High frequency data and realized volatility

Taro Kanatani

Graduate School of Economics, Kyoto University

Yoshida Honmachi, Sakyo-Ku, Kyoto 6068501, JAPAN

December 2004

Acknowledgements

Foremost, I would like to thank my advisor, Professor Kimio Morimune not only for his significant and helpful comments on my research but also for his patient and generous instruction throughout my four years undergraduate and five years graduate student life.

I also would like to thank Professor Yoshihiko Nishiyama for his constructive comments and suggestions as a member of my doctoral dissertation committee.

I am grateful to all people who gave me helpful comments and warm encouragements, especially, Masato Kagihara, Ryo Okui, Naoya Sueishi, Yoko Konishi, Qing-Feng Liu, and my fellow students at Graduate School of Economics, Kyoto University. I also have benefitted from many significant comments by seminar participants at Kyoto University, the 19th JAFEE, the Statistics Summer Seminar 2003, 2004, the Autumn Meeting of the Japanese

Economic Association 2003, 2004, the 11th KKKK, the 2004 Annual Meeting of the Japan Statistical Society, and the Australian Conference of Economists 2004.

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology(MEXT), Grant-in-Aid for 21st Century COE Program "Interfaces for Advanced Economic Analysis". All calculations made in this thesis are base on programs written by the author using the matrix language GAUSS.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and motivation

Since Black and Scholes (1973) established the theory of option pricing, *volatility*[1] has played an important role not only in the derivatives pricing but also in portfolio selection and risk management. Despite of the assumption of constant volatility in Black and Scholes (1973)[2], it is widely recognized that volatility changes over time, and other various stylized facts about volatility have been documented (see, e.g., Ghysels, Harvey, and Renault (1996) and Poon and Granger (2003)). These facts have motivated many academic researchers and practitioners to study the dynamics of volatility over the last

three decades. Starting with Engle (1982)'s autoregressive heteroskedasticity (ARCH) model, various discrete time models such as Bollerslev (1986)'s generalized ARCH, Nelson (1991)'s exponential ARCH, and stochastic volatility (SV) models have been proposed (see, e.g., Poon and Granger (2003)). On the other hand, volatility is often modeled as a parameterized diffusion coefficient of continuous time diffusion process and then the parameters are estimated via the maximum likelihood methods or general method of moments (see e.g., Lo (1988), Florens-Zmirou (1993), Sueishi (2004)). The link between continuous and discrete time parametric models has been explicitly demonstrated by Drost and Nijman (1993) and Drost and Werker (1996). This thesis, however, focuses on nonparametric estimation of volatility process rather than parametric modeling of volatility structure.

In principle, the more data we can use, the more accurate the estimate will be. However, we usually have the technological restriction on the amount of data. Recently this restriction on some kind of financial data has been removing by development of computer power and data recording systems. Those kind of data are called high-frequency data. Such high-frequency data lend the validity to the method based on quadratic variation formula, that is called as *realized volatility* in the finance and econometrics literature. We

concentrate on the *ex post* volatility measuring by these type of methods.

Because of the facility of handling, tick-by-tick (transaction) data, which inherently arrive in irregular time intervals, are usually transformed into regularly spaced data through a certain interpolation. However, that interpolation method reduces the number of data and introduces the bias. The bias is serious especially in cases of cross volatility measuring. We examine this problem by proposing a new framework building on the theory of quadratic variation.

1.2 The literature

This section briefly introduces some recent literatures related to realized volatility and high frequency finance. Barndorff-Nielsen and Shephard (2004) derives asymptotic distribution of realized volatility matrix — the sum of outer products of high frequency vectors of returns. Since their purpose is to provide the asymptotic distribution theory, they establish the theory for data observed at equally spaced time intervals.

Andersen, Bollerslev, Diebold, and Labys (2003) provide methods of realized volatility incorporated into lower frequency volatility models. For

example, Using intradaily observations for the Deutschemark/Dollar and Yen/Dollar spot exchange rates, they find that forecasts from a long memory Gaussian vector autoregression for the logarithmic daily volatilities perform admirably.

Foster and Nelson (1996) provide the asymptotic distribution theory of an estimator of the spot (not integrated) covariance. Since it is essentially more difficult to estimate instantaneous volatility than integrated one, Foster and Nelson (1996)' assumption is stronger than Barndorff-Nielsen and Shephard (2004). See Andeou and Ghysels (2002) for the relationship between the spot estimator of Foster and Nelson (1996) and the integrated one of Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2004).

While all of the theories mentioned above are built on the evenly sampled observations, Malliavin and Mancino (2002) proposed an estimator base on Fourier series analysis that is well suited for unevenly sampled observations, in other words, for tick-by-tick data without any data manipulation. One of the most important purpose to use tick-by-tick data is to avoid the interpolation bias. Although it appears that the Fourier estimator is not directly involved in the quadratic variation-like method, Kanatani (2004a)

and Kanatani (2004c) provide the explicit link between them.

1.3 Outline of the thesis

The outline of the rest of the thesis is as follows. Chapter 2 is the revised version of Kanatani (2004a). In this chapter, we derive the linear interpolation bias of realized volatility. To avoid the bias, the Fourier series estimator has been proposed by Malliavin and Mancino (2002). We examine the theoretical link between the Fourier estimator and realized volatility, and show that the latter is the most efficient estimator in the class of the former. In this chapter, we focus on the analysis of univariate process.

Chapter 3 is the revised version of Kanatani (2004c). In this chapter, we define an estimator of cross-volatility (conditional covariance between two asset returns) by weighted sum of products of two returns. This estimator nests Fourier series estimator of Malliavin and Mancino (2002) and realized volatility based on interpolated returns. Each estimation method is characterized by weight matrix. We derive MSE-minimizing weights and introduce a feasible example. Our method for measuring cross-volatility is well applicable to tick-by-tick data.

Chapter 4 is the revised version of Kanatani (2004b). This chapter focuses on measuring not integrated but spot volatility. We propose an iterative method for exponentially weighted rolling regression (EWRR), which was proved to be an optimal estimator of volatility by Foster and Nelson (1996). The method accelerates the numerical evaluation of EWRR under certain circumstances. An alternative to usual realized volatility is proposed for its application.

Notes

1 Throughout this thesis, we use the term "volatility" to reference both variance (not standard deviation) and covariance.

2 Hull and White (1987) modifies Black and Scholes (1973)'s option pricing formula for stochastic volatility.

Chapter 2

Estimation of univariate integrated volatility

Because of the facility of handling, tick-by-tick data, which are usually unevenly sampled, are transformed into regularly spaced data through a certain interpolation. However, that interpolation method reduces the number of data and introduces the bias. Through Monte Carlo simulations, Barucci and Renò (2002) demonstrates that linear interpolation introduces a downward bias to realized volatilities. In Section 2.2, we theoretically derives the linear interpolation bias of realized volatility. To avoid these problems of interpolation methods, Malliavin and Mancino (2002) proposed an estimator

based on Fourier series analysis that is well suited for unevenly sampled data. In Section 2.3, we derive a theoretical relationship between the Fourier series estimator and realized volatility. The latter is proved to be the most efficient estimator in the class of the former. In Section 2.4, we confirm our theory through a Monte Carlo simulation. Throughout this chapter, we restrict our attention to univariate setting. Multivariate situation will be studied in the next chapter.

2.1 Date generating process and observed time points

This chapter specifically addresses the following situation. Let p_t be a logarithmic asset price that is generated by diffusion:

$$dp(t) = \sigma(t)dW(t),\ 0 \leq t \leq T, \tag{2.1}$$

where $W(t)$ is a standard Brownian Motion and $\sigma(t)$ is a random time dependent function. That diffusion is observed at $(N+1)$ irregular time points:[1]

$$0 = t_0 < t_1 < \cdots < t_i < \cdots < t_N = T$$

We assume that every time difference (duration) is sufficiently small:

$$\lim_{N\to\infty} \sup_{i\geq 1} (t_i - t_{i-1}) = 0.$$

For purposes of simplification, we set the drift of diffusion as 0. This simplification is acceptable not only because it means an efficient market in financial economics, but also because, mathematically, the martingale component swamps the predictable portion over short time intervals. In such a situation, we study the nonparametric estimators of integrated volatility $\int_0^T \sigma^2(t)dt$. Because we make no hypothesis on the structure of the underlying probability space Ω, we can construct an auxiliary probability space X where we consider $\sigma(t)$ as a deterministic function, see Malliavin and Mancino (2002).

Throughout this chapter, E denotes the expectation on the probability space X.

2.2 Realized volatility from evenly spaced observations

In this section, we examine realized volatility from evenly spaced data. Unevenly sampled raw data are converted into evenly spaced data through in-

terpolation. We consider two interpolation methods for converting N raw data, $\{p(t_i)\}_{i=0}^{N}$, to m evenly spaced data, $\{q(jT/m)\}_{j=0}^{m}$:

$$q\left(\frac{jT}{m}\right) = \begin{cases} (1-\rho_j)\,p\left(t_j^-\right) + \rho_j p\left(t_j^+\right) & \text{linear interpolation} \\ p\left(t_j^-\right) & \text{previous-tick interpolation} \end{cases} \tag{2.2}$$

where

$$\rho_j = \frac{(jT/m) - t_j^-}{t_j^+ - t_j^-},$$

$$t_j^- = \max\{t_i : t_i \leq jT/m\},$$

$$t_j^+ = \min\{t_i : t_i \geq jT/m\},$$

and where $\max A$ and $\min A$ denote maximum and minimum elements of A, respectively.

Using the evenly spaced data series $\{q(jT/m)\}_{j=0}^{m}$, the volatility is measured by the following estimator.

$$\hat{\sigma}^2(m) = \sum_{j=1}^{m}\left(q\left(\frac{jT}{m}\right) - q\left(\frac{(j-1)T}{m}\right)\right)^2. \tag{2.3}$$

Whereas Barucci and Renò (2002) found through Monte Carlo simulation that linear interpolation procedures introduce bias into realized volatility

(2.3), we derive the theoretical linear interpolation bias as

$$-\frac{m}{T}\sum_{j=1}^{m}\left\{\rho_{j-1}\left(1-\rho_{j-1}\right)\left(t_{j-1}^{+}-t_{j-1}^{-}\right)+\rho_{j}\left(1-\rho_{j}\right)\left(t_{j}^{+}-t_{j}^{-}\right)\right\}\int_{(j-1)T/m}^{jT/m}\sigma^{2}\left(t\right)dt. \tag{2.4}$$

See Appendix A for derivation of (2.4). Note that the downward bias (2.4) is more pronounced: (a) when the time window $[0,T]$ is divided more finely (m is large); (b) when the interpolated time point is far from the observed time points ($|\rho_j - 1/2|$ is small); (c) in coarsely-sampled periods ($t_j^+ - t_j^-$ is large); or (d) in the volatile period ($\sigma^2(t)$ is large).

In the case of previous-tick interpolation, the realized volatility (2.3) is unbiased since the diffusion is observed at $t = 0$ and $t = T$ ($t_0 = 0$ and $t_N = T$).

2.3 Estimators using raw data

Malliavin and Mancino (2002) proposed a method based on Fourier series to use unevenly sampled data. In this section, we normalized the time window $[0,T]$ to $[0,2\pi]$. The Fourier estimator of integrated volatility $\int_0^T \sigma^2(t)dt$ is

given as

$$\hat{\sigma}_F^2 = \frac{\pi^2}{K}\sum_{k=1}^{K}(a_k^2(dp) + b_k^2(dp)), \tag{2.5}$$

where

$$a_k(dp) = \frac{1}{\pi}\int_0^{2\pi}\cos(kt)dp(t), \tag{2.6}$$

$$b_k(dp) = \frac{1}{\pi}\int_0^{2\pi}\sin(kt)dp(t), \tag{2.7}$$

and K is a large integer. In practice, we compute the integrals (2.6) through integration by parts, as

$$\begin{aligned} a_k(dp) &= \frac{1}{\pi}\int_0^{2\pi}\cos(kt)dp(t) \\ &= \frac{p\,(2\pi) - p\,(0)}{\pi} + \frac{1}{\pi}\int_0^{2\pi}\sin(kt)p(t)dt \\ &\approx \frac{p\,(2\pi) - p\,(0)}{\pi} + \frac{1}{\pi}\sum_{i=0}^{N-1}\left[\cos(kt_i) - \cos(kt_{i+1})\right]p(t_i) \end{aligned} \tag{2.8}$$

because the piecewise constant is valid under assumption $\lim_{N\to\infty}\sup_{i\geq 1}(t_i - t_{i-1}) = 0$. Similarly, we approximate (2.7) by

$$b_k(dp) \approx \frac{1}{\pi}\sum_{i=0}^{N-1}\left[\sin(kt_i) - \sin(kt_{i+1})\right]p(t_i). \tag{2.9}$$

Another method using unevenly sampled observations $\{p\,(t_i)\}_{i=0}^{N}$ is an estimator based on the quadratic variation formula:

$$\hat{\sigma}_R^2 = \sum_{i=1}^{N}\left\{\Delta p(t_i)\right\}^2, \tag{2.10}$$

where $\Delta p(t_i) = p(t_i) - p(t_{i-1})$. To distinguish (2.10) from (2.3), we refer to (2.10) as *raw data realized volatility*. The Fourier estimator (3.8) can be rewritten as

$$\hat{\sigma}_F^2 = \hat{\sigma}_R^2 + \sum_{i \neq j} \Delta p\left(t_i\right) \Delta p\left(t_j\right) w_{ij}, \tag{2.11}$$

where

$$w_{ij} = \frac{\sin \frac{(K+1)(t_i - t_j)}{2} \cos \frac{K(t_i - t_j)}{2}}{K \sin \frac{(t_i - t_j)}{2}}. \tag{2.12}$$

See Appendix B for the derivation. (2.11) and (2.12) imply that as $K \to \infty$, $\hat{\sigma}_F^2 \to \hat{\sigma}_R^2$. Because the second parts of (2.11) are uncorrelated to $\hat{\sigma}_R^2$,

$$\begin{aligned} V\left(\hat{\sigma}_F^2\right) &= V\left(\hat{\sigma}_R^2\right) + V\left(\sum_{i \neq j} \Delta p\left(t_i\right) \Delta p\left(t_j\right) w_{ij}\right) \\ &= V\left(\hat{\sigma}_R^2\right) + \sum_{i \neq j} \left\{ h \int_{t_{i-1}}^{t_i} \sigma^2\left(t\right) dt \right\} \left\{ \int_{t_{j-1}}^{t_j} \sigma^2\left(t\right) dt \right\} w_{ij}^2 \\ &> V\left(\hat{\sigma}_R^2\right). \end{aligned} \tag{2.13}$$

In other words, as $K \uparrow \infty$, $V\left(\hat{\sigma}_F^2\right) \downarrow V\left(\hat{\sigma}_R^2\right)$. That is to say, (2.10) is the most efficient estimator in the class of (3.8).

2.4 Monte Carlo study

We follow the Monte Carlo design of Barucci and Renò (2002) with little modification: we generate a proxy for continuous observation by discretizing

the following stochastic differential equations with a time step of one second:

$$d\log\sigma^2(t) = -k\log\sigma^2(t)\,dt + \gamma dW_\sigma(t)$$

$$dp(t) = \sigma(t)dW_p(t),\ 0 \le t \le T,$$

where W_σ and W_p are mutually independent standard Brownian motions, $k = 0.01, \gamma = 0.1$, and $T = 60 \times 60 \times 24$ seconds (s). Time differences are drawn from an exponential distribution with a mean of 45 s.[2] We compare the performances of estimators (2.3), (3.8), and (2.10). In calculations of (2.3), we set $m = 144,\ 288$, and 720, corresponding to so-called daily realized volatility based on 10-min, 5-min, and 2-min returns. Each return is computed by two interpolation methods in (2.2). In (3.8), we set $K = 10,\ 50,\ 100,\ 500$, and $[N/2]$ where $[\cdot]$ denotes a Gaussian symbol.[3] In our simulation, $[N/2]$ is expected to be around $60 \times 60 \times 24 \div (45 \times 2) = 960$.

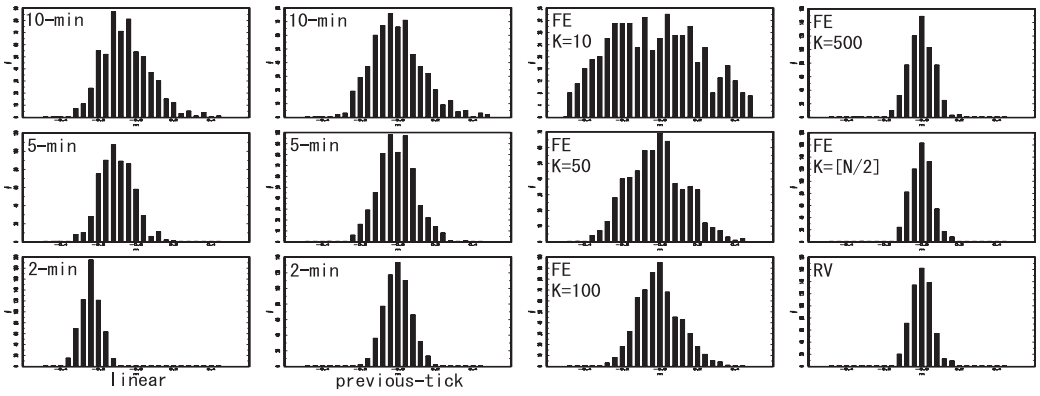
We performed 600 replications.

Figure 2.1 shows the distributions of normalized errors

$$\frac{\hat{\sigma}^2(m) - \int_0^T \sigma^2(t)dt}{\int_0^T \sigma^2(t)dt},\quad \frac{\hat{\sigma}_F^2 - \int_0^T \sigma^2(t)dt}{\int_0^T \sigma^2(t)dt},\ \text{and}\ \frac{\hat{\sigma}_R^2 - \int_0^T \sigma^2(t)dt}{\int_0^T \sigma^2(t)dt}. \qquad (2.14)$$

Table 2.1 reports the means and standard deviations of (2.14) from that set of 600 replications. Increasing the number of partitions m, we can reduce the variance of realized volatility. However, as stated in (2.4), in the case

Figure 2.1: Distribution of (2.14)



Note: 10-min, 5-min, and 2-min denote the estimators (2.3) with m = 144, 288, and 720, respectively, through linear and previous-tick interpolation. FE signifies the Fourier estimator (3.8) with $K = 10,\ 50,\ 100,\ 500,$ and $[N/2]$. RV denotes the raw data realized volatility (2.10). The distribution is computed with 600 'daily' replications.

of using linear interpolation, the downward bias increases. In contrast, the realized volatility is unbiased when using previous-tick interpolation. As stated in (2.13), Figure 2.1 and Table 2.1 show that as $K \uparrow \infty$, $V(\hat{\sigma}_F^2) \downarrow V(\hat{\sigma}_R^2)$.

Table 2.2 compares means of the theoretical linear interpolation bias (2.4) and measurement error of $\hat{\sigma}^2(m) - \int_0^T \sigma^2(t)dt$ from 600 replications. Both means are approximately consistent. We can verify the validity of (2.4) by

these results.

2.5 Summary

In this chapter we derived the linear interpolation bias of realized volatility. Results indicate that linear interpolation should not be used as the preparation for realized volatility calculations. The theoretical relationship between the Fourier series estimator proposed by Malliavin and Mancino (2002) and raw data realized volatility implies that the latter is the most efficient estimator in the class of the former. The result of this chapter will be generalized in the next chapter.

Notes

1 For the purpose of simplification, we set $t_0 = 0$ and $t_N = T$. These assumptions can be ignored if the number of observations is sufficiently large.

2 Although each duration is independent in our simulation, our method requires no assumption except that $\sup_{i \geq 1} (t_i - t_{i-1})$ is small. See Engle and Russell (1998) for the autoregressive time duration models. See e.g., Aït-Sahalia and Mykland (2003) for an example of the exponentially distributed duration.

3 $[N/2]$ is the so-called Nyquist frequency if observations are sampled evenly.

Table 2.1: Means and standard deviations of (2.14) from 600 'daily' replications

linear	10-min	-0.04734	(0.12293)
	5-min	-0.09763	(0.08937)
	2-min	-0.23911	(0.05152)
previous-tick	10-min	0.00109	(0.13139)
	5-min	-0.00092	(0.09864)
	2-min	-0.00017	(0.07086)
Fourier estimator	$K = 10$	0.00107	(0.33051)
	$K = 50$	-0.01371	(0.15253)
	$K = 100$	-0.00341	(0.11409)
	$K = 500$	-0.00252	(0.06538)
	$K = [N/2]$	-0.00055	(0.05730)
raw data realized volatility		-0.00094	(0.05460)

Note: Standard deviations are given in parentheses. 10-min, 5-min, and 2-min denote the estimators (2.3) with $m = 144$, 288, and 720, respectively, through two different interpolations in (2.2). Fourier estimators are computed with five different Ks. Raw data realized volatility denotes the estimator (2.10).

Table 2.2: Linear interpolation bias

	$\hat{\sigma}^2(m) - \int_0^T \sigma^2(t)dt$		bias (2.4)	
10-min	-5291.56	(13682.3)	-5436.57	(478.465)
5-min	-10858.1	(9953.48)	-10915.0	(720.160)
2-min	-26675.4	(5907.82)	-27360.9	(1464.48)

Note: Means of measurement errors of realized volatility (2.3) through linear interpolation procedures and the theoretical linear interpolation bias (2.4) from 600 replications. Standard deviations are given in parentheses.

Chapter 3

Estimation of integrated cross volatility

In this chapter we generalize the analysis to the multivariate settings and focus on the estimation of conditional covariance between two asset returns. The conditional covariance is referred to as *cross volatility* in financial literatures.

In Section 3.2 we define weighted realized volatility as a estimator of integrated cross volatility and show that it nests several estimators mentioned in the previous chapter. In Section 3.3 we derive the MSE-minimizing weight and provide a feasible example of it. Through a Monte Carlo simulation, we

examine our theory in Section 3.4. Section 3.5 summarizes this chapter and overviews future studies.

3.1 Data generating process and observations

We consider n-dimensional logarithmic price $p(t) = (p_1(t), \cdots, p_n(t))'$ which follows the stochastic differential equation:

$$dp(t) = \Sigma(t)\, dz(t), \quad 0 \leq t \leq T$$

where $\Sigma(t)$ is an $n \times n$ matrix $[\sigma_{ij}(t)]_{i,j=1,\cdots,n}$, and z is an $n \times 1$ vector of independent standard Brownian motions. We set the drift vector as 0, as well as the previous chapter.[1] We define the volatility matrix as

$$\Omega \equiv \Sigma\Sigma',$$

that is to say, cross volatility between ith and jth asset is denoted as the ij element of Ω:

$$\omega_{ij}(t) = \sum_{k=1}^{n} \sigma_{ik}(t)\, \sigma_{jk}(t).$$

Each ith asset price is observed at irregular time points $\{t_k^i\}_{k=0}^{N_i}$.[2] We just impose the assumption on the observation points that the time intervals are small: $\lim_{N_i \to \infty} \sup_{j \geq 1} \left(t_j^i - t_{j-1}^i\right) = 0$. Since we concentrate on the *ex post*

cross volatility measuring and do not make any hypothesis on the structure of the underlying probability space Ω, we can construct an auxiliary probability space X where we consider $\Sigma(t)$ as deterministic functions. See Malliavin and Mancino (2002). Through this paper, E denotes the expectation on the probability space X.

3.2 Weighted realized volatility

3.2.1 Representation

We define the estimator of $\int_0^T \omega_{ij}(t)\,dt$ as

$$\hat{\omega}_{ij} = \Delta p_i' W \Delta p_j = \sum_{k=1}^{N_i}\sum_{l=1}^{N_j} \Delta p_i\left(t_k^i\right)\Delta p_j\left(t_l^j\right) w_{kl} \tag{3.1}$$

where

$$\Delta p_i = \begin{pmatrix} p_i(t_1^i) - p_i(t_0^i) \\ \vdots \\ p_i(t_{Ni}^i) - p_i\left(t_{N_i-1}^i\right) \end{pmatrix}, W = \begin{pmatrix} w_{11} & \cdots & w_{1N_j} \\ \vdots & \ddots & \vdots \\ w_{N_i1} & \cdots & w_{N_iN_j} \end{pmatrix}.$$

We call (3.1) *weighted realized volatility*. (3.1) nests several estimators of the integrated volatility $\int_0^T \omega_{ij}(t)\,dt$. For example, if $w_{ij} = 1$ for any k, l,

$$\begin{aligned}
\hat{\omega}_{ij} &= \sum_{k=1}^{N_i}\sum_{l=1}^{N_j} \Delta p_i\left(t_k^i\right)\Delta p_j\left(t_l^j\right) \qquad (3.2)\\
&= \left\{\sum_{k=1}^{N_i}\Delta p_i\left(t_k^i\right)\right\}\left\{\sum_{l=1}^{N_j}\Delta p_j\left(t_l^j\right)\right\}\\
&= \left\{p_i\left(t_{Ni}\right) - p_i\left(t_0\right)\right\}\left\{p_j\left(t_{Nj}\right) - p_j\left(t_0\right)\right\}\\
&= \left\{p_i\left(T\right) - p_i\left(0\right)\right\}\left\{p_j\left(T\right) - p_j\left(0\right)\right\}
\end{aligned}$$

which is an unbiased but very noisy estimator of $\int_0^T \omega_{ij}(t)\,dt$. If the window $[0, T]$ is one day, (3.2) means that we measure daily (cross) volatility by using daily return, in other words, discarding all intradaily data of $\{p_i(t_k^i)\}_{k=1}^{N_i-1}$. In this manner, the weight matrix characterizes the data for measuring volatility. In order to understand this point, we look at an another example. In univariate settings, if $W = I_{N_i}$,

$$\hat{\omega}_{ii} = \sum_{k=1}^{N_i}\left(p_i\left(t_k^i\right) - p_i\left(t_{k-1}^i\right)\right)^2.$$

Note that this estimator uses all available observations, therefore, is expected to be less noisy. We discuss the multivariate version of this in the subsection 3.2.4. Through the following three subsections, the examples of (3.1) are discussed.

3.2.2 Interpolation and realized volatility

The raw data which are unevenly spaced, are converted to evenly spaced data in order to apply to the usual discrete time series analysis. Dacorogna, Gençay, Müller, Olsen, and Pictet (2001) introduces some interpolation methods including *linear interpolation* and *previous tick interpolation.*[3] When constructing M evenly spaced data $\{q\,(mT/M)\}_{j=0}^{M}$ from $\{p_i\,(t_k^i)\}_{k=0}^{N_i}$, those data manipulation is as follows:

$$q_i\left(\frac{mT}{M}\right) = \begin{cases} (1-\rho_m^i)\,p_i\,({}_*t_m^i) + \rho_m^i p_i\,({}^*t_m^i) & \text{linear interpolation} \\ p_i\,({}_*t_m^i) & \text{previous-tick interpolation} \end{cases} \tag{3.3}$$

where

$$\begin{aligned} \rho_m^i &= \frac{(mT/M) - {}_*t_m^i}{{}^*t_m^i - {}_*t_m^i}, \\ {}_*t_m^i &= \max\left\{t_k^i : t_k^i \le mT/M\right\}, \\ {}^*t_m^i &= \min\left\{t_k^i : t_k^i \ge mT/M\right\}, \end{aligned}$$

and where $\max A$ and $\min A$ denote maximum and minimum elements of A, respectively.

Using evenly spaced data of $\{q_i(mT/M)\}_{m=0}^{M}$ and $\{q_j(mT/M)\}_{m=0}^{M}$, the

integrated cross volatility $\int_0^T \omega_{ij}(t)dt$ is measured by the following estimator,

$$\hat{\omega}_{ij}(M) = \sum_{m=1}^{M} \left(q_i \left(\frac{mT}{M} \right) - q_i \left(\frac{(m-1)\,T}{M} \right) \right) \left(q_j \left(\frac{mT}{M} \right) - q_j \left(\frac{(m-1)\,T}{M} \right) \right). \tag{3.4}$$

In order to distinguish difference on the interpolation procedure, we introduce the notation of $\hat{\omega}_{ij}^{L}(M)$ and $\hat{\omega}_{ij}^{P}(M)$ for liner interpolation and previous-tick interpolation, respectively. As mentioned in the previous chapter, $\hat{\omega}_{ii}^{L}(M)$ has the downward bias. Barucci and Renò (2002) reports the linear interpolation bias through Monte Carlo simulations. Kanatani (2004a) calculates the theoretical bias. As we use higher and higher frequency data, the bias becomes more profound. Thus, the linear interpolation is not suitable for calculation of realized volatility.

On the other hand, $\hat{\omega}_{ii}^{P}(M)$ is unbiased. The bias of $\hat{\omega}_{ij}^{P}(M)$ is

$$\sum_{m=1}^{M} \int_{t_m^-}^{t_m^+} \omega_{ij}\,(t)\,dt \tag{3.5}$$

where

$$t_m^- = \min\left\{ \max\left\{ t_k^i : t_k^i \le mT/M \right\}, \max\left\{ t_l^j : t_l^j \le mT/M \right\} \right\},$$

$$t_m^+ = \max\left\{ \max\left\{ t_k^i : t_k^i \le mT/M \right\}, \max\left\{ t_l^j : t_l^j \le mT/M \right\} \right\}.$$

Notice that in the case of univariate volatility ($i = j$), for $t_m^- = t_m^+$, the realized volatility through previous tick interpolation is an unbiased estimator.

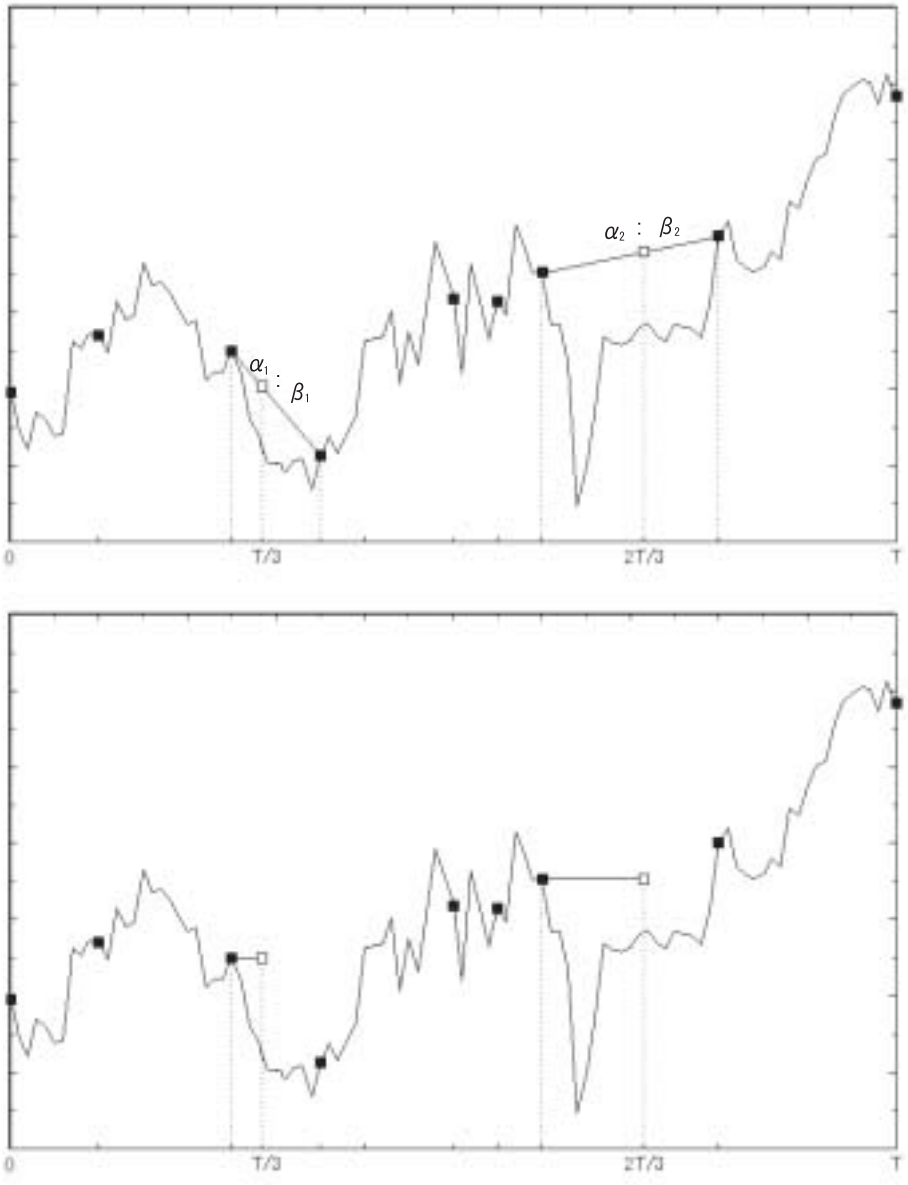
In order to show that the realized volatility (3.4) can be written by the expression of the weighted realized volatility (3.1), we shall present a simple example.

Example 1 *Let us consider a simple case as shown in Figure 3.1:* $M = 3, N_i = 8$. $\hat{\omega}_{ii}^{L}(M)$ *can be written by the form of weighted realized volatility (3.1) with the weight matrix:*

$$W = \begin{pmatrix} 1 & 1 & \alpha_1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & \alpha_1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_1 & \alpha_1 & \alpha_1^2+\beta_1^2 & \beta_1 & \beta_1 & \beta_1 & \beta_1\alpha_2 & 0 \\ 0 & 0 & \beta_1 & 1 & 1 & 1 & \alpha_2 & 0 \\ 0 & 0 & \beta_1 & 1 & 1 & 1 & \alpha_2 & 0 \\ 0 & 0 & \beta_1 & 1 & 1 & 1 & \alpha_2 & 0 \\ 0 & 0 & \beta_1\alpha_2 & \alpha_2 & \alpha_2 & \alpha_2 & \alpha_2^2+\beta_2^2 & \beta_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \beta_2 & 1 \end{pmatrix}. \tag{3.6}$$

See Appendix C for the detail derivation of (3.6). Since previous tick interpolation is a special case of the linear interpolation for $\alpha_m = 0$ *and* $\beta_m = 1$, $\hat{\omega}_{ii}^{P}(M)$ *can be written by the form of weighted realized volatility (3.1) with*

Figure 3.1: Linear interpolation and Previous-tick interpolation



Note: Linear interpolation (upper) and Previous-tick interpolation (lower).

Black and white squares denote observed and interpolated data respectively.

the weight matrix:

$$W = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}. \tag{3.7}$$

3.2.3 Fourier series estimator of Malliavin and Mancino (2002)

Malliavin and Mancino (2002) proposed a new method for measuring volatility by using Fourier series. The method is especially suitable for unevenly sampled observations. We prove that the Fourier series estimator can be written by the form of the weighted realized volatility. In this subsection, we normalized the time window $[0, T]$ to $[0, 2\pi]$.

$$\hat{\omega}_{ij}^{F} = \frac{\pi^2}{Q} \sum_{q=1}^{Q} (a_q(dp_i)a_q(dp_j) + b_q(dp_i)b_q(dp_j)) \tag{3.8}$$

where

$$a_q(dp_i) = \frac{1}{\pi}\int_0^{2\pi} \cos(qt) dp_i(t), \tag{3.9}$$

$$b_q(dp_i) = \frac{1}{\pi}\int_0^{2\pi} \sin(qt) dp_i(t), \tag{3.10}$$

and Q is a large integer. We will compute the integrals (3.9) and (3.10) through integration by parts:

$$\begin{aligned} a_q(dp_i) &= \frac{1}{\pi}\int_0^{2\pi} \cos(qt) dp(t) \\ &= \frac{p_i(2\pi) - p_i(0)}{\pi} + \frac{1}{\pi}\int_0^{2\pi} \sin(qt) p_i(t) dt \\ &\approx \frac{p_i(2\pi) - p_i(0)}{\pi} + \frac{1}{\pi}\sum_{j=0}^{N_i-1} \left[\cos(qt_j^i) - \cos(qt_{j+1}^i)\right] p_i(t_j^i) \end{aligned}$$

since the piecewise constant is valid under assumption $\lim_{N\to\infty} \sup_{j\geq 1} \left(t_j^i - t_{j-1}^i\right) = 0$. Similarly, we approximate (3.10) by

$$b_q(dp_i) \approx \frac{1}{\pi}\sum_{j=0}^{N-1} \left[\sin(qt_j^i) - \sin(qt_{j+1}^i)\right] p_i(t_j^i). \tag{3.11}$$

These approximation of the integrals is proved to be equivalent to setting the weight in (3.1) as

$$w_{kl} = \begin{cases} 1 & \text{if } t_k^i = t_l^j, \\ \dfrac{\sin\frac{(Q+1)\left(t_k^i - t_l^j\right)}{2} \cos\frac{Q\left(t_k^i - t_l^j\right)}{2}}{Q \sin\frac{\left(t_k^i - t_l^j\right)}{2}} & \text{otherwise} \end{cases}.$$

See the Appendix D. In the special case of univariate volatility ($i = j$), as we increase the number of Fourier coefficients ($Q \to \infty$), the weight matrix converges to identity matrix ($W \to I_{N_i}$). In the case of cross volatility ($i \neq j$), since transaction is usually nonsynchronous, $t_k^i - t_l^j$ has some width. Therefore, as $K \to \infty$, $w_{kl} \to 0$: $\hat{\omega}_{ij}^F \to 0$. Thus we should not increase the number of Fourier coefficients.

3.2.4 Raw data realized volatility

Another method for measuring integrated volatility using unevenly sampled observations $\{p_i(t_k^i)\}_{k=0}^{N_i}$ is

$$\hat{\omega}_{ii}^R = \sum_{i=1}^{N_i} \left\{\Delta p_i\left(t_k^i\right)\right\}^2$$

This estimator is also written by the form of weighted realized volatility with identity matrix I_{N_i}. As mentioned in the previous chapter, the relationship between raw data realized volatility and Fourier series estimator is as follows,

$$\hat{\omega}_{ii}^F \to \hat{\omega}_{ii}^R \text{ and } V(\hat{\omega}_{ii}^F) \downarrow V(\hat{\omega}_{ii}^R) \text{ as } Q \to \infty.$$

For measuring cross volatility, we extend the method using unevenly sampled observations $\{p_i(t_k^i)\}_{k=0}^{N_i}$ and $\{p_j\left(t_k^j\right)\}_{k=0}^{N_j}$:

$$\hat{\omega}_{ij}^R = \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \Delta p_i\left(t_k^i\right) \Delta p_j\left(t_l^j\right) I(A) \tag{3.12}$$

where $A = \left\{ \left(t_k^i, t_{k-1}^i \right) \cap \left(t_l^j, t_{l-1}^j \right) \neq \emptyset \right\}$ and $I(\cdot)$ denotes indicator function. We refer to (3.12) as *raw data realized (cross) volatility*. (3.12) is expressed by the form of weighted realized volatility with the weights:

$$w_{kl} = \begin{cases} 1 & \text{if } \left(t_k^i, t_{k-1}^i \right) \cap \left(t_l^j, t_{l-1}^j \right) \neq \emptyset, \\ 0 & \text{otherwise} \end{cases} .$$

Although all estimators of cross volatility mentioned above introduce the bias, this simple estimator is constructed to be unbiased.

3.3 Optimal weight

3.3.1 MSE-minimizing weight

In this subsection, we derive the optimal weight that minimizes the MSE of (3.1):

$$E\left(\hat{\omega}_{ij} - \int_0^T \omega_{ij}(t)\, dt \right)^2 = bias^2 + V\left(\hat{\omega}_{ij} \right).$$

We define the intersection interval as

$$I(k,l) \equiv \left(t_k^i, t_{k-1}^i \right) \cap \left(t_l^j, t_{l-1}^j \right).$$

We introduce a convenient notation for the element of weight matrix W as follows

$$\begin{cases} w_{kl}^{A} & \text{if } I(k,l) \neq \emptyset, \\ w_{kl}^{A^C} & \text{otherwise} \end{cases}.$$

The bias is given by

$$\begin{aligned} & E\left(\sum_{k=1}^{N_i}\sum_{l=1}^{N_j}\Delta p_i\left(t_k^i\right)\Delta p_j\left(t_l^j\right)w_{kl}\right) - \int_0^T \omega_{ij}\left(t\right)dt \\ &= E\left(\sum\int_{t_{k-1}}^{t_k} dp_i\left(t\right)\int_{t_{l-1}}^{t_l} dp_j\left(t\right)w_{kl}^{A}\right) - \int_0^T \omega_{ij}\left(t\right)dt \\ &= \sum\int_{I(k,l)} \omega_{ij}dt w_{kl}^{A} - \int_0^T \omega_{ij}\left(t\right)dt. \end{aligned} \tag{3.13}$$

Note that if $w_{kl}^{A} = 1$, the bias is zero. On the other hand, the variance is given by

$$\begin{aligned} & V\left(\hat{\omega}_{ij}\right) \\ &= \sum\left\{\left(\int_{I(k,l)}\omega_{ij}dt\right)^2 + \left(\int_{t_{k-1}}^{t_k}\omega_{ii}dt\right)\left(\int_{t_{l-1}}^{t_l}\omega_{jj}dt\right)\right\}\left(w_{kl}^{A}\right)^2 \\ &+ \sum\left(\int_{t_{k-1}}^{t_k}\omega_{ii}dt\right)\left(\int_{t_{l-1}}^{t_l}\omega_{jj}dt\right)\left(w_{kl}^{A^C}\right)^2 \end{aligned} \tag{3.14}$$

See the Appendix E. It is obvious that we should set $w_{kl}^{A^C} = 0$ in order to minimize the MSE because $\omega_{ii}(t)$ is nonnegative.

For example, compare the identity matrix with (3.7). Although diagonal element of these two are identical, (3.7) has some non-zero non-diagonal

elements ($w_{kl}^{A^C} \neq 0$). This means that variance of previous-tick realized volatility is larger than raw data realized volatility. As another example, remember the weight matrices of (3.8) and (3.12) in the case of univariate volatility. Both of them have the same diagonal elements $w_{kl}^{A} = 1$, while (3.8) have non-zero $w_{kl}^{A^C}$. Therefore, variance of (3.8) is larger than that of (3.12). As $Q \to \infty$, $w_{kl}^{A^C}$ of (3.8) goes to 0, then these two are almost same. See Kanatani (2004a).

In order to minimize the MSE, we set $w_{kl}^{A^C} = 0$ and then rewrite the MSE in matrix expression.

$$MSE = w'Dw + \left(x'\left(w - 1\right)\right)^2$$

where

$$w = \begin{pmatrix} w_{11}^A \\ \vdots \\ w_{kl}^A \\ \vdots \\ w_{N_iN_j}^A \end{pmatrix}, D = \begin{pmatrix} v_{11} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & v_{kl} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & v_{N_iN_j} \end{pmatrix},$$

$$x = \begin{pmatrix} \int_{I(1,1)} \omega_{ij} dt \\ \vdots \\ \int_{I(k,l)} \omega_{ij} dt \\ \vdots \\ \int_{I(N_i,N_j)} \omega_{ij} dt \end{pmatrix}, 1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

$$v_{kl} = \left(\int_{I(k,l)} \omega_{ij} dt \right)^2 + \left(\int_{t_{k-1}}^{t_k} \omega_{ii} dt \right) \left(\int_{t_{l-1}}^{t_l} \omega_{jj} dt \right).$$

Let

$$u_{kl} = \frac{\left(\int_{I(k,l)} \omega_{ij} dt \right)^2}{v_{kl}},$$

then we get the following theorem.

Theorem 2 *The MSE of (3.1) is globally minimized by the following weight:*

$$w_{kl}^A = \frac{\int_0^T \omega_{ij} dt \int_{I(k,l)} \omega_{ij} dt}{v_{kl} \{1 + \sum u_{kl}\}}. \tag{3.15}$$

The bias and variance obtained by using the optimal weight (3.15) are

$$\frac{-\int_0^T \omega_{ij} dt}{1 + \sum u_{kl}} \qquad \textit{and} \tag{3.16}$$

$$\left(\frac{\int_0^T \omega_{ij} dt}{1 + \sum u_{kl}} \right)^2 \sum u_{kl}, \tag{3.17}$$

respectively. The minimized MSE is

$$\left(\frac{\int_0^T \omega_{ij} dt}{1 + \sum u_{kl}} \right)^2 \left\{ 1 + \sum u_{kl} \right\}. \tag{3.18}$$

Proof. See Appendix F ■

In order to understand the property of the optimal weight, consider a special case of the individual volatility ($i = j$). Since

$$v_{kk} = 2 \left(\int_{t_{k-1}}^{t_k} \omega_{ij} dt \right)^2 \quad \text{and} \quad u_{kk} = \frac{1}{2},$$

W is an $N_i \times N_i$ diagonal matrix that has diagonal elements

$$w_{kk}^A = \frac{1}{N_i + 2} \frac{\int_0^T \omega_{ii} dt}{\int_{t_{k-1}}^{t_k} \omega_{ii} dt}.$$

This weight increases (decreases) when $\int_{t_{k-1}}^{t_k} \omega_{ii} dt$ decreases (increases). This fact implies that larger (smaller) weights are assigned in densely (coarsely) sampled periods and that smaller (larger) weights are assigned in volatile (less volatile) periods. The bias and variance are

$$\frac{-2}{N_i + 2} \int_0^T \omega_{ii}(t)\, dt \qquad \text{and}$$

$$\frac{2N_i}{(N_i + 2)^2} \left(\int_0^T \omega_{ii}(t)\, dt \right)^2,$$

respectively. The estimator is not unbiased, however, the bias shrinks at order $1/N_i$. The variance also shrinks at order $1/N_i$ in similar fashion to the variance of *realized variance* of Barndorff-Nielsen and Shephard (2004)[4].

3.3.2 An estimator of nuisance parameters

To construct the optimal weight of Theorem 2, we must estimate $\int_{I(k,l)} \omega_{ij} dt$. We call it *piecewise integrated volatilities* (PWIV). It is essentially difficult to estimate them. We give an example of unbiased estimators: we use

$$\Delta p_i \left(t_k^i\right) \Delta p_j \left(t_l^j\right) \text{ and } \left\{\Delta p_i \left(t_k^i\right)\right\}^2,$$

as estimators of $\int_{I(k,l)} \omega_{ij} dt$ and $\int_{t_{k-1}}^{t_k} \omega_{ii} dt$ respectively. We also need an estimate of $\int_0^T \omega_{ij}(t)\, dt$, in Monte Carlo study of next section, we use (3.12). By using these estimators to construct the weights, the weighted realized volatility (3.1) is equivalent to

$$\hat{\omega}_{ij}^N = \frac{N_{ij} \hat{\omega}_{ij}^R}{N_{ij} + 2} \tag{3.19}$$

where $N_{ij} = N_i + N_j - \sum I\left(\left\{t_k^i = t_l^j\right\}\right)$. We refer to (3.19) as naively weighted realized volatility. Although, there is little difference between $\hat{\omega}_{ij}^N$ and $\hat{\omega}_{ij}^R$ when N_{ij} is large, we find that $\hat{\omega}_{ij}^N$ slightly improves the MSE compared with $\hat{\omega}_{ij}^R$ in the Monte Carlo study of next section.

3.4 Monte Carlo study

We examine the above theory through a Monte Carlo study. Without loss of generality, we set the number of assets as two. We follow the Monte Carlo design of Barucci and Renò (2002) with some modification for multivariate setting: we generate proxy for continuous observations by discretizing following stochastic differential equations with a time step of one second,

$$\begin{pmatrix} dp_1(t) \\ dp_2(t) \end{pmatrix} = \begin{pmatrix} \sigma_{11}(t) & \sigma_{12}(t) \\ \sigma_{21}(t) & \sigma_{22}(t) \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}, \ 0 \leq t \leq T$$

$$d\sigma_{ij}(t) = \kappa_{ij}(\theta_{ij} - \sigma_{ij}(t))\, dt + \gamma_{ij} dW_{ij}(t), \ i, j = 1, 2.$$

where $\kappa_{ij} = 0.01$, $\theta_{ij} = 0.01$, and $\gamma_{ij} = 0.001$ for any i, j and $T = 60 \times 60 \times 24$ seconds. Time differences are drawn from an exponential distribution with mean 45 seconds for p_1 and 60 seconds for p_2:[5]

$$F\left(t_k^i - t_{k-1}^i\right) = 1 - \exp\left\{-\lambda_i \left(t_k^i - t_{k-1}^i\right)\right\}, \ i = 1, 2$$

where $F(\cdot)$ denotes a cumulative distribution function, $\lambda_1 = 1/45$ and $\lambda_2 = 1/60$.

We compared the performances of previous tick interpolation realized volatility $\hat{\omega}_{ij}^P(M)$, Fourier series estimator $\hat{\omega}_{ij}^F$, raw data realized volatility

$\hat{\omega}_{ij}^{R}$, and naively weighted realized volatility $\hat{\omega}_{ij}^{N}$. We also observed the performance of the estimator using the optimal weight. In calculations of previous tick interpolation realized volatility $\hat{\omega}_{ij}^{P}(M)$, we set $M = 144, 288$, and 720, corresponding to so-called daily realized volatility based on 10-min, 5-min and 2-min returns. In calculations of Fourier series estimator $\hat{\omega}_{ij}^{F}$, we set $Q = 10, 25, 50, 100, 250, 500, 750$, and 1000. We performed 100 replications.

Table 3.1 and 3.2 report the sample MSE and bias (in parenthesis) of each estimator from 100 replications:

$$\frac{1}{100}\sum_{r=1}^{100}\left(\hat{\omega}_{ij}^{(r)} - \int_0^T \omega_{ij}^{(r)}(t)\,dt\right)^2 \text{ and } \frac{1}{100}\sum_{r=1}^{100}\left(\hat{\omega}_{ij}^{(r)} - \int_0^T \omega_{ij}^{(r)}(t)\,dt\right),$$

where r denotes the number of replications.

Figure 3.2, 3.3, and 3.4 show the distribution of normalized errors of each estimator:

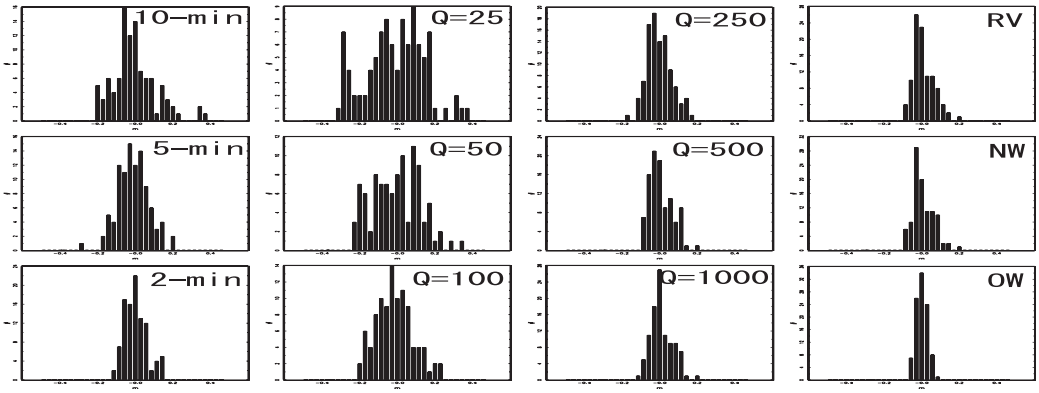
$$\frac{\hat{\omega}_{11} - \int_0^T \omega_{11}(t)dt}{\int_0^T \omega_{11}(t)dt},\ \frac{\hat{\omega}_{22} - \int_0^T \omega_{22}(t)dt}{\int_0^T \omega_{22}(t)dt},\ \text{and},\ \frac{\hat{\omega}_{12} - \int_0^T \omega_{12}(t)dt}{\int_0^T \omega_{12}(t)dt},$$

respectively.

Because 1st asset is more high-frequency sampled (average duration is 45 seconds) than 2nd asset (average duration is 60 seconds), each estimate of $\int_0^T \omega_{11}(t)\,dt$ is more accurate than that of $\int_0^T \omega_{22}(t)\,dt$.

Under our simulation design, the correlation between the 1st and 2nd

Figure 3.2: Distribution of normalized error (volatility of 1st asset)



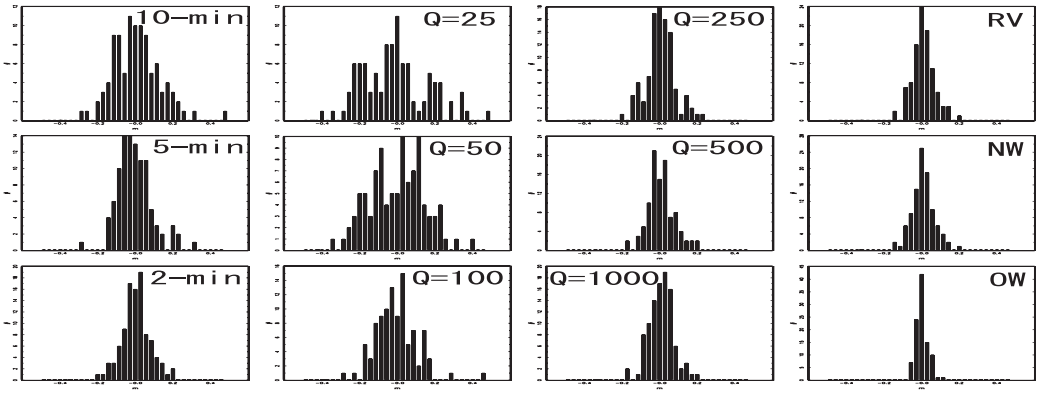
Note: 10-min, 5-min, and 2-min denote $\hat{\omega}_{11}^{P}(M)$ with $M = 144$, 288, and 720, respectively. "$Q =$" signifies the Fourier estimator $\hat{\omega}_{11}^{F}$ with $Q = 25,\ 50,\ 100,\ 250,\ 500$, and 1000. RV denotes the raw data realized volatility $\hat{\omega}_{11}^{R}$. NW denotes the naively weighted realized volatility $\hat{\omega}_{11}^{N}$. OW denotes the weighted realized volatility using the optimal weight. The distribution is computed with 100 'daily' replications.

asset is on average positive: $\omega_{12}(t)$ varies around a positive mean of 0.0002 because

$$\omega_{12}(t) = \sigma_{11}(t)\sigma_{21}(t) + \sigma_{12}(t)\sigma_{22}(t)$$

and each σ_{ij} has the mean of 0.01. As expected from the bias (3.5), the shorter the interpolation time intervals is, the more downward biased the

Figure 3.3: Distribution of normalized error (volatility of 2nd asset)

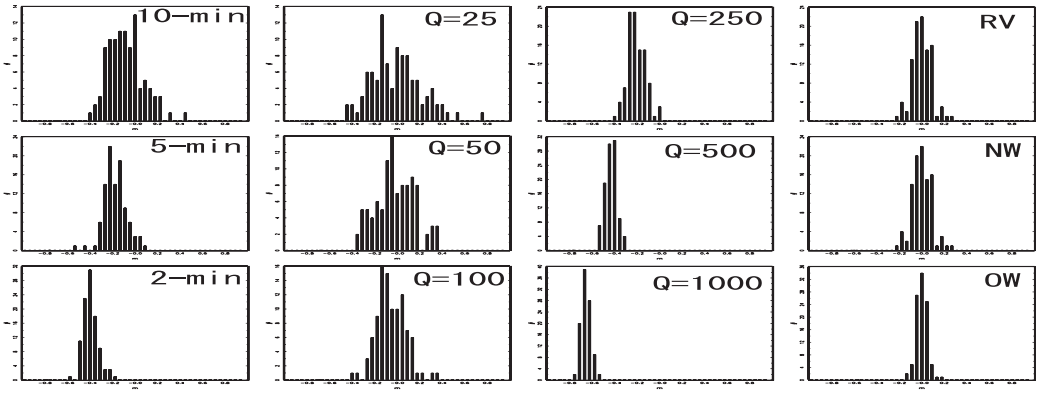


Note: 10-min, 5-min, and 2-min denote $\hat{\omega}^P_{22}(M)$ with $M = 144$, 288, and 720, respectively. "Q =" signifies the Fourier estimator $\hat{\omega}^F_{22}$ with $Q = 25$, 50, 100, 250, 500, and 1000. RV denotes the raw data realized volatility $\hat{\omega}^R_{22}$. NW denotes the naively weighted realized volatility $\hat{\omega}^N_{22}$. OW denotes the weighted realized volatility using the optimal weight. The distribution is computed with 100 'daily' replications.

previous tick interpolation realized cross volatility $\hat{\omega}^P_{12}$ is. On the other hand, as the partitions get finer and finer, $\hat{\omega}^P_{11}(M)$ and $\hat{\omega}^P_{22}(M)$ become more accurate. If $M \to \infty$ (in this case, $M = 60 \times 60 \times 24$), $\hat{\omega}^P_{11}(M)$ and $\hat{\omega}^P_{22}(M)$ are exactly consistent with $\hat{\omega}^R_{11}$ and $\hat{\omega}^R_{22}$, respectively.

This relationship between previous tick realized volatility and the number of partition is similar to that between Fourier series estimator and the number

Figure 3.4: Distribution of normalized error (cross volatility)



Note: 10-min, 5-min, and 2-min denote $\hat{\omega}_{12}^{P}(M)$ with $M = 144$, 288, and 720, respectively. "$Q =$" signifies the Fourier estimator $\hat{\omega}_{12}^{F}$ with $Q = 25,\ 50,\ 100,\ 250,\ 500$, and 1000. RV denotes the raw data realized volatility $\hat{\omega}_{12}^{R}$. NW denotes the naively weighted realized volatility $\hat{\omega}_{12}^{N}$. OW denotes the weighted realized volatility using the optimal weight. The distribution is computed with 100 'daily' replications.

of Fourier coefficients. As mentioned in 3.2.3, as $Q \to \infty$, $\hat{\omega}_{11}^{F}$, $\hat{\omega}_{22}^{F}$, and $\hat{\omega}_{12}^{F}$ go to $\hat{\omega}_{11}^{R}$, $\hat{\omega}_{22}^{R}$, and 0, respectively. We cannot find the optimal Q for Fourier estimator of cross volatility unless the we know true process of volatility.

Since (3.12) is unbiased estimator of cross volatility, the sample bias is very small. As expected from the link between naively weighted realized volatility $\hat{\omega}_{ij}^{N}$ and the raw data realized volatility $\hat{\omega}_{ij}^{R}$, although the former

is slightly more downward biased than the latter, the former has slightly smaller sample MSE than the latter.

The optimally weighed realized volatility is overwhelming the other method. The results of optimally weighted realized volatility show principal limit of the weighted realized volatility. One of the most important remaining works is to investigate the other feasible weighting schemes by using the framework of the optimal weight.

3.5 Summary

In this chapter we propose the definition of weighted realized volatility which nests various estimators and show some important examples. The definition is useful to make a comparative study on them. As a natural consequence, we derive the MSE-minimizing estimator in the class. To construct it, the estimates of optimal weights are required. We propose a feasible example of it. However, it is one of the remaining works to refine upon the feasible estimator. Another remaining work is the correction of interpolation bias. It is necessary when we can just obtain evenly spaced data which have already been interpolated.

Notes

1 This simplification is acceptable not only because it means an efficient market in financial economics, but also because, mathematically, the martingale component swamps the predictable portion over short time intervals.

2 For the purpose of simplification, we set $t_0^i = 0$ and $t_{N_i}^i = T$.

3 Dacorogna, Gençay, Müller, Olsen, and Pictet (2001) also introduces *next tick interpolation* which is analogous to previous tick interpolation.

4 Barndorff-Nielsen and Shephard (2004) studies properties of sum of squared returns in the case of evenly sampled observations. They refer to the sum of square returns realized variance and to square root of it as realized volatility.

5 Of course, our method allows the duration to be correlated or autocorrelated. See Engle and Russell (1998) for an autocorrelated duration model.

Table 3.1: Sample MSE from 100 'daily' replications

	$\int_0^T \omega_{11}(t)\,dt$	$\int_0^T \omega_{12}(t)\,dt$	$\int_0^T \omega_{22}(t)\,dt$
10-min	9.19	8.85	10.53
	(-0.10)	(-1.37)	(0.21)
5-min	4.80	12.27	5.79
	(-0.30)	(-3.05)	(-0.088)
2-min	2.38	47.68	3.69
	(0.033)	(-6.81)	(0.11)
FE	2.31	5.78	2.83
	(0.29)	(-0.93)	(2.83)
RV	2.10	2.17	2.57
	(0.21)	(-0.015)	(0.26)
NW	2.08	2.16	2.55
	(0.18)	(-0.026)	(0.22)
OW	0.58	0.77	0.70
	(0.068)	(-0.030)	(0.055)

Note: Sample biases are given in parentheses. 10-min: $\hat{\omega}_{ij}^P(144)$; 5-min: $\hat{\omega}_{ij}^P(288)$; 2-min: $\hat{\omega}_{ij}^P(720)$; FE: $\hat{\omega}_{11}^F$ and $\hat{\omega}_{22}^F$ with $Q = 1000$, $\hat{\omega}_{12}^F$ with $Q = 100$; NE: $\hat{\omega}_{ij}^N$; OW: weighted realized volatility using optimal weights.

Table 3.2: Sample MSE of Fourier estimators

Q	$\int_0^T \omega_{11}(t)\,dt$	$\int_0^T \omega_{12}(t)\,dt$	$\int_0^T \omega_{22}(t)\,dt$
10	65.24	46.06	61.60
	(-0.72)	(-1.18)	(-1.56)
25	18.40	14.64	21.94
	(-0.17)	(-0.17)	(-0.15)
50	10.78	8.75	13.92
	(-0.11)	(-0.24)	(-0.013)
100	6.47	5.78	7.71
	(-0.40)	(-0.93)	(-0.12)
250	2.98	14.45	4.01
	(-0.060)	(-3.56)	(0.13)
500	2.51	55.65	3.34
	(0.17)	(-7.40)	(0.22)
750	2.42	95.08	3.11
	(0.24)	(-9.71)	(0.21)
1000	2.31	127.10	2.83
	(0.29)	(-11.24)	(0.21)

Note: Sample biases are given in parentheses.

Chapter 4

Estimation of instantaneous volatility

Various stylized facts about asset return or its volatility can be expressed in state-space models that essentially consist of two stochastic differential equations: the observation equation and the state equation (see, e.g., Ghysels, Harvey, and Renault, 1996). As a diffusion is observed at shorter and shorter time intervals, its conditional variance at any instant can be approximated with greater accuracy by a simple flat-weight moving average of squared residuals. This is the theoretical basis for using the standard (flat-weight) rolling regression of squared residuals as an estimator of volatility in

the context of high-frequency data.

Foster and Nelson (1996) proved that exponentially weighted rolling regression (EWRR) minimizes the asymptotic variance of measurement error when the time interval is sufficiently small. However, in its application, flat-weight rolling regression (FWRR) was used because it can be calculated efficiently by the conventional iterative method. In this chapter, we propose a similar iterative method for EWRRs. An alternative to the usual realized volatility is proposed for its application.

4.1 Exponentially weighted rolling regression of Foster and Nelson (1996)

First we review optimal weighted rolling regression of Foster and Nelson (1996).[1] Let ${}_hX_t$ be a locally squared integrable semimartingale and adapted to the filtration $\{{}_h\mathscr{F}_t\}$, where $\{{}_h\mathscr{F}_t\}$ is increasing and right continuous; time is discrete such that $t = 0, h, 2h, \cdots, Nh$, where h and N denote the time interval and the number of available observations, respectively. In this note, we assume that the data generating process (DGP) is described by the following

state-space representation:

$$\Delta_h X_t = {}_h\mu_t h + \Delta_h M_t, \qquad E((\Delta_h M_t)^2|_h\mathscr{F}_{t-h}) = {}_h\Omega_t h, \tag{4.1}$$

$$\Delta_h \Omega_t = {}_h\lambda_t h + \Delta_h M_t^*, \qquad E((\Delta_h M_t^*)^2|_h\mathscr{F}_{t-2h}) = {}_h\Lambda_t h, \tag{4.2}$$

$$\Delta_h B_t = h^{-1/2}((\Delta_h M_t)^2 - {}_h\Omega_t h), \qquad E((\Delta_h B_t)^2|_h\mathscr{F}_{t-h}) = {}_h\theta_t h, \tag{4.3}$$

where Δ denotes the first order difference (e.g., $\Delta_h X_t = {}_hX_t - {}_hX_{t-h}$), ${}_hM_t$ and ${}_hM_t^*$ are local martingales with respect to ${}_h\mathscr{F}_{t-h}$ and ${}_h\mathscr{F}_{t-2h}$, ${}_h\mu_t$ and ${}_h\Omega_t$ are ${}_h\mathscr{F}_{t-h}$-measurable, and ${}_h\lambda_t$ and ${}_h\Lambda_t$ are ${}_h\mathscr{F}_{t-2h}$-measurable.

Difference equations in (4.1) and (4.2) are called the *observation equation* and *state equation*, respectively. ${}_h\Omega_t$ represents volatility when ${}_hX_t$ is the log of asset price. In (4.3), the *sampling error* $\Delta_h B_t$ is defined as the martingale difference. Note that ${}_h\theta_t/{}_h\Omega_t^2$ describes the conditional kurtosis of $\Delta_h M_t$ minus one because

$${}_h\theta_t = h^{-1}E((\Delta_h B_t)^2|_h\mathscr{F}_{t-h}) = E((\Delta_h M_t/\sqrt{h})^4|_h\mathscr{F}_{t-h}) - {}_h\Omega_t^2.$$

The estimator addressed in this study is the rolling regression of squared residuals

$${}_h\hat{\Omega}_t \equiv \sum_{s={}_hT_*(t)}^{{}_hT^*(t)} {}_hw_{s-t} z_s h, \quad z_t \equiv \frac{(\Delta_h X_t - {}_h\hat{\mu}_t h)^2}{h},$$

where ${}_hT_*(t)$ and ${}_hT^*(t)$ are the *start* and *end times* of the rolling regression, $\hat{\mu}_t$ is an estimation of μ_t, and $\sum_t {}_hw_t h = 1$. Furthermore, some additional

assumptions on DGP and weight are required for the following asymptotic results.[2]

Foster and Nelson (1996) derived the asymptotic distribution of the measurement error:

$$h^{-1/4}({}_h\hat{\Omega}_t - {}_h\Omega_t)|\mathscr{F}_{T_*} \overset{a}{\sim} N(0, {}_hC_{T_*}),$$

where

$${}_hC_{T_*} = {}_h\theta_{T_*}\sqrt{h}\sum_t {}_hw_t^2 h + \frac{{}_h\Lambda_{T_*}}{\sqrt{h}}\sum_t {}_h\Psi_t^2 h,$$

and

$${}_h\Psi_t = \begin{cases} \displaystyle\sum_{s=t+h,t+2h,\cdots}^{\infty} {}_hw_s h & \text{if } t \geq 0, \\ \displaystyle -\sum_{s=-\infty}^{t} {}_hw_s h & \text{if } t < 0. \end{cases}$$

For discussion in the next section, we display variances of EWRR and backward-looking FWRR:

$${}_hC_{T_*} = \begin{cases} \frac{1}{4}\left({}_h\theta_{T_*} a\sqrt{h} + \frac{{}_h\Lambda_{T_*}}{a\sqrt{h}}\right) & \text{if } {}_hw_{s-t} = \frac{a}{2}e^{-a|s-t|}, \\ \frac{{}_h\theta_{T_*}}{n\sqrt{h}} + \frac{{}_h\Lambda_{T_*} n\sqrt{h}}{3} & \text{if } {}_hw_{s-t} = \frac{1}{nh}\cdot I(\{s \in [t-nh, t]\}), \end{cases}$$

where $I(\cdot)$ denotes the indicator function.[3] Obviously, these variances are minimized when $a = \sqrt{{}_h\Lambda_{T_*}/{}_h\theta_{T_*}h}$ and $n = \sqrt{3{}_h\theta_{T_*}/{}_h\Lambda_{T_*}h}$, respectively.

Foster and Nelson (1996) proved that the EWRR with setting $a = \sqrt{{}_h\Lambda_{T_*}/{}_h\theta_{T_*}h}$ achieves the smallest variance in all weights. If ${}_hw_t$ is constant over time, FWRRs can be evaluated easily because recursive calculation is possible. For example, the backward-looking FWRR is written by the first-order difference equation:

$$ {}_h\hat{\Omega}_t = {}_h\hat{\Omega}_{t-h} + \frac{1}{nh} \cdot (z_{T^*} - z_{T_*-h}). $$

In fact, Foster and Nelson (1996) used the two-side FWRR in an empirical example and Monte Carlo simulation.

We propose a similar iterative method for the EWRR. To simplify the notation, we define the EWRR as

$$ \mathrm{EWRR}[z|a](t) = \sum_s \frac{a}{2} e^{-a|s-t|} z_s h, $$

and divide the EWRR into past and future portions as

$$ \mathrm{EWRR}[z|a](t) = P[z|a](t) + F[z|a](t), \tag{4.4} $$

where

$$ P[z|a](t) = \sum_{s \le t} \frac{a}{2} e^{a(s-t)} z_s h, \quad \text{and} $$
$$ F[z|a](t) = \sum_{s > t} \frac{a}{2} e^{-a(s-t)} z_s h. $$

Thereby, we can find the iterative rule in each process as

$$P[z|a](t) = e^{-ah}P[z|a](t-h) + \frac{a}{2}z_t h, \tag{4.5}$$

$$F[z|a](t) = e^{-ah}F[z|a](t+h) + \frac{a}{2}e^{-ah}z_{t+h} h. \tag{4.6}$$

In the same manner as for the flat-weight, if the weight function does not change (i.e., a is constant) over time, these recurrence formulas improve the efficiency of numerical evaluation. Using (4.5) and (4.6), two series of $\{P[z|a](t)\}_{t=0,h,2h,\cdots}^{Nh}$ and $\{F[z|a](t)\}_{t=Nh,Nh-h,Nh-2h,\cdots}^{0}$ are calculated forward and backward respectively. Then EWRR$[z|a](t)$ is completed by (4.4) at each t. As $N \to \infty$, the theoretical computational time with the method increases at order N, whereas that without the method increases at order N^2.

4.2 An application: Comparison with instantaneous realized volatility

We need estimates of ${}_h\theta_{T_*}$ and ${}_h\Lambda_{T_*}$ to use the optimal EWRR. However, it is burdensome to estimate them. Even under the simplifying assumptions that ${}_h\Lambda_t/{}_h\Omega_t^2$ and ${}_h\theta_t/{}_h\Omega_t^2$ are constant over time, they cannot be estimated

accurately, as explained in Foster and Nelson (1996). Instead of seeking the optimal estimator, we propose a practical usage of EWRR.

Realized volatility[4], which is often used as a proxy for true volatility to measure the performance of forecasting in empirical contexts, is defined as backward-looking FWRR,

$$ {}_h\hat{\Omega}_t = \sum_s \frac{1}{n_r} \cdot I(\{s \in [t - n_r h, t]\}) \cdot z_s, \tag{4.7} $$

where n_r is constant over time. Each researcher determines window length n_r by some method. In the context of Foster and Nelson (1996)'s theory, the estimator (4.7) implies that the researcher believes n_r to be the optimal $\sqrt{3_h\theta_{T_*}/{}_h\Lambda_{T_*}h}$ over time. This is equivalent to setting $\sqrt{{}_h\Lambda_{T_*}/{}_h\theta_{T_*}h} = \sqrt{3}/n_r h$. Variances of the asymptotic measurement error of EWRR$[z|\sqrt{3}/n_r h]$ and backward-looking FWRR (4.7) are

$$ \frac{\sqrt{3}}{4}\left(\frac{{}_h\theta_{T_*}}{n_r\sqrt{h}} + \frac{{}_h\Lambda_{T_*}n_r\sqrt{h}}{3}\right), \quad \text{and} \quad \frac{{}_h\theta_{T_*}}{n_r\sqrt{h}} + \frac{{}_h\Lambda_{T_*}n_r\sqrt{h}}{3}, \tag{4.8} $$

respectively. Therefore, the EWRR achieves $\sqrt{3}/4$ smaller measurement error variance than realized volatility at any t. Thereby, we expect EWRR to reduce mean squared error (MSE) by $\sqrt{3}/4$ compared to realized volatility.

We performed Monte Carlo simulation according to Foster and Nelson (1996) to confirm this; we generated 16,885 observations from the following

DGP,

$$\Delta \log \Omega_t = 0.0056 \cdot (-0.4246 - \log \Omega_{t-1}) + \sqrt{0.012} \cdot u_{2t}, \tag{4.9}$$

$$\Delta M_t = \sqrt{\Omega_t} \cdot u_{1t}, \tag{4.10}$$

where both u_{1t} and u_{2t} are mutually independent, $u_{1t} \sim$ i.i.d. standardized-t_{12}, and $u_{2t} \sim$ i.i.d.$N(0,1)$.[5]

(4.9) implies that $\log \Omega_t$ is conditionally homoskedastic. This is equivalent to the constancy of Λ_t/Ω_t^2, which is specified by 0.012 in this DGP. In (4.10), kurtosis of u_{1t} is assumed to be 3.75. This assumption means that $\theta_t/\Omega_t^2 = 2.75$ over time because θ_t/Ω_t^2 is conditional kurtosis of u_{1t} minus one. The constancy of Λ_t/Ω_t^2 and θ_t/Ω_t^2 implies that the optimal n_r is $\sqrt{3 \cdot 2.75/0.012}(\approx 26)$ over time.[6]

Table 4.1 reports means of the MSEs of realized volatility and the EWRR from 600 simulations along with ratios of two estimators' means of MSEs. Both estimators minimize MSEs at optimal n_r. As expected, the ratios are approximately $\sqrt{3}/4(\approx 0.433)$ near the optimal n_r. The ratios separate from 0.433 when n_r is far from 26. A n_r that is too small violates the assumption that the number of observations in the window must be sufficiently large to hold the asymptotic theory. On the other hand, a n_r that is too large violates the assumption that the window length must be sufficiently short to

maintain the parameter constancy.

Although the simplifying assumptions hold in the above example, (4.8) suggests that regardless of whether the assumptions hold or not (whether nuisance parameters can be estimated accurately or not), the measurement error variances ratio is always $\sqrt{3}/4$. This relation holds unless not-so-restrictive assumptions on DGP and weight (i.e., Foster and Nelson (1996), Assumptions A–D) are violated. We can say that the EWRR$[(Residual)^2|\sqrt{3}/n_r]$ should be used instead of the usual realized volatility with window length n_r in a broad range of situations.

Table 4.1 reports means of the MSEs of realized volatility and the EWRR from 600 simulations. Figure 4.1 shows one of the 600 simulations. The curve shape is similar to that of function (4.8).

4.3 Summary

Because of the use of the iterative method, EWRR is as tractable as FWRR. However, the optimal EWRR of Foster and Nelson (1996) requires estimates of nuisance parameters. Even under the simplifying assumptions, it is a nuisance problem to estimate these parameters. This note proposes a prac-

Table 4.1: Means of MSEs of realized volatility and EWRR

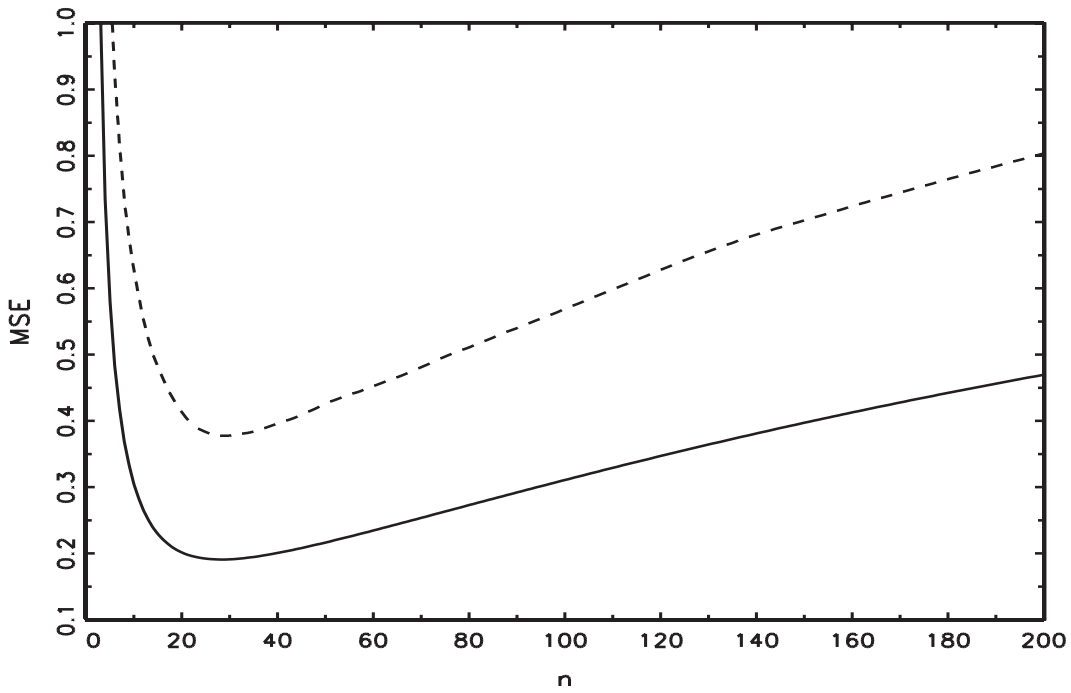
n_r	1	20	26	50	100
Realized volatility	12.456	0.768	0.724	0.805	1.146
	(108.154)	(0.706)	(0.644)	(0.723)	(1.138)
EWRR	8.557	0.354	0.338	0.395	0.603
	(7.241)	(0.258)	(0.237)	(0.273)	(0.478)
Ratio	0.687	0.461	0.467	0.491	0.526

Note: Realized volatility and EWRR are computed by

$$\frac{1}{n_r}\sum_{i=0}^{n_r-1} z_{t-i} \quad \text{and} \quad \frac{\sqrt{3}}{2n_r}\sum_{s} z_s \exp\left[-\frac{\sqrt{3}}{n_r}|s-t|\right],$$

where z_t is the squared residual at t. All the means are computed with 600 replications. Standard deviations are given in parentheses. The row of 'Ratio' shows ratios of the two estimators' means of MSEs at each n_r.

Figure 4.1: EWRR vs realized volatility



Note: MSE of $\hat{\Omega}_t$ by EWRR (solid line) and realized volatility (dashed line). One of the 600 simulations is drawn.

tical application of EWRR: an alternative to the usual realized volatility with window length n. EWRR$[(Residual)^2|\sqrt{3}/n]$ achieves $\sqrt{3}/4$ smaller measurement error variance than the realized volatility. This relation does not require overly restrictive assumptions. Therefore, instead of the realized volatility, we can use the EWRR in a wide range of situations as a more accurate and equally simple estimator.

Notes

1 For simplification, we restrict our study to scalar and diffusion processes.

2 See Foster and Nelson (1996) for these assumptions.

3 These can be verified easily by thinking of sums as integrals:

$$\sum_t {}_h w_t^2 h \cong \int_{-\infty}^{\infty} {}_h w_t^2 dt,$$
$$\sum_t {}_h \Psi_t^2 h \cong \int_0^{\infty} \left(\int_t^{\infty} {}_h w_s ds \right)^2 dt + \int_{-\infty}^{0} \left(\int_{-\infty}^{t} {}_h w_s ds \right)^2 dt.$$

4 Although we studied realized volatility for integrated volatility in the previous chapters, now we define realized volatility for spot volatility. The definition of spot realized volatility follows Dacorogna, Gençay, Müller, Olsen, and Pictet (2001).

5 The prefix $h(=1)$ is dropped for the remainder of this section.

6 According to French, Schwert, and Stambaugh (1987), this seems to be reasonable in reference to U.S. stock prices.

Appendix A

Linear interpolation bias

Each interpolated point is written as

$$q\left(\frac{jT}{m}\right) = (1-\rho_j)\, p\left(t_j^-\right) + \rho_j p\left(t_j^+\right) = (1-\rho_j)\int_0^{t_j^-} dp\,(t) + \rho_j \int_0^{t_j^+} dp\,(t)\,.$$

See Figure A.1. Using this, we get

$$\begin{aligned}
&E\left[q\left(\frac{jT}{m}\right) - q\left(\frac{(j-1)\,T}{m}\right)\right]^2 \\
&= \left[\rho_j^2 \int_0^{t_j^+} \sigma^2\,(t)\,dt + \left(1-\rho_j^2\right)\int_0^{t_j^-} \sigma^2\,(t)\,dt\right] \\
&- \left[\left(1-(1-\rho_{j-1})^2\right)\int_0^{t_{j-1}^+} \sigma^2\,(t)\,dt + \left(1-\rho_{j-1}^2\right)\int_0^{t_{j-1}^-} \sigma^2\,(t)\,dt\right].
\end{aligned}$$

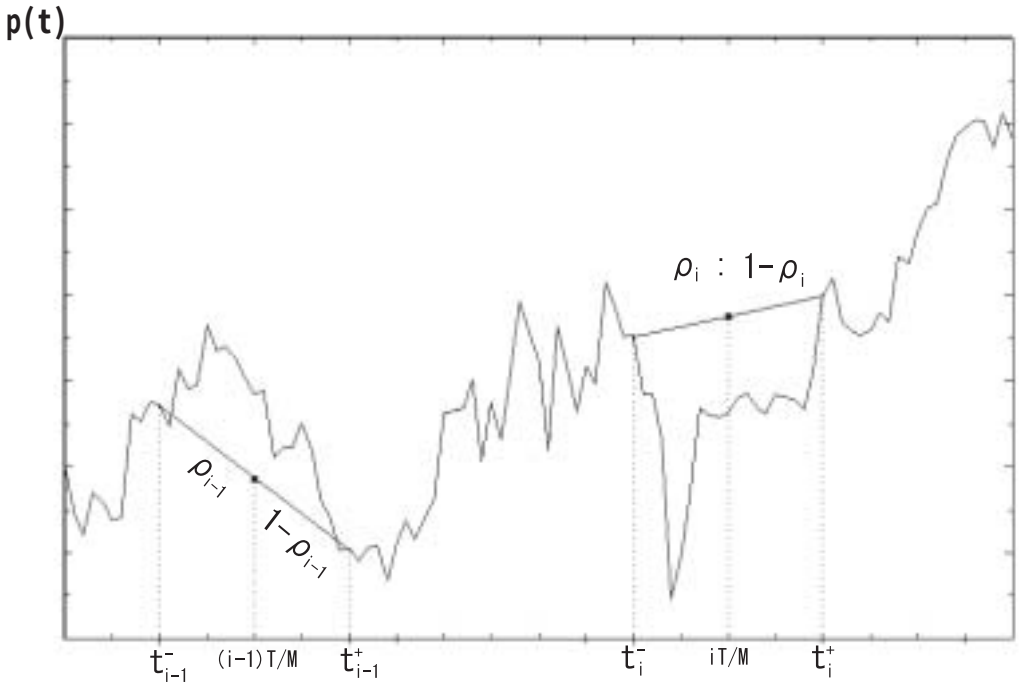


Figure A.1: Linear interpolation

Each ith term of the bias $E[\hat{\sigma}^2(M)] - \int_0^T \sigma^2(t)dt$ is

$$E\left[\left\{q\left(\frac{iT}{M}\right) - q\left(\frac{(i-1)T}{M}\right)\right\}^2\right] - \int_{(i-1)T/M}^{iT/M} \sigma^2(t)dt$$
$$= A - B - (C - D)$$

where

$$A = \rho_i^2 \int_0^{t_i^+} \sigma^2(t)dt + (1-\rho_i^2)\int_0^{t_i^-} \sigma^2(t)dt$$
$$B = (1-(1-\rho_{i-1})^2)\int_0^{t_{i-1}^+} \sigma^2(t)dt + (1-\rho_{i-1})^2 \int_0^{t_{i-1}^-} \sigma^2(t)dt$$
$$C = \int_0^{iT/M} \sigma^2(t)dt$$
$$D = \int_0^{(i-1)T/M} \sigma^2(t)dt$$

See Figure A.2. Then calculating

$$E\left[\hat{\sigma}^2(m)\right] - \int_0^T \sigma^2(t)\,dt$$
$$= \sum_{j=1}^{m} E\left[q\left(\frac{jT}{m}\right) - q\left(\frac{(j-1)T}{m}\right)\right]^2 - \int_{(j-1)T/m}^{jT/m} \sigma^2(t)\,dt$$

we obtain (2.4).

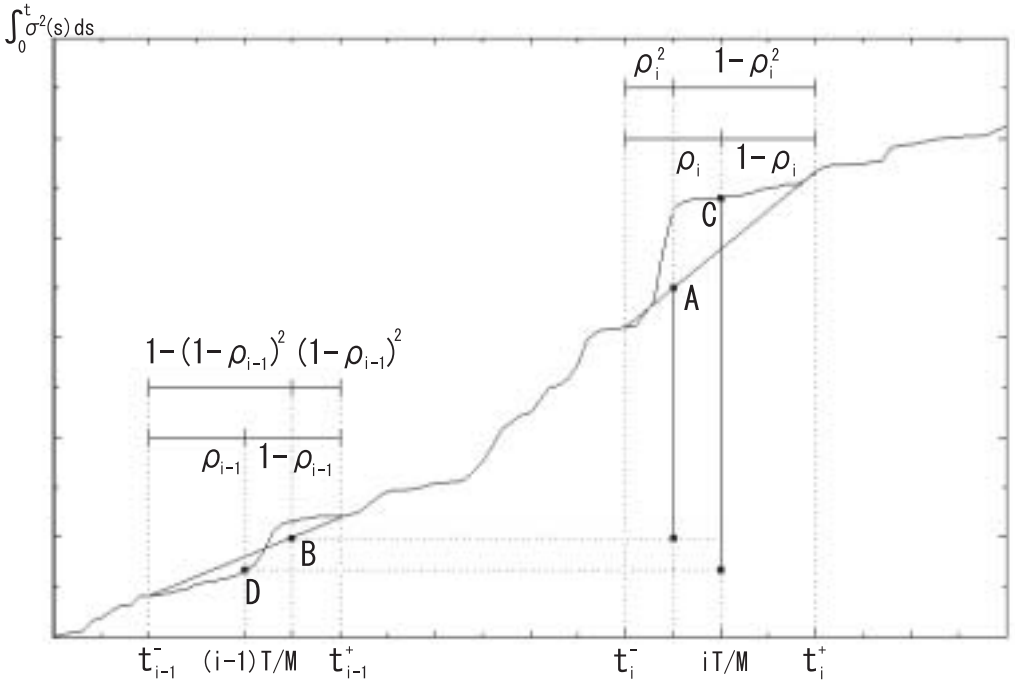


Figure A.2: Linear interpolation bias

Appendix B

Relationship between Fourier estimator and raw data realized volatility

(2.8) and (3.11) are rewritten as

$$a_k(dp) = \frac{1}{\pi}\sum_{i=1}^{N} \cos kt_i \Delta p(t_i) \tag{B.1}$$

$$b_k(dp) = \frac{1}{\pi}\sum_{i=1}^{N} \sin kt_i \Delta p(t_i) \tag{B.2}$$

respectively. Using (B.1), (B.2), and the addition theorem,

$$\begin{aligned}
\hat{\sigma}_F^2 &= \frac{\pi^2}{K}\sum_{k=1}^{K}(a_k^2(dp)+b_k^2(dp_i)) \\
&= \frac{1}{K}\sum_{k=1}^{K}\left(\left\{\sum_{i=1}^{N}\cos kt_i\Delta p(t_i)\right\}^2+\left\{\sum_{i=1}^{N}\sin kt_i\Delta p(t_i)\right\}^2\right) \\
&= \frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta p\left(t_i\right)\Delta p\left(t_j\right)\left\{\cos kt_i\cos kt_j+\sin kt_i\sin kt_j\right\} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\Delta p\left(t_i\right)\Delta p\left(t_j\right)\left\{\frac{\sum_{k=1}^{K}\cos k\left(t_i-t_j\right)}{K}\right\}.
\end{aligned}$$

Since

$$\sum_{k=1}^{K}\cos kx=\begin{cases} K & \text{if } x=0 \\ \frac{\sin\frac{(K+1)x}{2}\cos\frac{Kx}{2}}{\sin\frac{x}{2}} & \text{otherwise}\end{cases}$$

we get (2.11).

Appendix C

Weight matrix of $\hat{\omega}_{ii}^{L}$

Using

$$\alpha_m + \beta_m = 1, \qquad \text{and}$$

$$p_i(t_k^i) = p_i(t_0) + \sum_{l=1}^{k} \Delta p_i(t_l^i)$$

we obtain

$$
\begin{aligned}
&\hat{\omega}_{ii}^{L}(3) \\
&= \left(q_i\left(\frac{T}{3}\right) - q_i\left(0\right) \right)^2 + \left(q_i\left(\frac{2T}{3}\right) - q_i\left(\frac{T}{3}\right) \right)^2 + \left(q_i\left(T\right) - q_i\left(\frac{2T}{3}\right) \right)^2 \\
&= \left\{ \alpha_1 p_i(t_3^i) + \beta_1 p_i(t_2^i) - p_i(t_0) \right\}^2 \\
&+ \left\{ \alpha_2 p_i(t_7^i) + \beta_2 p_i(t_6^i) - \alpha_1 p_i(t_3^i) - \beta_1 p_i(t_2^i) \right\}^2 \\
&+ \left\{ p_i(t_8^i) - \alpha_2 p_i(t_7^i) - \beta_2 p_i(t_6^i) \right\}^2 \\
&= \left\{ \sum_{k=1}^{2} \Delta p_i(t_k^i) + \alpha_1 \Delta p_i(t_3^i) \right\}^2 \\
&+ \left\{ \beta_1 \Delta p_i(t_3^i) + \sum_{k=3}^{6} \Delta p_i(t_k^i) + \alpha_2 \Delta p_i(t_7^i) \right\}^2 \\
&+ \left\{ \beta_2 \Delta p_i(t_7^i) + \Delta p_i(t_8^i) \right\}^2 . \qquad \text{(C.1)}
\end{aligned}
$$

Each coefficient of $\Delta p_i(t_k^i)\Delta p_i(t_l^i)$ in (pre-wm-pol) is equivalent to the kl element of (3.7).

Appendix D

Weighted realized volatility representation of Fourier estimator

Fourier coefficients of $a_q(dp_i)$ and $b_q(dp_i)$ are approximated by

$$a_q(dp_i) \approx \frac{1}{\pi} \sum_{k=1}^{N_i} \cos qt_k^i \Delta p_i(t_k^i)$$
$$b_q(dp_i) \approx \frac{1}{\pi} \sum_{k=1}^{N_i} \sin qt_k^i \Delta p_i(t_k^i),$$

respectively. By these approximates and the additional theorem,

$$
\begin{aligned}
\hat{\omega}_{ij}^{F} &= \frac{\pi^2}{Q}\sum_{q=1}^{Q}(a_q(dp_i)a_q(dp_j)+b_q(dp_i)b_q(dp_j)) \\
&= \frac{1}{Q}\sum_{q=1}^{Q}\left\{\sum_{k=1}^{N_i}\cos(qt_k^i)\Delta p_i(t_k^i)\sum_{l=1}^{N_j}\cos(qt_l^j)\Delta p_j(t_l^j)\right\} \\
&+ \frac{1}{Q}\sum_{q=1}^{Q}\left\{\sum_{k=1}^{N_i}\sin(qt_k^i)\Delta p_i(t_k^i)\sum_{l=1}^{N_j}\sin(qt_l^j)\Delta p_j(t_l^j)\right\} \\
&= \frac{1}{Q}\sum_{q=1}^{Q}\left\{\sum_{k=1}^{N_i}\sum_{l=1}^{N_j}\left\{\cos(qt_k^i)\cos(qt_l^j)+\sin(qt_k^i)\sin(qt_l^j)\right\}\Delta p_i(t_k^i)\Delta p_j(t_l^j)\right\} \\
&= \frac{1}{Q}\sum_{q=1}^{Q}\left\{\sum_{k=1}^{N_i}\sum_{l=1}^{N_j}\cos q(t_k^i-t_l^j)\Delta p_i(t_k^i)\Delta p_j(t_l^j)\right\} \\
&= \sum_{k=1}^{N_i}\sum_{l=1}^{N_j}\Delta p_i\left(t_k^i\right)\Delta p_i\left(t_l^j\right)\left\{\frac{\sum_{q=1}^{Q}\cos q\left(t_k^i-t_l^j\right)}{Q}\right\}.
\end{aligned}
$$

Since

$$
\sum_{q=1}^{Q}\cos qx = \begin{cases} Q & \text{if } x=0 \\ \frac{\sin\frac{(Q+1)x}{2}\cos\frac{Qx}{2}}{\sin\frac{x}{2}} & \text{otherwise} \end{cases},
$$

we get the desired result.

Appendix E

Variance of $\hat{\omega}_{ij}$

Using

$$
\begin{aligned}
& E\left(\Delta p_i\left(t_k^i\right)\Delta p_j\left(t_l^j\right)w_{kl}^A\right)^2 \\
& = \left(w_{kl}^A\right)^2\left\{2\left(\int_{I(k,l)}\omega_{ij}dt\right)^2+\left(\int_{t_{k-1}}^{t_k}\omega_{ii}dt\right)\left(\int_{t_{l-1}}^{t_l}\omega_{jj}dt\right)\right\},
\end{aligned}
$$

$$
\begin{aligned}
&V\left(\hat{\omega}_{ij}\right) \\
&= V\left(\sum_{A} \Delta p_i\left(t_k^i\right) \Delta p_j\left(t_l^j\right) w_{kl} + \sum_{A^C} \Delta p_i\left(t_k^i\right) \Delta p_j\left(t_l^j\right) w_{kl}\right) \\
&= E\left(\sum_{A} \Delta p_i\left(t_k^i\right) \Delta p_j\left(t_l^j\right) w_{kl}\right)^2 - \left\{E\left(\sum_{A} \Delta p_i\left(t_k^i\right) \Delta p_j\left(t_l^j\right) w_{kl}\right)\right\}^2 \\
&+ E\left(\sum_{A^C} \Delta p_i\left(t_k^i\right) \Delta p_j\left(t_l^j\right) w_{kl}\right)^2 \\
&= \sum \left(w_{kl}^A\right)^2 \left\{\left(\int_{I(k,l)} \omega_{ij} dt\right)^2 + \left(\int_{t_{k-1}}^{t_k} \omega_{ii} dt\right)\left(\int_{t_{l-1}}^{t_l} \omega_{jj} dt\right)\right\} \\
&+ \left\{\sum w_{kl}^A \int_{I(k,l)} \omega_{ij} dt\right\}^2 - \left\{\sum w_{kl}^A \int_{I(k,l)} \omega_{ij} dt\right\}^2 \\
&+ \sum \left(\int_{t_{k-1}}^{t_k} \omega_{ii} dt\right)\left(\int_{t_{l-1}}^{t_l} \omega_{jj} dt\right)\left(w_{kl}^{A^C}\right)^2 \\
&= \sum \left(w_{kl}^A\right)^2 \left\{\left(\int_{I(k,l)} \omega_{ij} dt\right)^2 + \left(\int_{t_{k-1}}^{t_k} \omega_{ii} dt\right)\left(\int_{t_{l-1}}^{t_l} \omega_{jj} dt\right)\right\} \\
&+ \sum \left(w_{kl}^{A^C}\right)^2 \left(\int_{t_{k-1}}^{t_k} \omega_{ii} dt\right)\left(\int_{t_{l-1}}^{t_l} \omega_{jj} dt\right).
\end{aligned}
$$

Appendix F

Proof of Theorem 2

The first order condition is

$$\frac{\partial MSE}{\partial w} = 2Dw + 2xx'w - 2xx'1,$$

then we get

$$
\begin{aligned}
w &= \left(D + xx'\right)^{-1} xx'1 \\
&= \left(D^{-1} - \frac{1}{1 + x'D^{-1}x} D^{-1}xx'D^{-1}\right) xx'1 \\
&= \begin{pmatrix}
\frac{\int_0^T \omega_{ij}dt \int_{I(1,1)} \omega_{ij}dt}{v_{11}\{1+\sum u_{kl}\}} \\
\vdots \\
\frac{\int_0^T \omega_{ij}dt \int_{I(k,l)} \omega_{ij}dt}{v_{kl}\{1+\sum u_{kl}\}} \\
\vdots \\
\frac{\int_0^T \omega_{ij}dt \int_{I\left(N_i,N_j\right)} \omega_{ij}dt}{v_{N_iN_j}\{1+\sum u_{kl}\}}
\end{pmatrix}.
\end{aligned}
$$

The second equality follows from the updating formula. See e.g., Greene (1999). The second order derivative matrix is

$$\frac{\partial^2 MSE}{\partial w \partial w'} = 2D + 2xx'.$$

This matrix is positive definite. Substituting the optimal weight into (3.13) and (3.14), we obtain (3.16) and (3.17), respectively.

References

Aït-Sahalia, Y., and P. A. Mykland (2003): "The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions," *Econometrica*, 71, 483–549.

Andeou, E., and E. Ghysels (2002): "Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation, and Empirical Results," *Journal of Business and Economic Statistics*, 20, 363–376.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Barndorff-Nielsen, O. E., and N. Shephard (2004): "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Re-

gression, and Correlation in Financial Economics," *Econometrica*, 72, 885–925.

Barucci, E., and R. Renò (2002): "On Measuring Volatility of Diffusion Processes with High Frequency Data," *Economics Letters*, 74, 371–378.

Black, F., and M. Scholes (1973): "The Pricing of Options and Corporate liabilities," *Journal of Political Economy*, 81, 637–654.

Bollerslev, T. (1986): "Generalized Auotregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–328.

Dacorogna, M. M., R. Gençay, U. Müller, R. B. Olsen, and O. V. Pictet (2001): *An Introduction to High-Frequency Finance*. Academic Press.

Drost, F. C., and T. E. Nijman (1993): "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909–927.

Drost, F. C., and B. J. Werker (1996): "Closing the GARCH Gaps: Continuous Time GARCH Modeling," *Journal of Econometrics*, 74, 31–57.

Engle, R. F. (1982): "Autoregressive Conditional Heterosickedasticity with

Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1008.

Engle, R. F., and J. R. Russell (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Florens-Zmirou, D. (1993): "On Estimating the Diffusion Coefficient from Discrete Observations," *Journal of Applied Probability*, 30, 790–804.

Foster, D. P., and D. B. Nelson (1996): "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64, 139–174.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.

Ghysels, E., A. C. Harvey, and E. Renault (1996): "Stochastic Volatility," in *Handbook of Statistics*, ed. by G. S. Maddala, and C. R. Rao, vol. 14. North Holland, Amsterdam.

Greene, W. H. (1999): *Econometric Analysis: Forth Edition*. Prentice Hall.

Hull, J., and A. White (1987): "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, 42, 281–300.

Kanatani, T. (2004a): "Integrated Volatility Measuring from Unevenly Sampled Observations," *Economics Bulletin*, 3(36), 1–8.

——— (2004b): "Iterative Method for Exponentially Weighted Rolling Regression," *Finance Research Letters*, 1, 196–201.

——— (2004c): "Optimally Weighted Realized Volatility," *CAEA Discussion Paper*, (26).

Lo, A. W. (1988): "Maximum Likelihood Estimation of Generalized Ito Process with Discretely Sampled Data," *Econometric Theory*, 4, 231–247.

Malliavin, P., and M. E. Mancino (2002): "Fourier Series Method for Measurement of Multivariate Volatilities," *Finance and Stochastics*, 6, 49–61.

Nelson, D. B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Poon, S.-H., and C. W. J. Granger (2003): "Forecasting Volatility in Financial Markets: A Review," *Journal of Economic Literature*, 41, 478–539.

Sueishi, N. (2004): "Estimation of Discretely Sampled Diffusion Processes," In *Proceedings of the 11th Kansai Econometrics Workshop*, Kobe University, January 10, 2004, 1, 5–27.