_l(\boldsymbol{\theta}_{u_0})\nabla_j \mu_l(\boldsymbol{\theta}_{u_0})}{\tilde{h}_{l,0}(u_0,\boldsymbol{\theta}_{u_0})} + \frac{1}{2}\frac{\nabla_i \tilde{h}_{l,0}(u_0,\boldsymbol{\theta}_{u_0})\nabla_j \tilde{h}_{l,0}(u_0,\boldsymbol{\theta}_{u_0})}{\tilde{h}_{l,0}(u_0,\boldsymbol{\theta}_{u_0})^2}\right)\\
K_{ij}(u_0) &= \sum_{l=1}^{m} E\left(\frac{\nabla_i \mu_l(\boldsymbol{\theta}_{u_0})\nabla_j \mu_l(\boldsymbol{\theta}_{u_0})}{\tilde{h}_{l,0}(u_0,\boldsymbol{\theta}_{u_0})}\right)\\
&\quad + \frac{1}{4}\sum_{l_1,l_2=1}^{m}\left\{E\left(\frac{\nabla_i \tilde{h}_{l_1,0}(u_0,\boldsymbol{\theta}_{u_0})\nabla_j \tilde{h}_{l_2,0}(u_0,\boldsymbol{\theta}_{u_0})}{\tilde{h}_{l_1,0}(u_0,\boldsymbol{\theta}_{u_0})\tilde{h}_{l_2,0}(u_0,\boldsymbol{\theta}_{u_0})}\right) Cov\left(\epsilon_{l_1,0}^2,\epsilon_{l_2,0}^2\right)\right.\\
&\quad + 2E\left(\frac{\nabla_i \tilde{h}_{l_1,0}(u_0,\boldsymbol{\theta}_{u_0})\nabla_j \mu_{l_2}(\boldsymbol{\theta}_{u_0})}{\tilde{h}_{l_1,0}(u_0,\boldsymbol{\theta}_{u_0})\tilde{h}_{l_2,0}(u_0,\boldsymbol{\theta}_{u_0})^{1/2}}\right) Cov\left(\epsilon_{l_1,0}^2,\epsilon_{l_2,0}\right)\\
&\quad \left. + 2E\left(\frac{\nabla_i \mu_{l_1}(\boldsymbol{\theta}_{u_0})\nabla_j \tilde{h}_{l_2,0}(u_0,\boldsymbol{\theta}_{u_0})}{\tilde{h}_{l_1,0}(u_0,\boldsymbol{\theta}_{u_0})^{1/2}\tilde{h}_{l_2,0}(u_0,\boldsymbol{\theta}_{u_0})}\right) Cov\left(\epsilon_{l_1,0},\epsilon_{l_2,0}^2\right)\right\}.
\end{aligned}$$

The following theorem corresponds to Theorem 3 of Dahlhaus and Rao (2006).

Theorem 1. *Suppose $\{\boldsymbol{X}_{t,N} : t = 1,\ldots,N\}$ is a tvARCH(p) process which satisfies Assumption 1 and W is a kernel function of bounded variation with $\int_{-1/2}^{1/2} W(x)dx = 1$ and $\int_{-1/2}^{1/2} xW(x)dx = 0$. Then, if $|u_0 - t_0/N| < 1/N$ and $b = O(N^{-1/5})$,*

$$\sqrt{bN}(\hat{\boldsymbol{\theta}}_{t_0/N} - \boldsymbol{\theta}_{u_0}) \xrightarrow{d} N(\boldsymbol{B}_L(u_0), \Sigma(u_0)^{-1}K(u_0)\Sigma(u_0)^{-1}),$$

where L is a constant value satisfied with $L = bN^{1/5}$ and $\boldsymbol{B}_L(u_0) = \lim_{N\to\infty} L^{5/2}\boldsymbol{B}(u_0)$.

We have shown that the bias can be explained in terms of the derivatives of the tvARCH process. Furthermore, we have proved asymptotic normality of the estimator. This derivative enables us to study more precisely the nonstationary behavior of the process. Theorem 1 leads us to take an optimal bandwidth b_{opt} based on minimization the mean squared error.

Remark 1. *Under the conditions of Theorem 1, the mean squared error $MSE(\hat{\boldsymbol{\theta}}_{t_0/N}) = E|\hat{\boldsymbol{\theta}}_{t_0/N} - \boldsymbol{\theta}_{u_0}|^2$ is minimized by*

$$b_{opt} = N^{-\frac{1}{5}}\frac{tr\left\{\Sigma(u_0)^{-1}K(u_0)\Sigma(u_0)^{-1}\right\}}{|\boldsymbol{B}(u_0)|^2}.$$

According to Dahlhaus and Rao (2006), since $\tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta})$ is the (weighted) likelihood of the stationary approximation, magnitude of the bias, that is

$$\boldsymbol{B}(u_0) = \lim_{N\to\infty} E\left[\sqrt{bN}\Sigma(u_0)^{-1}\nabla\left(\mathcal{L}_{t_0,N}(\boldsymbol{\theta}_{u_0}) - \tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta}_{u_0})\right)\right],$$

depends on degree of the nonstationarity which implies the difference between $\mathcal{L}_{t_0,N}(\boldsymbol{\theta})$ and $\tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta})$. In other words, $\boldsymbol{B}(u_0)$ describes how different between the tvARCH (p) process $\{\boldsymbol{X}_{t,N}\}$ (i.e., nonstationary process) and the locally approximated ARCH (p) process $\{\tilde{\boldsymbol{X}}_t(\boldsymbol{\theta}_{u_0})\}$ (i.e., stationary process) because the difference $|\left(\mathcal{L}_{t_0,N}(\boldsymbol{\theta}) - \tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta})\right)|$ for any $\boldsymbol{\theta} \in Int(\Theta)$ tends to be 0 as the time-varying parameters $\{\boldsymbol{\theta}_{t/N}\}$ of $\{\boldsymbol{X}_{t,N}\}$ are close to a constant vector $\boldsymbol{\theta}_{u_0}$. Since the bias $\boldsymbol{B}(u_0)$ depends on degree of the nonstationarity, the optimal choice of the bandwidth (of the segment length) also depends on the degree of stationarity of the process. Because b_{opt} depends on the asymptotic bias and variance, it is not applicable directly in the actual real data analysis, so that one idea is to start with preliminary estimators $\boldsymbol{\mu}(\hat{\boldsymbol{\theta}}_{u_0}), \tilde{\boldsymbol{H}}_0(u_0,\hat{\boldsymbol{\theta}}_{u_0})$ for $\boldsymbol{\mu}(\boldsymbol{\theta}_{u_0}), \tilde{\boldsymbol{H}}_0(u_0,\boldsymbol{\theta}_{u_0})$, respectively. Then we can calculate their derivatives numerically, and plug them in $\Sigma(u_0), K(u_0)$ and $\boldsymbol{B}(u_0)$.

Next, we construct unbiased estimator of $\boldsymbol{\theta}_{N/N}$ by use of resampling. Let $\mathcal{X}_N =$

$\{\boldsymbol{X}_{1,N}, \ldots, \boldsymbol{X}_{N,N}\}$ be observations from a model described as (1) and (2). Based on $\mathcal{X}_N$, we can construct $\{\hat{\boldsymbol{\theta}}_{t/N}\}_{t=1,\ldots,N}$ by (6). Then, the error $\boldsymbol{\epsilon}_t$ are recovered by

$$\hat{\boldsymbol{\epsilon}}_t \equiv D_{t,N}^{-1}(\hat{\boldsymbol{\theta}}_{t/N})\{\boldsymbol{X}_{t,N} - \boldsymbol{\mu}(\hat{\boldsymbol{\theta}}_{t/N})\} \quad (t = p+1, \cdots, N).$$

Let $G_N(\cdot)$ denote the empirical distribution which put mass $1/(N-p)$ at each $\hat{\boldsymbol{\epsilon}}_t$. Let $F_N^*(\boldsymbol{x}) = G_N(\sigma_{\boldsymbol{\epsilon}}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{\epsilon}}))$ where $\boldsymbol{x} \in \mathbb{R}^m$, $\bar{\boldsymbol{\epsilon}} = 1/(N-p)\sum_{t=p+1}^N \hat{\boldsymbol{\epsilon}}_t$ and $\sigma_{\boldsymbol{\epsilon}}^2 = 1/(N-p)\sum_{t=p+1}^N (\hat{\boldsymbol{\epsilon}}_t - \bar{\boldsymbol{\epsilon}})(\hat{\boldsymbol{\epsilon}}_t - \bar{\boldsymbol{\epsilon}})'$. Let $\{\boldsymbol{\epsilon}_t^*\}$ be i.i.d. observations from $F_N^*(\cdot)$. Given $\{\boldsymbol{\epsilon}_t^*\}$, we generate $\{\tilde{\boldsymbol{X}}_t^*(1)\}$ by

$$\tilde{\boldsymbol{X}}_t^*(1) = \boldsymbol{\mu}(\hat{\boldsymbol{\theta}}_{N/N}) + \tilde{D}_t^*(1, \hat{\boldsymbol{\theta}}_{N/N})\boldsymbol{\epsilon}_t^*, \tag{8}$$

where $\tilde{D}_t^*(1, \hat{\boldsymbol{\theta}}_{N/N}) = diag(\tilde{h}_{1,t}^*(1, \hat{\boldsymbol{\theta}}_{N/N})^{1/2}, \ldots, \tilde{h}_{m,t}^*(1, \hat{\boldsymbol{\theta}}_{N/N})^{1/2})$. Here $\tilde{\boldsymbol{H}}_t^*(1, \hat{\boldsymbol{\theta}}_{N/N}) = (\tilde{h}_{1,t}^*(1, \hat{\boldsymbol{\theta}}_{N/N}), \ldots, \tilde{h}_{m,t}^*(1, \hat{\boldsymbol{\theta}}_{N/N}))'$ are defined by

$$\tilde{\boldsymbol{H}}_t^*(1, \hat{\boldsymbol{\theta}}_{N/N}) = \boldsymbol{U}(\hat{\boldsymbol{\theta}}_{N/N}) + \sum_{j=1}^p A_j(\hat{\boldsymbol{\theta}}_{N/N})\tilde{\boldsymbol{Y}}_{t-j}^{*2}(1, \hat{\boldsymbol{\theta}}_{N/N})$$

with $\tilde{\boldsymbol{Y}}_{t-j}^{*2}(1, \hat{\boldsymbol{\theta}}_{N/N}) = (\tilde{Y}_{1,t-j}^*(1, \hat{\boldsymbol{\theta}}_{N/N})^2, \ldots, \tilde{Y}_{m,t-j}^*(1, \hat{\boldsymbol{\theta}}_{N/N})^2)'$ and

$$\tilde{Y}_{i,t}^*(1, \hat{\boldsymbol{\theta}}_{N/N}) = \begin{cases} \tilde{X}_{i,t}^*(1) - \mu_i(\hat{\boldsymbol{\theta}}_{N/N}) & if\ t > -T \\ 0 & if\ t \le -T \end{cases}.$$

¿From the above procedure, we need to draw $\boldsymbol{\epsilon}_t^*$ for $-T \le t \le N$. If it is necessary to remove suspected edge effects, we may treat T as a sufficiently large integer. Throughout this paper, $(*)$ implies that we are dealing with the bootstrap quantity, hence distribution P^* and expectation E^* etc. are taken under $\{\boldsymbol{\epsilon}_t^*\} \sim i.i.d.\ F_N^*$ given $\mathcal{X}_N = \{\boldsymbol{X}_{1,N}, \ldots, \boldsymbol{X}_{N,N}\}$.

By using this model, we introduce a resampled estimator of the parameter $\boldsymbol{\theta}_{N/N}$, that is,

$$\boldsymbol{\theta}_{N/N}^* = \arg\min_{\boldsymbol{\theta} \in \Theta} \tilde{\mathcal{L}}_N^*(1, \boldsymbol{\theta}) \tag{9}$$

where

$$\tilde{\mathcal{L}}_N^*(1, \boldsymbol{\theta}) = \sum_{k=p+1}^N \frac{1}{bN} W\left(\frac{N-k}{bN}\right) \tilde{l}_k^*(1, \boldsymbol{\theta})$$

and

$$\begin{aligned} \tilde{l}_k^*(1, \boldsymbol{\theta}) \quad &= \quad \frac{1}{2}\Big\{ \log\det\left(\tilde{D}_k^*(1, \boldsymbol{\theta})^2\right) \\ &\quad + \left(\tilde{\boldsymbol{X}}_k^*(1) - \boldsymbol{\mu}(\boldsymbol{\theta})\right)' \tilde{D}_k^*(1, \boldsymbol{\theta})^{-2} \left(\tilde{\boldsymbol{X}}_k^*(1) - \boldsymbol{\mu}(\boldsymbol{\theta})\right)\Big\}. \end{aligned}$$

Then, we have the following result.

Theorem 2. *Suppose $\{\tilde{\boldsymbol{X}}_t^*(1) : t = 1, \ldots, N\}$ is generated by (8) satisfying Assumption 1. Suppose also that W is a kernel function of bounded variation with $\int_{-1/2}^{1/2} W(x)dx = 1$, $\int_{-1/2}^{1/2} xW(x)dx = 0$ and $b = O(N^{-1/5})$. Then,*

$$\sqrt{bN}(\boldsymbol{\theta}_{N/N}^* - \boldsymbol{\theta}_{N/N}) \xrightarrow{d^*} N(\mathbf{0}, \Sigma(1)^{-1}K(1)\Sigma(1)^{-1}).$$

3. Optimal portfolios

In this section, we propose an optimal portfolio weight estimator when the observed return process is written as (1) and (2). We construct an optimal portfolio weight estimator for the stationary ARCH process associated with the tvARCH(p) process at time point $1(= N/N)$.

Suppose that (pseudo) return process is described by (3) and (4) at time point 1. Then, the mean vector and variance matrix are written by

$$E(\tilde{\boldsymbol{X}}_t(1)) = \boldsymbol{\mu}(\boldsymbol{\theta}_{N/N}) \tag{10}$$

$$V(\tilde{\boldsymbol{X}}_t(1)) = \prod_{j=1}^{p}(I_m - A_j(\boldsymbol{\theta}_{N/N}))^{-1}\tilde{\boldsymbol{U}}(\boldsymbol{\theta}_{N/N}) \equiv V(\boldsymbol{\theta}_{N/N}) \tag{11}$$

where $\tilde{\boldsymbol{U}} = diag(U_1, \ldots, U_m)$. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)'$ be the vector of portfolio weights. Then, the return of portfolio at time t is $\tilde{\boldsymbol{X}}_t(1)'\boldsymbol{\omega}$, and the expectation and variance are, respectively, given by $\boldsymbol{\mu}(\boldsymbol{\theta}_{N/N})'\boldsymbol{\omega}$, $\boldsymbol{\omega}'V(\boldsymbol{\theta}_{N/N})\boldsymbol{\omega}$. Optimal portfolio weights have been proposed by various criteria. They are expressed as a function $g(\boldsymbol{\mu}, V)$ of $\boldsymbol{\mu}$ and V. (See Shiraishi and Taniguchi (2008)).

Since the mean vector and variance matrix are parameterized by $\boldsymbol{\theta}$, we can express the optimal portfolio weight function as $g = g(\boldsymbol{\theta}) = g(\boldsymbol{\mu}_{\boldsymbol{\theta}}, V_{\boldsymbol{\theta}})$. Also, it should be noted that since the vector of portfolio weight $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)'$ satisfies the restriction $\boldsymbol{e}'\boldsymbol{\omega} = 1$, where $\boldsymbol{e} = (1, \ldots, 1)'$, we have only to estimate the subvector $(\omega_1, \ldots, \omega_{m-1})'$. Hence we assume that the function g is $(m-1)$-dimensional, i.e.,

$$g : \boldsymbol{\theta} \to \mathbb{R}^{m-1}. \tag{12}$$

For g given by (12) we impose the following.

Assumption 2. *The function $g(\boldsymbol{\theta})$ is continuously differentiable.*

Then we have the following result.

Theorem 3. *Suppose $\{\tilde{\boldsymbol{X}}_t^*(1) : t = 1, \ldots, N\}$ is generated by (8) satisfying Assumption 1. Suppose also that W is a kernel function of bounded variation with $\int_{-1/2}^{1/2} W(x)dx = 1$, $\int_{-1/2}^{1/2} xW(x)dx = 0$ and $b = O(N^{-1/5})$. Then, under Assumption 2, we have*

$$\sqrt{bN}(g(\boldsymbol{\theta}^*_{N/N}) - g(\boldsymbol{\theta}_{N/N})) \xrightarrow{d^*} N(\boldsymbol{0}, \nabla g(\boldsymbol{\theta}_{N/N})\Sigma(1)^{-1}K(1)\Sigma(1)^{-1}\nabla g(\boldsymbol{\theta}_{N/N})').$$

In what follows, we examine our approach numerically. Here, we discuss a global asset allocation problem where Japanese capital must be allocated to "U.S. dollar", "Australian dollar" and "Euro", respectively. Based on the daily log-returns for these exchange rates, we construct the mean-variance optimal portfolios. The data are from Jan 1st, 2007 to Jun 1st, 2007.

Suppose now the return process $\{\boldsymbol{X}_{t,N} = (X_{1,t,N}, X_{2,t,N}, X_{3,t,N})'; t = 1, \ldots, N,\ N = 2, \ldots, 100\}$ is the following tvARCH(1) process ;

$$\boldsymbol{X}_{t,N} = \boldsymbol{\theta}^{(1)}_{t/N} + D_{t,N}(\boldsymbol{\theta}_{t/N})\boldsymbol{\epsilon}_t$$

where $D_{t,N} = diag(h^{1/2}_{1,t,N}, h^{1/2}_{2,t,N}, h^{1/2}_{3,t,N})$, and $\boldsymbol{H}_{t,N} = (h_{1,t,N}, h_{2,t,N}, h_{3,t,N})'$ is defined by

$$\boldsymbol{H}_{t,N}(\boldsymbol{\theta}_{t/N}) = \boldsymbol{\theta}^{(2)}_{t/N} + \boldsymbol{\theta}^{(3)}_{t/N}\boldsymbol{Y}^2_{t-1,N}(\boldsymbol{\theta}^{(1)}_{t/N}),$$

with $\boldsymbol{\theta}_{t/N} = (\boldsymbol{\theta}^{(1)'}_{t/N}, \boldsymbol{\theta}^{(2)'}_{t/N}, vec(\boldsymbol{\theta}^{(3)}_{t/N})')'$, $\boldsymbol{\theta}^{(i)}_{t/N} = (\theta^{(i)}_{1,t/N}, \theta^{(i)}_{2,t/N}, \theta^{(i)}_{3,t/N})'$ for $i = 1, 2$, $\boldsymbol{\theta}^{(3)}_{t/N} = (\theta^{(3)}_{ij,t/N})_{i,j=1,2,3}$ and $\boldsymbol{Y}^2_{t,N} = ((X_{1,t,N} - \theta^{(1)}_{1,t/N})^2, (X_{2,t,N} - \theta^{(1)}_{2,t/N})^2, (X_{3,t,N} - \theta^{(1)}_{3,t/N})^2)'$. By using (6) and (9), we construct Gaussian quasi maximum likelihood estimator $\boldsymbol{\theta}_{N/N}$ (GQMLE), and resampled Gaussian quasi maximum likelihood estimator $\boldsymbol{\theta}^*_{N/N}$ (rGQMLE), respectively. Then, two types of optimal portfolio weight estimator $g(\boldsymbol{\theta}_{N/N})$ and $g(\boldsymbol{\theta}^*_{N/N})$ are constructed from (10) and (11). Here, we consider the following optimal portfolio problem:

$$\begin{cases} \min_{\boldsymbol{w}} \boldsymbol{w}' V(\boldsymbol{\theta}) \boldsymbol{w} \\ \textit{subject to } \boldsymbol{\mu}(\boldsymbol{\theta})'\boldsymbol{w} = \mu_P \textit{ and } \mathbf{e}'\boldsymbol{w} = 1 \end{cases}.$$

Then, optimal portfolio weight $g(\boldsymbol{\theta})$ is written as

$$g(\boldsymbol{\theta}) = \frac{1}{\sigma_{\mu\mu} \cdot \sigma_{ee} - (\sigma_{\mu e})^2} \{(\sigma_{ee} \cdot \mu_P - \sigma_{\mu e})\boldsymbol{\mu}(\boldsymbol{\theta}) - (\sigma_{\mu e} - \sigma_{\mu\mu})\boldsymbol{e}\}$$

where $\sigma_{\mu\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})' V(\boldsymbol{\theta})^{-1} \boldsymbol{\mu}(\boldsymbol{\theta}), \sigma_{\mu e} = \boldsymbol{\mu}(\boldsymbol{\theta})' V(\boldsymbol{\theta})^{-1} \boldsymbol{e}$ and $\sigma_{ee} = \mathbf{e}' V(\boldsymbol{\theta})^{-1} \boldsymbol{e}$.

Figure 1 shows the portfolio returns ($= \boldsymbol{\alpha}'_N \boldsymbol{X}_{N+1,N+1}$) for

- GQMLE (i.e. $\boldsymbol{\alpha}_N = g(\boldsymbol{\theta}_{N/N})$),
- rGQMLE (i.e. $\boldsymbol{\alpha}_N = g(\boldsymbol{\theta}^*_{N/N})$),
- SM&SV (i.e. $\boldsymbol{\alpha}_N = g(\textit{sample mean},\ \textit{sample variance})$),

under $\mu_P = 0.0001, \ldots, 0.0010$ and $N = 2, \ldots, 100$.

Figure 1 is about here.

It is easy to see that the variations for GQLME and rGQMLE are smaller than those for SM&SV. This symptom leads us our optimal portfolio estimators are low-risk in view of the variation. In Table 1, we can see sample means of the portfolio returns for $N = 2, \ldots, 100$ and $\mu_P = 0.0001, \ldots, 0.0010$.

Table 1 is about here.

Obviously, the sample means for GQLME and rGQMLE are larger than those for SM&SV. Compare GQMLE with rGQMLE, those for rGQMLE are quite close to the target returns rather than those for GQMLE, which shows eventually the effect of bias adjustment. Finally, we show sample mean squares errors of the portfolio returns for $N = 2, \ldots, 100$ and $\mu_P = 0.0001, \ldots, 0.0010$ in Table 2.

Table 2 is about here.

For almost all target returns, the MSEs for GQLME and rGQMLE are smaller than those for SM&SV. Although those for rGQMLE are relatively larger than those for GQMLE, the spreads are not so wide.

Summarizing the numerical study, our proposed optimal portfolio estimators are obviously attractive rather than traditional one. Furthermore, we obtained the large effect of bias adjustment by use of resampling.

4. Proofs

This section provides the proofs of Theorems and Lemmas.

Proof of Lemma 1. Let $\eta_t = diag(\epsilon^2_{1,t}, \ldots, \epsilon^2_{m,t})$. Then, from Volterra expansions (e.g.,Giraitis et al.(2000)), it follows that

$$\begin{aligned} \boldsymbol{Y}^2_{t,N}(\boldsymbol{\theta}_{t/N}) &= \eta_t \boldsymbol{U}(\boldsymbol{\theta}_{t/N}) + \sum_{k=1}^{\infty} m_{t,N}(k) \\ \tilde{\boldsymbol{Y}}^2_t(u_0, \boldsymbol{\theta}_{u_0}) &= \eta_t \boldsymbol{U}(\boldsymbol{\theta}_{u_0}) + \sum_{k=1}^{\infty} \tilde{m}_t(u_0, k) \end{aligned}$$

where

$$\begin{aligned} m_{t,N}(k) &= \eta_t \sum_{j_1,\ldots,j_k=1}^{p} \left\{ \prod_{r=1}^{k} A_{j_r}(\boldsymbol{\theta}_{(t-\sum_{s=1}^{r-1} j_s)/N}) \eta_{t-\sum_{s'=1}^{r} j_{s'}} \right\} \boldsymbol{U}(\boldsymbol{\theta}_{(t-\sum_{s''=1}^{k} j_{s''})/N}) \\ \tilde{m}_t(u_0,k) &= \eta_t \sum_{j_1,\ldots,j_k=1}^{p} \left\{ \prod_{r=1}^{k} A_{j_r}(\boldsymbol{\theta}_{u_0}) \eta_{t-\sum_{s'=1}^{r} j_{s'}} \right\} \boldsymbol{U}(\boldsymbol{\theta}_{u_0}). \end{aligned}$$

Then, in an analogous manner to the proof of theorem 1 of Dahlhaus and Rao (2006), we have

$$\begin{aligned} E\left|\boldsymbol{Y}_{t,N}^2(\boldsymbol{\theta}_{t/N}) - \tilde{\boldsymbol{Y}}_t^2(t/N,\boldsymbol{\theta}_{t/N})\right| &= O\left(\frac{1}{N}\right) \\ E\left|\boldsymbol{Y}_{t,N}^2(\boldsymbol{\theta}_{t/N}) - \tilde{\boldsymbol{Y}}_t^2(t/N,\boldsymbol{\theta}_{t/N})\right| &= O\left(\left|\frac{t}{N} - u_0\right|\right) \end{aligned}$$

from Assumption 1. Therefore, we have shown Lemma 1 by using the triangle inequality and Markov's inequality.

To prove Theorem 1, we need the following lemmas. The following lemma corresponds to Lemma 1 of Dahlhaus and Rao (2006).

Lemma 2. *Suppose* $\{\boldsymbol{X}_{t,N} : t = 1,\ldots,N\}$ *is a tvARCH(p) process which satisfies Assumption 1. Then, if* $|u_0 - t_0/N| < 1/N$,

$$\sup_{u\in(0,1]} \left|\nabla^2\mathcal{L}_{t_0,N}(\boldsymbol{\theta}_u) - \nabla^2\mathcal{L}(u_0,\boldsymbol{\theta}_u)\right| \xrightarrow{p} 0,$$

where $b \to 0, bN \to \infty$ *as* $N \to \infty$.

Proof of Lemma 2. We write

$$\begin{aligned} & \nabla_{ij}\mathcal{L}_{t_0,N}(\boldsymbol{\theta}_u) - \nabla_{ij}\mathcal{L}(u_0,\boldsymbol{\theta}_u) \\ =\ & \Big(\nabla_{ij}\mathcal{L}_{t_0,N}(\boldsymbol{\theta}_u) - \nabla_{ij}\tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta}_u)\Big) + \Big(\nabla_{ij}\tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta}_u) - \nabla_{ij}\mathcal{L}(u_0,\boldsymbol{\theta}_u)\Big) \\ \equiv\ & R_1 + R_2 \quad (say). \end{aligned}$$

¿From Lemma 1 and Assumption 1, we have

$$\begin{aligned} |h_{l,t_0,N}(\boldsymbol{\theta}_u) - \tilde{h}_{l,t_0}(u_0,\boldsymbol{\theta}_u)| &= O_p\left(\sum_{k=1}^{p}\left|\frac{t_0-k}{N} - u_0\right| + \frac{1}{N}\right), \\ |\nabla_i h_{l,t_0,N}(\boldsymbol{\theta}_u) - \nabla_i\tilde{h}_{l,t_0}(u_0,\boldsymbol{\theta}_u)| &= O_p\left(\sum_{k=1}^{p}\left|\frac{t_0}{N} - u_0\right| + \frac{1}{N}\right) \end{aligned}$$

for $l = 1,\ldots,m$. Hence,

$$R_1 = \sum_{k=p+1}^{N} \frac{1}{bN} W\left(\frac{t_0-k}{bN}\right) \Big\{\nabla_{ij} l_{k,N}(\boldsymbol{\theta}_u) - \nabla_{ij}\tilde{l}_k(u_0,\boldsymbol{\theta}_u)\Big\} = o_p(1).$$

On the other hand, from Dahlhaus and Rao (2006) Lemma A.2,

$$R_2 \le \left|\nabla_{ij}\tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta}_u) - \nabla_{ij}\mathcal{L}(u_0,\boldsymbol{\theta}_u)\right| = o_p(1).$$

The proof now follows from this observation.

The following lemma corresponds to Proposition 2 of Dahlhaus and Rao (2006).

Lemma 3. *Suppose* $\{\boldsymbol{X}_{t,N} : t = 1,\ldots,N\}$ *is a tvARCH(p) process which satisfies Assumption 1. Then, if* $|u_0 - t_0/N| < 1/N$,

$$\sqrt{bN}\nabla\tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta}_{u_0}) \xrightarrow{d} N(\boldsymbol{0}, K(u_0))$$

where $b \to 0, bN \to \infty$ *as* $N \to \infty$.

Proof of Lemma 3. Since $\nabla\tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta}_{u_0})$ is the weighted sum of martingale differences, the result follows from Fullar (1996) Theorem 5.3.4.

The following lemma corresponds to Theorem 3 of Dahlhaus and Rao (2006).

Lemma 4. *Suppose $\{\boldsymbol{X}_{t,N} : t = 1,\dots,N\}$ is a tvARCH(p) process which satisfies Assumption 1. Then, if $|u_0 - t_0/N| < 1/N$ and $b = O(N^{-1/5})$,*

$$\sqrt{bN}\Sigma(u_0)^{-1}\nabla\mathcal{B}_{t_0,N}(\boldsymbol{\theta}_{u_0}) \xrightarrow{p} \boldsymbol{B}(u_0)$$

where $\mathcal{B}_{t_0,N}(\boldsymbol{\theta}) := \mathcal{L}_{t_0,N}(\boldsymbol{\theta}) - \tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta})$.

Proof of Lemma 4. ¿From the definition of $\mathcal{B}_{t_0,N}(\boldsymbol{\theta}_{u_0})$, we can write

$$\nabla_i\mathcal{B}_{t_0,N}(\boldsymbol{\theta}_{u_0}) = \sum_{k=p+1}^{N}\frac{1}{bN}W\left(\frac{t_0-k}{bN}\right)\left\{\nabla_i\tilde{l}_k\left(\frac{k}{N},\boldsymbol{\theta}_{u_0}\right) - \nabla_i\tilde{l}_k(u_0,\boldsymbol{\theta}_{u_0}) + r_{k,N}\right\}$$

where $r_{k,N} = \nabla_i l_{k,N}(\boldsymbol{\theta}_{u_0}) - \nabla_i\tilde{l}_k(\frac{k}{N},\boldsymbol{\theta}_{u_0})$.
Taking Taylor expansion of $\nabla_i\tilde{l}_k(u,\boldsymbol{\theta}_{u_0})$ about $u = u_0$, we have

$$\begin{aligned}
&\nabla_i\tilde{l}_k\left(\frac{k}{N},\boldsymbol{\theta}_{u_0}\right) - \nabla_i\tilde{l}_k(u_0,\boldsymbol{\theta}_{u_0})\\
=\ &\left(\frac{k}{N}-u_0\right)\frac{\partial\nabla_i\tilde{l}_k(u,\boldsymbol{\theta}_{u_0})}{\partial u}\bigg\rfloor_{u=u_0} + \frac{1}{2}\left(\frac{k}{N}-u_0\right)^2\frac{\partial^2\nabla_i\tilde{l}_k(u,\boldsymbol{\theta}_{u_0})}{\partial u^2}\bigg\rfloor_{u=u_0}\\
&+\frac{1}{6}\left(\frac{k}{N}-u_0\right)^3\frac{\partial^3\nabla_i\tilde{l}_k(u,\boldsymbol{\theta}_{u_0})}{\partial u^3}\bigg\rfloor_{u=\tilde{U}_k},\\
\equiv\ &\left(\frac{k}{N}-u_0\right)\tilde{a}_{i,k} + \frac{1}{2}\left(\frac{k}{N}-u_0\right)^2\tilde{b}_{i,k} + \frac{1}{6}\left(\frac{k}{N}-u_0\right)^3\tilde{c}_{i,k} \quad (say)
\end{aligned}$$

where the random variable $\tilde{U}_k \in (0,1]$. Similarly to Dahlhaus and Rao (2006) Lemma A.7.-A.10. and Corollary A.2., if Assumption 1 is satisfied, it is shown that

$$E\left(\left|\frac{\partial^s\nabla_i\tilde{l}_k(u,\boldsymbol{\theta}_{u_0})}{\partial u^s}\right|\right) = O(1) \quad for\ s=1,2,3,\ u\in(0,1],$$

$$\sum_{l=0}^{\infty}\left|Cov\left(\frac{\partial^s\nabla_i\tilde{l}_k(u,\boldsymbol{\theta}_{u_0})}{\partial u^s},\frac{\partial^s\nabla_i\tilde{l}_{k+l}(u,\boldsymbol{\theta}_{u_0})}{\partial u^s}\right)\right| = O(1)\ for\ s=1,2,\ u\in(0,1].$$

Hence, if $b = O(N^{-1/5})$, it is easy to see that

$$\begin{aligned}
&\sum_{k=p+1}^{N}\frac{1}{bN}W\left(\frac{t_0-k}{bN}\right)\left\{\nabla_i\tilde{l}_k\left(\frac{k}{N},\boldsymbol{\theta}_{u_0}\right) - \nabla_i\tilde{l}_k(u_0,\boldsymbol{\theta}_{u_0})\right\}\\
&= \frac{b^2}{2}\int_{-1/2}^{1/2}x^2W(x)dxE(\tilde{b}_{i,k}) + o_p\left(\frac{1}{\sqrt{bN}}\right).
\end{aligned} \tag{13}$$

On the other hand, from Lemma 1, it follows that

$$\sum_{k=p+1}^{N}\frac{1}{bN}W\left(\frac{t_0-k}{bN}\right)r_{k,N} = O_p\left(\frac{1}{N}\right). \tag{14}$$

Therefore, from (13) and (14), we obtain

$$\nabla_i\mathcal{B}_{t_0,N}(\boldsymbol{\theta}_{u_0}) = \frac{b^2}{2}\int_{-1/2}^{1/2}x^2W(x)dxE(\tilde{b}_{i,k}) + o_p\left(\frac{1}{\sqrt{bN}}\right).$$

Since

$$\boldsymbol{B}(u_0)_i = \Sigma(u_0)^{-1}\int_{-1/2}^{1/2}x^2W(x)dxE(\tilde{b}_{i,k}),$$

the proof is completed under $b = O(N^{-1/5})$.
Proof of Theorem 1. ¿From the usual Taylor expansion, we can see that

$$\mathbf{0} = \nabla\mathcal{L}_{t_0,N}(\boldsymbol{\theta}_{u_0}) + \nabla^2\mathcal{L}_{t_0,N}(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{t_0/N} - \boldsymbol{\theta}_{u_0})$$

with $\bar{\boldsymbol{\theta}}$ between $\hat{\boldsymbol{\theta}}_{t_0/N}$ and $\boldsymbol{\theta}_{u_0}$. By using Lemmas 2-4, we have

$$\begin{aligned}
\sqrt{bN}(\hat{\boldsymbol{\theta}}_{t_0/N} - \boldsymbol{\theta}_{u_0}) &= \left\{-\nabla^2\mathcal{L}_{t_0,N}(\bar{\boldsymbol{\theta}})\right\}^{-1}\sqrt{bN}\nabla\mathcal{L}_{t_0,N}(\boldsymbol{\theta}_{u_0}) \\
&= \Sigma(u_0)^{-1}\sqrt{bN}\nabla\mathcal{L}_{t_0,N}(\boldsymbol{\theta}_{u_0}) + o_p(1) \quad (\textit{by Lemma } 2) \\
&= \Sigma(u_0)^{-1}\sqrt{bN}\nabla\tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta}_{u_0}) + \sqrt{bN}\Sigma(u_0)^{-1}\nabla\mathcal{B}_{t_0,N}(\boldsymbol{\theta}_{u_0}) + o_p(1) \\
&= \Sigma(u_0)^{-1}\sqrt{bN}\nabla\tilde{\mathcal{L}}_N(u_0,\boldsymbol{\theta}_{u_0}) + \boldsymbol{B}_L(u_0) + o_p(1) \\
&\qquad (\textit{by Lemma } 4) \\
&\xrightarrow{d} N\left(\boldsymbol{B}_L(u_0), \Sigma(u_0)^{-1}K(u_0)\Sigma(u_0)^{-1}\right). \quad (\textit{by Lemma } 3)
\end{aligned}$$

Proof of Theorem 2. Similarly to Lemma 2 and Lemma 3, we obtain that

$$\nabla^2\tilde{\mathcal{L}}_N^*(1,\boldsymbol{\theta}_{N/N}^*) \xrightarrow{p^*} \nabla^2\mathcal{L}^*(1,\boldsymbol{\theta}_{N/N}) = -\Sigma^*(1) \tag{15}$$

$$\sqrt{bN}\nabla\tilde{\mathcal{L}}_N^*(1,\boldsymbol{\theta}_{N/N}^*) \xrightarrow{d^*} N(\mathbf{0}, K^*(1)) \tag{16}$$

where $\Sigma^*(1)$ and $K^*(1)$ are resampling version of $\Sigma(1)$ and $K(1)$. Therefore, we have

$$\begin{aligned}
\sqrt{bN}(\boldsymbol{\theta}_{N/N}^* - \boldsymbol{\theta}_{N/N}) &= \left\{-\nabla^2\tilde{\mathcal{L}}_N^*(1,\bar{\boldsymbol{\theta}})\right\}^{-1}\sqrt{bN}\nabla\tilde{\mathcal{L}}_N^*(1,\boldsymbol{\theta}_{N/N}) \\
&= \Sigma^*(1)^{-1}\sqrt{bN}\nabla\tilde{\mathcal{L}}_N^*(1,\boldsymbol{\theta}_{N/N}) + o_p(1) \quad (\textit{by } (15)) \\
&\xrightarrow{d^*} N\left(\mathbf{0}, \Sigma^*(1)^{-1}K^*(1)\Sigma^*(1)^{-1}\right). \quad (\textit{by } (16))
\end{aligned}$$

Hence, the proof is completed if it is shown that

$$\Sigma^*(1)^{-1}K^*(1)\Sigma^*(1)^{-1} \xrightarrow{p} \Sigma(1)^{-1}K(1)\Sigma(1)^{-1}.$$

Now note that from the definition of $\hat{\boldsymbol{\epsilon}}_t$, it may be deduced that

$$\begin{aligned}
\hat{\boldsymbol{\epsilon}}_t &= D_{t,N}(\hat{\boldsymbol{\theta}}_{t/N})^{-1}\{\boldsymbol{X}_{t,N} - \boldsymbol{\mu}(\hat{\boldsymbol{\theta}}_{t/N})\} \\
&= D_{t,N}(\boldsymbol{\theta}_{t/N})^{-1}\{\boldsymbol{X}_{t,N} - \boldsymbol{\mu}(\boldsymbol{\theta}_{t/N})\} \\
&\quad + \{D_{t,N}(\hat{\boldsymbol{\theta}}_{t/N})^{-1} - D_{t,N}(\boldsymbol{\theta}_{t/N})^{-1}\}\{\boldsymbol{X}_{t,N} - \boldsymbol{\mu}(\boldsymbol{\theta}_{t/N})\} \\
&\quad + D_{t,N}(\hat{\boldsymbol{\theta}}_{t/N})^{-1}\{\boldsymbol{\mu}(\boldsymbol{\theta}_{t/N}) - \boldsymbol{\mu}(\hat{\boldsymbol{\theta}}_{t/N})\} \\
&= \boldsymbol{\epsilon}_t + o_p(1),
\end{aligned}$$

which implies $\tilde{\boldsymbol{\epsilon}}_t \equiv \sigma_{\boldsymbol{\epsilon}}^{-1}(\hat{\boldsymbol{\epsilon}}_t - \bar{\boldsymbol{\epsilon}}) = \boldsymbol{\epsilon}_t + o_p(1)$. Using the above result and the ergodic theorem by Stout(1974), for any measurable function $f(\cdot)$ satisfying $E|f(\epsilon_{a_1,t},\ldots,\epsilon_{a_k,t})| < \infty$, we obtain that

$$\begin{aligned}
E^*f(\epsilon_{a_1,t}^*,\ldots,\epsilon_{a_k,t}^*) &= \frac{1}{N-p}\sum_{t=1}^{N-p} f(\tilde{\epsilon}_{a_1,t},\ldots,\tilde{\epsilon}_{a_k,t}) \\
&= \frac{1}{N-p}\sum_{t=1}^{N-p} f(\epsilon_{a_1,t},\ldots,\epsilon_{a_k,t}) + o_p(1) \\
&= Ef(\epsilon_{a_1,t},\ldots,\epsilon_{a_k,t}) + o_p(1),
\end{aligned}$$

which implies

$$\Sigma^*(1)^{-1}K^*(1)\Sigma^*(1)^{-1} \xrightarrow{p} \Sigma(1)^{-1}K(1)\Sigma(1)^{-1}.$$

The proof now follows from this observation.
Proof of Theorem 3. The proof follows from Theorem 2 and the $\delta-$*method* (e.g.,Proposition 6.4.3 of Brockwell and Davis (1991)).

Acknowledgement

The author would like to thank for Professor Masanobu Taniguchi of Waseda University for his encouragement, guidance, and advice for this paper. He is also grateful for referee for his helpful comments and advices.

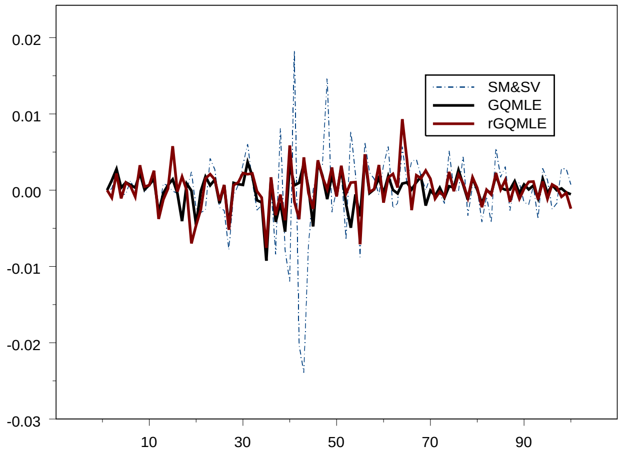

Figure 1. Time series plots of portfolio return for SM&SV portfolio, GQMLE portfolio and rGQMLE portfolio

Table 1. Sample Mean of portfolio return for SM&SV portfolio, GQMLE portfolio and rGQMLE portfolio

Target Return	SM&SV	GQMLE	rGQMLE
0.0001	-0.0000493	0.0000442	0.0003009
0.0002	-0.0000704	0.0000394	0.0002921
0.0003	-0.0001079	0.0000347	0.0002833
0.0004	-0.0001253	0.0000299	0.0002745
0.0005	-0.0001685	0.0000252	0.0002657
0.0006	-0.0002217	0.0000204	0.0002569
0.0007	-0.0002234	0.0000156	0.0002481
0.0008	-0.0002554	0.0000109	0.0002393
0.0009	-0.0003253	0.0000061	0.0002305
0.0010	-0.0003003	0.0000014	0.0002216

References

[1] Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*. New York: Springer.

[2] Dahlhaus, R. and Rao, S, S. (2006) Statistical Inference for Time-varying ARCH processes. *The Annals of Statistics*. **34** 1075-1114.

[3] Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*. **50**, 987-1007.

[4] Fuller, W. A. (1996) *Introduction to Statistical Time Series*, second edition. New York: Willy.

[5] Giraitis, L., Kokoszka, P., and Leipus, R. (2000) Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory*. **16**(1), 3-22.

TABLE 2. Mean Squares Error of portfolio return for SM&SV portfolio, GQMLE portfolio and rGQMLE portfolio

Target Return	SM&SV	GQMLE	rGQMLE
0.0001	0.0000057	0.0000057	0.0000064
0.0002	0.0000092	0.0000053	0.0000054
0.0003	0.0000141	0.0000049	0.0000051
0.0004	0.0000206	0.0000046	0.0000056
0.0005	0.0000285	0.0000045	0.0000068
0.0006	0.0000379	0.0000043	0.0000088
0.0007	0.0000488	0.0000043	0.0000114
0.0008	0.0000611	0.0000044	0.0000148
0.0009	0.0000751	0.0000045	0.0000189
0.0010	0.0000903	0.0000048	0.0000238

[6] Hall, P. and Yao, Q. (2003) Inference in ARCH and GARCH Models with Heavy-Tailed Errors. *Econometrica.* **71**,285-317.
[7] Shiraishi, H. (2005) Statistical Estimation of Optimal Portfolios for Gaussian Dependent Returns of Assets. *Scientiae Mathematicae Japonicae.* **62**, 439-449.
[8] Shiraishi, H. (2010) Resampling Procedure to Construct Estimation Error Efficient Portfolios for Stationary Returns of Assets. *Journal of The Japan Statistical Society.* **40**, 189-206.
[9] Shiraishi, H. and Taniguchi, M. (2006) Statistical Estimation of Optimal Portfolios for Locally Stationary Returns of Assets. *International Journal of Theoretical and Applied Finance.* **10**, 129-154.
[10] Shiraishi, H. and Taniguchi, M. (2008) Statistical Estimation of Optimal Portfolios for non-Gaussian Dependent Returns of Assets. *Journal of Forecasting.* **27**, 193-215.
[11] Stout, W. F. (1974) *Almost Sure Convergence.* New York: Academic Press.

commmunicated by *Masanobu Taniguchi*

HIROSHI SHIRAISHI
LABORATORY OF MATHEMATICS
JIKEI UNIVERSITY SCHOOL OF MEDICINE
8-3-1 KOKURYO, CHOFU-SI
TOKYO 182-8570
JAPAN
E-mail: hiroshi.shiraishi@jikei.ac.jp