Economics Letters 115 (2012) 334–337

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Efficient realized variance, regression coefficient, and correlation coefficient under different sampling frequencies[☆]

Dong Wan Shin [*], Sangun Park

Department of Statistics, Ewha University, Seoul 120-750, Republic of Korea
Department of Applied Statistics, Yonsei University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 21 January 2010
Received in revised form
19 December 2011
Accepted 20 December 2011
Available online 30 December 2011

JEL classification:
C13
C22

Keywords:
Efficiency
Factoring likelihood
Realized variance

ABSTRACT

Efficiency of the realized variance of an asset is improved by taking advantage of another asset whose return is cross-sectionally correlated with that of the asset and is less sensitive to market microstructure noises permitting higher frequency sampling than the original asset.

1. Introduction

Recently, realized variances attracted much attention as estimators of the latent conditional variances of financial assets. A lot of papers have appeared in the literature. For example, Andersen and Bollerslev (1998) estimated conditional variance of ex post daily foreign exchange by daily realized variance based on 5-min sampling, i.e., sum of 288 squared 5-min returns. Good reviews were provided by Poon and Granger (2003), Barndorff-Nielsen and Shephard (2007), McAleer and Medeiros (2008), and others.

The five-minute frequency is a common choice which is regarded as the highest possible frequency permitting no significant market microstructure noise due to the asynchronous trading, bid-ask spread, infrequent trading, and others. In order to improve the realized variance for an asset, say asset 1, it is worth finding another asset, say asset 2, whose return is cross-sectionally correlated with the return of the original asset and is subject to smaller market microstructure noises. Asset 2 can be observed more frequently than the original asset, providing extra information for improving the realized variance of asset 1.

This observation leads us to study estimation of the realized variances of two asset processes under two different sampling frequencies. We adopt the factoring likelihood method of Anderson (1957) which is also given in Little and Rubin (2002). In this method, from the high frequency data set of asset 2, the realized variance of asset 2 is computed and, from the low frequency combined data set for (asset 1, asset 2), the realized regression coefficient and the realized variance are computed for the residual process in the regression of return of asset 1 on return of asset 2. These three estimators are asymptotically efficient. Improved realized variance of asset 1 is computed from these three efficient estimators. Moreover, the realized covariance, the realized correlation coefficient, and the realized regression coefficients are also constructed, which are more efficient than those based on low frequency combined data set.

Limiting distributions of the improved estimators are derived, which are normal distributions. Asymptotic efficiencies of the estimators as well as finite sample efficiencies are investigated, revealing that considerable efficiency gains are attained for highly correlated cases.

2. Estimators

We consider a two dimensional continuous stochastic volatility semimartingale $\mathbf{y}^*(t) = (y_1(t), y_2(t))'$ defined by

$$\mathbf{y}^*(t) = \alpha^*(t) + \mathbf{m}^*(t),$$

[☆] This research was supported by the National Research Foundation of Korea (NRF-2009-0084772, NRF-2009-0070618).

[*] Correspondence to: Department Statistics, Ewha Womans University, Seoul 120-750, Republic of Korea. Tel.: +82 1 3277 2614; fax: +82 1 3277 3607, +82 2 3277 3607.

E-mail address: shindw@ewha.ac.kr (D.W. Shin).

0165-1765/$ – see front matter
doi:10.1016/j.econlet.2011.12.113

where $\alpha^*(t)$ is a process with a finite variation path,

$$\mathbf{m}^*(t) = \int_0^t \Theta(u) d\mathbf{w}(u),$$

Θ is an instantaneous or spot covolatility process having elements that are all cadlag, and $\mathbf{w}(t)$ is a two-dimensional standard Brownian motion.

Quadratic variation process is defined by

$$[\mathbf{y}^*](t) = \text{plim}_{M\to\infty} \sum_{j=1}^{M} \{\mathbf{y}^*(t_{j+1}) - \mathbf{y}^*(t_j)\}\{\mathbf{y}^*(t_{j+1}) - \mathbf{y}^*(t_j)\}'$$

for any sequence of partitions $0 = t_0 < t_1 < \cdots < t_M = t$ such that $\sup_j(t_{j+1} - t_j) \to 0$ as $M \to \infty$. Under some mild conditions,

$$[\mathbf{y}^*](t) = \int_0^t \Sigma(u) du,$$

where $\Sigma(t)$ is the instantaneous or spot covariance matrix process defined by

$$\Sigma(t) = \Theta(t)\Theta'(t).$$

Let h be the length of a period. Then the quadratic variation of period i is

$$[\mathbf{y}^*]_i = [\mathbf{y}^*](ih) - [\mathbf{y}^*]((i-1)h) = \int_{(i-1)h}^{ih} \Sigma(u) du.$$

The i-th period quadratic variation is usually estimated from M-equally spaced observations of the period. The j-th intra-h return for the i-th period is

$$(\mathbf{y}_{jM})_i = \mathbf{y}^*((i-1)h + jh/M) - \mathbf{y}^*((i-1)h + (j-1)h/M),$$
$$j = 1, \ldots, M.$$

The i-th period realized variance matrix defined by

$$[\mathbf{y}^*_M]_i = \sum_{j=1}^{M} (\mathbf{y}_{jM})_i (\mathbf{y}_{jM})'_i$$

is a consistent estimator of $[\mathbf{y}^*]_i$ as $M \to \infty$.

We assume that y^*_2 is more frequently observed than y^*_1, say $K = \tau M$ times per period, where $\tau \geq 1$ is a positive integer. The j-th intra-h return of y_2 for the i-th period is

$$(y_{2jK})_i = y^*_2((i-1)h + jh/K) - y^*_2((i-1)h + (j-1)h/K),$$
$$j = 1, \ldots, K.$$

We will say that y^*_1 is observed by "M-sampling" and y^*_2 is observed by "K-sampling". Note that $M < K$.

Let $(\sigma_{k\ell})_i$ be the (k, ℓ)-element of $[\mathbf{y}^*]_i$, $k, \ell = 1, 2$. Let $(\hat{\sigma}_{k\ell M})_i = \sum_{j=1}^{M} (y_{kjM})_i (y_{\ell jM})_i$ be the (k, ℓ)-element of $[\mathbf{y}^*_M]_i$, $k, \ell = 1, 2$. Let $(\hat{\sigma}_{22K})_i = [y^*_{2K}]_i = \sum_{j=1}^{K} (y_{2jK})_i^2$. We note that $(\hat{\sigma}_{22M})_i$ is not efficient for $(\sigma_{22})_i$ because $(\hat{\sigma}_{22K})_i$ is a better estimator. We will show that $(\hat{\sigma}_{11M})_i$ is not efficient by constructing a more efficient estimator.

In order to improve $(\hat{\sigma}_{11M})_i$, we adopt the factoring likelihood approach of Anderson (1957). The point is that the returns of asset 2 for times at which asset 1 is not observed, being cross-sectionally correlated with the unobserved returns of asset 1, can improve the realized variance of asset 1.

From the high-frequency data set of y_2, the marginal actual variance $(\sigma_{22})_i$ is efficiently estimated by $(\hat{\sigma}_{22K})_i$. From the low-frequency combined data set of (y_1, y_2), the following conditional quantities

$$(\beta_{12})_i = (\sigma_{12})_i/(\sigma_{22})_i, \qquad (\sigma_{11.2})_i = (\sigma_{11})_i - (\sigma_{12})_i^2/(\sigma_{22})_i$$

are estimated by

$$(\hat{\beta}_{12M})_i = (\hat{\sigma}_{12M})_i/(\hat{\sigma}_{22M})_i,$$
$$(\hat{\sigma}_{11.2M})_i = (\hat{\sigma}_{11M})_i - (\hat{\sigma}_{12M})_i^2/(\hat{\sigma}_{22M})_i.$$

All the other estimators are constructed from these three estimators $(\hat{\theta})_i = [(\hat{\sigma}_{22K})_i, (\hat{\beta}_{12M})_i, (\hat{\sigma}_{11.2M})_i]$. Following the approach of Anderson (1957), we can show that, in case of constant instantaneous covariance matrix, the estimator $(\hat{\theta})_i$ is the maximum likelihood estimator of $[(\sigma_{22})_i, (\beta_{12})_i, (\sigma_{11.2})_i]$ and hence is asymptotically efficient. Now, an efficient estimator of $(\sigma_{11})_i$ is obtained from the relation

$$(\sigma_{11})_i = (\sigma_{11.2})_i + (\beta_{12})_i^2(\sigma_{22})_i$$

as given by

$$(\hat{\sigma}_{11})_i = (\hat{\sigma}_{11.2M})_i + (\hat{\beta}_{12M})_i^2(\hat{\sigma}_{22K})_i.$$

Also, the actual covariance, the actual regression coefficient and the actual correlation coefficient is given by

$$(\sigma_{12})_i = (\beta_{12})_i(\sigma_{22})_i,$$
$$(\beta_{21})_i = (\sigma_{12})_i/(\sigma_{11})_i = (\beta_{12})_i(\sigma_{22})_i/(\sigma_{11})_i,$$
$$(\rho)_i = (\sigma_{12})_i/\{(\sigma_{22})_i(\sigma_{11})_i\}^{1/2} = (\beta_{12})_i\sqrt{(\sigma_{22})_i/(\sigma_{11})_i}$$

are efficiently estimated by

$$(\hat{\sigma}_{12})_i = (\hat{\beta}_{12M})_i(\hat{\sigma}_{22K})_i,$$
$$(\hat{\beta}_{21})_i = (\hat{\beta}_{12M})_i(\hat{\sigma}_{22K})_i/(\hat{\sigma}_{11})_i,$$
$$(\hat{\rho})_i = (\hat{\beta}_{12M})_i\sqrt{(\hat{\sigma}_{22K})_i/(\hat{\sigma}_{11})_i},$$

respectively. In the following theorem, we state limiting distributions of the improved estimators.

Theorem 1. *Assume that* (i) α^* *and* Σ *are jointly independent of* $\mathbf{w}$, (ii) $\delta^{-1}\int_{(i-1)h+(j-1)\delta}^{(i-1)h+j\delta} \Sigma_{kk}(u)du$, $k = 1, 2$ *are bounded away from* 0 *and* ∞ *uniformly in* j *and* δ, (iii) *for* $k = 1, 2$, *the mean process* $\alpha^* = (\alpha^*_1, \alpha^*_2)'$ *satisfy that, as* $\delta \to 0$,

$$\delta^{-3/4} \max_{1\le j\le M} |\alpha^*_k((i-1)h + j\delta) - \alpha^*_k((i-1)h + (j-1)\delta)|$$
$$= o(1), \quad k = 1, 2.$$

Then as $M \to \infty$, *conditionally on the path of* α^* *and* Σ, *we have*

$$(h/M)^{-1/2}\{(\hat{\sigma}_{11})_i - (\sigma_{11})_i\} \xrightarrow{d} N[0, 2(\sigma_{11})_i^2\{(\rho)_i^4/\tau + 1 - (\rho)_i^4\}],$$
$$(h/M)^{-1/2}\{(\hat{\sigma}_{12})_i - (\sigma_{12})_i\}$$
$$\xrightarrow{d} N[0, (\sigma_{11})_i(\sigma_{22})_i\{2(\rho)_i^2/\tau + 1 - (\rho)_i^2\}],$$
$$(h/M)^{-1/2}\{(\hat{\beta}_{21})_i - (\beta_{21})_i\}$$
$$\xrightarrow{d} N[0, ((\sigma_{22})_i/(\sigma_{11})_i)(1 - (\rho)_i^2)$$
$$\times \{2(1 - (\rho_i)^2)(\rho)_i^2(1 + \tau^{-1}) + (1 - 2(\rho)_i^2)^2\}],$$
$$(h/M)^{-1/2}\{(\hat{\rho})_i - (\rho)_i\}$$
$$\xrightarrow{d} N[0, 2^{-1}\{1 - (\rho)_i^2\}\{(\rho_i)^2\tau^{-1} + 2 - (\rho)_i^2\}],$$

where $\xrightarrow{d}$ *denotes convergence in distribution.*

Note that, in the case of $\tau = 1$, Theorem 1 reduces to the limiting distribution of $(\hat{\sigma}_{11M})_i$, $(\hat{\sigma}_{12M})_i$, $(\hat{\beta}_{21M})_i$, $(\hat{\rho}_M)_i$ based on M-sampling established by Barndorff-Nielsen and Shephard (2004).