CHAPTER FOURTEEN

Likelihood-Based Volatility Estimators in the Presence of Market Microstructure Noise

YACINE AÏT-SAHALIA and DACHENG XIU

14.1 Introduction

Driven by the need for accurate measurement of financial risk using intraday data, high frequency econometric methods have been evolving rapidly, bringing into focus a range of issues that were otherwise unobservable or, in many cases, irrelevant from the perspective of daily and weekly data or lower frequency data. The huge amount of intraday tick-by-tick data provides rich and timely information regarding fluctuations of traded assets and their comovements, which may yield more accurate measurements of volatility and covariance over relatively short horizons than estimates based on years of historical data at low frequency.

However, statistical inference with high frequency data presents many challenges. Closer scrutiny of the data reveals the pervasive presence of market microstructure noise, including frictions such as the existence of bid-ask spreads and bounces, the discreteness of price changes, the price impact of some transactions, and informational effects, all of which add volatility to the observed price process. When measuring correlation, the fact that the two assets may not trade or otherwise be observed at exactly the same times, known as *observation*

Handbook of Volatility Models and Their Applications, First Edition.
Edited by Luc Bauwens, Christian Hafner, and Sebastien Laurent.

asynchronicity, is another issue that may distort the covariance and correlation estimates, unlike the synchronous daily and lower frequency data.

For example, the popular realized volatility estimator, that is, the sum of squared log-returns,[1] diverges when the sampling frequency increases to beyond approximately every 5 min, a fact that was well recognized empirically in the realized volatility literature in the form of "signature plots" (Andersen et al., 2000) and was first analyzed theoretically in Aït-Sahalia et al. (2005) in the presence of market microstructure noise. Noise clearly has the potential to generate an increase of the realized volatility, and, in turn, bias the correlation estimates, since the realized volatility appears in the denominator when calculating the correlation. When it comes to estimating the correlation between two assets' returns, the estimate is known to be biased toward 0 as the sampling interval progressively shrinks. This puzzling phenomenon is known as the *Epps effect* after Epps (1979). In view of this, we start in the univariate case with a decomposition of the observed transaction log-price, Y, into the sum of an unobservable efficient log-price, X, and a noise component due to the frictions induced by the trading process, ε:

$$Y_t = X_t + \varepsilon_t, \tag{14.1}$$

where the efficient log-price process follows a general Itô process:

$$\mathrm{d}X_t = \mu_t \, \mathrm{d}t + \sigma_t \, \mathrm{d}W_t \tag{14.2}$$

and W is a Brownian motion. The goal is to disentangle the quadratic variation of the efficient price process, $\int_0^T \sigma_t^2 \, \mathrm{d}t$, on a fixed time interval $[0, T]$, from the variance of the noisy observations, $E[\varepsilon^2]$, using high frequency discrete observations on Y. A pair of assets (X_{1t}, X_{2t}) can be modeled in the same way when considering correlation estimation.

Initially, the literature focused on sampling sparsely to address the issue of noise; even though the data may be available every few seconds, one would sample every 15 min or so as a means of limiting the damaging impact of the noise. More recently, however, the focus has shifted toward developing noise-robust statistics. For instance, such estimators in the univariate variance case include two scales realized volatility of Zhang et al. (2005), the first consistent estimator for integrated volatility in the presence of noise, multiscale realized volatility, a modification that achieves the best possible rate of convergence proposed by Zhang (2006), realized kernels by Barndorff-Nielsen et al. (2008), and the preaveraging approach by Jacod et al. (2009), both of which contain sets of nonparametric estimators that can also achieve the best convergence rate. In terms of covariance estimator, Zhang (2011) proposes a consistent two scales realized covariance estimator using the previous tick method that is capable of dealing with asynchronous and noisy data. Barndorff-Nielsen et al. (2010) suggest multivariate realized kernels with a refresh time synchronization scheme to provide a consistent and semidefinite estimator of the covariance matrix,

[1] Which should therefore be more accurately termed *realized variance*.

while Kinnebrock and Podolskij (2008) proposed a multivariate preaveraging estimator. Related works also include, among others, Hayashi and Yoshidam (2005), Hansen and Lunde (2006b), Li and Mykland (2007), Kalnina and Linton (2008), Bandi and Russell (2008), Audrino and Corsi (2010), Aït-Sahalia et al. (2011), Zhang et al. (2011), and Kalnina (2011). A model consisting of a pure rounding error has been studied by Gloter and Jacod (2000).

In this chapter, we review one particular class of methods that have been developed to address these issues. The class of method we review are based on maximum-likelihood estimators (MLEs) and have been proposed and analyzed in Aït-Sahalia et al. (2005), Xiu (2010), and Aït-Sahalia et al. (2010). Likelihood-based methods, when available, are often the privileged parametric type of inference method for an econometrician, because of their statistical efficiency and ease of implementation. But, especially when vast quantities of high frequency observations are available, it is tempting to conduct inference that is nonparametric in nature, including, in particular, stochastic volatility of an unrestricted form, and this might seem to invalidate maximum-likelihood based on a parametric volatility structure, in fact a constant volatility parameter, and a Gaussian distribution for the noise term. We see that this is not the case, and that maximum-likelihood based on a constant volatility parameter is in fact a robust estimation method that produces consistent and even rate-efficient estimators in the cases of essentially arbitrary stochastic volatility and noise distributions.

We are also interested in estimating consistently the variance of the noise, which can be regarded as a measure of the liquidity of the market, or the quality of the trade execution in a given exchange or market structure. Measuring the impact of the bid-ask spread dates back to as early as Roll (1984). A recent study by Aït-Sahalia and Yu (2009) employ the likelihood-based noise estimators to decompose the transaction prices of NYSE stocks into a fundamental component and a microstructure noise component, and relates the two components to observable financial characteristics of these stocks and, in particular, to different observable measures of their liquidity.

This chapter is organized as follows: Section 14.2 discusses the volatility estimators, including both the constant volatility and the stochastic volatility case. Section 14.3 extends the previous method to covariance estimation. Section 14.4 illustrates the implementation of the method by providing an empirical application to measuring the volatility and correlation of stock and commodity financial returns. Section 14.5 concludes.

14.2 Volatility Estimation

14.2.1 CONSTANT VOLATILITY AND GAUSSIAN NOISE CASE: MLE

We first consider the benchmark case where volatility is a constant, and where the noise is Gaussian, which permits the straightforward use of likelihood methods. Even though the constant volatility assumption sounds implausible in practice, we shall see later that the estimator constructed under this assumption retains desirable properties in a stochastic volatility environment, and where the noise is not necessarily Gaussian.

For now, under these unrealistic assumptions, the efficient log-price X satisfies $\mathrm{d}X_t = \sigma \mathrm{d}W_t$. If there were no market microstructure noise, that is, $\varepsilon \equiv 0$, the log-returns $R_i = Y_{\tau_i} - Y_{\tau_{i-1}}$ would be i.i.d. $N(0, \sigma^2 \Delta)$. The MLE for σ^2 then coincides with the realized variance estimator of the process,

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^{n} R_i^2. \tag{14.3}$$

Moreover, the estimator achieves the optimal convergence rate $n^{1/2}$, where $n = T/\Delta$, with the following central limit result:

$$n^{\frac{1}{2}}(\hat{\sigma}^2 - \sigma_0^2) \xrightarrow{\mathcal{L}} N(0, 2\sigma^4). \tag{14.4}$$

When the observations are contaminated by i.i.d. noise $\varepsilon's$ with mean 0 and variance a^2, the log-returns R_i's now exhibit the structure of an MA(1) process, since

$$R_i = \sigma \left(W_{\tau_i} - W_{\tau_{i-1}} \right) + \varepsilon_{\tau_i} - \varepsilon_{\tau_{i-1}} \equiv u_i + \eta u_{i-1}, \tag{14.5}$$

where the u's are mean 0 and variance γ^2 with

$$\gamma^2(1+\eta^2) = \mathrm{Var}[R_i] = \sigma^2 \Delta + 2a^2 \tag{14.6}$$

$$\gamma^2 \eta = \mathrm{Cov}(R_i, R_{i-1}) = -a^2. \tag{14.7}$$

If we assume that the noise distribution is Gaussian, then the log-likelihood function for the vector of observed log-returns $R = [R_1, \ldots, R_n]'$ is

$$l(\sigma^2, a^2) = -\frac{1}{2} \log det(\Omega) - \frac{n}{2} \log(2\pi) - \frac{1}{2} R' \Omega^{-1} R, \tag{14.8}$$

where

$$\Omega = \begin{pmatrix} \sigma^2\Delta + 2a^2 & -a^2 & 0 & \cdots & 0 \\ -a^2 & \sigma^2\Delta + 2a^2 & -a^2 & \ddots & \vdots \\ 0 & -a^2 & \sigma^2\Delta + 2a^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a^2 \\ 0 & \cdots & 0 & -a^2 & \sigma^2\Delta + 2a^2 \end{pmatrix}. \tag{14.9}$$

The MLE $(\hat{\sigma}^2, \hat{a}^2)$ is consistent with different rates of convergence for its volatility part and noise part:

$$\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2 - \sigma_0^2) \\ n^{\frac{1}{2}}(\hat{a}^2 - a_0^2) \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 8a_0\sigma_0^3 T^{-\frac{1}{2}} & 0 \\ 0 & 2a_0^4 + \mathrm{Cum}_4[\varepsilon] \end{pmatrix} \right). \tag{14.10}$$

As ε has mean 0, its fourth cumulant can be written as

$$\text{Cum}_4[\varepsilon] = E[\varepsilon^4] - 3(E[\varepsilon^2])^2. \tag{14.11}$$

In the special case where ε is Normally distributed, $\text{Cum}_4[\varepsilon] = 0$.

14.2.2 ROBUSTNESS TO NON-GAUSSIAN NOISE

A key result here is that Equation 14.10 holds even if the noise term ε is *not* Normally distributed, as long as the noise is still i.i.d. with mean 0 and variance a^2. This is because the estimator using the log-likelihood function $l(\sigma^2, a^2)$ in Equation 14.8 is now a generalized method of moments (GMM) estimator, using the scores $\dot{l}_{\sigma^2}$ and $\dot{l}_{a^2}$ as moment functions.

Since the expected values of $\dot{l}_{\sigma^2}$ and $\dot{l}_{a^2}$ only depend on the first- and second-order moment structure of the log-returns, R, which is unchanged by the absence of normality, the moment functions are unbiased

$$E_{\text{true}}[\dot{l}_{\sigma^2}] = E_{\text{true}}[\dot{l}_{a^2}] = 0,$$

where "true" denotes the expected value computed under the true distribution of the Y's (where the ε's are not necessarily Gaussian).

Hence the estimator $(\hat{\sigma}^2, \hat{a}^2)$ based on these moment functions remains consistent. The effect of misspecification therefore lies in their asymptotic variance. By using the cumulants of the distribution of ε, we express this asymptotic variance in terms of deviations from normality.

We see from Equation 14.10 that the estimator $(\hat{\sigma}^2, \hat{a}^2)$ is consistent and its asymptotic variance is given by

$$\text{AVAR}_{\text{true}}(\hat{\sigma}^2, \hat{a}^2) = \text{AVAR}_{\text{normal}}(\hat{\sigma}^2, \hat{a}^2) + \text{Cum}_4\,[\varepsilon] \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

where $\text{AVAR}_{\text{normal}}(\hat{\sigma}^2, \hat{a}^2)$ is the asymptotic variance in the case where the distribution of ε is Normal, whereas $\text{AVAR}_{\text{true}}(\hat{\sigma}^2, \hat{a}^2)$ is the asymptotic variance under the true distribution of ε whatever that may be. That is, $\text{AVAR}_{\text{normal}}(\hat{\sigma}^2, \hat{a}^2)$ coincides with $\text{AVAR}_{\text{true}}(\hat{\sigma}^2, \hat{a}^2)$ for all but the (a^2, a^2) term.

So, not only do we retain consistency of the estimator but we in fact also retain rate efficiency for the estimation of both (σ^2, a^2) and actual efficiency for the estimation of σ^2. Aït-Sahalia et al. (2005) show how to interpret this in terms of the profile likelihood and the second Bartlett identity.

14.2.3 IMPLEMENTING MAXIMUM LIKELIHOOD

As can be seen from Equation 14.10, the optimal rate of convergence for volatility estimation turns out to be $n^{1/4}$ in the presence of noise, as opposed to $n^{1/2}$ in its absence (recall Equation 14.4). To be precise, it is not whether noise is present or not that matters, but whether it is incorporated or not in the estimation. The

MLE based on Equation 14.8 would still produce a rate $n^{1/4}$ for σ^2 even if ε were identically 0, as long as we actually did as if ε could have been present.

Given a typical 6.5 h trading day, if we model the noise and sample as high as every 1 s, we have the normalization constant $n^{1/4}$ around 12.4, whereas sampling every 5 min and ignoring the noise lead to a normalization constant $n^{1/2}$ that approximately equals 8.8. It appears that the gain in efficiency is marginal when sampling at highest frequency. However, if we further compare the asymptotic variances, $8a_0\sigma_0^3 T^{-1/2}$ is much smaller than $2\sigma_0^4$, since the standard deviation a_0 of the microstructure noise is usually very small in practice.

Implementing the MLE is more convenient than it appears. In fact, the likelihood function for the observed log-returns can be expressed in the following computationally efficient form, as a function of the transformed parameters (γ^2, η) by triangularizing the matrix Ω:

$$l(\eta, \gamma^2) = -\frac{1}{2}\sum_{i=1}^{N} \ln(2\pi d_i) - \frac{1}{2}\sum_{i=1}^{N} \frac{\tilde{Y}_i^2}{d_i}, \tag{14.12}$$

where

$$d_i = \gamma^2 \frac{1+\eta^2+\cdots+\eta^{2i}}{1+\eta^2+\cdots+\eta^{2(i-1)}},$$

and the $\tilde{Y}_i' s$ are obtained recursively as $\tilde{Y}_1 = Y_1$ and for $i = 2, \ldots, N$:

$$\tilde{Y}_i = Y_i - \frac{\eta\left(1+\eta^2+\cdots+\eta^{2(i-2)}\right)}{1+\eta^2+\cdots+\eta^{2(i-1)}}\tilde{Y}_{i-1}.$$

This algorithm avoids the brute-force computation of the inverse of the variance–covariance matrix Ω^{-1}, and hence significantly accelerates the optimization procedure in practice.

14.2.4 ROBUSTNESS TO STOCHASTIC VOLATILITY: QMLE

What happens to this MLE if the volatility is in fact not constant but is instead either deterministic and time-varying or stochastic? There is of course a vast theoretical literature on nonconstant volatility models initiated by Engle (1982a) and Bollerslev (1986), and empirical studies have documented a U-shaped intraday volatility pattern (Wood et al., 1985; Andersen and Bollerslev, 1997b; Boudt et al., 2011) and an implied volatility "smile" or "smirk" (Jackwerth and Rubinstein, 1996; Aït-Sahalia and Lo, 1998). Under such circumstances, simulation studies by Aït-Sahalia and Yu (2009) and Gatheral and Oomen (2010) suggest that the MLE defined above may perform well in practice as an estimator of the integrated volatility of the process, $T^{-1}\int_0^T \sigma_t^2\, dt$ instead of the constant σ^2. Intuitively, the conjecture that the estimator remains consistent is plausible in that when volatility becomes stochastic, the integrated volatility, the parameter of interest, happens to be the average of the volatility process, which is expected to be a legitimate candidate as an estimator.

In the absence of microstructure noise, the MLE has a closed-form (Eq. 14.3), which is of course an ideal estimator even with a nonparametric stochastic volatility model. However, the consistency of the MLE is no longer straightforward in the presence of noise, because there may not be a closed form available for this estimator. Its asymptotic variance is far more complicated because of heteroskedasticity and autocorrelation, as mentioned by Hansen et al. (2008).

Theoretically, the MLE under this new setting can be regarded as a quasi-MLE constructed under misspecified assumptions such as constant volatility, zero drift, and Gaussian microstructure noise. For this reason, we give the MLE an alias Quasi-Maximum Likelihood Estimator (QMLE) in such situation, in order to emphasize model misspecification and keep the notation in line with the classic results of likelihood-based estimation under misspecified models (White, 1982; Domowitz and White, 1982). In view of this, the consistency of the QMLE can be proved by extending the theory to cases where random parameters are allowed.

As shown by Xiu (2010), it turns out to be possible to establish consistency and derive the following central limit theorem (where $\mathcal{L}-\mathrm{s}$ denotes stable convergence in law, a stronger form of convergence than the usual convergence in law) for the QMLE:

$$\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2 - \frac{1}{T}\int_0^T \sigma_t^2\,\mathrm{d}t) \\ n^{1/2}(\hat{a}^2 - a_0^2) \end{pmatrix} \xrightarrow{\mathcal{L}-\mathrm{s}} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, V\right),$$

where

$$V = \begin{pmatrix} \frac{5a_0\int_0^T \sigma_t^4\,\mathrm{d}t}{T(\int_0^T \sigma_t^2\,\mathrm{d}t)^{\frac{1}{2}}} + \frac{3(\int_0^T \sigma_t^2\,\mathrm{d}t)^{\frac{3}{2}}a_0}{T^2} & 0 \\ 0 & 2a_0^4 + \mathrm{Cum}_4\,[\varepsilon] \end{pmatrix}.$$

Stochastic volatility therefore does not affect the efficiency of the noise variance estimator, which remains the same as in the previous case. Although the volatility estimator may not achieve the optimal efficiency except in the constant volatility case, it achieves the optimal rate of convergence $n^{1/4}$ for σ^2 and $n^{1/2}$ for a^2.

An intuitive way to understand how the QMLE $(\hat{\sigma}^2, \hat{a}^2)$ works theoretically is to rewrite it as an iterative quadratic estimator:

$$\hat{\sigma}^2 T = R' W_1(\hat{\sigma}^2, \hat{a}^2) R \tag{14.13}$$

$$\hat{a}^2 = R' W_2(\hat{\sigma}^2, \hat{a}^2) R. \tag{14.14}$$

The weighting matrices satisfy:

$$W_1(\sigma^2, a^2) = \frac{n \cdot tr(\Omega^{-2}\Lambda) \cdot \Omega^{-1}\Lambda\Omega^{-1} - n \cdot tr(\Omega^{-2}\Lambda^2) \cdot \Omega^{-2}}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2}) \cdot tr(\Omega^{-2}\Lambda^2)} \tag{14.15}$$

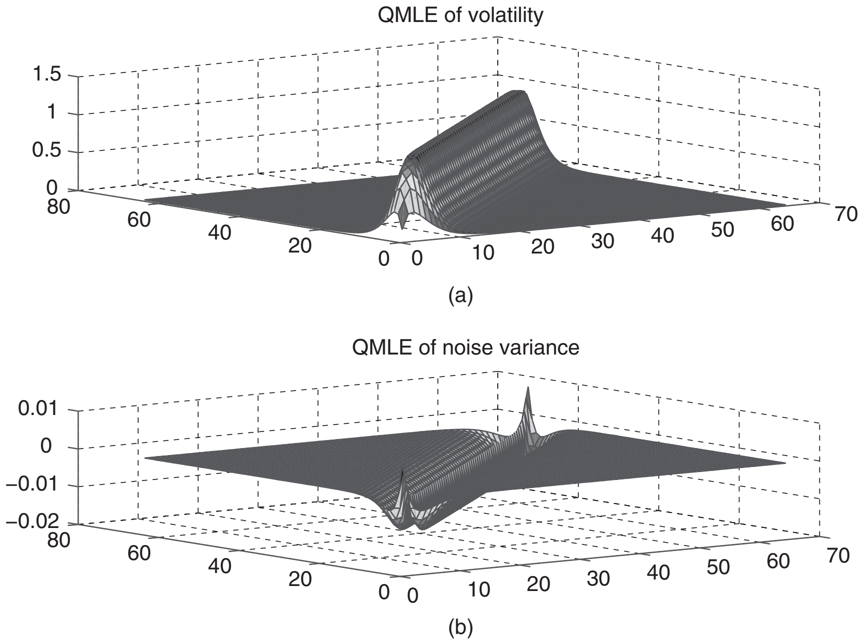
$$W_2(\sigma^2, a^2) = \frac{tr(\Omega^{-2}\Lambda) \cdot \Omega^{-2} - tr(\Omega^{-2}) \cdot \Omega^{-1}\Lambda\Omega^{-1}}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2}) \cdot tr(\Omega^{-2}\Lambda^2)}, \tag{14.16}$$

where Ω is given by Equation 14.9, and

$$\Lambda = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}.$$

Also, $W_1(\sigma^2, a^2)$ and $W_2(\sigma^2, a^2)$ depend on σ^2 and a^2 only through $\lambda^2 = a^2/(\sigma^2 T)$. Figure 14.1 plots the weighting matrices.

The quadratic representation also sheds light on the estimation procedure. Unlike a nonparametric estimator, the QMLE is fully parametric without any tuning parameters. Nevertheless, it seems reasonable (in view of the following comparison with Realized Kernels) to regard $\hat{\lambda} \cdot n^{\frac{1}{2}}$ as the "bandwidth" of the QMLE, which is automatically updated by the optimization algorithm, or more intuitively, by iterating Equations 14.13 and 14.14. Therefore, it is natural to construct a one-step alternative for the QMLE, which, instead of running nonlinear optimization, employs a consistent plug-in of $\hat{\lambda}$ for Equations 14.13 and 14.14.



(a)

(b)

FIGURE 14.1 **The weighting matrices of the quadratic representation of the QMLE.**

14.2.5 COMPARISON WITH OTHER ESTIMATORS

How does the QMLE of volatility compare with other nonparametric estimators, such as realized kernels, given by Barndorff-Nielsen et al. (2008)? Realized kernels include a series of nonparametric estimators designed for volatility estimation in the presence of noise. Flat-top realized kernels with kernel weight $k(\cdot)$ take on the following form:

$$K(Y_\tau) = \gamma_0(Y_\tau) + \sum_{h=1}^{n-1} k\left(\frac{h-1}{H}\right)(\gamma_h(Y_\tau) + \gamma_{-h}(Y_\tau)),$$

where the hth sample autocovariance function is

$$\gamma_h(Y_\tau) = \sum_{j=1}^{n}(Y_{\tau_j} - Y_{\tau_{j-1}})(Y_{\tau_{j-h}} - Y_{\tau_{j-h-1}}).$$

A major drawback of realized kernels is that they require a number of out-of-period intraday returns because of the construction of the autocovariance estimator $\gamma_h(Y_\tau)$. For this reason, a feasible finite-lag realized kernel is constructed using a feasible autocovariance estimator and its quadratic representation is

$$K(\tilde{X}_\tau) = Y'WY, \tag{14.17}$$

where W is determined by the kernel $k(\cdot)$ and bandwidth H.

$$\begin{aligned} W_{i,i} &= 1_{\{1+H \le i \le n-H\}} \\ W_{i,j} &= k\left(\frac{|i-j|-1}{H}\right) \cdot 1_{\{1 \le |i-j| \le H\}} \cdot 1_{\{1+H \le j \le n-H\}}. \end{aligned}$$

Infinite-lag kernels, which have nonzero weights on every autocovariance function, are not implementable empirically, although in theory with appropriate bandwidth, they may achieve the optimal efficiency among all kernel estimators. The following exponential kernel is the optimal infinite-lag kernel:

$$k_{opt}(x) = (1+x)\mathrm{e}^{-x}.$$

In view of the quadratic representation, the QMLE is asymptotically equivalent to the optimal kernel with implicit bandwidth:

$$H = \hat{\lambda} \cdot n^{1/2} = a_0 \left(\int_0^T \sigma_t^2 \, \mathrm{d}t\right)^{-\frac{1}{2}} n^{\frac{1}{2}}.$$

In other words, for any $K = n^{1/2+\delta}$, $0 < \delta < \frac{1}{2}$, and any $K \le i,j \le n-K$, we have

$$W_{1,i,j}(\sigma^2, a^2) \approx k_{\mathrm{opt}}\left(\frac{|i-j|}{\lambda \cdot n^{1/2}}\right).$$

Therefore, the QMLE, in some sense, implements the exponential optimal kernel except for the implicit bandwidth, which is suboptimal. In addition, its weighting matrix $W_1(\sigma^2, a^2)$ is approximately a symmetric Toeplitz matrix with equal weight along the diagonal, barring the boundary. The weights on the boundary along the diagonal decrease gradually, in contrast, with a discontinuous cut-off boundary for most kernel estimators, which may lead to a better finite sample performance for the QMLE.

In regard to asymptotic efficiency, realized kernels can achieve the same optimal convergence rate as the QMLE with appropriate kernels and bandwidths. When volatility is constant, the asymptotic variance of finite-lag kernels can only approximate the parametric variance bound, which, by contrast, can be obtained by the QMLE and the optimal kernel. If volatility is stochastic, the relative efficiency of the QMLE and realized kernels depends on the extent of heteroskedasticity, as measured by $\rho = \int_0^T \sigma_u^2 \,\mathrm{d}u / \sqrt{T \int_0^T \sigma_u^4 \,\mathrm{d}u}$. Apparently, the QMLE tends to be more favorable than finite-lag kernels as ρ becomes larger, whereas realized kernels are better when ρ is small. Intuitively, the smaller ρ is, the further the misspecified model deviates from the truth.

14.2.6 RANDOM SAMPLING AND NON-I.I.D. NOISE

If the sampling intervals between two consecutive observations are random, but i.i.d., and independent of the price process, we may pretend that the data are regularly spaced, and employ the same estimator as before. In fact, this estimator can be regarded as the pretend fixed MLE discussed in Aït-Sahalia and Mykland (2003). In light of this, the full information maximum likelihood or integrated out MLE may be preferred, if information concerning the distribution of the random sampling intervals is available.

If the microstructure noise exhibits autocorrelation such as an MA(1) structure, it may be better to divide the whole sample into two subsamples such that the noises within each subsample are uncorrelated, apply the QMLE to each subsample and aggregate the estimates. Such method is parsimonious, and easy to implement, as opposed to performing the maximum-likelihood estimation using the whole sample with one more parameter.

14.3 Covariance Estimation

We now extend the previous results to covariance estimation with a two-dimensional log-price process $\boldsymbol{X}_t = (X_{1t}, X_{2t})$, discretely observed over the interval $[0, T]$. The latent log-price processes satisfy

$$\mathrm{d}X_{it} = \mu_{it}\,\mathrm{d}t + \sigma_{it}\,\mathrm{d}W_{it},$$

with $E(\mathrm{d}W_{1t} \cdot \mathrm{d}W_{2t}) = \rho_t\,\mathrm{d}t$. Suppose that the observations are recorded at times $0 = t_{i,0} \le t_{i,1} \le t_{i,2} \le \cdots \le t_{i,n_i} = T$, respectively, where $i = 1, 2$. As in the univariate case, one can only observe $Y_{i,t}$, contaminated by an additive error $\varepsilon_{i,t}$, associated at each observation point. The noise $\boldsymbol{\varepsilon}_t$ is an i.i.d. two-dimensional vector with mean 0, diagonal covariance matrix $\boldsymbol{\Theta}$, and has a finite fourth moment.

On the basis of the identity that expresses covariance in terms of variances, Aït-Sahalia et al. (2010) proposed the following covariance and correlation estimators:

$$\widehat{\text{Cov}}(Y_1, Y_2) = \frac{1}{4}\left(\widehat{\text{Var}}(Y_1 + Y_2) - \widehat{\text{Var}}(Y_1 - Y_2)\right) \tag{14.18}$$

$$\widehat{\text{Cov}}(Y_1, Y_2) = \frac{\widehat{\text{Cov}}(Y_1, Y_2)}{\sqrt{\widehat{\text{Var}}(Y_1)}\sqrt{\widehat{\text{Var}}(Y_2)}}, \tag{14.19}$$

where $\widehat{\text{Cov}}(\cdot, \cdot)$ is the covariance estimator, $\widehat{\text{Var}}(\cdot, \cdot)$ denotes the QMLE in the univariate case, and $\cdot$ indicates the data we are actually using.

In order to compute the prices $Y_1 + Y_2$ and $Y_1 - Y_2$, we need the two assets to be synchronically traded. Nevertheless, this is not the case in practice, at least for high frequency financial data. Instead, high frequency transactions for two assets occur at times that are not synchrone. This practical issue may induce a large bias for the estimation, and may be (at least partly) responsible for the Epps effect. The remaining question is what kind of data synchronization procedure one should use. Clearly, if we apply the QMLE to estimate the diagonal elements in the covariance matrix, it would be better to cross out a small number of data points rather than adding more through an interpolation method, because the former strategy may suffer from efficiency loss, while the latter one may result in inconsistency due to the change in the autocorrelation structure.

We define a *generalized sampling time*, which we then use to propose a general synchronization scheme. A sequence of time points $\{\tau_0, \tau_1, \tau_2, \ldots, \tau_n\}$ is said to be the generalized sampling time for a collection of M assets, if they form a partition of the time interval $[0, T]$, and there exists at least one observation for each asset between consecutive τ_is. In addition, the time intervals, $\{\Delta_j = \tau_j - \tau_{j-1}, 1 \leq j \leq n\}$, satisfy $\sup_i \Delta_i \xrightarrow{\text{P}} 0$, as n increases to ∞.

The generalized synchronization method is then built on the generalized sampling time by selecting an arbitrary observation $Y_{i,\check{t}_j}$ for the ith asset between the time interval $(\tau_{j-1}, \tau_j]$. The synchronized data sets are, therefore, $\{Y^{\tau}_{i,\tau_j}, 1 \leq i \leq M, 1 \leq j \leq n\}$ such that $Y^{\tau}_{i,\tau_j} = Y_{i,\check{t}_j}$.

The concept of generalized synchronization method is more general than that of the Previous Tick approach discussed in Zhang (2011), and the Refresh Time scheme proposed by Barndorff-Nielsen et al. (2010), namely, the Replace All scheme in deB. Harris et al. (1995).

More precisely, if we require $\{\tau_j\}$ to be equally spaced on $[0, T]$, and the previous tick for each asset before τ_j to be selected, we are back to the Previous Tick approach. Or, if we choose τ_j recursively as

$$\tau_{j+1} = \max_{1 \leq i \leq M} \{t_{i, N_i(\tau_j)+1}\},$$

where $\tau_1 = \max\{t_{1,1}, t_{2,1}, \ldots, t_{M,1}\}$ and $N_i(t)$ measures the number of observations for asset i before time t, and if we select those ticks that occur right before or at τ_js, we return to the Refresh Time scheme. In both cases, the previous ticks of the assets, if needed, are regarded as if they were observed at the sampling time τ_js. By contrast, we advocate choosing an arbitrary tick for each asset within each interval. In practice, it may happen that the order of consecutive ticks is not recorded correctly. Because our synchronization method has no requirement on tick selection, the estimator is robust to data misplacement error, as long as these misplaced data points are within the same sampling intervals.

It is apparent that the Refresh Time scheme is highly dependent on the relatively illiquid asset. On the one hand, the number of synchronized pairs is smaller than the number of observations of this asset, inducing an inevitable loss of data for the other asset. More importantly, it is very likely that the Refresh Time points are determined by the occurrence of the relatively more illiquid asset, rendering the selected observations of the other asset always ahead of the corresponding illiquid asset. This hidden effect may induce some additional bias in the estimation.

Alternatively, we can design the synchronization scheme requiring each asset to lead in turn. For example, take two assets. If we require the first asset to lead, we choose $\tau_1 = t_{2,N_2(t_{1,1})+1}$. Recursively,

$$\tau_i = t_{2,N_2(t_{1,N_1(\tau_{i-1})+1})+1}.$$

Literally, it means that right after τ_{i-1}, we find the first observation of Y_{1t}, which should happen at $t_{1,N_1(\tau_{i-1})+1}$, and then the next generalized sampling time is defined to be the point when the first Y_{2t} is observed right after $t_{1,N_1(\tau_{i-1})+1}$. In this case, at all sampling time points, the second asset would always have records. The previous tick of the first asset, if needed, is regarded as if it were observed a bit later at the sampling time. Hence, in the synchronized pairs, the first asset always leads the second.

If the generalized sampling time $\{\tau_j\}$ is independent of the price process, the volatility process and the noise, and the time intervals, $\{\Delta_j = \tau_j - \tau_{j-1}, 1 \leq j \leq n\}$, are i.i.d. with mean $\bar{\Delta}$, then replacing the idealized data with the products of the generalized sampling time maintains the consistency and rate efficiency of the estimators:

$$\widehat{\mathrm{Cov}}(Y_1^\tau, Y_2^\tau) - 1/T\int_0^T \rho_t\sigma_{1t}\sigma_{2t} = O_P(\bar{\Delta}^{\frac{1}{4}})$$

$$\widehat{\mathrm{Corr}}(Y_1^\tau, Y_2^\tau) - \frac{\int_0^T \rho_t\sigma_{1t}\sigma_{2t}\,\mathrm{d}t}{\sqrt{\int_0^T \sigma_{1t}^2\,\mathrm{d}t}\sqrt{\int_0^T \sigma_{2t}^2\,\mathrm{d}t}} = O_P(\bar{\Delta}^{\frac{1}{4}}).$$

In other words, the proposed estimators are robust with respect to asynchronous data.

14.4 Empirical Application: Correlation between Stock and Commodity Futures

To assess the empirical relevance of these likelihood-based estimators, we estimate the volatilities of S&P 500 futures and crude oil futures as well as their correlation. The S&P 500 futures are traded on the Chicago Mercantile Exchange (CME) from 8:30 AM to 3:15 PM central time. The regular-size contracts are liquidly traded via the open outcry market, whereas the electronic market is in charge of mini-contracts. The crude oil futures used to be the most popular energy contract in the New York Mercantile Exchange (NYMEX), which has become part of the CME group recently. Since June 2006, the 24-h electronic market for crude oil has started to take over the open outcry market. Nevertheless, the most active trading period is between 9:00 AM and 2:30 PM Eastern Time (after February 2007), when the open outcry market is open. We only consider the liquid S&P500 futures traded via open outcry and the crude oil futures traded electronically. It is plausible to assume that the two microstructure noise terms, reflecting markets with little cross-trading or cross-arbitrage, are uncorrelated across the two markets. The sample period ranges from February 1, 2007 to December 18, 2009, therefore including the height of the financial crisis during the fall of 2008 and the winter of 2008–2009.

Figure 14.2 plots the daily realized volatility for S&P 500 and crude oil, and Figure 14.3 plots their daily correlation, estimated from high frequency data over the sampling period. The results shows that crude oil, which incidentally is the primary component of most commodity indices, is more volatile than the

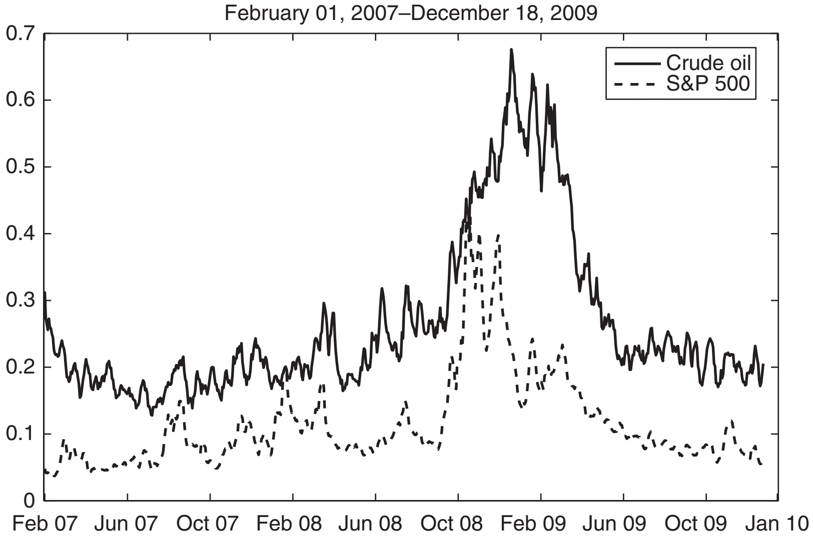


FIGURE 14.2 **The 5-day moving average of the daily annualized volatilities of the S&P 500 and crude oil futures.**

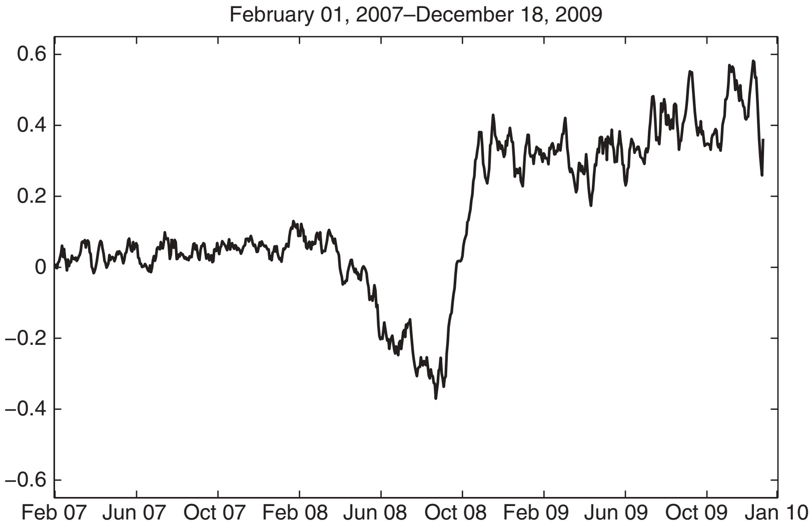


FIGURE 14.3 The 5-day moving average of the daily correlation between the S&P 500 and crude oil futures.

aggregate stock market, confirming that the commodity markets are generally riskier than the equity markets. With commodities representing a relatively new trading avenue for investors, as an asset class, the price of an individual commodity is not only simply determined by its own supply and demand but also by overall portfolio considerations, which may lead to larger correlations between the aggregate financial markets for the standard asset classes such as U.S. equities and commodity markets (Tang and Xiong (2010)). We find that the correlation is time-varying and may sometimes behave differently for a short period. For instance, at the beginning of the financial crisis, commodities appeared to be considered as a relatively safe hedge against the stock market downturn, hence the returns of the two markets were negatively correlated. Shortly, after October 2010, however, which may be viewed as after the heightened phase of the crisis, both markets started to move at the same pace, leading to a positive correlation of the returns. Although such short-term pattern is important to investors, it may not be captured, at least punctually, using lower frequency data, since years of historical data may dilute the short-term abnormality.

14.5 Conclusion

This chapter reviews the likelihood-based parametric methods designed to estimate the volatility and covariance of asset returns in the presence of market microstructure noise. Compared to nonparametric estimators, the major advantage of such estimators is their convenience and robustness; there is no need to choose any tuning parameters such as kernel functions, bandwidths, or number of subsamples. Moreover, being parametric in this circumstance does not lead to loss of robustness or generality. As we have seen, these estimators are consistent and achieve the optimal rate of convergence even in situations where their "construction hypotheses" of constant volatility and Gaussian noise are no longer satisfied, and in practice have good finite sample performance. All these features make them good choices for empirical applications.

Acknowledgments

Financial support from the NSF under grants DMS-0532370 (Aït-Sahalia) is gratefully acknowledged.