A Bias-Corrected Rate-Optimal Estimator of the Integrated Covariance of Security Returns with Serially Dependent Noise

(Job Market Paper)

Shinsuke Ikeda*

Boston University

November 15, 2009

Abstract

A bias-corrected non-parametric estimator of the integrated covariance matrix of security returns is proposed. It is constructed using a linear combination of two realized kernels. In addition to its simplicity, my estimator has desirable statistical properties including consistency, asymptotic normality and the best parametric rate of convergence in the presence of serially dependent microstructure noise. A transformation using the Cholesky decomposition guarantees that the estimator is positive semi-definite in finite samples without altering its asymptotic properties. This transformation allows me to show that non-linear functions of the estimated matrix, such as the hedge ratio, enjoy the same asymptotic properties. Simulations show that my estimator has much smaller bias than currently available estimators with only a mild increase in variance, so that the overall mean squared error is also smaller. In an application to intra-daily data of S&P 500 spot and futures prices, my method delivers estimates of hedge ratios with an 80% reduction in bias assessed in relation to the daily benchmark compared with the realized kernel estimator.

*e-mail: sikeda@bu.edu. I am grateful to Zhongjun Qu, Pierre Perron and Ivan Fernandez-Val for their advice and suggestions. A part of this article was presented at Econometric Society North American Summer Meeting in Boston and Far Eastern and South Asia Meeting in Tokyo, 2009. I wish to gratefully acknowledge the financial support of the Nakajima Foundation.
1 Introduction

The integrated covariance is a multivariate and stochastic generalization of the univariate constant volatility parameter in continuous time finance models. It is defined as the integral of a possibly time-varying covariance matrix of multiple security returns to summarize their fluctuations occurring over a certain period of time. Important financial quantities that depend on it include the no-arbitrage price of an option with stochastic volatility, optimal portfolio allocation, hedge ratios, market betas, and correlations between several security prices. Four complications make its estimation challenging. First, one only has discrete data to estimate this integral which is defined in a continuous time setting. For any estimator, the discretization error always creates an extra asymptotic variance component given the frequency of available data. Second, observed prices may be contaminated by market microstructure noise, reflecting market imperfection and measurement errors. This complication hinders one from using data recorded at very high frequency because the cumulative effect of the noise may become dominant in the estimates. Third, such noise may exhibit unknown serial dependence. This complication invalidates the parametric approach because possibly misspecified dependence of the noise may interfere with an otherwise consistent estimator of the integrated covariance. Fourth, non-linear functions of an estimated matrix may behave poorly. This is a serious problem in practice because many, if not most, of the previously mentioned financial quantities of interest depend on the integrated covariance matrix in a non-linear way. Therefore, a desirable estimator of the integrated covariance should deal with these four complications at the same time.

Recently, Barndorff-Nielsen, Hansen, Lunde and Shephard (BNHLS 2008b) developed the first non-parametric estimator of the integrated covariance in the presence of serially dependent microstructure noise. This estimator is called a realized kernel because it is constructed as a kernel-weighted sum of sample autocovariance matrices of return vectors. Realized kernels can deal with unknown serial dependence in the microstructure noise by implementing a data-dependent increase of the number of autocovariances used in estimation. Unfortunately, the limiting distribution of this estimator has an asymptotic bias caused by the long-run covariance of the microstructure noise. It reduces the estimator’s convergence rate relative to the best parametric rate. Moreover, it leads to strong bias for non-linear functions of the estimate in finite samples. This is a crucial flaw of the realized kernel estimator for the analysis of the financial quantities mentioned above.

To fix all of the above complications at the same time, this paper proposes a new non-parametric estimator of the integrated covariance matrix with built-in asymptotic bias correction. A key idea behind the construction of this estimator is to reformulate the problem as a joint estimation of the integrated covariance and the long-run covariance of the noise. A joint estimation of these two variations within the realized kernel framework induces the new estimator of the integrated covariance as a particular linear combination of two different realized kernels. Therefore, I call it a “Two Scale Realized Kernel” estimator. The linear
combination defining the estimator automatically eliminates the asymptotic biases of the two realized kernels. Because of this built-in asymptotic bias correction, the limiting distribution of my estimator is centered at the true integrated covariance matrix. This result is intuitive because I tried to estimate the integrated covariance and the long-run covariance matrix of the noise jointly so that the latter is no more a nuisance parameter for the former.

The absence of asymptotic bias in my estimator makes its convergence rate faster than that of the realized kernel because the number of autocovariances used in estimation can be designed to minimize the asymptotic variance. In particular, my estimator achieves the $n^{1/4}$-consistency, where $n$ is the number of observations within a day. This is the the best attainable rate as Gloter and Jacod (2001) established for a volatility estimator in the presence of noise. Therefore, my procedure delivers the first rate-optimal non-parametric estimator of the integrated covariance matrix in the presence of serially dependent microstructure noise.

My estimator is different from another bias corrected estimator of Zhang et. al. (2005). The former is based on a twice continuously differentiable kernel, while the latter is implicitly based on a kernel with kinks. The smoothness of the underlying kernels determines the orders of the asymptotic variance and the convergence rates as well. In fact, the estimator of Zhang et. al. (2005) can achieve only $n^{1/6}$-consistency.

To obtain estimates of non-linear functions of the integrated covariance, I need to ensure that the estimator is positive semi-definite. Although the estimator on its own is not guaranteed to be positive semi-definite in finite samples, a simple transformation using the Cholesky decomposition can be applied to it so that the transformed version is positive semi-definite. Moreover, the transformed version is equivalent to the original version in terms of the asymptotic properties. This positive semi-definite version allows me to apply the delta method to show that the desirable properties of the estimator also hold for non-linear functions of it. Therefore, the procedure delivers the first $n^{1/4}$-consistent estimators of non-linear functions of the integrated covariance matrix in the presence of microstructure noise. Moreover, the limiting distributions are correctly centered at the true values. The absence of asymptotic biases in non-linear functions of my estimator means that the procedure in this paper can be more suited for the estimation of financial quantities represented by non-linear functions of the integrated covariance than currently available procedures.

A simulation study confirms that non-linear functions based on the proposed procedure have a smaller bias in finite samples with a mild increase in variance than those based on the realized kernel. In particular, the overall mean squared error of my estimator is often smaller than that of the realized kernel. This result is encouraging because a bias correction often causes a severe variance inflation, leading to a much larger mean squared error. Moreover, my estimator turns out to be very robust to serial as well as cross-sectional dependence in the noise. This property makes my estimator attractive for the analysis of intra-daily data recorded at very high frequency with strong dependence in observed returns.

In the empirical section, I estimate the dynamic hedge ratio between S&P 500 spot and futures prices day by day, using their intra-daily data. The daily benchmark is specified by an
unconditional time series average of 22-day rolling-window OLS estimates of the hedge ratio using daily returns. In terms of the bias and root mean squared errors assessed in relation to this benchmark, the two scale realized kernel outperform the other methods such as the OLS, the realized covariance and the realized kernel. Specifically, my estimator achieves a larger than 80% reduction in bias compared with the realized kernel estimator. Furthermore, the other considered estimators using intra-daily data suffer from strong downward bias and higher root mean squared errors than the OLS estimator using only daily data. This result shows the potential importance of bias correction when estimating non-linear functions of the integrated covariance matrix using high frequency data.

The structure of this paper is as follows. Section 2 introduces a list of assumptions. It also defines the estimator, states the main theorem and provides some remarks on the issue of bandwidth selection. Section 3 proposes a new method to guarantee positive semi-definiteness of my estimator in finite samples. Section 4 considers a representative non-linear function of the estimator and derives its asymptotic properties. Section 5 studies the finite sample properties of the estimator by simulations. Section 6 applies the method to estimate the hedge ratio between the S&P 500 index spot and futures. After concluding the paper, all mathematical derivations are provided in an analytical appendix.

2 The Multivariate Two Scale Realized Kernel Estimator

In this section I define the multivariate two scale realized kernel estimator and derive its asymptotic properties. In what follows, capital Greek letters will typically denote the squared multivariate analogs of the lower-case denoted univariate counterparts.

2.1 Assumptions and Definitions

A period of trading is normalized on a [0, 1] continuum. The leading example is the daytime trading hours corresponding to a range from 9:30 AM (t = 0) to 4:00 PM (t = 1). All random variables are defined on some filtered probability space (\(\mathcal{E}, \mathcal{F}, \mathcal{F}, P\)) where \(\mathcal{F} = (\mathcal{F}_t)_{t \in [0,1]}\) is a non-decreasing and t-right continuous family of sub \(\sigma\)-fields of \(\mathcal{F}\) for the information filtration. For more precise conditions imposed on this space, see Karatzas and Shreve (1991, Chapter 1). The setup follows those in BNHLS (2008a, 2008b), Zhang et. al. (2005), Ait-Sahalia et. al. (2008) and Jacod (2008)\(^1\). All the integrals in this paper are defined element-wise, and all the integrands satisfy appropriate integrability conditions.

Assumption 1 (Data Generating Processes of Efficient Log Prices and Noise)

\((D-1)\) For \(t \in [0,1]\), the d-dimensional vector of efficient log prices \(\ln P^*_t\) follows a continuous

\(^1\)Special thanks to Jean Jacod for kindly providing his forthcoming manuscript.
time, continuous path Brownian semi-martingale process of the form

$$\ln P_t^* = \ln P_0^* + \int_0^t b_v dv + \int_0^t \Sigma^{1/2}(v)dW_v$$

where \( W_v \in \mathbb{R}^k \) is a vector of independent standard Brownian motions, \( b_v \in \mathbb{R}^d \) is a \( \mathbb{F} \)-predictable stochastic process, and \( \Sigma^{1/2} \in \mathbb{R}^{d \times k} \) is also a stochastic process. For \( b \) and \( \Sigma^{1/2} \), \( \exists \Lambda > 0 \ \forall (t, \omega) \in [0, 1] \times \mathcal{E}: \| b_t(\omega) \| \leq \Lambda \) and \( \| \Sigma^{1/2}(t, \omega) \| \leq \Lambda \).

(D-2) For \( t \in [0, 1] \), \( \text{vec}(\Sigma^{1/2}(t)) \in \mathbb{R}^{dk \times 1} \) follows a continuous time, continuous path Brownian semi-martingale process of the form

$$\text{vec}(\Sigma^{1/2}(t)) = \text{vec}(\Sigma^{1/2}(0)) + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\Sigma}^{1/2}_s dW_s \quad t \in [0, 1]$$

where \( \text{vec}(\Sigma^{1/2}(t)) \) is the stack of all column vectors of \( \Sigma^{1/2}(t) \), \( \tilde{b}_s \in \mathbb{R}^{dk \times 1} \) is \( \mathbb{F} \)-predictable, \( \tilde{\Sigma}^{1/2}_s \in \mathbb{R}^{dk \times k} \) is \( \mathbb{F} \)-adapted, and \( \exists \Lambda > 0 \ \forall (t, \omega) \in [0, 1] \times \mathcal{E}: \| \tilde{b}_t(\omega) \| \leq \Lambda \) and \( \| \tilde{\Sigma}^{1/2}_t(\omega) \| \leq \Lambda \). \( s \mapsto \tilde{b}_s \) and \( s \mapsto \tilde{\Sigma}^{1/2}_s \) are left continuous with right limits.

(D-3) The raw discrete observations \( \{ \ln P_{\tau_i} \}_{i=1, \ldots, n-1+2m_n} \) are given by

$$\ln P_{\tau_i} = \ln P_{\tau_i}^* + U_i, \quad i = 1, \ldots, n - 1 + 2m_n$$

The data with end-point jittering are obtained by setting \( \ln P_{\tau_i} = \ln P_{\tau_i + m_n}, \quad i = 1, 2, \ldots, n - 1, \) and

$$\ln P_{t_0} = m_n^{-1} \sum_{j=1}^{m_n} \ln P_{\tau_j} \quad \text{and} \quad \ln P_{t_n} = m_n^{-1} \sum_{j=1}^{m_n} \ln P_{\tau_{n-m_n+j}}.$$

Finally, \( t_i - t_{i-1} = 1/n, \ \forall i = 1 \ldots n \).

(D-4) \( U_i \in \mathbb{R}^d, \ i = 0, 1, \ldots, n \) is a zero-mean, strictly stationary and \( \alpha \)-mixing sequence of random variables with moments of all orders and exponentially decaying mixing coefficients. The \( h \)-th order autocovariance matrix of \( U_i, \ \Omega(h) \), is defined as \( \Omega(h) = E[U_iU_{i-h}'] \) for \( h \geq 0 \) and \( \Omega(-h) = \Omega'(h) \). It satisfies \( \sum_{h \in \mathbb{Z}} |h| \| \Omega(h) \| < \infty \) where \( \| \cdot \| \) is the Euclidean norm of a vector or matrix. Given \( \kappa_{pqrk} \) as the fourth order cumulant of \( p, q, r, s \)-th elements of \( U_i, \sum_{j,k,l \in \mathbb{Z}} \kappa_{pqrs}(0, j, k, l) < \infty, \ \forall p, q, r, s \in \{1, \ldots, d\} \). Finally, \( U_0 \sim O_p(1) \).

(D-5) \( \ln P^* \) and \( U \) are independent.

(D-6) The integrated covariance matrix \( \int_0^1 \Sigma(v)dv \) is positive definite.

(D-1) and (D-2) mean that the efficient log price vector and the instantaneous volatility evolve continuously, ruling out the possibility of jumps. The other conditions about \( b, \Sigma^{1/2}, \)
and $\Sigma_{1}^{1/2}$ follow Jacod (2008, (1.7) and (SH)). The same Brownian motions can drive both \( \ln P^\ast \) and \( \Sigma \). Therefore, these assumptions allow for leverage effect, i.e. negative correlation between returns and volatility innovations. The Brownian semi-martingale assumption on the efficient log price and its volatility matrix implies the following important result: given \( \ln P_{t_i}^\ast - \ln P_{t_{i-1}}^\ast \) a vector of efficient returns over the interval \([t_{i-1}, t_i] \),

\[
\rho \lim_{n \to \infty} \sum_{i=1}^{n} (\ln P_{t_i}^\ast - \ln P_{t_{i-1}}^\ast) (\ln P_{t_i}^\ast - \ln P_{t_{i-1}}^\ast)' = \int_{0}^{1} \Sigma(v)dv
\]

I call \( \int_{0}^{1} \Sigma(v)dv \) the integrated covariance matrix, the variable of interest in this paper. This quantity is also called the ex-post covariation or the quadratic variation in the literature. The equation (1) states that if \( (\ln P_{t_i}^\ast)_{i=0,1,...,n} \) were available, the integrated covariance could be estimated as a sum of outer products of efficient return vectors. Barndorff-Nielsen and Shephard (2004) show that in (1), the convergence occurs at the rate \( n^{1/2} \) under such an assumption.

In practice, however, \( \ln P_{t_i}^\ast \) is not directly observable. The observed log price vector is usually a mixture of the efficient log price vector and the measurement error \( U_i \). This is formally stated in the assumption (D-3). It follows the theory of market microstructure, eliciting the term market microstructure noise for \( U_i \in \mathbb{R}^d \). See Hasbrouck (2007) for an accessible reference. I follow BNHLS (2008b) for the end-point jittering: the initial and the final observations, \( \ln P_{t_0} \) and \( \ln P_{t_n} \), are defined by the averages of the first and last \( m_n \) raw observations, respectively. This device is necessary to obtain the consistency of the estimator. A strong assumption in (D-3) is that the sampling times \( t_i \) are common to all securities. I assume that the researchers have already synchronized the data of multiple security prices through, e.g., the refresh time sampling of BNHLS (2008b) and this synchronized data is subject to the next assumption.

The key assumption in this paper is (D-4). The main part follows Andrews (1991, Assumption A). See Andrews (1991, equation 3.1) for the definition of the cumulant function \( \kappa \). The assumption \( \sum_{h \in \mathbb{Z}} |h| \| \Omega(h) \| < \infty \) guarantees the existence of the long-run covariance matrix of \( U_i \):

\[
\Omega \equiv \sum_{h \in \mathbb{Z}} \Omega(h).
\]


In this paper I explicitly assume (D-5) to derive the asymptotic covariance matrix of my estimator. However, its consistency does not rely on (D-5).

(D-6) is required for continuous time finance models because one needs to invert the integrated covariance matrix to construct, e.g., the optimal portfolio allocations or the Sharpe ratios.
Definition 1 (A Class of Kernel Windows)

\( \mathcal{K} \) is a set of functions \( k : \mathbb{R} \to [-1, 1] \) such that for all \( k(x) \in \mathcal{K} \), \( k(x) \) is twice continuously differentiable, \( \forall x \in \mathbb{R}, k(x) = k(-x), \forall x \in \mathbb{R}, k(0) = 1, k(1)(0) = 0, k(2)(0) < 0 \), \( k^{(j)} := \int_0^\infty [k^{(j)}(x)]^2 dx < \infty \), \( j = 0, 1, 2 \), and \( \int_\infty^- k(x)e^{-ix\lambda} dx \geq 0 \), \( \forall \lambda \in \mathbb{R} \). Here, \( k^{(j)}(x) := \partial^j k(x)/\partial x^j, \ j = 0, 1, 2 \).

This class of kernel windows follows BNHLS (2008b). There are two important characteristics of the kernel windows in this class. First, they are twice continuously differentiable. This property excludes the kinked Bartlett window from \( \mathcal{K} \). Second, they satisfy \( \int_\infty^- k(x)e^{-ix\lambda} dx \geq 0 \), \( \forall \lambda \in \mathbb{R} \). This condition guarantees the positive semi-definiteness of a weighted sum of autocovariance matrices of any covariance stationary process using \( k \in \mathcal{K} \) as a weight function—see Andrews (1991, p823). All those conditions are satisfied by, e.g. the Parzen window:

\[
k(x) = \begin{cases} 
1 - 6x^2 + 6|x|^3 & |x| \in [0, 1/2] \\
2(1 - |x|^3) & |x| \in [1/2, 1] \\
0 & \text{otherwise}
\end{cases}
\]

Definition 2 (Realized Kernels)

For \( k \in \mathcal{K} \), the multivariate realized kernel is defined by

\[
RK(G_n) = \sum_{h=-(n-1)}^{n-1} k \left( \frac{h}{G_n + 1} \right) \Gamma(h)
\]

where \( G_n \) is a data-dependent bandwidth, \( \Gamma(h) \) is the \( h \)-th order sample autocovariance matrix defined by

\[
\Gamma(h) = \sum_{i=|h|+1}^n (\ln P_i - \ln P_{i-h})(\ln P_{i-h} - \ln P_{i-h-1})' \quad h \geq 0,
\]

and \( \Gamma(-h) = \Gamma'(h) \), the transpose of \( \Gamma(h) \).

Note that \( \Gamma(h) \) is slightly different from the \( h \)-th order sample autocovariance matrix of return vectors because there is no scaling by the sample size, \( 1/n \). The meaning of the lack of scaling would become clear if I focused on \( \Gamma(0) \). \( \Gamma(0) \) is referred to as the realized covariance in the literature because it is a finite sample analog of the integrated covariance defined in [1]. It can be represented as follows:

\[
\Gamma(0) = \sum_{i=1}^n (\ln P^*_i - \ln P^*_{i-1})(\ln P^*_i - \ln P^*_{i-1})' + \sum_{i=1}^n (U_i - U_{i-1})(U_i - U_{i-1})' + C
\]

where \( C \) is a collection of cross terms. If I divided \( \Gamma \) by the sample size, then [1] would imply \( n^{-1} \sum_{i=1}^n (\ln P^*_i - \ln P^*_{i-1})(\ln P^*_i - \ln P^*_{i-1})' \xrightarrow{p} 0 \), a practically useless result. On the other hand, the sample covariance matrix of the first differenced noise vector satisfies \( n^{-1} \sum_{i=1}^n (U_i - U_{i-1})(U_i - U_{i-1})' \xrightarrow{p} 2\Omega(0) - \Omega(1) - \Omega'(1) \) (see Voev and Lunde (2007)).
Therefore, the lack of scaling makes this matrix diverge. In this way, the integrated covariance of the efficient returns and the covariance matrix of the noise vectors have different rates of convergence.

**Definition 3** (Matrix-Valued Normal Variable)

\( X \in \mathbb{R}^{d \times k} \) is a normally distributed matrix-valued random variable for \( A \in \mathbb{R}^{d \times k}, B \in \mathbb{R}^{dk \times dk} \), denoted by \( X \sim N(A, B) \), if

1. \( \text{vec}(X) \) follows a \( dk \times 1 \)-variate normal distribution,
2. \( E[\text{vec}(X)] = \text{vec}(A) \) and
3. for any four vectors \( a \in \mathbb{R}^d, b \in \mathbb{R}^k, p \in \mathbb{R}^d \) and \( q \in \mathbb{R}^k \),
\[
\text{Cov}(a'Xb, p'Xq) = v_{ab}Bv_{pq}
\]
where \( v_{ab} = \text{vec}([ab' + ba']/2) \in \mathbb{R}^{dk \times 1} \).

This definition follows Dinh and Nguyen (1994) and BNHLS (2008b). I use \( MN(0, B) \) to stand for a matrix-valued mixture-of-normals distribution with an ex-ante random covariance matrix \( B \).

### 2.2 Asymptotic Properties of Realized Kernels

The next lemma summarizes the asymptotic properties of the realized kernel estimator.

**Lemma 1** (Asymptotic Characterization of the Realized Kernels)

Let \( G_n \) be a bandwidth associated with the kernel window \( k \in K \). Under Assumption 1,

\[
\mathcal{R} \mathcal{K}(G_n) = d \int_0^1 \Sigma(v)dv + [k^{(2)}(0)] n G_n^{-2} \Omega + Z_{n, G_n} + \phi_n
\]

where

\[
Z_{n, G_n} \sim MN(0, n^{-1} G_n\mathcal{D} + n G_n^{-3} \mathcal{M} + G_n^{-1} \mathcal{C}), \quad (4)
\]

\[
\mathcal{D} = 4k^{(00)} \int_0^1 \Sigma(v) \otimes \Sigma(v)dv \quad (5)
\]

\[
\mathcal{M} = 4k^{(22)} \Omega \otimes \Omega \quad (6)
\]

\[
\mathcal{C} = 4k^{(11)} \left\{ \Omega \otimes \int_0^1 \Sigma(v)dv + \int_0^1 \Sigma(v)dv \otimes \Omega \right\} \quad (7)
\]

\( \phi_n \) is a collection of terms including the end points of the sample and remainders of the asymptotic covariance components. \( k^{(jj)} = \int_0^\infty [k^{(jj)}(x)]^2 dx \), \( j = 0, 1, 2 \), and \( \otimes \) is the Kronecker product. \( MN \) in (4) stands for the mixture-of-normals distribution with zero-mean and the ex-ante random covariance matrix.

*(Proof)* See the appendix.
Lemma 1 gives the first- and second-order characterization of the multivariate realized kernel estimator of BNHLS (2008b). They did not derive the explicit forms of \( M \) and \( C \), the exact order of \( nG_n^{-3} \) in front of \( M \), and the asymptotic normality of \( Z_{n,G_n} \) in (4).

The asymptotic covariance matrix of \( Z_{n,G_n} \) highlights the complications when estimating the integrated covariance matrix within the realized kernel framework. It has the component \( D \) due to the discretization error because the realized kernel estimator is based on the sample autocovariation matrices of the discrete observations. It has the component \( M \) due to the \( \mathcal{M} \)icrostructure noise because the discrete observations may be contaminated by the noise. The matrix \( M \) depends on the long-run covariance matrix of the noise because the noise may have serial as well as cross-sectional dependence. Another component \( C \) due to the \( \mathcal{C} \)ross product of the efficient returns and the noise appears in the asymptotic covariance because the sample autocovariation matrices are constructed by outer products of the observed returns.

Moreover, Lemma 1 suggests that the long run covariance of the noise \( \Omega \) may have a first order effect on the estimation of the integrated covariance \( \int_0^1 \Sigma(v)dv \). The asymptotic bias \( |k^{(2)}|nG_n^{-2}\Omega \) needs to be eliminated to obtain a consistent estimator of \( \int_0^1 \Sigma(v)dv \). BNHLS (2008b) deal with this issue by setting the bandwidth \( G_n \) proportional to \( n^g \) with \( g > 1/2 \). In this case, the bias term shrinks with the sample size \( n \), disappearing as \( n \to \infty \). Unfortunately, setting \( g > 1/2 \) makes \( n^{-1}G_nD \) the dominant term in the asymptotic variance, leading to a bias-variance trade off. The squared bias and variance are asymptotically balanced when \( G_n \) is proportional to \( n^{3/5} \), leading to the convergence rate \( n^{1/5} \). Since the parametric estimator of Gloter and Jacod (2001) with IID Gaussian noise achieves the convergence rate \( n^{1/4} \), there is a loss of efficiency using the realized kernel framework. Finally, the bias appears in the limiting distribution because of the bias-variance trade-off. When the realized kernel estimator is transformed non-linearly for the purpose of financial applications, this bias is also transformed non-linearly. The transformed bias may cause a poor behavior of a non-linear function of the estimated matrix in finite samples, as reported by BNHLS (2008b). Suboptimal properties of the realized kernel estimator stem from the nuisance parameter \( \Omega \).

2.3 Motivation for the New Estimators

A key idea to break the chain of the bias-variance trade-off is to reformulate the problem as a joint estimation of the integrated covariance matrix and the long-run covariance matrix of the noise. What if I strive to estimate these two unknown quantities jointly within the realized kernel framework? For that purpose, let me provide two realized kernels \( RK(H_n) \) and \( RK(G_n) \) with different bandwidths \( H_n \neq G_n \). Apart from \( \phi_n \), the asymptotic characterization of the individual realized kernel (4) implies the following joint asymptotic characterization of the two realized kernels: for the \( d \times d \) identity matrix \( I \),

\[
\begin{bmatrix}
RK(H_n) \\
RK(G_n)
\end{bmatrix} \sim \begin{bmatrix}
I & [k^{(2)}(0)|nH_n^{-2}I] \\
I & [k^{(2)}(0)|nG_n^{-2}I]
\end{bmatrix} \begin{bmatrix}
f_0^1 \Sigma(v)dv \\
\Omega
\end{bmatrix} + \begin{bmatrix}
Z_{n,H_n} \\
Z_{n,G_n}
\end{bmatrix}
\]

(8)
We know the left hand side: both $RK(H_n)$ and $RK(G_n)$ are constructed using the actual data based on Definition 2 with our choice of a kernel function $k \in K$ and bandwidths. We know the block-matrix $X_n$: $I$ is the identity matrix; $|k^{(2)}(0)|nH_n^{-2}$ and $|k^{(2)}(0)|nG_n^{-2}$ are based on the selected kernel window, the bandwidths and the sample size. The unknown quantities are the integrated covariance $\int_0^1 \Sigma(v)dv$ and the long-run covariance of the noise $\Omega$. For a large sample size and large bandwidths, we know that distributions of $Z_{n,H_n}$ and $Z_{n,G_n}$ are zero-mean mixture-of-normals. Therefore, we can interpret $\Xi$ as a linear regression model with two matrix-valued observations. This interpretation delivers a simple yet important insight. Given $\text{det}(X_n) = |k^{(2)}(0)|nG_n^{-2}(1 - G_n^2H_n^{-2}) \neq 0$ when $G_n \neq H_n$, one can invert $X_n$ to obtain the following expression:

$$
\begin{bmatrix}
I & |k^{(2)}(0)|nH_n^{-2}I \\
I & |k^{(2)}(0)|nG_n^{-2}I
\end{bmatrix}^{-1}
\begin{bmatrix}
RK(H_n) \\
RK(G_n)
\end{bmatrix}
\overset{d}{\sim}
\begin{bmatrix}
\int_0^1 \Sigma(v)dv \\
\Omega
\end{bmatrix}
+ X_n^{-1}
\begin{bmatrix}
Z_{n,H_n} \\
Z_{n,G_n}
\end{bmatrix}
$$

This expression suggests that the left hand side, which is also known to us completely, can be a joint estimator of the two unknown quantities $\int_0^1 \Sigma(v)dv$ and $\Omega$. A straightforward calculation shows that the left hand side is a stack of two $d \times d$ matrices given by

$$
TSRK(G_n, H_n) := (1 - G_n^2H_n^{-2})^{-2} \{RK(H_n) - G_n^2H_n^{-2}RK(G_n)\} 
$$

$$
TS\Omega(G_n, H_n) := (1 - G_n^2H_n^{-2})^{-2}(|k^{(2)}(0)|nG_n^{-2})^{-2} \{-RK(H_n) + RK(G_n)\}. \quad (9)
$$

The first $d \times d$ matrix denoted by $TSRK$ is a new non-parametric estimator of the integrated covariance. I specifically call it the “Two Scale Realized Kernel” estimator because it is a particular linear combination of two realized kernel estimators with different bandwidths, $H_n \neq G_n$. Note that the coefficients of the linear combination automatically eliminate the asymptotic biases of the individual realized kernels due to the long-run covariance matrix of the noise. This mechanism is quite intuitive because I tried to estimate the integrated covariance and the long-run covariance matrix of the noise jointly so that the latter is no more a nuisance parameter for the former. My joint approach also delivers the second $d \times d$ matrix denoted by $TS\Omega$ as a new non-parametric estimator of the long-run covariance matrix of the noise.

### 2.4 Asymptotic Properties of New Estimators

Note that two bandwidths are treated symmetrically in the definition of new estimators, (9) and (10). Therefore, given $G_n = bn^g$ and $H_n = cn^h$ with $b, c > 0$, it is enough to state the asymptotic result for $g \leq h$. The following is the main theorem of this paper.

**Theorem 1 (Multivariate Two Scale Realized Kernels)**

Let $G_n = bn^g$, $H_n = cn^h$, $r_n = G_n/H_n$ and $m_n$ be proportional to $n^\nu$ where $m_n$ is introduced in (D). Suppose $b, c > 0$, $0 < g \leq h$, $r_n \neq 0, 1$ and Assumption 1 holds. Define $TSRK(G_n, H_n)$ and $TS\Omega(G_n, H_n)$ by (9) and (10).
1. (Joint Consistency)

For \( \max\{g, 1/3\} < h < \min\{3 - 4g, 1\} \) or \( 1/3 < g = h < 3/5 \) and \( \nu \in (0, 1) \), as \( n \to \infty \),

\[
\begin{bmatrix}
TSRK(G_n, H_n) \\
TS\Omega(G_n, H_n)
\end{bmatrix} \overset{p}{\to} \begin{bmatrix}
\int_0^1 \Sigma(v)dv \\
\Omega
\end{bmatrix}
\]

2. (Joint Asymptotic Normality)

Suppose \( h = 1/2 \) and \( \nu > 1/4 \) additionally. Then, as \( n \to \infty \),

\[
\begin{bmatrix}
n^{1/4} \left( TSRK - \int_0^1 \Sigma(v)dv \right) \\
n^{1-\eta/2} \left( TS\Omega - \Omega \right)
\end{bmatrix} \overset{d}{\to} MN \left( 0, \lim_{n \to \infty} V(c, r_n) \right)
\]

where \( O \) is the \( 2d \times d \) matrix of zeros and \( V(c, r) \) is a \( 2d^2 \times 2d^2 \) matrix given by

\[
V(c, r) = \Phi^{(0)}(r) \otimes \{cD\} + \Phi^{(2)}(r) \otimes \{c^{-3}M\} + \Phi^{(1)}(r) \otimes \{c^{-1}C\}
\]

The \((1, 1)\) elements of \( \Phi^{(i)}(r) \in \mathbb{R}^{2 \times 2}, i = 0, 1, 2 \) are given by

\[
\Phi^{(0)}_{11}(r) = \frac{1 - 2r^2\delta^{(0)}(r) + r^5}{1 - 2r^2 + r^4},
\]

\[
\Phi^{(2)}_{11}(r) = \frac{1 - 2\delta^{(2)}(r) + r}{1 - 2r^2 + r^4},
\]

\[
\Phi^{(1)}_{11}(r) = \frac{1 - 2r\delta^{(1)}(r) + r^3}{1 - 2r^2 + r^4}.
\]

where \( \delta^{(i)}(r) := \left( \int_0^\infty k^{(i)}(x)^2dx \right)^{-1} \int_0^\infty k^{(i)}(x)k^{(i)}(x/r)dx < \infty, i = 0, 1, 2 \). The full expressions of \( \Phi^{(i)}(r) \) are given explicitly in the appendix.

(Proof) See the Appendix.

2.5 Discussion of Theorem 1

This theorem highlights the many important features of my proposed method.

First, to the best of my knowledge, Theorem 1 is the first joint asymptotic result for the non-parametric estimation of the efficient variation and the noisy variation. Moreover, the joint asymptotic distribution is correctly centered at the true values, \( \int_0^1 \Sigma(v)dv \) and \( \Omega \). The bias-free joint distribution may potentially be useful for an efficient inference on my estimator or higher order refinement.

Second, the elimination of the asymptotic biases of the individual realized kernels allows one to select a growth rate of the bandwidth \( H_n \) according to the trade-off between three variance components instead of the trade-off between bias and variance. When \( g < h \), the asymptotic covariance of the two scale realized kernel estimator of the integrated covariance

11
is given by

$$n^{-1}H_n D + nH_n^{-3}M + H_n^{-1}C$$

(11)

where $D$, $M$ and $C$ are given by (11). These three components are balanced if $H_n$ is proportional to $n^{1/2}$, leading to the asymptotic covariance of order $n^{-1/2}$. This is why my estimator achieves the convergence rate $n^{1/4}$. Since $n^{1/4}$ is the best parametric convergence rate for a volatility estimator in the presence of noise, my method delivers the first rate-optimal non-parametric estimator of the integrated covariance matrix in the presence of serially dependent noise. The same argument also applies when $g = h$.

Third, my estimator is different from another bias-corrected non-parametric estimator of Zhang et al. (2005) in terms of the smoothness of underlying kernel windows. My estimator is based on a twice continuously differentiable kernel window $k \in \mathcal{K}$ in Definition 1. On the other hand, BNHLS (2007) show that the estimator of Zhang et al. (2005) is implicitly based on the kinked Bartlett window. The difference in the smoothness of underlying kernel windows leads to the difference in the orders of the asymptotic variance, further leading to the difference in the convergence rates. In fact, the estimator of Zhang et al. (2005) is only $n^{1/6}$-consistent. Moreover, there is no joint limiting distribution for the efficient variation and the noisy variation because these two variations are estimated separately. Ait-Sahalia et al. (2008) also propose a rate-optimal multi-scale estimator. However, it is derived only for a univariate case. Moreover, as Ait-Sahalia et al. (2008, Section 6) explained, we have to pay the cost of higher complexity to use it. In contrast, the construction of my estimator is transparent and relatively straightforward.

Fourth, the asymptotic covariance matrix of my estimator is tractable. If one is interested in a feasible selection of the bandwidth based on the asymptotic covariance, then the consistent estimates of $\Omega$ and $\int \Sigma(v) \otimes \Sigma(v) dv$ are not essential, as emphasized by BNHLS (2008c). In that case, one may often approximate the latter by $\int \Sigma(v) dv \otimes \int \Sigma(v) dv$. Then, all one needs to know are $\int \Sigma(v) dv$, $\Omega$ and $\Phi^{(i)}_{11}(r)$ where $r = \lim_{n \to \infty} G_n / H_n$. Pilot estimates of $\int \Sigma(v) dv$ and $\Omega$ are assessed easily by several methods. The computation of $\Phi^{(i)}_{11}(r)$ is straightforward because $\delta$ can be approximated to an arbitrary degree of accuracy using numerical methods. The following figure plots the numerical evaluation of the function $\Phi^{(i)}_{11}(r)$ using the Parzen window for $r \in \{0, .01, .02, \ldots, .99\}$. [Figure here]

The figure shows $\Phi^{(i)}_{11}(r) \geq 1$ and $\lim_{r \to 0} \Phi^{(i)}_{11}(r) = 1$. Therefore, if the two bandwidths have different growth rates $g < h = 1/2$ so that $r_n = G_n / H_n \to 0$, the asymptotic covariance matrix of $TSRK(G_n,H_n)$ reduces to

$$\lim_{n \to \infty} V_{1,1}(c, r_n) = cD + c^{-3}M + c^{-1}C$$

(12)
If $g = h = 1/2$, $r_n = b/c$ is a constant so that

$$\lim_{n \to \infty} V_{1,1}(c, r_n) = \Phi_{11}^{(0)}(r)cD + \Phi_{11}^{(2)}(r)c^{-3}M + \Phi_{11}^{(1)}(r)c^{-1}C.$$  \hspace{1cm} \text{(13)}$$

Even if $\lim_{n \to \infty} r_n = 0$, the inference based on (13) with $r = r_n$ is more conservative than the one based on (12).

Fifth, this theorem allows a wider admissible range of the growth rates of the bandwidths compared with the previous results such as those of BNHLS (2008b), Zhang et. al. (2005) and Ait-Sahalia et. al. (2008). As already mentioned, it is possible to select the two bandwidths $G_n = bn^g$, $H_n = cn^h$ with the same growth rates $g = h = 1/2$ but with different proportionality factors $b \neq c$ to guarantee $G_n \neq H_n$. Such pair for the bandwidths still delivers the best parametric rate of convergence $n^{1/4}$.

Finally, $\nu > 1/4$ is theoretically important to obtain the limiting distribution. This specific order is necessary to eliminate the bias caused by the microstructure noise on the end points. In practice, however, the terms due to the end points of the sample do not contribute to the finite sample bias and variance so much—see BNHLS (2008b, Section 6.5).

2.6 Bandwidth Selection

The growth rate of $H_n = cn^h$ was selected as $h = 1/2$ for the trade-off between three variance components given by (11). However, $c$ in $H_n = cn^{1/2}$ and $G_n$ are still undetermined. Let me discuss the selection of $c$ at first. A good candidate of $c$ is given by a minimizer of each diagonal element of the asymptotic covariance matrix as a candidate for a single proportionality factor. Focusing on the diagonal elements is justified by the Cauchy-Schwartz inequality: any covariance between two random variables is smaller than the product of the individual standard deviations.

Suppose $g < 1/2$. In this case, the asymptotic covariance is given by (12). Given the conventional approximation $\int_0^1 \Sigma(v) \otimes \Sigma(v) dv \approx \int_0^1 \Sigma(v) dv \otimes \int_0^1 \Sigma(v) dv$ and the Parzen window for the construction of realized kernels, the minimizer of $i$-th diagonal element is given by

$$c_i^* = 4.78 R_{ii}^{1/2}$$  \hspace{1cm} \text{(14)}$$

where $R_{ii} = \Omega_{ii}/\int_0^1 \Sigma_{ii}(v) dv$ is the ratio of the $i$-th diagonal elements of $\Omega$ and $\int_0^1 \Sigma(v) dv$ and the number 4.78 comes from the characteristics of the Parzen window. See the appendix for the derivation of (14).

Next, I need to estimate $R_{ii} = \Omega_{ii}/\int_0^1 \Sigma_{ii}(v) dv$ by some feasible method. Following BNHLS (2008b), I estimate $\int_0^1 \Sigma_{ii}(v) dv$ by a subsampling-based estimator using the 15-minute returns:

$$SUB_{ii} = \frac{1}{15} \sum_{j=0}^{14} \left[ \sum_{k=j+15,j+30,j+45,...} (\ln P_t^{(i)} - \ln P_{t-k}^{(i)})^2 \right]$$  \hspace{1cm} \text{(15)}$$
where $\ln P_{i_k}^{(i)}$ is the $k$-th observation of the $i$-th security’s log price, $k = 0, 1, \ldots, n$. We have fifteen different realized variance estimates depending on the fifteen different initial points. Their average gives the estimate of $i$-th security’s integrated variance. The choice of the 15-minute returns is based on the fact that the effect of microstructure noise is less pronounced at this frequency. In this case, the subsampling-based method gives a reasonably good first-stage estimate because the variance is not large by using all data while the bias caused by the microstructure noise is not so severe. Then, I estimate $\Omega_{ii}$ by

$$\hat{\Omega}_{ii} = RK_{ii}(n^{1/2})/12$$

i.e. the $i$-th diagonal element of $RK(n^{1/2})/12$. This is a special case of the long-run covariance matrix estimator $RK(A_n)/(|k^{(2)}(0)|nA_n^{-2})$ proposed by Ikeda (2009) with $A_n = n^{1/2}$ and $|k^{(2)}(0)| = 12$ for the Parzen window. This estimator is positively biased, but it is necessary to offset the positive bias of the subsampling-based estimator. Moreover, the upward bias in an estimate of $R_{ii}$ implies a conservative choice of the bandwidth. As BNHLS (2008c) noted, it is better to err on the side of a large bandwidth. Another justification is that a larger value of the bandwidth makes the estimator more robust to serial dependence in the noise.

The ratio of (15) and (16) is a feasible version of $R_{ii}$. Plugging it in (14), I obtain a feasible minimizer of the $i$-th diagonal element of the asymptotic covariance matrix of my estimator. Finally, take the maximum of such minimizers as a single proportionality factor for $H_n = cn^{1/2}$.

If the two bandwidths $G_n = bn^g$ and $H_n = cn^h$ have the same growth rate, i.e. $g = 1/2$, the asymptotic covariance matrix of my estimator is given by (13). Even in this case, I stick to the rule for the case of $g < 1/2$. It does not fully optimize the diagonal elements of the asymptotic covariance matrix due to the negligence of the dependence of $\Phi_{11}^{(i)}(r)$ on $c$ through $r = b/c$. However, $\Phi_{11}^{(i)}(r)$ is highly non-linear and no closed form solution is available.

Unfortunately, the first order asymptotic result does not dictate any optimality criterion for the selection of another bandwidth $G_n = bn^g$ except the admissible range of the growth rate $0 < g \leq 1/2$. When $G_n$ is much smaller than $H_n$, a portion of the finite sample bias may still be large. On the other hand, the previous asymptotic variance expression indicates that $r_n = G_nH_n^{-1}$ is significantly above zero for a large $G_n$ and therefore $\Phi_{11}^{(i)}(r_n) > 1$, causing a variance inflation in finite samples. Therefore, $G_n$ is subject to a higher order trade-off. Given the strong downward bias of non-linear functions of the realized kernel in finite samples reported by BNHLS (2008b), I recommend to use a large $G_n$, especially $G_n = n^{1/2}$. 
A Recommended Rule

The following is the recommended rule of the bandwidth selection for the $TSRK(G_n, H_n)$ estimator, based on the Parzen window.

\begin{align*}
G_n &= n^{1/2} \\
H_n &= c^* n^{1/2} \\
c^* &= 4.78 \max_{i=1,\ldots,d} \{ \hat{R}_{ii}^{1/2} \} \\
\hat{R}_{ii} &= \hat{\Omega}_{ii} / \text{SUB}_{ii} \\
\hat{\Omega}_{ii} &= RK_{ii}(n^{1/2})/12
\end{align*}

Since the growth rates of the two bandwidths are the same, the asymptotic covariance involves the multiplicative factors $\Phi^{(i)}_{11}(r)$ where $r = G_n/H_n = 1/c^*$ is now a constant. Therefore, whenever following this recommendation, one should always use the expression of the asymptotic covariance in (13).

3 Positive Semi-Definite Correction for Two Scale Estimators

To estimate non-linear functions of the integrated covariance, it is crucial to ensure positive semi-definiteness of my estimator in finite samples. Although my estimator on its own is not guaranteed to be positive semi-definite in finite samples, there is a simple transformation to fix this problem. The implementation is straightforward, and the proof of the asymptotic equivalence with and without that correction is very simple.

My two scale realized kernel estimator can be rewritten as

\[ TSRK(G_n, H_n) = \hat{R}_n - \hat{S}_n \]

where $\hat{R}_n = (1 - G_n^2 H_n^{-2})^{-1} RK(H_n)$ and $\hat{S}_n = (1 - G_n^2 H_n^{-2})^{-1} G_n^2 H_n^{-2} RK(G_n)$. By the symmetry of the estimator with respect to the two bandwidths $G_n$ and $H_n$, I assume $G_n < H_n$ without loss of generality. This means that $\hat{R}_n$ and $\hat{S}_n$ are always positive semi-definite given $k \in K$ in Definition 1. Suppose additionally that $RK(H_n)$ is positive definite. This is a very weak assumption in practice if the number of securities is not so large. The case where it is not satisfied would be a day with only two trades at the same price, for instance. Since $\hat{R}_n$ is then positive definite, it has a non-singular Cholesky decomposition such that $\hat{R}_n = \Lambda_n \Lambda_n'$, where $\Lambda_n$ is a lower triangular matrix with strictly positive diagonal elements. Therefore, I can transform $TSRK(G_n, H_n)$ into the following sandwich form:

\[ TSRK(G_n, H_n) = \hat{R}_n - \hat{S}_n = \Lambda_n \Lambda_n' - \hat{S}_n = \Lambda_n (I - \Lambda_n^{-1} \hat{S}_n \Lambda_n^{-1}) \Lambda_n' \equiv \Lambda_n (I - \hat{X}_n) \Lambda_n' \]
where \( \hat{X}_n \equiv \Lambda_n^{-1} \hat{S}_n \Lambda_n^{-1} \) is positive semi-definite. If \( I - \hat{X}_n \) is positive semi-definite, my estimator is positive semi-definite as well. Notice the following relation for the matrix power series:

\[
I - \hat{X}_n = \left( \lim_{J \to \infty} \sum_{j=0}^{J-1} \hat{X}_n^j \right)^{-1}, \quad \hat{X}_n^0 = I
\]  

(23)

provided \( \max_{i=1 \ldots d} |\lambda_i| < 1 \) where \( \lambda_i \) is \( i \)-th eigenvalue of \( \hat{X}_n \) - see Abadir and Magnus (2005, p249). (23) is not a statistical asymptotic relation; rather, one should fix \( \hat{X}_n \) and increase \( J \) independently of \( n \). This relation holds for \( X = p \lim_{n \to \infty} \hat{X}_n \) because \( X \) is positive semi-definite - see the appendix. On the other hand, it may not hold in finite samples because \( \max_{i=1 \ldots d} |\lambda_i| < 1 \) may be violated. However, if the summation in (23) stops at a finite number of terms, say \( J_n \), then \( (\sum_{j=0}^{J_n-1} \hat{X}_n^j)^{-1} \) is positive definite since \( \hat{X}_n \) is positive semi-definite and \( \hat{X}_n^0 = I \). Therefore, a positive definite version of \( TSRK(G_n, H_n) \) is defined as follows:

\[
TSRK^+(G_n, H_n, J_n) = \Lambda_n \left( \sum_{j=0}^{J_n-1} \hat{X}_n^j \right)^{-1} \Lambda_n'
\]  

(24)

Since \( (\sum_{j=0}^{J_n-1} \hat{X}_n^j)^{-1} \) is positive definite, \( TSRK^+(G_n, H_n, J_n) \) is positive definite given the non-singularity of \( \Lambda_n \). To enjoy the statistical asymptotic properties of the original version in this transformed version, \( J_n \) should grow at some appropriate rate with \( n \) so that when \( \hat{X}_n \overset{p}{\to} X \), \( (\sum_{j=0}^{J_n-1} \hat{X}_n^j)^{-1} \overset{p}{\to} I - X \), which is positive semi-definite.

Now I can prove the following simple yet important result.

**Proposition 1 (Positive Semi-Definite Correction)**

Suppose \( RK(H_n) \) is positive definite. Let \( TSRK^+(G_n, H_n, J_n) \) be the positive semi-definite version of \( TSRK(G_n, H_n) \) defined in (24). Under the assumptions of Theorem 1,2,

\[
TSRK^+(G_n, H_n, J_n) = TSRK(G_n, H_n) + o_p(n^{-1/4}) \quad \text{as} \quad \frac{1}{n} + \frac{\ln n}{J_n} \to 0.
\]

(Proof) See the Appendix.

Therefore, \( TSRK^+(G_n, H_n, J_n) \) is always positive definite in finite samples and is asymptotically equivalent to \( TSRK(G_n, H_n) \) given \( RK(H_n) \) being positive definite and \( J_n \) growing at an appropriately slow rate. Any positive power of the sample size \( J_n = n^\theta, \theta > 0 \) works because \( n^\theta = e^{\theta \ln n} \) grows faster than \( \ln n \). Although no specific rule to select \( \theta \) exists, I can give the following informal argument: setting \( \theta \) too large for \( J_n = n^\theta \) may cause a near-singularity of \( \sum_{j=0}^{J_n-1} \hat{X}_n^j \) in numerical inversion when the condition required for the formula (23) does not hold, which is fairly consistent with the ill-definedness of \( I - \hat{X}_n \) as the inverse of the limit of \( \sum_{j=0}^{J_n-1} \hat{X}_n^j \). On the other hand, setting \( \theta \) too small may make \( TSRK^+(G_n, H_n, J_n) \) different
from $TSRK(G_n, H_n)$ when it is positive. I use $\theta = 1/2$, namely $J_n = \sqrt{n}$ throughout the simulation as well as empirical application. Although this rule is ad-hoc, the original version and the transformed version are almost identical if the former is positive. In the rest of the paper, I always assume $J_n = \sqrt{n}$ and the positive definiteness of $RK(H_n)$ whenever I use the transformed version. Therefore, I omit $J_n$ from the argument of $TSRK^+$. 

4 Nonlinear Functions of the Integrated Covariance Matrix

In many financial applications, it is important to guarantee precise estimation of a non-linear function of the integrated covariance matrix. To apply a non-linear transformation, the estimated matrix needs to be positive semi-definite. For instance, the correlation estimate may be greater than one in absolute value if the matrix does not satisfy this property. Therefore, I use the positive semi-definite version of my two scale realized kernel estimator, denoted by $TSRK^+$. 

The main examples of non-linear functions used in this paper are the high frequency regression coefficient and the high frequency correlation. Let me focus on two securities, 1 and 2. The high frequency regression coefficient of the second asset’s returns onto the first asset’s returns is defined by

$$
\beta_{12} = \left( \int_0^1 \Sigma_{11}(v)dv \right)^{-1} \int_0^1 \Sigma_{12}(v)dv
$$

This quantity represents the day-by-day market beta of the second asset if the first asset is the market portfolio because the expected returns are typically negligible for such a short period of time. The financial importance of this ratio in the high frequency setting is highlighted by, among others, Andersen et. al. (2006). They use the realized covariance to estimate this ratio and therefore the associated estimate is called the realized beta. An alternative is given by replacing the numerator and the denominator by the corresponding elements of my estimator:

$$
\hat{\beta}_{12}^{(TS)}(G_n, H_n) = \frac{TSRK^+_2(G_n, H_n)}{TSRK^+_1(G_n, H_n)}
$$

I call $\hat{\beta}_{12}^{(TS)}$ the $TSRK^+$-based beta.

The same argument applies to the high-frequency correlation between the first asset’s returns and the second asset’s returns, defined by

$$
\rho_{12} = \left( \int_0^1 \Sigma_{11}(v)dv \cdot \int_0^1 \Sigma_{22}(v)dv \right)^{-1/2} \int_0^1 \Sigma_{12}(v)dv
$$

Then, the $TSRK^+$-based correlation is defined by

$$
\hat{\rho}_{12}^{(TS)}(G_n, H_n) = \left( TSRK^+_1 \cdot TSRK^+_2 \right)^{-1/2} TSRK^+_1
$$
Proposition 2 (Limiting Distributions of the $T SRK^+$-based Beta and Correlation)

Suppose $\int_0^1 \Sigma_{12}(v)dv$ is nonzero, $G_n = bn^a$, $H_n = cn^b$, $0 < g \leq h$, $r_n = G_n/H_n$ and all assumptions in Theorem 1 hold. Then, as $n \to \infty$,

$$n^{1/4} \left( \hat{\beta}_{12}^{(TS)}(G_n, H_n) - \beta_{12} \right) \overset{d}{\to} N \left( 0, \lim_{n \to \infty} V_{\beta_{12}}(c, r) \right),$$

$$n^{1/4} \left( \hat{\rho}_{12}^{(TS)}(G_n, H_n) - \rho_{12} \right) \overset{d}{\to} N \left( 0, \lim_{n \to \infty} V_{\rho_{12}}(c, r) \right),$$

where the asymptotic variances of $T SRK^+$-based beta and correlation are respectively given by

$$V_{\beta_{12}}(c, r) = \Phi^{(0)}_{11}(r)cD_{\beta_{12}} + \Phi^{(2)}_{11}(r)c^{-3}M_{\beta_{12}} + \Phi^{(1)}_{11}(r)c^{-1}C_{\beta_{12}}, \quad \text{and}$$

$$V_{\rho_{12}}(c, r) = \Phi^{(0)}_{11}(r)cD_{\rho_{12}} + \Phi^{(2)}_{11}(r)c^{-3}M_{\rho_{12}} + \Phi^{(1)}_{11}(r)c^{-1}C_{\rho_{12}}.$$

$\Phi^{(i)}_{11}$ are identical to those given in Theorem 1. For the concrete expression of each component of the asymptotic variances and the proof of this claim, see the appendix.

Proposition 2 is the first $n^{1/4}$-consistent result regarding non-linear functions of the integrated covariance matrix in the presence of microstructure noise. The limiting distributions are correctly centered at the true values. Therefore, the biases in the non-linear functions of my estimator should be small. This asymptotic prediction highlights the theoretical goal of this paper, because it means that my procedure can be more suited for the estimation of financial quantities represented by non-linear functions of the integrated covariance than currently available procedures. In fact, BNHLS (2008b, Example 1) shows the presence of bias in the limiting distribution of the realized kernel-based estimator of $\beta_{12}$.

5 Simulation Studies

In this section I study a finite sample performance of the positive semi-definite version of my estimator by simulations. There are two goals in this simulation. First, I need to compare the performance of non-linear functions of my estimator with those of the realized kernel estimator. In particular, I need to study the effect of the serial as well as cross-sectional dependence of the noise on my estimator. Second, I need to study the effect of the bandwidth selection on the performance of my estimator because the first order asymptotic result does not dictate any optimality criterion for the selection of the bandwidth $G_n$. Since I recommended to use a large $G_n$ for the bias reduction, I need to study the possible variance inflation associated with it.

The simulation design follows BNHLS (2008a,b). One trading period of 6.5 hours is normalized on $[0, 1]$ continuum so that 1 second corresponds to the point $1/23400$. Given three mutually independent Brownian motion processes $B^{(1)}$, $B^{(2)}$ and $W$, the efficient log prices of each security, $\ln P_t^{(i)}$, $i = 1, 2$, follows a 2-factor stochastic volatility model: for
\[ t \in [0,1], \]
\[
d\ln P^*_t(i) = \mu dt + \sigma^{(i)}_t \left\{ \phi dB^{(i)}_t + \sqrt{1-\phi^2}dW_t \right\} \quad (25)\]
\[
\sigma^{(i)}_t = \exp(b_0 + b_1 \tau^{(i)}_t) \quad (26)\]
\[
d\tau^{(i)}_t = \alpha \tau^{(i)}_t dt + dB^{(i)}_t \quad (27)\]
\[
\ln P_{tk}^{(i)} = \ln P_t^{(i)} + U^{(i)}_k \quad (28)\]
\[
t_k = 60k/23400, \quad k = 1, 2, \ldots, 390 \quad (29)\]

\( B^{(i)} \) is a factor specific to \( i \)-th security, while \( W \) is a common factor for both securities. The parameter values I use are \( \mu = .03 \), \( b_1 = .125 \), \( \alpha = -.025 \) and \( b_0 = b_1^2/(2\alpha) \) and \( \phi = -.3 \), following Hansen and Lunde (2006) and BNHLS (2008b). Note that \( \phi \) controls the degree of the leverage effect, namely the correlation between the returns and the volatility innovations.

I generate 23400 data points from discretized versions of (25), (26) and (27), then pick up every sixtyth to construct minute-by-minute data. Therefore, the effective sample size is \( n = 390 \). Each discrete observation is contaminated by the market microstructure noise \( U^{(i)}_k \).

Given \( \epsilon^{(i)}_k \) a serially independent white noise, the microstructure noise is specified by
\[
U^{(1)}_k = \theta_1 U^{(1)}_{k-1} + \epsilon^{(1)}_k, \quad U^{(2)}_k = \epsilon^{(2)}_k + \theta_2 \epsilon^{(2)}_{k-1}. \]

I study two pairs of \( \theta \)'s. For the first case, \( \theta_1 = \theta_2 = 0 \) so that two noise components are serially independent. For the second case, \( \theta = -.7 \) and \( \theta_2 = .5 \), following Ikeda (2009). The \((j,l)\) element of the long-run covariance matrix of the noise is given by
\[
\Omega_{jl} = .005 \left( \frac{1}{23,400} \sum_{i=1}^{23,400} \Sigma_{j,j}(t_i) \Sigma_{j,j}(t_i) \right)^{1/2} \cdot \zeta_{jl} \]

where the noise-to-signal ratio .005 is based on a preliminary analysis of the actual data of S&P 500 index futures prices. \( \zeta_{jl} \) is either +1, 0 or −1. While \( \zeta_{11} = \zeta_{22} = 1 \), \( \zeta_{12} \) may be zero or negative one. The number of replications is 10,000.

I compare three estimators.

- The subsampling-based realized covariance using the 15-minute sparse sampling, as defined by (15). I denote this estimator by \( SUB \). This is a multivariate version of the second best estimator defined in Zhang et. al. (2005). It gives a pilot estimate of the integrated covariance matrix in selecting the bandwidths of kernel-based estimators in BNHLS (2008b) and in this paper as well.

- The realized kernel of BNHLS (2008b) using the Parzen window and the asymptotic mean-squared-error optimal bandwidth, \( RK(fn^{3/5}) \). The proportionality factor \( f \) is
selected by the following rule as recommended by BNHLS (2008c):

\[ f = 3.51 \max \left\{ \left( \frac{\hat{\omega}^2_{11}}{SUB_{11}} \right)^{1/5}, \left( \frac{\hat{\omega}^2_{22}}{SUB_{22}} \right)^{1/5} \right\} \]  

(30)

where \( \hat{\omega}^2_{jj} \) is the subsampling-based estimator of the short-run variance of \( j \)-th noise using 2-minute sparse sampling, and the number 3.51 comes from the characteristics of the Parzen window. \( SUB_{jj} \) is the \( j \)-th diagonal element of the previous \( SUB \) estimate.

- The two scale realized kernel estimator using the Parzen window, \( TSRK^+(G_n, H_n) \). \( H_n = cn^{1/2} \) is determined by following (18). Specifically,

\[ c = 4.78 \max \left\{ \left( \frac{RK_{11}(n^{1/2})}{12SUB_{11}} \right)^{1/2}, \left( \frac{RK_{22}(n^{1/2})}{12SUB_{22}} \right)^{1/2} \right\} \]  

(31)

where \( RK_{jj}(n^{1/2}) \) and \( SUB_{jj} \) are the \( j \)-th diagonal elements of the realized kernel estimator with the bandwidth \( n^{1/2} \) and the previous subsampling-based estimator, respectively. I try \( G_n = n^g \), \( g = 1/5, 1/4, 1/3 \) and 1/2. I also impose the restriction \( G_n + 1 \leq H_n \) so that \( G_n \neq H_n \).

The bias and root mean squared error are in percentages of the true values. For instance, if the true value of \( \beta_{12} \) is 1 and the relative bias of an estimator is \(-10\), it means that the estimator generates the estimates of \( \hat{\beta}_{12} = .9 \) on average.

Table 1 reports the results for the pair of two independent IID noises.

The general features of the simulation results in Table 1 can be summarized as follows.

First, my method successfully reduces the bias in the estimates of all quantities. In particular, the individual integrated variances are estimated quite accurately, leading to small biases in the three non-linear functions as well. On the other hand, the other estimators are not as good for these quantities. The relatively large biases for individual integrated variances are transformed to the dominant biases in the root mean squared errors of the three non-linear functions. All three estimators do decent jobs for the estimation of the cross variation \( \int \Sigma_{12}(v)dv \), because there is no cross sectional dependence between the two noise processes.

Second, the gain in bias reduction outweighs the cost in variance inflation for the non-linear functions of my estimator, often leading to a smaller root mean squared error than that of the realized kernel estimator. Therefore, Table 1 confirms the asymptotic prediction from Proposition 2 that the biases of non-linear functions of my estimator are small, with an extra benefit of the smaller root mean squared errors in this case. The web appendix includes the case where the realized kernel estimator outperforms the two scale realized kernel estimator.
in terms of the root mean squared error\(^2\). Even in that case, however, the deterioration of the mean squared error is modest and the bias is always smaller than that of the former. Although the estimates of the individual elements of the matrix are less good in terms of the root mean squared error, Ikeda (2009) reported that the two scale realized kernel also generates better estimates of the univariate integrated variance than the realized kernel for a reasonably large sample size. Moreover, the first order bias correction would be more important than the second-order variance inflation in the context of high frequency data analysis because it is possible to fit a cross-daily model such as ARMA to the day-by-day raw estimates for smoothing, as proposed by Andersen et. al. (2006) or BNHLS (2008b).

Third, the correlation based on my method is very accurate. Given the strong downward biases of the correlation estimates based on the realized kernel reported by BNHLS (2008b, Sections 4 and 5) and reconfirmed by my simulations, my estimator is a preferred choice for the analysis of day-by-day correlation using high frequency data.

Fourth, the results for different growth rates \( g \) of \( G_n = n^g \) are mixed. Typically a larger \( g \) leads to a smaller bias and a larger root mean squared error of non-linear functions. Whenever we use a large \( g \), therefore, we should always multiply the estimates of asymptotic covariance components by \( \Phi_{11}(r_n) \) with \( r_n = G_n/H_n \) introduced in (1) as a finite sample correction for a conservative inference.

Table 2 summarizes the results for the pair of AR and MA noises. My estimator turns out to be very robust to serial dependence in the noise. On the other hand, the other estimators are plagued by the upward bias in the estimates of the integrated variance of the first security, \( \int \Sigma_{11}(v)dv \), due to the negative AR(1) noise. It leads to the strong downward biases of non-linear functions of it.

Finally, I study the effect of the cross-sectional dependence in the noise. In this case, the off-diagonal elements of the long-run covariance matrix of the noise is given by

\[
\Omega_{12} = -.005 \left( \frac{1}{23,400} \sum_{i=1}^{23,400} \Sigma_{1,1}(t_i) \Sigma_{1,2}(t_i) \right)^{1/2},
\]

leading to a negative cross-sectional dependence between two noises. Table 3 reveals that my estimator is quite robust to the dependence of this type. On the other hand, the estimates of the cross variation \( \int_0^1 \Sigma_{12}(v)dv \) based on the other two are quite sensitive to it, leading to large downward biases in the estimates of the three non-linear functions.

In sum, my estimator performs excellently for the non-linear functions of the integrated covariance matrix. In particular, it is notably less biased and more robust to the dependent

\(^2\)See [http://people.bu.edu/sikeda](http://people.bu.edu/sikeda)
noise than the other considered estimators. Ait-Sahalia et. al. (2008) reported a clear evidence of the serial dependence of negative AR(1) type in the noise associated with the transaction prices of Microsoft and Intel, while Ubukata and Oya (2008) reported a non-trivial pattern of the cross autocorrelation in the noise. Therefore, my method seems suited for an application to actual data of transaction prices.

6 Application: Dynamic Hedge Ratio Between S&P 500 Spot and Futures Prices

In this section I apply my multivariate two scale realized kernel method to the estimation of the dynamic hedge ratio between efficient transaction prices of the S&P500 index spot and futures using their intra-daily one minute prices. After cleaning the raw data, the final sample covers a period from 01-Aug-1997 to 02-Mar-2007 of total 2332 days. For a detailed description of the cleaning procedure and additional results in this section, see the web appendix.

The hedge ratio at the end of the day-$\tau$ is given by

$$\beta(\tau) = \left( \int_{9:35(\tau)}^{16:00(\tau)} \Sigma_{ff}(v) dv \right)^{-1} \left( \int_{9:35(\tau)}^{16:00(\tau)} \Sigma_{sf}(v) dv \right)$$

where $\tau = 1, \ldots, 2332$ corresponds to 01/Aug/1997, . . . , 02/Mar/2007, and the integrals are taken with respect to the intra-daily trading time from 9:35 to 16:00 EST of day-$\tau$. $\Sigma_{ff}(v)$ and $\Sigma_{sf}(v)$ are the intra-daily instantaneous variance of the efficient futures returns and the covariance between the efficient spot returns and the efficient futures returns, respectively. I initiate the sampling of intra-daily prices at 9:35 because the spot market in the first few minutes after the opening seems to be in a process of discovering the efficient price relative to the futures price. Since my interest is in the hedge ratio between the efficient prices of the spot and futures, I exclude the first five minutes from the sample. Including the data from 9:30 to 9:34 does not change the subsequent conclusion, however. The sampling ends at 16:00, the closing time of the spot market. Therefore, I have $n = 385$ data points for each day.

I estimate the hedge ratio using only daily data as a benchmark. For $\tau = 23, 24, \ldots, 2332$, let $R_f(\tau)$ and $R_s(\tau)$ be the day-$\tau$ open-to-close returns of the futures and spot contracts, respectively. Then, the 22-day rolling-window OLS estimates of the hedge ratios are given by

$$\bar{\beta}_{OLS}(\tau) = \left( \sum_{j=1}^{22} R_f(\tau - j)^2 \right)^{-1} \sum_{j=1}^{22} R_f(\tau - j) R_s(\tau - j)$$

\footnote{Special thanks to Pierre Perron for kindly providing the latter dataset. They can be purchased at \url{http://www.grainmarketresearch.com} \url{http://people.bu.edu/sikeda}}

22
Next, I estimate the integrated covariance matrix day by day using the vector of the one-minute spot returns and the one-minute futures returns. I use the following three estimators:

- the realized covariance using the five-minute returns, which is widely used in the literature;
- the realized kernel \( RK(gn^{3/5}) \); and
- the positive semi-definite version of the two scale realized kernel \( TSRK^{+}(n^{1/2}, cn^{1/2}) \).

The proportionality factors of kernel-based estimators, \( f \) and \( c \), are selected according to the rules given by (30) and (31). Next, I construct a raw estimate of the hedge ratio \( \tilde{\beta}(\tau) \) in (32) by the ratio of the corresponding elements of the estimated matrix, denoted by \( \tilde{\beta}_x(\tau) \) where \( x = RCOV, RK \) and \( TS \) stand for the realized covariance, the realized kernel and the two scale realized kernel. To compare these estimates with the previous daily benchmark, I take the 22-day rolling-window averages of these raw estimates of the hedge ratios: for \( \tau = 23, 24, \ldots, 2332 \),

\[
\tilde{\beta}_x(\tau) = \frac{1}{22} \sum_{j=1}^{22} \tilde{\beta}_x(\tau - j),
\]

Figure 2 shows these four trajectories, \( \tilde{\beta}_x(\tau) \), \( x = RCOV, RK, TS \) and OLS.

Three features emerge from the figure.

First, the trajectory based on my method stays around the same level as the trajectory using daily data. This suggests that my estimator successfully maintains the correct level because the trajectory using only daily data should be relatively free of bias due to any intra-daily complications such as the microstructure noise. On the other hand, the trajectories based on the realized covariance and the realized kernel are well below the other two trajectories, suggesting the strong downward biases of these two estimators.

Second, the trajectory based on my method seems less volatile than the one based on OLS method. This is not surprising because the former uses more data than the latter. However, such comparison with the OLS-based trajectory is not possible for the other two trajectories because of the large differences in levels.

Third, the trajectories based on the realized covariance and the realized kernel increase gradually and have a jump in the middle of the sample. These patterns seem to be matched with the change of the contemporaneous correlation between the spot returns and the futures returns over time, as shown in Figure 3. On the other hand, the trajectory based on my estimator does not exhibit such a pattern. If these patterns were generated by the noise, it would suggest the robustness of my estimator to dependence structure of the noise.
The next table summarizes the relative bias and relative root mean squared error of these smoothed versions $\bar{\beta}_x(\tau)$, based on the OLS, the realized covariance (RCOV), the realized kernel (RK) and the two scale realized kernel (TS). All numbers are assessed relative to the unconditional time series average of the OLS-based estimates. Therefore, this exercise automatically gives a great advantage to the OLS method. In fact, the realized covariance and the realized kernel are beaten by the OLS method. However, my method delivers a smaller relative root mean squared error than the OLS method.

<table>
<thead>
<tr>
<th>%</th>
<th>OLS</th>
<th>RCOV</th>
<th>RK</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.00</td>
<td>-26.74</td>
<td>-9.93</td>
<td>-1.26</td>
</tr>
<tr>
<td>Rmse</td>
<td>4.64</td>
<td>32.35</td>
<td>13.76</td>
<td>2.65</td>
</tr>
</tbody>
</table>

Table: Bias and Root Mean Squared Error (Rmse) relative to the unconditional time-series average of OLS-based estimates.

This table also suggests the relative advantage of my estimator over the other estimators. In particular, my estimator achieves a larger than 80% reduction in bias compared with the realized kernel estimator: $100 \times (-9.93 - (-1.26))/(-9.93) \approx 87%$.

We should be very careful in interpreting these results because of the approximate nature of the daily benchmark and the lack of inference procedure for these rolling-window smoothed versions. However, this exercise suggests that the bias correction in the realized kernel framework can provide a notably large statistical gain in estimating a non-linear function of the integrated covariance matrix.

7 Conclusion

This paper proposed a bias correction for the multivariate realized kernel estimator of the integrated covariance matrix of multiple security returns in the presence of market microstructure noise with unknown serial dependence. The bias correction is indispensable for a clean estimation of non-linear functions of the integrated covariance matrix such as the hedge ratio. The proposed estimator is given by a particular linear combination of two distinctive realized kernel estimators. The estimator has many desirable asymptotic properties: consistency, asymptotic normality, the correctly centered asymptotic distribution and tractable asymptotic variance, the fastest convergence rate, and all mentioned properties hold for non-linear functions of it given an easy-to-implement transformation to guarantee the positive semi-definiteness of it in finite samples. According to simulation studies, the estimates of these non-linear functions based on existing approaches are flawed by a serious downward bias while those based on my approach are more accurate with relatively smaller bias, often leading to smaller root mean squared errors as well. In an application to an actual minute-by-minute dataset of S&P500 index spot and futures prices, I found that the theoretical
prediction also applies in practice as well: compared with the realized kernel estimator, my method delivers a larger than 80% reduction in bias measured from a daily benchmark.

There are several future directions. The hedge ratio analyzed in this paper is just one of possibly numerous examples of non-linear functions of the integrated covariance. Among those, the optimal portfolio allocation and the option price are particularly important quantities for financial risk management. The impact of bias in estimates of the integrated covariance matrix on the performance of the associated risk management is an urgent topic to be explored. Another interesting direction is to consider an application of $TS\Omega$ estimator of the long-run covariance matrix of the noise. It is potentially useful for a non-parametric analysis of the market microstructure. For example, a specification testing of a particular microstructure model may be possible.

One major theoretical issue I did not cover is the non-synchronous trading. Multiple security prices are recorded irregularly without synchronization. In Assumption 1, I assumed that the researchers have already applied a sort of the synchronization of the data, such as the refresh time sampling of BNHLS (2008b). However, some intra-daily data lack the precise time stamps of the transaction. It is not trivial how to synchronize the data in this case. Another issue I did not cover is the interval estimation. Since I derived the explicit asymptotic variance of the estimator, an extensive analysis of the accuracy of inference should be conducted. Finally, it is theoretically as well as practically important to find an objective way to pin-down the growth rate of the first bandwidth $G_n$. I plan to visit these issues soon in separate papers.
References


Barndorff-Nielsen, O.E. and N. Shephard (2004). Econometric analysis of realised covari-
ation: high frequency based covariance, regression and correlation in financial economics. 
Econometrica 72, 885-925.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2006). Subsampling Re-
alized Kernels. SSRN.

http://www.finance.ox.ac.uk/file_links/finecon_papers/2006fe06.pdf

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2008a). Designing realized 
kernels to measure the ex-post variation of equity prices in the presence of noise, forthcoming, 
Econometrica.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2008b). Multivariate Re-
alized Kernels: consistent positive semi-definite estimators of the covariation of equity prices 
with noise and non-synchronous trading. Unpublished paper. SSRN.


Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2008c). Realized Kernels 
in Practice. SSRN.


Biometrika 56, p 375-390.


Econometric Reviews 27, 79-111.


http://people.bu.edu/sikeda/tsrk.pdf


[http://personal.lse.ac.uk/KALNINA/jmp.pdf](http://personal.lse.ac.uk/KALNINA/jmp.pdf)


### Table 1. Relative Bias and Relative Root Mean Squared Error (Rmse).

Noise-to-signal ratio = .005 for IID and IID noises.

- $\int \Sigma_{ij}$: $(i,j)$ element of the integrated covariance matrix.
- $\beta_{ij}$: the high frequency regression of $j$-th asset’s returns onto $i$-th asset’s returns.
- $\rho_{ij}$: the high frequency correlation between the two assets’ returns.
- **SUB**: the estimates based on the subsampling-based realized covariance, a generalization of the “second-best” estimator of Zhang et. al. (2005).
- **RK**: the estimates based on the realized kernel of Barndorff-Nielsen et. al. (2008b).
- $TS^g$: the estimates based on my $TSRK^+(G_n, H_n)$ with $H_n = c^* n^{1/2}$ and $G_n = r^g$, $g = 1/5, 1/4, 1/3, 1/2$. See Section 2.6 in the main text for the rule to select $c^*$.
- All numbers are in percentages of the true values.

<table>
<thead>
<tr>
<th></th>
<th>Bias</th>
<th>SUB</th>
<th>RK</th>
<th>$TS^{1/5}$</th>
<th>$TS^{1/4}$</th>
<th>$TS^{1/3}$</th>
<th>$TS^{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\int \Sigma_{11}$</td>
<td>20.91</td>
<td>6.88</td>
<td>1.39</td>
<td>1.15</td>
<td>0.79</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>$\int \Sigma_{12}$</td>
<td>21.02</td>
<td>6.89</td>
<td>1.45</td>
<td>2.12</td>
<td>0.85</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>$\int \Sigma_{12}$</td>
<td>-4.13</td>
<td>-1.51</td>
<td>-0.68</td>
<td>-0.68</td>
<td>-0.70</td>
<td>-0.83</td>
<td></td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-21.30</td>
<td>-7.86</td>
<td>-1.95</td>
<td>-1.71</td>
<td>-1.35</td>
<td>-0.96</td>
<td></td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>-21.35</td>
<td>-7.83</td>
<td>-1.97</td>
<td>-1.73</td>
<td>-1.36</td>
<td>-1.19</td>
<td></td>
</tr>
<tr>
<td>$\rho_{12}$</td>
<td>-21.48</td>
<td>-8.12</td>
<td>-2.34</td>
<td>-2.11</td>
<td>-1.78</td>
<td>-1.68</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Rmse</th>
<th>SUB</th>
<th>RK</th>
<th>$TS^{1/5}$</th>
<th>$TS^{1/4}$</th>
<th>$TS^{1/3}$</th>
<th>$TS^{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\int \Sigma_{11}$</td>
<td>30.69</td>
<td>24.07</td>
<td>25.62</td>
<td>25.90</td>
<td>26.94</td>
<td>33.15</td>
<td></td>
</tr>
<tr>
<td>$\int \Sigma_{22}$</td>
<td>30.90</td>
<td>24.31</td>
<td>25.80</td>
<td>26.08</td>
<td>27.10</td>
<td>33.07</td>
<td></td>
</tr>
<tr>
<td>$\int \Sigma_{12}$</td>
<td>23.67</td>
<td>22.99</td>
<td>26.54</td>
<td>26.85</td>
<td>27.94</td>
<td>34.48</td>
<td></td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>22.72</td>
<td>11.46</td>
<td>9.95</td>
<td>10.05</td>
<td>10.45</td>
<td>12.38</td>
<td></td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>22.74</td>
<td>11.41</td>
<td>9.93</td>
<td>10.03</td>
<td>10.44</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td>$\rho_{12}$</td>
<td>22.33</td>
<td>9.23</td>
<td>5.21</td>
<td>5.15</td>
<td>5.18</td>
<td>6.04</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Relative Bias and Relative Root Mean Squared Error (Rmse).

Noise-to-signal ratio = .005 for AR and MA noises.

- $\int \Sigma_{ij}$: $(i,j)$ element of the integrated covariance matrix.
- $\beta_{ij}$: the high frequency regression of $j$-th asset’s returns onto $i$-th asset’s returns.
- $\rho_{ij}$: the high frequency correlation between the two assets’ returns.
- $\text{SUB}$: the estimates based on the subsampling-based realized covariance, a generalization of the “second-best” estimator of Zhang et. al. (2005).
- $\text{RK}$: the estimates based on the realized kernel of Barndorff-Nielsen et. al. (2008b).
- $TS^g$: the estimates based on my $TSRK^+(G_n, H_n)$ with $H_n = c^* n^{1/2}$ and $G_n = n^g$, $g = 1/5, 1/4, 1/3, 1/2$. See Section 2.6 in the main text for the rule to select $c^*$.
- All numbers are in percentages of the true values.

<table>
<thead>
<tr>
<th>Bias</th>
<th>$\text{SUB}$</th>
<th>$\text{RK}$</th>
<th>$TS^{1/5}$</th>
<th>$TS^{1/4}$</th>
<th>$TS^{1/3}$</th>
<th>$TS^{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\int \Sigma_{11}$</td>
<td>138.44</td>
<td>10.40</td>
<td>1.64</td>
<td>1.25</td>
<td>1.83</td>
<td>1.98</td>
</tr>
<tr>
<td>$\int \Sigma_{22}$</td>
<td>9.82</td>
<td>5.92</td>
<td>1.72</td>
<td>1.27</td>
<td>0.60</td>
<td>-0.15</td>
</tr>
<tr>
<td>$\int \Sigma_{12}$</td>
<td>-4.19</td>
<td>-1.09</td>
<td>-0.69</td>
<td>-0.70</td>
<td>-0.72</td>
<td>-0.92</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-59.97</td>
<td>-10.47</td>
<td>-2.08</td>
<td>-1.67</td>
<td>-2.28</td>
<td>-2.69</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>-13.14</td>
<td>-6.65</td>
<td>-2.30</td>
<td>-1.84</td>
<td>-1.16</td>
<td>-0.64</td>
</tr>
<tr>
<td>$\rho_{12}$</td>
<td>-41.25</td>
<td>-8.86</td>
<td>-2.58</td>
<td>-2.16</td>
<td>-2.16</td>
<td>-2.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rmse</th>
<th>$\text{SUB}$</th>
<th>$\text{RK}$</th>
<th>$TS^{1/5}$</th>
<th>$TS^{1/4}$</th>
<th>$TS^{1/3}$</th>
<th>$TS^{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\int \Sigma_{11}$</td>
<td>141.99</td>
<td>25.79</td>
<td>26.07</td>
<td>26.34</td>
<td>27.43</td>
<td>33.99</td>
</tr>
<tr>
<td>$\int \Sigma_{22}$</td>
<td>24.60</td>
<td>23.84</td>
<td>25.67</td>
<td>25.94</td>
<td>26.94</td>
<td>33.16</td>
</tr>
<tr>
<td>$\int \Sigma_{12}$</td>
<td>23.73</td>
<td>23.38</td>
<td>26.54</td>
<td>26.85</td>
<td>27.96</td>
<td>34.79</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>60.41</td>
<td>13.62</td>
<td>10.50</td>
<td>10.59</td>
<td>11.02</td>
<td>13.07</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>15.51</td>
<td>10.79</td>
<td>10.08</td>
<td>10.16</td>
<td>10.58</td>
<td>12.89</td>
</tr>
<tr>
<td>$\rho_{12}$</td>
<td>41.83</td>
<td>10.11</td>
<td>5.59</td>
<td>5.44</td>
<td>5.63</td>
<td>6.66</td>
</tr>
<tr>
<td></td>
<td>SUB</td>
<td>RK</td>
<td>TS(^{1/5})</td>
<td>TS(^{1/4})</td>
<td>TS(^{1/3})</td>
<td>TS(^{1/2})</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>----</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>(\int \Sigma_{11})</td>
<td>20.91</td>
<td>6.87</td>
<td>1.39</td>
<td>1.16</td>
<td>0.80</td>
<td>0.28</td>
</tr>
<tr>
<td>(\int \Sigma_{22})</td>
<td>21.03</td>
<td>6.92</td>
<td>1.48</td>
<td>1.24</td>
<td>0.88</td>
<td>0.51</td>
</tr>
<tr>
<td>(\int \Sigma_{12})</td>
<td>-16.64</td>
<td>-5.60</td>
<td>-1.67</td>
<td>-1.56</td>
<td>-1.41</td>
<td>-1.37</td>
</tr>
<tr>
<td>(\beta_{12})</td>
<td>-31.94</td>
<td>-11.69</td>
<td>-2.90</td>
<td>-2.54</td>
<td>-2.00</td>
<td>-1.46</td>
</tr>
<tr>
<td>(\beta_{21})</td>
<td>-32.00</td>
<td>-11.69</td>
<td>-2.95</td>
<td>-2.58</td>
<td>-2.04</td>
<td>-1.71</td>
</tr>
<tr>
<td>(\rho_{12})</td>
<td>-32.11</td>
<td>-11.98</td>
<td>-3.33</td>
<td>-2.98</td>
<td>-2.48</td>
<td>-2.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SUB</th>
<th>RK</th>
<th>TS(^{1/5})</th>
<th>TS(^{1/4})</th>
<th>TS(^{1/3})</th>
<th>TS(^{1/2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\int \Sigma_{11})</td>
<td>30.70</td>
<td>24.07</td>
<td>25.62</td>
<td>25.91</td>
<td>26.95</td>
<td>33.16</td>
</tr>
<tr>
<td>(\int \Sigma_{22})</td>
<td>30.93</td>
<td>24.33</td>
<td>25.79</td>
<td>26.08</td>
<td>27.09</td>
<td>33.05</td>
</tr>
<tr>
<td>(\int \Sigma_{12})</td>
<td>28.55</td>
<td>23.04</td>
<td>26.43</td>
<td>26.74</td>
<td>27.82</td>
<td>34.38</td>
</tr>
<tr>
<td>(\beta_{12})</td>
<td>33.23</td>
<td>14.64</td>
<td>10.71</td>
<td>10.77</td>
<td>11.11</td>
<td>12.87</td>
</tr>
<tr>
<td>(\beta_{21})</td>
<td>33.25</td>
<td>14.60</td>
<td>10.68</td>
<td>10.74</td>
<td>11.10</td>
<td>13.03</td>
</tr>
<tr>
<td>(\rho_{12})</td>
<td>33.08</td>
<td>13.01</td>
<td>6.21</td>
<td>6.05</td>
<td>5.97</td>
<td>6.63</td>
</tr>
</tbody>
</table>

Table 3. Relative Bias and Relative Root Mean Squared Error (Rmse). Noise-to-signal ratio = .005 for IID noises, negative cross-dependence.

- \(\int \Sigma_{ij}\): \((i, j)\) element of the integrated covariance matrix.
- \(\beta_{ij}\): the high frequency regression of \(j\)-th asset’s returns onto \(i\)-th asset’s returns.
- \(\rho_{ij}\): the high frequency correlation between the two assets’ returns.
- \(SUB\): the estimates based on the subsampling-based realized covariance, a generalization of the “second-best” estimator of Zhang et. al. (2005).
- \(RK\): the estimates based on the realized kernel of Barndorff-Nielsen et. al. (2008b).
- \(TS^g\): the estimates based on my TSRK\(^+\)(\(G_n, H_n\)) with \(H_n = c^* n^{1/2}\) and \(G_n = n^g\), \(g = 1/5, 1/4, 1/3, 1/2\). See Section 2.6 in the main text for the rule to select \(c^*\).
- All numbers are in percentages of the true values.
Figure 1: $\Phi_{11}^{(i)}(r)$ functions for $r = 0, .01, \ldots, .99$, $i = 0, 1, 2$. $\Phi_{11}^{(i)}(r)$ are the multiplicative correction factors for the asymptotic covariance matrix of $TSRK(G,H)$ estimator, where $r = G/H$ is the ratio of two bandwidths. The upward-sloping curve is the $\Phi_{11}^{(0)}(r)$ associated with the discretization error. The thin curve at the bottom is the $\Phi_{11}^{(1)}(r)$ associated with the cross product, and the dashed curve is the $\Phi_{11}^{(2)}(r)$ associated with the microstructure noise. We can see that $\Phi_{11}^{(0)}(0) = 1$ and $\Phi_{11}^{(i)}(r) \geq 1$. All curves are based on the Parzen window. Below is the definition of these functions:

\[
\begin{align*}
\Phi_{11}^{(0)}(r) & = \frac{1 - 2r^2\delta^{(00)}(r) + r^5}{1 - 2r^2 + r^4}, \\
\Phi_{11}^{(1)}(r) & = \frac{1 - 2r\delta^{(11)}(r) + r^3}{1 - 2r^2 + r^4}, \\
\Phi_{11}^{(2)}(r) & = \frac{1 - 2\delta^{(22)}(r) + r}{1 - 2r^2 + r^4}.
\end{align*}
\]
Figure 2: Dynamic Hedge Ratio Estimates: Rolling-Window Smoothed Trajectories of Intra-Daily Estimates and Rolling-OLS Estimates using daily data. The wild trajectory around the $\beta = 1$ line is the OLS based trajectory using only open-to-close returns. The thick stable line around it is the trajectory based on my $TSRK^+(n^{1/2}, cn^{1/2})$ estimator. We also have the realized kernel-based trajectory and 5 minute realized covariance-based trajectory. The last two trajectories are systematically below the first two trajectories, suggesting the serious downward bias because the open-to-close returns are expected to be free of bias due to the intra-daily microstructure noise. Notice that even after the transition point of 2000 to 2001, the the trajectory based on my estimator is consistently above the realized kernel-based trajectory.
Figure 3: Time series plots of the cross autocorrelations at each lag of the futures returns with the current spot returns. The number $k$ on top of each panel means that it is the time series plot of the cross autocorrelation of the current spot returns and the futures returns at the $k$-th lag. The futures returns clearly lead the spot returns while no evidence for the opposite case. The jump in the middle of the sample corresponds to the transition from the year 2000 to 2001. These particular patterns of the cross autocorrelations explain the gradual increase and the jump in the trajectories of the dynamic hedge ratios based on the realized covariance and the realized kernels. In particular, notice that the shape of the contemporaneous cross correlation, $k = 0$, is very similar to the trajectories of the hedge ratio estimates based on the realized covariance and the realized kernel in Figure 2.
A Analytical Appendix

A.1 Proof of Lemma

For two \( \mathbb{R}^d \)-valued discrete-time processes \( A = (A_t)_{i=0,1,...,n} \) and \( B = (B_t)_{i=0,1,...,n} \), the realized kernel of the product of the \( A \) and \( B \) is defined by

\[
RK_{AB}(H_n) = \sum_{h=-(n-1)}^{n-1} k(h/(H_n + 1)) \Gamma_{AB}(h)
\]

where \( k \in K \) in Definition

\[
\Gamma_{AB}(h) := \begin{cases} 
\sum_{i=h+1}^{n} (A_{t_i} - A_{t_{i-1}}) (B_{t_{i-h}} - B_{t_{i-h-1}})' & h \geq 0 \\
\sum_{i=h+1}^{n} (A_{t_{i+h}} - A_{t_{i+h-1}}) (B_{t_i} - B_{t_{i-1}})' & h < 0
\end{cases}
\]

Using this notation, the realized kernel based on the observed log-price vectors \( X = (\ln P_t) \) is represented by

\[
RK_{XX}(H_n) = RK_{YY}(H_n) + RK_{UU}(H_n) + RK_{UY}(H_n) + RK_{UY}(H_n)
\]

where \( Y = (\ln P_t^*)_{i=0,1,...,n} \) is the efficient log price vectors and \( U = (U_t)_{i=0,1,...,n} \) is the market microstructure noise.

In the following, I divide the proof into five steps. The first step reminds the definition of the end-point jittering. The second step is to reconfirm the asymptotic normality of \( RK_{YY}(H_n) \) established by BNHLS (2008b). The third step derives the asymptotic variance and the limiting distribution of the \( RK_{UU}(H_n) \). The fourth step establishes the asymptotic results for \( RK_{UY}(H_n) + RK_{UY}(H_n) \). Everything is combined in the fifth step.

Step 1: End-of-the-sample Jittering

Following BNHLS (2008b), I introduce the jittering of end points as follows: given the raw observations \( \{\ln P_{t_j}\}_{j=1,...,n-1+2m_n} \), set \( \ln P_{t_0} = \ln P_{t_{n-1+2m_n}} \) and \( \ln P_{t_n} = \ln P_{t_{n-1+2m_n}} \) for \( j = 1, 2, \ldots, n-1 \),

\[
\ln P_{t_0} = m_n^{-1} \sum_{j=1}^{m_n} \ln P_{t_j} \quad \text{and} \quad \ln P_{t_n} = m_n^{-1} \sum_{j=1}^{m_n} \ln P_{t_{n-m+n+j}}.
\]

Therefore, we take the average of the first and last \( m_n \) raw observations to define the initial observation \( \ln P_{t_0} \) and the last observation \( \ln P_{t_n} \), respectively. This device is necessary to obtain the consistency of the estimator because it is based on the sample autocovariation matrices without scaling \( 1/n \).
Step 2: Asymptotic Normality of $RK_{YY}(H_n)$

This is basically given by BNHLS (2008b, Lemma A.1) as follows:

$$\sqrt{\frac{n}{H_n}} \left( RK_{YY}(H_n) - \int_0^1 \Sigma(v)dv \right) \xrightarrow{d} MN \left( 0, 4k^{(00)} \int_0^1 \Sigma(v) \otimes \Sigma(v)dv \right)$$

where I used the assumption of the equally spaced and synchronized observations $t_i - t_{i-1} = 1/n > 0$ so that the asymptotic covariance matrix is a multi-variate generalization of the so called integrated quarticity. Here, $\xrightarrow{d}$ should be understood as the $\sigma(\ln P^*)$-stable convergence in law where $\sigma(\ln P^*)$ is the information set generated by the whole history of the efficient log prices by the end of the day: see BNHLS (2008a, Appendix 1). This result is deduced from

$$RK_{YY}(H_n) = E[RK_{YY}(H_n)] + \{Var[RK_{YY}(H_n)]\}^{1/2} \cdot \{Var[RK_{YY}(H_n)]\}^{-1/2} \{RK_{YY}(H_n) - E[RK_{YY}(H_n)]\}$$

$$= \int_0^1 \Sigma(v)dv + \left\{ n^{-1}H_n \cdot 4k^{(00)} \int_0^1 \Sigma(v) \otimes \Sigma(v)dv \right\}^{1/2} \cdot Z_{disc} + O_p(n^{-1/2}) + o_p(n^{-1/2}H_n^{1/2})$$

where $O_p(n^{-1/2})$ is the rate derived for the realized covariance estimator in the noise-free environment by Barndorff-Nielsen and Shephard (2004), which is clearly dominated by $o_p(n^{-1/2}H_n^{1/2})$ as long as $H_n$ is proportional to $n^h$, $h > 0$. $Z_{disc}$ is a matrix-valued standard normal variate generated by the discretization error. Note that the $\sigma(\ln P^*)$-stable convergence in law implies that this $Z_{disc}$ and the Brownian motion in Definition are independent: see Jacod (2008, 3.10) and Mykland and Zhang (2006, Proposition 1). Note also that since there is no serial dependence in the efficient returns, this asymptotic result does not have anything regarding the endpoints of the sample.

Step 3: Asymptotic Normality of $RK_{UU}(H_n)$

As previously,

$$RK_{UU}(H_n) = E[RK_{UU}(H_n)] + \{Var[RK_{UU}(H_n)]\}^{1/2} \cdot \{Var[RK_{UU}(H_n)]\}^{-1/2} \{RK_{UU}(H_n) - E[RK_{UU}(H_n)]\}$$

Therefore, I need to evaluate the mean, variance and the orders of remainder terms.

Lemma 2 For a large sample size $n$,

$$E[RK_{UU}(H_n)] = |k^{(2)}(0)| nH_n^{-2} \Omega + o(nH_n^{-5/2}) + O(m_n^{-1})$$

$$Var[RK_{UU}(H_n)] = nH_n^{-3} \cdot 4k^{(22)} \Omega \otimes \Omega + o(nH_n^{-3}) + O(H_n^{-1}m_n^{-1})$$

where $k^{(22)} = \int_0^\infty (k^{(2)}(x))^2 dx$, $\Omega$ is the long-run variance covariance matrix of $U$, $Var(\cdot | \ln P^*)$ is the ex-post variance given the whole history of the trajectory of efficient log prices, and $m_n$ is the number of observations in the ends of the samples for the end-point jittering as introduced above.
In this case, 

\[ \gamma \int_0^\infty \exists \text{ of the autocovariance matrices of} \]

ing exponentially fast. This in turn implies the exponentially fast decay of the absolute values

and

where

Let me first discuss about \( \delta \) and \( \epsilon \) and for some \( a_0 \)

\[ \text{Next, let me evaluate} \]

(Proof) The key expressions in BNHLS (2008b, (A.3) and A.5) together imply

\[ RK_{UU}(H_n) = \gamma_n + \delta_n \]

and for some \( \epsilon_{H_n} \in (0, (H_n + 1)^{-1}) \),

\[ a_h = \left\{ \begin{array}{ll}
- \left( \frac{k(h+1)/(H_n+1) - k(h)/(H_n+1)}{(H_n+1)^{-1}} \right) & h = 0 \\
\frac{-k(2)\epsilon_{H_n}}{(H_n+1)^{-1}} - \frac{k(h)/(H_n+1) - k((h-1)/(H_n+1))}{(H_n+1)^{-1}} & h \neq 0 
\end{array} \right. \]

and \( \delta_n \) is a collection of terms depending on the end points of the sample, given by

\[ \delta_n = \frac{1}{2} \left\{ Z_0 - k \left( \frac{n-1}{H_n+1} \right) Z_n \right\} + \sum_{h=1}^{n-1} \left\{ k \left( \frac{h}{H_n+1} \right) - k \left( \frac{h-1}{H_n+1} \right) \right\} Z_h, \]

\[ Z_h = U_0 U_{h}^T + U_h U_{0}^T + U_n U_{n-h}^T + U_{n-h} U_{n}^T. \]

Let me first discuss about \( \delta_n \). Under the conditions that \( \sum_{h \in \mathbb{Z}} |h| \Omega(h) \ll \infty \) and \( \int_0^\infty |k'(x)|^2 dx < \infty, E[\delta_n] \sim O(m^{-1}) \) follow by BNHLS (2008b, Lemma A.4).

Next, let me evaluate \( \gamma_n \), the main part of \( RK_{UU}(H_n) \). Notice that \( a_h \sim -k^{(2)}(x_h) \), \( \exists x_h \in [(h-1)(H_n+1)^{-1}, (h+1)(H_n+1)^{-1}] \). Let me define

\[ \beta(x) = k^{(2)}(x)/k^{(2)}(0) \]

and \( \beta_h := \beta(h/H_n) \). Note that \( a_h/a_0 \sim \beta(x_h) \) and \( \sum_{h=-\infty}^{n-1} \beta(x_h)^2 H^{-1} \rightarrow \int_{-\infty}^\infty |\beta(x)|^2 dx \).

In this case, \( \gamma_n \) is further decomposed into three parts:

\[ \gamma_n = \gamma_n(1) + \gamma_n(2) \]

where

\[ \gamma_n(1) = nH_n^{-2}a_0 \left\{ \sum_{i=1}^{n-2} U_i U_i^T + \sum_{h=1}^{\sqrt{n}} \beta(x_h) \left( n-1 \sum_{i=h+1}^{n-2} U_i U_{i-h}^T + n-1 \sum_{i=h+1}^{n-2} U_{i-h} U_i^T \right) \right\} \]

and

\[ \gamma_n(2) = nH_n^{-2}a_0 \sum_{h>\sqrt{n}} \beta(x_h) \left( n-1 \sum_{i=h+1}^{n-2} U_i U_{i-h}^T + n-1 \sum_{i=h+1}^{n-2} U_{i-h} U_i^T \right) \]

Let me evaluate \( \gamma_n(2) \) at first. Recall \( U_i \) is \( \alpha \)-mixing with the \( \alpha \)-mixing coefficients decaying exponentially fast. This in turn implies the exponentially fast decay of the absolute values of the autocovariance matrices of \( U_i \); see Davidson (2002, Corollary 14.3). In particular, for
\( \rho \in (0, 1) \),
\[
\Omega(h) = C \rho^h = C \exp (\ln \rho \cdot h) \to 0 \quad \text{as} \quad h \to \infty
\]
because \( \ln \rho < 0 \). In this case, as \( n \to \infty \),
\[
E[\| \gamma_n(2) \|] \leq \int_{-\infty}^{\infty} |\beta(x)| dx \cdot n H_n^{-1} \sum_{|h| > \sqrt{H_n}} |k(2)(h/N_n)| H_n^{-1} \cdot \sup_{h > \sqrt{H_n}} \| \Omega(h) \| \to 0
\]
extponentially fast.

Secondly, the small order term \( o(nH_n^{-5/2}) \) in Lemma 2 is due to the approximation error of \( a_h \) by \(-k(2)(0)\). The corresponding rate is given by \( o(nH_n^{-2}) \) in BNHLS (2008b, Lemma A.5), but it is actually faster than that. For instance, the Parzen window \( k(x) \) has a finite Holder coefficient
\[
\sup_{|x|, |y| < 1/2} |(x - y)^{-1}(k(2)(x) - k(2)(y))| < \infty.
\]
For \( H_n > 4 \), \( \sqrt{H_n} < H_n/2 \). In addition, given the continuity of \( k(2)(x) \), \( a_h = k(2)(x_h) \) for some \( x_h \in \left( \begin{array}{c}
\frac{|h|-1}{H_n+1}, \frac{|h|+1}{H_n+1}
\end{array} \right) \) and \( 1 < |h| < \sqrt{H_n} \) by the intermediate value theorem. Therefore, given \( |h| < \sqrt{H_n} \),
\[
a_h(1 - (|h| + 1)/n) + k(2)(0) = -k(2)(x_h) + k(2)(0) + O(H_n^{1/2}n^{-1})
\]
This means
\[
\sqrt{H_n} \cdot \sup_{|h| \leq \sqrt{H_n}} |a_h(1 - (|h| + 1)/n) + k(2)(0)| \leq \sup_{|h| \leq \sqrt{H_n}} \left| \frac{k(2)(x_h) - k(2)(0)}{1/H_n} \right| \frac{1}{\sqrt{H_n}} + O(H_n n^{-1}) \to 0
\]
given \( H_n \) is proportional to \( n^{\beta} \), \( \beta < 1 \) because then \( O(H_n n^{-1}) \sim o(1) \). Therefore, \( a_h \sim -k(2)(0) + o(H_n^{-1/2}) \). In conjunction with \( nH_n^{-2} \) in \( \gamma_n(1) \) and \( \gamma_n(2) \), we have \( o(nH_n^{-5/2}) \). This means that \( E[\gamma_n(2)] \) is always dominated by \( E[\gamma_n(1)] \). Combining everything together, I have
\[
E[RK_{UU}(H_n)] = E[\gamma_n(1)] + E[\gamma_n(2)] + E[\delta_n] = |k(2)(0)|nH_n^{-2}\Omega + o(nH_n^{-5/2}) + O(m_n^{-1})
\]
Let me move on to the evaluation of the variance of \( RK_{UU}(H_n) \). Firstly, BNHLS (2008b, A.5) establish the variance of \( \delta_n \) as \( \text{Var}(\delta_n) \sim O(H_n^{-1}m_n^{-1}) \).

Secondly, let \( \alpha(h) \) be the \( h \)-th \( \alpha \)-mixing coefficient of \( U_i \). Then,
\[
\text{Var}[\| \gamma_n(2) \|] \leq n^2H_n^{-4}\alpha_0^2 \sum_{|h|, |i| > \sqrt{H_n} + |j|, |j| > |i|} \beta(x_h) \beta(x_l)n^{-2} \| Cov(U_i U_{i+h}, U_j U_{j+i}) \| \leq C \cdot n^2H_n^{-4}(n - \sqrt{H_n}) \sup_{|h| > \sqrt{H_n}} \alpha^2(h) \to 0
\]

38
exponentially fast.

Thirdly, the variance of $\gamma_n(1)$ is derived as follows: given $k(x) \in K$ in Definition 11 $\beta(x) \in K_1$ in Andrews (1991) because

1. $\beta: \mathbb{R} \to [-1, 1]$ because $|k^{(2)}(x)|$ attains the maximum at zero for typical windows.
2. $\beta(0) = 1$ by construction.
3. $\beta(x) = \beta(-x)$ because $k^{(2)}(x)$ is an even function.
4. $\int_{-\infty}^{\infty} \beta(x)^2 dx \propto \int_{-\infty}^{\infty} k^{(2)}(x)^2 dx < \infty$ given $k^{(22)} < \infty$ and
5. $\beta(x)$ is continuous because $k \in K$ in Definition 11.

Since $U_t$ is $\alpha$-mixing with strict stationarity, the third claim in Hannan (1970, Theorem 9) for the case of $\lambda_1 = \pm \lambda_2 = \lambda = 0$ implies that the asymptotic covariance between $(j, l)$ element and $(j', l')$ element of $\gamma_n(1)$, denoted by $\gamma_{n,jl}(1)$ and $\gamma_{n,j'l'}(1)$ respectively, is given by

$$ Cov\left[\gamma_{n,jl}(1), \gamma_{n,j'l'}(1)\right] = nH_n^{-3} \cdot 2k^{(22)}(\Omega_{jj'}\Omega_{ll'} + \Omega_{jl'}\Omega_{j'l'}) + o(nH_n^{-3}) $$

where $f_{jl}$ is the $(j, l)$ element of the spectral density matrix of $(U_t)$ at zero frequency. Using (i) $2\pi f_{ll}(0) = \Omega$ is the long-run variance-covariance matrix of the noise, (ii) $\int_{-\infty}^{\infty} \beta(x)^2 dx = 2 \int_0^{\infty} \beta(x)^2 dx$, (iii) $\beta(x) = k^{(2)}(x) / k^{(2)}(0)$ and (iv) $k^{(22)} := \int_0^{\infty} [k^{(2)}(x)]^2 dx$, the above expression reduces to

$$ Cov\left[\gamma_{n,jl}(1), \gamma_{n,j'l'}(1)\right] = nH_n^{-3} \cdot 2k^{(22)}(\Omega_{jj'}\Omega_{ll'} + \Omega_{jl'}\Omega_{j'l'}) + o(nH_n^{-3}) $$

On the other hand, given

$$ a = 1_j, b = 1_l, c = 1_{j'}, d = 1_{l'}, $$

the straightforward calculation gives

$$ 2 \cdot v_{ab}(\Omega \otimes \Omega) v_{cd} = \Omega_{jj'}\Omega_{ll'} + \Omega_{jl'}\Omega_{j'l'} $$

where $v_{ab} = vec([ab^t + ba^t]/2)$ as introduced in Definition 3. The right hand side is exactly the expression we had previously. By the convention of the matrix-valued normal variate, this establishes

$$ Var\left[RK_{UU}(H_n)\right] = nH_n^{-3} \cdot 4k^{(22)} \otimes \Omega + o(nH_n^{-3}) + O(H_n^{-1}m_n^{-1}) $$

Q.E.D.
Given this lemma, the asymptotic normality of $RK_{UU}(H_n)$ is derived as follows. Notice that for $\forall \lambda \in [-\pi, \pi]$, the evenness of $k^{(2)}(x)$ implies that

$$K(\lambda) = (2\pi)^{-1} \int_{-\infty}^{\infty} \frac{k^{(2)}(x)}{k^{(2)}(0)} e^{ix\lambda} dx = \pi^{-1} \int_{0}^{1} \frac{k^{(2)}(x)}{k^{(2)}(0)} \cos(\lambda x) dx$$

Therefore,

- $|K(\lambda)| \leq \pi^{-1} \int_{0}^{1} \frac{|k^{(2)}(x)|}{k^{(2)}(0)} \cos(\lambda x) dx < \infty$ because $\int_{0}^{1} |k^{(2)}(x)| dx \leq (\int_{0}^{1} [k^{(2)}(x)]^2 dx)^{1/2} < \infty$ by Cauchy-Schwartz inequality and $k^{(22)} = \int_{0}^{\infty} [k^{(2)}(x)]^2 dx < \infty$ since $k \in \mathcal{K}$.
- $K(\lambda)$ is symmetric about zero because $\cos(x)$ is an even function.
- $K'(\lambda) = \pi^{-1} \int_{0}^{1} \frac{k^{(2)}(x)}{k^{(2)}(0)} (-x) \sin(\lambda x) dx \leq \pi^{-1} \int_{0}^{1} \frac{k^{(2)}(x)}{k^{(2)}(0)} dx < \infty$ given $k^{(22)} < \infty$.
- Finally, $\int_{-\pi}^{\pi} K(\lambda) d\lambda = \pi^{-1} \int_{0}^{1} \frac{k^{(2)}(x)}{k^{(2)}(0)} \sin(\pi x) dx \leq \int_{0}^{1} \frac{k^{(2)}(x)}{k^{(2)}(0)} dx < \infty$ because $\sin(\pi x) / x$ is also even and $\sin(\pi x) / x \in [0, \pi]$, $\forall x \in [0, 1]$. Therefore, I can always normalize $K(\lambda)$ such that $\int_{-\pi}^{\pi} K(\lambda) d\lambda = 1$.

Given strict stationarity of $U_i$ with the existence of all moments, and let $m_n$ be proportional to $n^\mu$ where $\mu > (3h - 1)/2$, Brillinger (1969, Theorem 6.2) implies

$$\text{Var}[RK_{UU}(H_n)]^{-1/2} \cdot \{RK_{UU}(H_n) - E[RK_{UU}(H_n)]\} \xrightarrow{d} N(0, I)$$

where $I$ is the $d^2 \times d^2$ identity matrix. Alternatively, I can rewrite this as

$$RK_{UU}(H_n) = nH_n^{-2} \frac{k^{(2)}(0)}{k^{(2)}(0)} \Omega + nH_n^{-3} \cdot k^{(22)} \Omega \otimes \Omega \right)^{1/2} \mathcal{Z}_{\text{noise}} + o(nH_n^{-5/2}) + O(m_n^{-1}) + o(n^{1/2}H_n^{-3/2}) + O(H_n^{-1/2}m_n^{-1/2})$$

where $\mathcal{Z}_{\text{noise}}$ is the matrix-valued standard normal variate caused by the microstructure noise. It is independent of the Brownian motion in the data generating process.

**Step 4: Asymptotic Normality of $RK_{UY}(H_n) + RK_{UY}(H_n)$**

Clearly $E[RK_{UY}(H_n) + RK_{UY}(H_n)] = 0$.

**Lemma 3** For a large sample size $n$,

$$\text{Var}(RK_{UY}(H_n) + RK_{UY}(H_n)| \ln P^*) = H_n^{-1} \cdot 4k^{(11)} \left\{ \Omega \otimes \int \Sigma(v) dv + \int_{0}^{1} \Sigma(v) dv \otimes \Omega \right\} + O_p(H_n^{-1/2}n^{-1}) + O_p(H_n n^{-1} m_n^{-1}) + o_p(H_n^{-1})$$

where $k^{(22)} = \int_{0}^{\infty} [k^{(2)}(x)]^2 dx$, $\Omega$ is the long-run variance covariance matrix of $U$, $\text{Var}(-| \ln P^*)$ is the ex-post variance given the whole history of the trajectory of efficient log prices, and $m_n$ is the number of observations in the ends of the samples for the end-point jittering as introduced above.
The mean of $\Theta$ is zero. For any $a, b, p, q \in \mathbb{R}^d$, the covariance of $\Theta_n(1)$ is given by

$$Var[a'\Theta_n(1)b \cdot p/\Theta_n(1)q| \ln P^*] = \sum_h k_h^2 \left\{ E \left[ a' E[U_0U_0'| \ln P^*] p \cdot b' y_{h+1} y_{h+1}' | q \cdot b' y_{h+1} y_{h+1}' | P^* \right] + E \left[ b' E[U_0U_0'| \ln P^*] q \cdot a' y_{h+1} y_{h+1}' | P^* \right] \right\} + E \left[ b' E[U_0U_0'| \ln P^*] q \cdot a' y_{h+1} y_{h+1}' | P^* \right] + E \left[ b' E[U_0U_0'| \ln P^*] q \cdot a' y_{h+1} y_{h+1}' | P^* \right]$$

$$\sim O(H_n n^{-1} m_n^{-1})$$
where \( E[U_0 U_0'] \sim O(m_n^{-1}) \) by BNHLS (2008b, Lemma), \( E[y_h y_h'] \leq \Lambda^2 n^{-1} \) for \( \Lambda \) as the uniform upper bound of \( \Sigma^{1/2} \) introduced in Assumption (D\( \Box \)) and the drift component is of lower order \( (E[\int_{t_i-1}^{t_i} b_i dv_i'] [\mathcal{F}_{t_i-1}] \leq (\Lambda n)^{-2} \sim O(n^{-2}) \), and \( \sum_{h=1}^n k_h^2 = \sum_{h=1}^n H_h^2 = O(H_n) \) given \( k^{00} = \int_0^\infty |k(x)|^2 dx < \infty \) as in Definition \( \Box \) Exactly the same argument applies to \( \Theta_n(2) \).

The first and fourth terms of the decomposition of \( RKYU(H_n) \) can be rewritten as

\[
\Theta_{YU}(H_n) := \sum_{h=0}^{n-1} (k_h - k_{h-1}) \sum_{i=0}^n y_i U_{i-h} + \sum_{h=-n}^{-1} (k|h| - k|h|-1) \sum_{i=|h|+1}^n y_i+h U_i' \\
= H_n^{-1} \sum_{h>0}^{n-1} \lambda_h \left\{ \sum_{i=h+1}^n y_i U_{i-h} + \sum_{i=h+1}^n y_{i-1} U_{i-1} \right\}
\]

where \( \lambda_h = \left[ k \left( \frac{h}{H_n+1} \right) - k \left( \frac{h-1}{H_n+1} \right) \right] / (H_n^{-1}) \) is the sample analogue of the first derivative of \( k \). Then, \( \forall a, b \in \mathbb{R}^d \setminus \{0\}, \)

\[
a' \Theta_{YU}(H_n)b \\
\sim H_n^{-1} \left\{ \sum_{h>0}^{n-1} \lambda_h \left( \sum_{i=h+1}^n a'y_i U_{i-h}b + \sum_{i=|h|+1}^n \lambda_h a' U_{i+h}y_i b \right) \right\}
\]

because \( a'U \) and \( y'b \) are scalars. One can write down a similar expression for \( \Theta_{YU}(H_n) := \sum_{h=0}^{n-1} (k_h - k_{h-1}) \sum_{i=h+1}^n U_i y_i' + \sum_{h=-n}^{-1} (k|h| - k|h|-1) \sum_{i=|h|+1}^n U_{i-h}y_i' \) as well. Then,

\[
H_nE \left[ a' \left( \Theta_{YU}(H_n) + \Theta_{YU}(H_n) b \cdot c' \left( \Theta_{YU}(H_n) + \Theta_{YU}(H_n) \right) \right) dY \right] \\
= 2H_n^{-1} \left\{ \sum_{h,l \geq 0} \lambda_h \lambda_l a' \sum_{i=j=|h|+1}^n y_i y_j' c \cdot b' \Omega(l - h) d + \sum_{h,l \geq 0} \lambda_l \lambda_j a' \sum_{i=j=|l|+1}^n y_i y_j' c \cdot b' \Omega(l - h) c \right\}
\]

because terms for \( (h, l) \) such that with \( |h - l| > \sqrt{H_n} \) decays exponentially fast, \( \sum_{j=|h| \vee |l|+1}^n y_j y_j' = \sum_{j=1}^n y_j y_j' + O_p(H_n^{1/2} n^{-1}) \) given \( |h| \) and \( |l| \) are less than \( H_n^{1/2} \) because \( y_j y_j' \sim O_p(n^{-1}) \) (there are at
most $\sqrt{H_n}$ such terms), and

$$H_n^{-1} \sum_{|h-l|<H_n^2} \lambda_h \lambda_l H_n^{-1} \Omega(l-h) = H_n^{-1} \sum_h \lambda_h^2 \sum_{l:|h-l|<H_n^2} \Omega(h-l) + O(H_n^{-1}) = \int_0^\infty [k^{(1)}(x)]^2 dx \cdot \sum_k \Omega(k) + o(1)$$

where $\lambda_h = \lambda_h + O(H_n^{-1})$ and $H_n^{-1} \sum_h \lambda_h^2 = \int_0^\infty [k^{(1)}(x)]^2 dx + o(1)$. Moreover, we applied the following rules:

$$a' \cdot A \cdot b' \cdot B \cdot c' = a' \cdot b' \cdot B \cdot c'$$

where $v_{ab} = vcc((ab' + ba')/2)$- see BNHLS (2008b, Lemma A.1). Therefore,

$$E[a'[RK_{YU}(H_n) + RK_{UY}(H_n)]b \cdot c'[RK_{YU}(H_n) + RK_{UY}(H_n)]] \ln P^*$$

$$\sim v_{ab} \left[H_n^{-1} k^{(1)} \left(\int_0^1 \Sigma(v)dv \otimes \Omega + \Omega \otimes \int_0^1 \Sigma(v)dv \right)\right] v_{cd} + O_p(H_n^{-1/2} n^{-1}) + O_p(H_n^{-1} m_n^{-1}) + o_p(H_n^{-1})$$

\[Q.E.D.\]

The asymptotic normality of $RK_{YU}(H_n) + RK_{UY}(H_n)$ is derived as follows: suppose $d = 1$, i.e. a univariate case. I can represent $\Upsilon_{YU}(H_n)$ as

$$H_n^{-1} \left\{ \int_0^1 \Sigma_{v} dv \right\}^{1/2} \left\{ \sum_{h=0}^{n-1} \lambda_h \sum_{i=h+1}^n \left\{ \int_0^1 \Sigma_{v} dv \right\}^{-1/2} y_i \right\} U_{t-h} + \sum_{h=0}^{n-1} \lambda_h \sum_{i=h+1}^n U_{t+i} \left\{ \int_0^1 \Sigma_{v} dv \right\}^{-1/2} y_i \right\}$$

where $y_i = \ln P_{t-i} - \ln P_{t-1}$. A similar expressoin is obtained for $\Upsilon_{UY}(H_n)$. Following BNHLS (2008a, Proof of Theorem 1), define $\bar{y}_i \equiv \left\{ \int_0^1 \Sigma_{v} dv \right\}^{-1/2} y_i$ and represent

$$\Upsilon_{YU}(H_n) + \Upsilon_{UY}(H_n) = H_n^{-1} \left\{ \int_0^1 \Sigma_{v} dv \right\}^{1/2} \sum_{k=1}^n |\psi_k| \cdot \sum_{j=1}^n \left[ \sum_{k=1}^n \frac{\bar{y}_j}{|\psi_k|} \right] U_j$$

where $\psi_k$ is a linear combination of $\lambda$ and $\bar{y}$ given by

$$\psi_k = \lambda_0 \bar{y}_k + 2 \lambda_1 \bar{y}_{k+1} + \cdots + 2 \lambda_{n-k} \bar{y}_n \quad k = 1, \ldots, n - 1,$$

$$\psi_n = \lambda_0.$$ Conditionally on $\ln P^*$, $\omega_k := \psi_k / \sum_{j=1}^n |\psi_j|$ are constant satisfying $|\omega_k| \leq 1$ and $\sum_{k=1}^n |\omega_k| = 1$. Then,

- $\sum_{k=1}^n |\omega_k| < \text{constant by construction}$,
- $\max\{|\omega_k|, k = 0 \ldots n\} \ln P^* \sim O\left( \sum_{k=1}^n \omega_k^2 \right)$ because $\max\{|\omega_k|, k = 0 \ldots n\} \ln P^* \leq 1$ and $\sum_{k=1}^n \omega_k^2 \leq n$ and therefore $1 \sim O(n)$ trivially.
- Finally, we have $\sum_{k=1}^n \omega_k^2 \ln P^* \sim O(\text{var}\{\sum_{k=1}^n \omega_k U_k \ln P^*\})$ by $\sum_{k=1}^n \omega_k^2 \leq n$ and

$$\text{var} \left( \sum_{k=1}^n \omega_k U_k \right) = \sum_{k=1}^n \omega_k^2 \Omega(0) + 2 \sum_{h=1}^{n-1} (n-h) |\omega_k| |\omega_k-h| \Omega(h) \leq n \sum_{h=-1}^{n-1} |\Omega(h)| \sim O(n)$$
Therefore, (A1), (A2) and (A3) in Roussas et. al. (1992) are satisfied for $\omega_k$ and $U_k$ corresponding to $w_{ni}(x)$ and $\xi_k$ in that article. Suppose $p$ and $q$ in Roussas et. al. (1992) are designed such that $p \propto n^a$, $a < -1/2$ and $q \propto n^b$, $b < -5/2$. Given $U_i \sim$ strictly stationary strong mixing process with the mixing coefficient decaying exponentially fast and $U_0 \sim O_p(1)$ given in Assumption 1, all of the conditions required for Theorem 2.1 in Roussas et. al. (1992) are satisfied. It delivers the asymptotic normality of $\sum_{k=1}^{n} \omega_k U_k$ and therefore of $\Upsilon_{YU}(H) + \Upsilon_{UY}(H)$.

For a multivariate case, take $a'(\Upsilon_{YU}(H) + \Upsilon_{UY}(H))a$, $\forall a \in \mathbb{R}^d$. It reduces to the following expression: $\sum_{k=0}^{n} \zeta_k \cdot a'U_k$, where $\zeta$ is a univariate weight depending on $a, \gamma, \bar{y}_i$. By defining $a'U_k$ as another univariate stationary strong-mixing sequence, the previous argument delivers the asymptotic normality of this quadratic form. Note that the central limit effect in $\Upsilon_{YU}(H) + \Upsilon_{UY}(H)$ is driven by the vector-valued noise process $U_i$ conditional on $\ln P^*$. Therefore, this Cramer-Wold type argument is enough for the multivariate asymptotic normality. The overall asymptotic covariance of $\Upsilon_{YU}(H_n) + \Upsilon_{UY}(H_n)$ has already been established in Lemma 3. Alternatively, one can represent $RK_{YU}(H_n) + RK_{UY}(H_n)$ as follows:

$$RK_{YU}(H_n) + RK_{UY}(H_n) = \left\{ \left( H_n^{-1} \cdot 4k^{(\text{II})} \right) \left( \Omega \otimes \int_0^1 \Sigma(v) dv + \int_0^1 \Sigma(v) dv \otimes \Omega \right) \right\}^{1/2} \cdot Z_{\text{cross}} + O_p(H_n^{-1/2})$$

where $Z_{\text{cross}}$ is the matrix-valued standard normal variate. It is independent of the Brownian motion driving the data generating process of the efficient log prices and the stochastic volatility as for $Z_{\text{disc}}$ derived previously.

**A.1.1 The final expression**

Combining everything together, I obtain the following key characterization of the realized kernel:

$$RK(H_n) \overset{d}{\sim} \int_0^1 \Sigma(v) dv + |k^{(\text{II})}(0)|nH_n^{-2}\Omega + \left\{ H_n^{-1/2}D \right\}^{1/2} \cdot Z_{\text{disc}} + \left\{ nH_n^{-3/2}M \right\}^{1/2} \cdot Z_{\text{noise}} + \left\{ H_n^{-1}C \right\}^{1/2} \cdot Z_{\text{cross}} + \phi_n$$

where $Z$’s are matrix valued standard normal variates induced by the asymptotic normality of each component due to the discretization error ($Z_{\text{disc}}$), the microstructure noise ($Z_{\text{noise}}$) and their cross product ($Z_{\text{cross}}$). Let me emphasize that the symbol $\overset{d}{\sim}$ in this characterization stands for the asymptotic equivalence in distribution, in the sense of the stable convergence in law: see BNHLS (2008a, Appendix A). Since these $Z$’s are independent of the Brownian motion driving the efficient log prices $\ln P^*$ and the stochastic volatility $\Sigma^{1/2}$ (see Jacod (2008, 3.10) and Mykland and Zhang (2006, Proposition 1)), Lemma 1 and Proposition 5 in BNHLS (2008a) allows us to combine these asymptotic covariance components into $Z_{n,H_n}$ in
the following expression:

\[
\begin{align*}
 RK(H_n) & \sim \int_0^1 \Sigma(v) dv + |k^{(2)}(0)| n H_n^{-2} \Omega + Z_{n,G_n} + \phi_n \\
 Z_{n,G_n} & := MN \left( O, H_n n^{-1} \mathcal{D} + n H_n^{-3} \mathcal{M} + H_n^{-1} \mathcal{C} \right), \\
 \mathcal{D} & := 4 k^{(00)} \int_0^1 \Sigma(v) \otimes \Sigma(v) dv, \\
 \mathcal{M} & := 4 k^{(22)} \Omega \otimes \Omega, \\
 \mathcal{C} & := 4 k^{(11)} \left\{ \Omega \otimes \int \Sigma(v) dv + \int_0^1 \Sigma(v) dv \otimes \Omega \right\}, \\
 \phi_n & = o(n H_n^{-5/2}) + O(m_n^{-1}) + O(H_n^{-1/2} m_n^{-1/2}) + O(H_n^{1/2} n^{-1/2} m_n^{-1/2}) \\
 & \quad + o_p(H_n^{1/2} n^{-1/2}) + o(n^{1/2} H_n^{-3/2}) + o_p(H_n^{-1/2}).
\end{align*}
\]

Note that \( o_p(H_n^{-1/4} n^{-1/2}) \) in the asymptotic variance of \( RK_{1U}(H_n) + RK_{UY}(H_n) \) is dominated by \( o_p(H_n^{1/2} n^{-1/2}) \).

Q.E.D.

Later we will see that if \( m_n \) is proportional to \( n^\nu, \nu > 0 \), it does not have any effect on the consistency of my estimator. If \( m_n = n^\nu, \nu > (3h - 1)/2 \) for \( H_n = cn^h \), it does not have any effect on the limiting distribution.

A.2 Proof of Theorem

The proof is divided into five steps. Step 1 derives the joint characterization of two realized kernels. Step 2 establishes the asymptotic covariance matrix of these two realized kernels. Step 3 mentions the asymptotic results for the individual two scale realized kernel estimator. Step 4 derives the asymptotic results for the individual \( TS \Omega \) estimator. Finally, Step 5 derives the joint asymptotic result for these estimators.

Step 1: The Joint Characterization of two Realized Kernels

The following simple lemma is the key to consider a different set of realized kernels in a unified way. Since different realized kernels can be constructed by different kernel windows, say \( k_A, k_B \in K \), I will make it explicit by assigning the subscripts \( A \) and \( B \).

**Lemma 4** *(Joint Asymptotic Characterization of two realized kernels)*

For \( k_A, k_B \in K \), define

\[
\begin{align*}
 RK_A(H_n) & = \sum_{h=-(n-1)}^{n-1} k_A \left( \frac{h}{H_n + 1} \right) \Gamma(h) \\
 RK_B(G_n) & = \sum_{h=-(n-1)}^{n-1} k_B \left( \frac{h}{G_n + 1} \right) \Gamma(h)
\end{align*}
\]
and stack them to form a \((2d \times d)\) matrix:
\[
V := \begin{bmatrix} RK_A(H_n) \\ RK_B(G_n) \end{bmatrix}.
\]

Finally, define the ratio of two bandwidths \( r_n := G_n H_n^{-1} \).

1. The leading terms of the mean of \( V \) are given by
\[
\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \left( \int_0^1 \Sigma(v)dv \right) + \begin{bmatrix} |k_A^{(2)}(0)| n H_n^{-2} \\ |k_B^{(2)}(0)| n G_n^{-2} \end{bmatrix} \otimes \Omega
\]

2. The leading terms of the asymptotic variance covariance matrix of \( V \) are given by the following \((2d^2 \times 2d^2)\) matrix:
\[
\frac{H_n}{n} \cdot P \otimes D_A + \frac{n}{H_n^3} \cdot Q \otimes M_A + \frac{1}{H_n} \cdot R \otimes C_A
\]

where \( P, Q \) and \( R \) are \((2 \times 2)\) symmetric matrix of the following forms:
\[
P = \begin{bmatrix} 1 & \delta^{(00)}(r_n) \\ \bullet & r_n \omega^{(00)} \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & r_n^{-2} \delta^{(22)}(r_n) \\ \bullet & r_n^{-3} \omega^{(22)} \end{bmatrix}, \quad R = \begin{bmatrix} 1 & r_n^{-1} \delta^{(i)}(r_n) \\ \bullet & r_n^{-1} \omega^{(ii)} \end{bmatrix}
\]

where
- \( \omega^{(ii)} = k_B^{(ii)} / k_A^{(ii)} \),
- \( k_{AB}^{(ii)}(r) = \int_0^\infty k_A^{(i)}(x) k_B^{(i)}(x/r) dx \), and
- \( \delta^{(i)}(r_n) = k_{AB}^{(ii)}(r_n) / k_A^{(ii)} \),

3. The numerator of \( \delta^{(i)}(r_n) \) behaves approximately as follows:
\[
\int_0^\infty k_A^{(i)}(x) k_B^{(i)}(x/r_n) dx \sim \begin{cases} r_n k_A^{(i)}(0) \int_0^\infty k_B^{(i)}(x) dx & r_n \to 0 \\ \int_0^\infty k_A(x) k_B(x/r) dx & r_n = r \\ k_B^{(i)}(0) \int_0^\infty k_A^{(i)}(x) dx & r_n \to \infty \end{cases}
\]

where \( r > 0 \) is some constant.

(Proof) 1. is obvious by BNHLS (2008b), Lemma A.5. For 2., the proof of the \((1,1)\) block diagonal element is just a repeat of the previous lemma. For the \((2,2)\) block, it is also straightforward. For instance, the asymptotic variance covariance matrix of \( RK_{B,Y Y}(G_n) \) is given by
\[
4k_B^{(00)} n^{-1} G_n \int_0^1 \Sigma(v) \otimes \Sigma(v) dv = 4k_A^{(00)} n^{-1} H_n \int_0^1 \Sigma(v) \otimes \Sigma(v) dv \cdot \frac{G_n k_B^{(00)}}{H_n k_A^{(00)}}.
\]

The last two multiplicative factors are nothing but \( r_n \omega^{(00)} \).
For the off-diagonal block, recall that any linear combination of the realized kernels is another realized kernel. In particular, given $Y = \ln P^*$,

$$Var(RK_{A,YY}(H_n), RK_{B,YY}(G_n)|Y)$$

$$= Var(RK_{A,YY}(H_n)|Y) + Var(RK_{B,YY}(G_n)|Y) + 2Cov(RK_{A,YY}(H_n), RK_{B,YY}(G_n)|Y)$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{h=-(n-1)}^{n-1} \left\{ k_A \left( \frac{h}{H_n + 1} \right) + k_B \left( \frac{h}{G_n + 1} \right) \right\} \Gamma(h) |Y$$

If I treat $k_A + k_B$ as another kernel window, the right hand side is asymptotically characterized by

$$2 \lim_{n \to \infty} \frac{1}{n} \sum_{h=-(n-1)}^{n-1} \left\{ k_A \left( \frac{h}{H_n + 1} \right) + k_B \left( \frac{h}{G_n + 1} \right) \right\} \frac{1}{n} \cdot v'_{ab} \left( \int_0^1 \Sigma(s) \otimes \Sigma(s) ds \right) v_{cd}$$

With $G_n = r_n H_n$,

$$\sum_{h=-(n-1)}^{n-1} \left\{ k_A \left( \frac{h}{H_n + 1} \right) + k_B \left( \frac{h}{G_n + 1} \right) \right\} \frac{1}{n} = \frac{H_n}{n} \left[ \int_{-\infty}^{\infty} k_A^2(x) dx + r_n \int_{-\infty}^{\infty} k_B^2(x) dx + 2 \int_{-\infty}^{\infty} k_A(x) k_B(x/r_n) dx \right]$$

$$\sim \frac{H_n}{n} \left[ \int_{0}^{\infty} k_A^2(x) dx + r_n \int_{0}^{\infty} k_B^2(x) dx + 2 \int_{0}^{\infty} k_A(x) k_B(x/r_n) dx \right]$$

$$= 2 \frac{H_n}{n} \left\{ k_A^{(00)} + k_B^{(00)} + 2k_{AB}^{(00)}(r_n) \right\}$$

As a result, the right hand side becomes

$$4k_A^{(00)} \{ \bullet \} + r \cdot 4k_B^{(00)} \{ \bullet \} + 8k_{AB}^{(00)}(r_n) \cdot \{ \bullet \}$$

where $\{ \bullet \} = v'_{ab} \left( \frac{H_n}{n} \int_0^1 \Sigma(s) \otimes \Sigma(s) ds \right) v_{cd}$. On the other hand, the left hand side is

$$4k_A^{(00)} \{ \bullet \} + r \cdot 4k_B^{(00)} \{ \bullet \} + 2Cov_{\Sigma}(RK_{A,YY}(H_n), RK_{B,YY}(G_n))$$

By matching both sides, I have

$$Cov_{\Sigma}(RK_{A,YY}(H_n), RK_{B,YY}(G_n)) = 4k_{AB}^{(00)}(r_n) \cdot v'_{ab} \left( \int_0^1 \Sigma(s) \otimes \Sigma(s) ds \right) v_{cd} \cdot \frac{H_n}{n}$$
As a result, the asymptotic variance covariance matrix of \( V = [RK_A(H_n), RK_B(G_n)]' \) due to the discretization error is given by

\[
\begin{bmatrix}
4k_A^{(00)}\{\bullet\} & 4k_{AB}^{(00)}\{\bullet\} \\
4k_{AB}^{(00)}(r_n)\{\bullet\} & r_n \cdot 4k_B^{(00)}\{\bullet\}
\end{bmatrix}

= \left[ I \otimes \nu_{ab} \right]' \left[ \begin{bmatrix}
\frac{1}{\delta^{(00)}(r_n)} \\
r_n \cdot \delta^{(00)}(00)
\end{bmatrix} \otimes \left\{ 4k_A^{(00)} \int_0^1 \Sigma(s) \otimes \Sigma(s) ds \cdot \frac{H_n}{n} \right\} [I \otimes \nu_{cd}] \right]
\]

For the off-diagonal block of the asymptotic variance-covariance matrix due to the microstructure noise, I can apply the previous Lemma \( \mathcal{L} \) in the following way: consider the variance covariance matrix of \( RK_{A,UU}(H_n) + RK_{B,UU}(G_n) \).

\[
RK_{A,UU}(H_n) + RK_{B,UU}(G_n)
\]

\[
= nH^{-2}|k_A^{(2)}(0)| \sum_{h=-(n-1)}^{n-1} \beta_A(h/H_n) \Gamma(h) + nG^{-2}|k_B^{(2)}(0)| \sum_{h=-(n-1)}^{n-1} \beta_A(h/G_n) \Gamma(h)
\]

\[
= nH^{-2}|k_A^{(2)}(0)| (1 + \kappa^{-1}r_n^{-2}) \sum_{h=-(n-1)}^{n-1} \beta(h/H, r_n) \Gamma(h)
\]

where I defined \( \kappa = |k_A^{(2)}(0)/k_B^{(2)}(0)| \) and

\[
\beta(h/H_n, r) = \frac{\beta_A(h/H_n) + \kappa^{-1}r_n^{-2}\beta_B(h/[rH_n])}{1 + \kappa^{-1}r_n^{-2}}.
\]

\( \beta(x, r) \) satisfies (i) \( \beta(x, r) \in [-1, 1] \), (ii) \( \beta(0, r) = 1 \), (iii) \( \beta(x, r) \) is even with respect to \( x \), and (iv) \( \int_{-\infty}^{\infty} \beta^2(x, r) dx < \infty \) because

\[
\int_{-\infty}^{\infty} \beta^2(x, r) dx
\]

\[
= (1 + \kappa^{-1}r_n^{-2})^{-2} \left\{ \int_{-\infty}^{\infty} \beta_A^2(x) dx + \kappa^{-2}r_n^{-4} \int_{-\infty}^{\infty} \beta_B^2(x/r) dx + 2\kappa^{-1}r_n^{-2} \int_{-\infty}^{\infty} k_A(x) k_B(x/r) dx \right\}
\]

\[
\propto \left\{ \int_{-\infty}^{\infty} \beta_A^2(x) dx + \kappa^{-2}r_n^{-3} \int_{-\infty}^{\infty} \beta_B^2(x) dx + 2\kappa^{-1}r_n^{-2} \int_{-\infty}^{\infty} k_A(x) k_B(x/r) dx \right\}
\]

\[
\leq \left\{ \int_{-\infty}^{\infty} \beta_A^2(x) dx + \kappa^{-2}r_n^{-3} \int_{-\infty}^{\infty} \beta_B^2(x) dx + 2\kappa^{-1}r_n^{-2/3} \sqrt{\int_{-\infty}^{\infty} k_A^2(x) dx} \sqrt{\int_{-\infty}^{\infty} k_B^2(x) dx} \right\}
\]

\[
< \infty
\]

where I applied the change of variable formula, the Cauchy-Schwarz inequality, and \( \int \beta_A^2(x) dx \propto \int [k_A^{(2)}(x)]^2 \propto \int [k_A^{(2)}(x)]^2 < \infty \). (Notice that in the expression below I would use the exact formula in the third line and not the inequality evaluation in the fourth line.) Therefore, \( \beta(x, r) \) as a function
of $x$ is in the class $\mathcal{K}_1$ of Andrews (1991). By applying the previous result,

$$\text{Var}(RK_{AU}(H_n) + RK_{BV}(G_n))$$

$$= \text{Var} \left( nH_n^{-2}k_A^{(2)}(0)|1 + \kappa^{-1}r_n^{-2} \sum_{h=-(n-1)}^{n-1} \beta(h/H_n, r_n)\Gamma(h) \right)$$

$$= \left\{ nH_n^{-2}k_A^{(2)}(0)|1 + \kappa^{-1}r_n^{-2} \right\}^2 \cdot \int_{-\infty}^{\infty} [\beta(x, r)^2] dx \cdot 2\Omega \otimes \Omega \cdot H_n r_n^{-1}$$

$$= nH_n^{-3} \cdot 4k_A^{(22)} \Omega \otimes \Omega \left\{ 1 + r_n^{-2}\omega^{(22)} + 2r_n^{-1}\delta^{(22)}(r_n) \right\}$$

On the other hand, I have

$$\text{Var}(RK_{AU}(H_n) + RK_{BV}(G_n))$$

$$= \text{Var}(RK_{AU}(G_n)) + \text{Var}(RK_{BV}(G_n)) + 2\text{Cov}(RK_{AU}(H_n), RK_{BV}(G_n))$$

$$= nH_n^{-3}4k_A^{(22)} \Omega \otimes \Omega + nH_n^{-3}4k_A^{(22)} \Omega \otimes \Omega \cdot \kappa^{-1}r_n^{-3} + 2\text{Cov}(RK_{AU}(H_n), RK_{BV}(G_n))$$

By matching terms,

$$\text{Cov}(RK_{AU}(H_n), RK_{BV}(G_n)) = r_n^{-2}\delta^{(22)}(r_n) \cdot nH_n^{-3} \cdot 4k_A^{(22)} \Omega \otimes \Omega.$$ 

Finally, the cross term due to the cross products are evaluated as follows:

$$RK_{AU}(H_n) + RK_{BV}(G_n) = H_n^{-1} \sum_h \gamma(h/H_n)\Gamma(h) + G_n^{-1} \sum_h \gamma(h/G_n)\Gamma(h)$$

$$= H_n^{-1} \left\{ \sum_h \gamma(h/H_n)\Gamma(h) + r_n^{-1} \sum_h \gamma(h/G_n)\Gamma(h) \right\}$$

$$= H_n^{-1} \left\{ \gamma(h/H_n) + r_n^{-1}\gamma(h/G_n) \right\} \Gamma(h)$$

Therefore, we just need to replace $\int_0^\infty [k_A^{(1)}(x)]^2 dx$ by

$$\int_0^\infty [k_A^{(1)}(x) + r_n^{-1}k_B^{(1)}(x/r_n)]^2 dx$$

$$= \int_0^\infty [k_A^{(1)}(x)]^2 dx + r_n^{-2} \int_0^\infty [k_B^{(1)}(x/r_n)]^2 dx + 2r_n^{-1} \int_0^\infty k_A^{(1)}(x)k_B^{(1)}(x/r_n) dx$$

$$= \int_0^\infty [k_A^{(1)}(x)]^2 dx + r_n^{-1} \int_0^\infty [k_B^{(1)}(x)]^2 dx + 2r_n^{-1} \int_0^\infty k_A^{(1)}(x)k_B^{(1)}(x/r_n) dx$$

where I applied the change-of-variable formula. The last term corresponds to

$$2\text{Cov}(RK_{AU}(H_n) + R_{AU}(H_n), RK_{BV}(G_n) + R_{BV}(G_n)).$$

Combining all expressions above, we establish the claim 2. in the above Lemma.
The claim-3. is also straightforward once we realize that when \( r \to \infty \),
\[
\lim_{r \to \infty} \int_0^\infty k_A^{(i)}(x)k_B^{(i)}(x/r)dx = \int_0^\infty k_A^{(i)}(x) \left( \lim_{r \to \infty} k_B^{(i)}(x/r) \right)dx = k_B^{(i)}(0) \int_0^\infty k_A^{(i)}(x)dx
\]
because the right hand side is bounded. For \( r_n \to 0 \) as \( n \to \infty \), we can apply the change-of-variable formula for \( y = x/r_n \) \((dx = r_n dy)\) so that \( \delta^{(i)}(r_n) = r_n \int_0^\infty k_B(y)^{(i)}k_A^{(i)}(r_ny)dy \) and the integral can be evaluated exactly the same way as above, with \( A \) and \( B \) interchanged.

**Step 2: Asymptotic Covariances of the Joint Estimator**

The joint estimator defined in the main theorem corresponds to

\[
(\Theta \otimes I) \left[ \begin{array}{c} RK_A(H_n) \\ RK_B(G_n) \end{array} \right]
\]

where \( \kappa := |k_A^{(2)}(0)/k_B^{(2)}(0)|, \alpha := 1 - \kappa r_n^2, \beta = |k_B^{(2)}(0)|nG_n^{-2} \) and

\[
\Theta = \alpha^{-1} \left[ \begin{array}{cc} 1 & -\kappa r_n^2 \\ -\beta^{-1} & \beta^{-1} \end{array} \right]
\]

Note that \( \Theta \otimes I = X_n^{-1} \) in the main text for \( k_A = k_B \). Therefore, its asymptotic variance covariance matrix is given by

\[
\frac{H_n}{n} \cdot (\Theta P \Theta') \otimes D_A + \frac{n}{H_n^3} (\Theta Q \Theta') \otimes M_A + \frac{1}{H_n} (\Theta R \Theta') \otimes C_A
\]

where

\[
\Theta P \Theta' = \alpha^{-2} \left[ \begin{array}{cc} 1 & -\kappa r_n^2 \\ -\beta^{-1} & \beta^{-1} \end{array} \right] \left[ \begin{array}{cc} 1 & \delta_{AB}^{(00)}(r_n) \\ \delta_{AB}^{(00)}(r_n) & \omega^{(00)}r_n \end{array} \right] \left[ \begin{array}{cc} 1 & -\beta^{-1} \\ -\kappa r_n^2 & \beta^{-1} \end{array} \right]
\]

\[
\Theta Q \Theta' = \alpha^{-2} \left[ \begin{array}{cc} 1 & -\kappa r_n^2 \\ -\beta^{-1} & -\beta^{-1} \end{array} \right] \left[ \begin{array}{cc} 1 & \delta_{AB}^{(22)}(r_n) \\ \delta_{AB}^{(22)}(r_n) & \omega^{(22)}r_n^{-3} \end{array} \right] \left[ \begin{array}{cc} 1 & -\beta^{-1} \\ -\kappa r_n^2 & \beta^{-1} \end{array} \right]
\]
Step 3: Marginal Asymptotics for TSRK

1. When $g < h$ so that $r_n \to 0$, $\alpha = 1 - \kappa r_n \sim 1$ and $\delta^{(i)}(r_n) \sim O(r_n) \sim o(1)$. In this case, the asymptotic variance covariance matrix of TSRK reduces to the following form:

$$
\frac{H_n}{n} D_A + \frac{n}{H_n^3} M_A + \frac{1}{H_n} C_A
$$

In conjunction with $g > 0$ and $g < h$, $0 < \max\{g, 1/3\} < h < 1$ and $\nu > 0$ deliver the consistency because $O(m_n^{-1}) = O(n^{-\nu}) \sim o(1)$. For asymptotic normality, $H_n \propto n^{1/2}$ delivers the rate-optimal trade-off. Therefore, given $0 < g < h = 1/2$ and $\nu > 1/4$,

$$
n^{1/4} \left( \text{TSRK}(G_n, cn^{1/2}) - \int_0^1 \Sigma(v) dv \right) \xrightarrow{L} MN \left( 0, cD_A + c^{-3} M_A + c^{-1} C_A \right)
$$

because $O(n^{1/4} m_n^{-1}) \sim o(1)$.

2. When $g = h$ so that $r_n = r$: constant. If $r = 0$, then TSRK reduces to RK. If one wants to establish the rate of convergence as fast as possible, one must select $H \propto n^{3/5}$ to attain the $n^{1/5}$-consistency with the bias in the limiting distribution. See BNHLS (2008b). If $r > 0$, the asymptotic variance covariance matrix of TSRK is

$$
\frac{H_n}{n} \cdot D\Phi^{(0)}_{11}(r) + \frac{n}{H_n^3} \cdot M\Phi^{(2)}_{11}(r) + \frac{1}{H_n} \cdot M\Phi^{(1)}_{11}(r)
$$

where

$$
\Phi^{(0)}_{11}(r) = \frac{1 - 2\kappa \delta^{(00)}(r)r^2 + \kappa^2 \omega^{(00)} r^5}{1 - 2\kappa r^2 + \kappa^2 r^4},
$$

$$
\Phi^{(2)}_{11}(r) = \frac{1 - 2\kappa \delta^{(22)}(r) + \kappa^2 \omega^{(22)} r}{1 - 2\kappa r^2 + \kappa^2 r^4},
$$

$$
\Phi^{(1)}_{11}(r) = \frac{1 - 2\kappa \delta^{(11)}(r)r + \kappa^2 \omega^{(11)} r^3}{1 - 2\kappa r^2 + \kappa^2 r^4}.
$$

In this case, the growth rate of $H_n$ is exactly the same as before, namely we need $h \in (1/3, 1)$ for consistency and $h = 1/2$ for the asymptotic normality. Given $r_n = r$: constant, this means that $g = h \in (1/3, 1)$ and $\nu > 0$ for consistency and $g = h = 1/2$ with $\nu > 1/4$ for the asymptotic normality. Since $r_n = r > 0$ is fixed, the asymptotic variance is larger than the previous case by the multiplicative factor $\Phi^{(i)}_{11}(r) > 1$. If $k_A = k_B$, $\kappa = \omega^{(ii)} = 1$ and $\delta^{(i)}(r) = \int_0^\infty k^{(i)}(x)k^{(i)}(x/r)dx$ so that $\Phi^{(i)}_{11}$'s reduce to the
Step 4: Marginal Asymptotics for $TS\Omega$

1. When $g < h$ so that $r_n \to 0$. The leading term of the asymptotic variance covariance matrix of $TS\Omega$ is

$$|k_B^{(2)} (0)|^{-2} n^{-2} G_n^4 \left\{ \frac{H_n}{n} D_A + \frac{n}{H_n^3} \omega^{(22)} r_n^{-3} M_A + \frac{1}{H_n} \omega^{(11)} r_n^{-1} C_A \right\}$$

$$\sim |k_B^{(2)} (0)|^{-2} \left\{ \frac{G_n^4 H_n}{n^3} D_A + \frac{G_n}{n} M_B + \frac{G_n^3}{n^2} C_B \right\}$$

Note that in this case the order of the bias due to the end points is given by $O(n^{-1} G_n^2 m_n^{-1})$. Therefore, for the consistency with $g < h$, I need $4g + h - 3 < 0$, $g - 1 < 0$ and $3g - 2 < 0$ in conjunction with $\nu > 2g - 1$. These conditions reduce to $g < h < (3 - 4g) \wedge 1$ and $\nu > 2g - 1$. Given these conditions, it is straightforward to show that

- For $g < h < -3g + 2$, the second component is dominant in the asymptotic covariance. Therefore, the asymptotic covariance is given by

$$|k_B^{(2)} (0)|^{-2} b M_B \cdot n^{g - 1}.$$

The bias due to the end points is eliminated if $\nu > (3g - 1)/2$.

- For $g < h = -3g + 2$, the first and second components are balanced. The asymptotic variance is given by

$$|k_B^{(2)} (0)|^{-2} \{b^4 c D_A + b M_B \} \cdot n^{g - 1}.$$

The bias due to the end points is eliminated if $\nu > (3g - 1)/2$.

- For $-3g + 2 < h < (3 - 4g) \wedge 1$, the first component is dominant. The asymptotic variance is given by

$$|k_B^{(2)} (0)|^{-2} b^4 c D_A \cdot n^{4g + h - 3}.$$

The bias due to the end points is eliminated if $\nu > (3 - h)/2$.

2. When $g = h$ so that $r_n = r = bc^{-1} > 0$ is a constant. The leading term of the asymptotic variance covariance matrix of $TS\Omega$ is

$$n^{-2} G_n^4 \left\{ \frac{H_n}{n} D_A \Phi_{22}^{(0)} (r) + \frac{n}{H_n^3} M_A \Phi_{22}^{(2)} (r) + \frac{1}{H_n} C_A \Phi_{22}^{(1)} (r) \right\}$$

$$\sim \left\{ n^{5h-3} c D_A \Phi_{22}^{(0)} (r) + n^{h-1} c^{-3} M_A \Phi_{22}^{(2)} (r) + n^{3h-2} c^{-1} C_A \Phi_{22}^{(1)} (r) \right\}$$
where

\[
\Phi^{(0)}_{22}(r) = \frac{b^4(1 - 2\delta^{(0)}(r) + \omega^{(0)}(r))}{|k_B^{(2)}(0)|^2(1 - 2\kappa r^2 + \kappa^2 r^4)}, \\
\Phi^{(2)}_{22}(r) = \frac{b^4(1 - 2\delta^{(0)}(r)r^{-2} + \omega^{(22)}(r)^{-3})}{|k_B^{(2)}(0)|^2(1 - 2\kappa r^2 + \kappa^2 r^4)}, \\
\Phi^{(1)}_{22}(r) = \frac{b^4(1 - 2\delta^{(0)}(r)r^{-1} + \omega^{(0)}(r)^{-1})}{|k_B^{(2)}(0)|^2(1 - 2\kappa r^2 + \kappa^2 r^4)}.
\]

All terms are \(o(1)\) if \(g = h < 3/5\) and \(\nu > 2g - 1\).

For asymptotic normality, it is straightforward to show that

- For \(1/2 < g = h < 3/5\), the first component is dominant. The asymptotic variance is then given by
  
  \[|k_B^{(2)}(0)|^{-2}(1 - \kappa r^2)^{-2}b^4cDA\Phi^{(0)}_{22}(r) \cdot n^{5h-3}.\]

  The bias due to the end points is eliminated if \(\nu > (1 - h)/2\).

- For \(0 < g = h < 1/2\), the second component is dominant. The asymptotic variance is given by
  
  \[|k_B^{(2)}(0)|^{-2}(1 - \kappa r^2)^{-2}b^4\left\{c^{-3}MA\Phi^{(2)}_{22}(r)\right\} \cdot n^{h-1}.\]

  Similarly, the end effect is negligible when \(\nu > (3h - 1)/2\).

- For \(g = h = 1/2\), the three components are balanced. The asymptotic variance is given by
  
  \[|k_B^{(2)}(0)|^{-2}(1 - \kappa r^2)^{-2}b^4\left\{cDA\Phi^{(0)}_{22}(r) + c^{-3}MA\Phi^{(2)}_{22}(r) + c^{-1}CA\Phi^{(1)}_{22}(r)\right\} \cdot n^{-1/2}.\]

  Similarly, \(\nu > 1/4\) is sufficient to eliminate the bias due to the end points.

**Step 5: Joint Asymptotics**

**Joint Consistency.**

1. When \(g < h\), \(TSRK\) and \(TS\Omega\) are consistent if \(\max\{g, 1/3\} < h < 1\) with \(\nu > 0\) and \(g < h < \min\{3 - 4g, 1\}\) with \(\nu > 2g - 1\), respectively. Therefore, by taking their intersection, we have \(\max\{g, 1/3\} < h < \min\{3 - 4g, 1\}\) with \(\nu > \max\{2g - 1, 0\}\) for joint consistency.

2. When \(g = h\), \(TSRK\) and \(TS\Omega\) are consistent if \(1/3 < g = h < 1\) and \(g = h < 3/5\), respectively. The condition to eliminate the end points is the same. Therefore, by taking their intersection, I have \(1/3 < g = h < 3/5\) with \(\nu > \max\{2g - 1, 0\}\) for joint consistency.
Joint Asymptotic Normality.

A rate-optimal joint asymptotic normality is given when either \( g < h = 1/2 \) or \( g = h = 1/2 \) holds. In this case, \( \nu > 1/4 \) eliminates the bias due to the end points.

1. For \( g < h = 1/2 \), \( r_n \to 0 \). In this case, since \((1-\nu r_n^2)^{-2} \sim 1 \) and \( \delta^{(i)} \sim r_n k_A^{(i)}(0) \int_0^\infty k_B^{(i)}(x) dx \), the leading terms of the asymptotic covariance block is given by

\[
\frac{-G_n^2}{|k_A^{(2)}(0)|n} \left\{ D_A \frac{H_n}{n} + M_A \left( k^{(2)} - k_A^{(2)}(0) \int_0^\infty k_B^{(i)}(x) dx \right) r_n^{-1} \frac{n}{H_n^2} + C_A \left( 1 - k_A^{(1)}(0) \int_0^\infty k_B^{(i)}(x) dx \right) \frac{1}{H_n} \right\}
\]

Since I have \( g < h = 1/2 \), the second term is dominant with the order \( O_p(n^{-1/2}) \). By multiplying \( n^{1/4} n^{(1-g)/2} \), the final order is \( O_p(n^{(g-1/2)/2}) \sim o_p(1) \).

2. When \( g = h = 1/2 \) so that \( r_n = r = bc^{-1} > 0 \): constant. In this case, all multiplicative terms are constant. All \( H_n n^{-1} \), \( nH_n^{-3} \) and \( H_n^{-1} \) are now proportional to \( n^{-1/2} \) given \( h = 1/2 \). Multiplying by \( n^{1/4} n^{1/4} \), this is still fixed. This means that when \( g = h = 1/2 \), the off-diagonal block does not disappear. The resulting expression of the limiting covariance block is given by

\[
cD_A \Phi_{12}^{(0)}(r) + c^{-3} M_A \Phi_{12}^{(2)}(r) + c^{-1} C_A \Phi_{12}^{(1)}(r)
\]

where

\[
\Phi_{12}^{(0)}(r) = \frac{-b^2 (1 - \delta^{(00)}(r) - \kappa \delta^{(00)}(0)r^2 + \kappa \omega^{(00)}r^3)}{|k_B^{(2)}(0)| (1 - 2 \nu r^2 + \nu^2 r^4)}
\]

\[
\Phi_{12}^{(2)}(r) = \frac{-b^2 (1 - \delta^{(22)}(r) r^{-2} - \kappa \delta^{(22)}(0)r^{-2} + \kappa \omega^{(22)} r^{-1})}{|k_B^{(2)}(0)| (1 - 2 \nu r^2 + \nu^2 r^4)}
\]

\[
\Phi_{12}^{(1)}(r) = \frac{-b^2 (1 - \delta^{(11)}(r) r^{-1} - \kappa \delta^{(11)}(0)r^{-1} + \kappa \omega^{(11)} r)}{|k_B^{(2)}(0)| (1 - 2 \nu r^2 + \nu^2 r^4)}
\]

Finally, I can prove consistency of \( TSRK(G_n, H_n) \) even if I did not assume \( P^* \perp U \) as follows: first, the bias expression of the realized kernel is independent of this assumption. For the asymptotic variance covariance matrix, the same argument as in BNHLS (2008b). Proof of Theorem 1 implies that the orders of the possible covariance terms between \( RK_{YY}, RK_{UU} \) and \( RK_{UY} + RK_{UY} \) are bounded by the largest order of \( O(\sqrt{H_n n^{-1} \cdot H_n^{-3}}) \), \( O(\sqrt{nH_n^{-3} \cdot H_n^{-1}}) \), and \( O(\sqrt{H_n n^{-1} \cdot H_n^{-1}}) \). These amount to \( O(H_n^{-1}) \), \( O(n^{1/2} H_n^{-2}) \) and \( O(n^{-1/2}) \). Obviously all three terms are \( o(1) \) given \( 0 < \max\{g, 1/3\} < h < 1 \).

Q.E.D.
A.3 Proof of Equation (14) for the optimal proportionality factor

When \( g < h \), the asymptotic variance of a \( TSRK^+ \) takes a following form:

\[
cD + c^{-3}M + c^{-1}C.
\]

**Lemma 5** (Optimal Choice of the Proportionality Factor for \( TSRK^+(G_n, H_n) \))

Suppose \( G_n = bn^g \), \( H_n = cn^h \) and \( g < h \). Then, \( c^* \) minimizing the \( i \)-th diagonal element of the asymptotic variance of a \( TSRK^+ \)-based integrated covariance matrix, beta or correlation, is given by

\[
c^*_i = \sqrt{\frac{C_{ii}}{2D_{ii}} \cdot \left( 1 + \sqrt{1 + \frac{12D_{ii}M_{ii}}{C_{ii}^2}} \right)}
\]

In particular, under the assumption that \( \int_0^1 \Sigma(v) \otimes \Sigma(v) dv = \int_0^1 \Sigma(v) dv \otimes \int_0^1 \Sigma(v) dv \), the minimizer of the \( i \)-th diagonal element of the asymptotic variance covariance matrix of \( TSRK \) is given by

\[
c^*_i \approx \sqrt{\left( k(11)/k(00) \right) \left( 1 + \sqrt{1 + 3[k(00)k(22)]/(k(11))^2} \cdot R_{kk}R_{hh}^{-1} \right)} \cdot \sqrt{R_{hh}}
\]

where, for \( h \) and \( k \) such that \( i = (h - 1)d + k \), \( h, k = 1 \ldots d \), \( R_{hh} = \Omega_{hh}/\int_0^1 \Sigma_{hh}(v) dv \).

(Proof) Just solve the first order condition.

The above expression has the term \( R_{kk}R_{hh}^{-1} \) inside the double root. This stems from the fact that some of the diagonal elements of \( \Omega \otimes \Omega \) are given by the product of different diagonal elements of \( \Omega \). For instance, for \( d = 2 \), the diagonal elements of \( \Omega \otimes \Omega \) are given by \( \Omega_{11}, \Omega_{11}\Omega_{22} \) (duplicate) and \( \Omega_{22} \). However, in practice this term does not play a significant role if we select a proportionality factor by the maximum of \( c^*_i \)'s such that \( R_{kk}R_{hh}^{-1} = 1 \), \( i = (h - 1)d + k, h, k = 1 \ldots d \). If I focused on the diagonal elements such that \( R_{kk} = R_{hh} \), and if I used the Parzen window, the expression above reduces to \( c^*_i \) stated in Section 2.4 because \( k(00) = .269, k(11) = 1.50 \) and \( k(22) = 24 \).

A.4 Proof of Proposition [1]

It is sufficient to prove \( P := n^{1/4}(TSRK^+(G_n, H_n) - TSRK(G_n, H_n)) \sim o_p(1) \).

\[
P = n^{1/4} \hat{\Lambda}_n \left( \sum_{j=0}^{J_n-1} \hat{X}_n^{j} \right)^{-1} (I - \hat{X}_n) \hat{\Lambda}_n^{'}
\]

\[
= n^{1/4} \hat{\Lambda}_n \left( \sum_{j=0}^{J_n-1} \hat{X}_n^{j} \right)^{-1} \left( I - \sum_{j=0}^{J_n-1} \hat{X}_n^{j}(I - \hat{X}_n) \right) \hat{\Lambda}_n^{'}
\]

\[
= \hat{\Lambda}_n \left( \sum_{j=0}^{J_n-1} \hat{X}_n^{j} \right)^{-1} \left( n^{1/4} X_n^{J_n} \right) \hat{\Lambda}_n^{'}
\]

55
Since $TSRK(G_n, H_n) - \int_0^1 \Sigma(v)dv \sim O_p(n^{-1/4})$, we have

$$TSRK(G_n, H_n) = \hat{\Lambda}_n (I - \hat{X}_n) \hat{\Lambda}_n^\prime \xrightarrow{p} \int_0^1 \Sigma(v)dv.$$  

where the right hand side is positive definite. Since $\hat{\Lambda}_n$ is non-singular, this implies $I - \hat{X}_n$ is positive definite asymptotically, or equivalently $\rho(\hat{X}_n) < 1$ asymptotically where $\rho(X_n) = \max_{i=1..k} |\lambda_i|$ is the spectral radius of $X_n$, and $\sum_{j=0}^{J_n-1} \hat{X}_n^j$ is finite asymptotically. These relations hold for $X$, i.e. $\rho(X) < 1$ and $(\sum_{j=0}^{\infty} X^j)^{-1} = I - X$ is positive definite.

Note that $\hat{\Lambda}_n \xrightarrow{p} \Lambda_n$, non-singularity of $\hat{\Lambda}_n$ and $\Lambda$, and continuous mapping theorem imply $\hat{\Lambda}_n^{-1} = \Lambda^{-1} + o_p(1)$. Since $\hat{S}_n = + o_p(n^{-b})$, $b > 0$, $\hat{S}_n = S + o_p(1)$. Combining everything together, we have

$$\hat{X}_n = \hat{\Lambda}_n^{-1} \hat{S}_n \hat{\Lambda}_n^\prime^{-1} = (\Lambda^{-1} + o_p(1)) (S + o_p(1)) \left( \Lambda^\prime^{-1} + o_p(1) \right) = \Lambda^{-1} S \Lambda^\prime^{-1} + o_p(1) = X + o_p(1)$$

Since $\rho(X) < 1$, $\sum_{j=0}^{\infty} X^j = (I - X)^{-1}$ is bounded and positive semi-definite. By the dominated convergence theorem and continuous mapping theorem,

$$\left( \sum_{j=0}^{J_n-1} \hat{X}_n^j \right)^{-1} \xrightarrow{p} I - X > O$$

so that $\sum_{j=0}^{J_n-1} \hat{X}_n^j \sim O_p(1)$. Given $\Lambda_n \Lambda_n^\prime = \lim A_n$, the order of $P$ is determined by $n^{\alpha} \hat{X}_n^{J_n}$, which in turn is determined by $n^{\alpha} \{\rho(\hat{X}_n)\}^{J_n}$ because the eigenvalue of $X^{J_n}$ is given by the $J_n$-th power of the eigenvalue of $X$. By exponentiating,

$$n^{\alpha} \{\rho(\hat{X}_n)\}^{J_n} = \exp \left( J_n \{\alpha J_n^{-1} \ln n + \ln \rho(\hat{X}_n) \} \right)$$

The eigenvalues are given by the zeros of a polynomial equation with the coefficients given by the trace and determinant of the matrix under consideration. Trace and determinant operators are continuous. The zeros of polynomial are continuous with respect to its coefficients. Combining everything together, the continuous mapping theorem implies that $\ln \rho(\hat{X}_n) \xrightarrow{p} \ln \rho(X) < 0$. By letting $J_n$ grow at the rate faster than $\ln n$, the first term in the exponent becomes $o_p(1)$-negligible. Since $\ln \rho(X) < 0$, the exponent tends to be $-\infty$ and therefore $n^{\alpha} \{\rho(\hat{X}_n)\}^{J_n} \sim o_p(1)$.

Q.E.D.
A.5 Proof of Proposition 2

The proof is divided into four steps. Step 1 focuses on the $2 \times 2$ sub-matrix of the two scale realized kernel estimator and derives the asymptotic covariance matrix of the $vech$ part it. Step 2 calculates the asymptotic variance of the $TSRK^+$-based beta. Step 3 derives the asymptotic variance of the $TSRK^+$-based correlation. Step 4 establishes the asymptotic normality of these non-linear functions of the $TSRK^+$ estimator.

**Step 1: Asymptotic Covariance of $vech(X^{(ij)})$**

Let me define $X = n^{1/4}(TSRK^+(G_n, H_n) - \int_0^1 \Sigma(v)dv)$, and $X^{(ij)}$ as the sub-matrix of $X$ composed of $(i, i)$, $(i, j)$, $(j, j)$ and $(j, i)$ elements (clockwise from the northwest element of $X$) as $X^{(ij)} \in \mathbb{R}^{2 \times 2}$. I can represent the diagonal and upper (or lower) off-diagonal elements of $X^{(ij)}$ as follows:

$$vech(X^{(ij)}) = \begin{pmatrix} X_{ii} \\ X_{ij} \\ X_{jj} \end{pmatrix} = \begin{pmatrix} e_i^t X e_i \\ e_i^t X e_j \\ e_j^t X e_j \end{pmatrix},$$

where $e_i$ is the $i$-th unit vector, i.e. the column vector such that the $i$-th element is 1 and and any other elements are zeros.

In the following I will state the result for $g < h$. The case for $g = h$ is almost identical except that we need a multiplicative correction $\Phi^{(i)}$. Given the rule of calculating the covariance between any two elements of the matrix-valued random variable introduced in Definition 3 the asymptotic variance of $vech(X^{(ij)})$ is represented in terms of the asymptotic covariance matrix $B$ as follows:

$$\begin{pmatrix} v_i^t B v_{ii} & v_i^t B v_{ij} & v_i^t B v_{jj} \\ v_j^t B v_{ii} & v_j^t B v_{ij} & v_j^t B v_{jj} \\ v_{jj}^t B v_{ii} & v_{jj}^t B v_{ij} & v_{jj}^t B v_{jj} \end{pmatrix} = \begin{pmatrix} v_i^t \\ v_j^t \\ v_{jj}^t \end{pmatrix} B(v_{ii}, v_{ij}, v_{jj}) = v^t B v$$

where $v = (v_{ii}, v_{ij}, v_{jj}) \in \mathbb{R}^{4 \times 3}$ and $v_{ij} = vec([e_i e_j^t + e_j e_i^t]/2) \in \mathbb{R}^d$.

Since the asymptotic covariance of $X = n^{1/4}(TSRK^+(G_n, H_n) - \int_0^1 \Sigma(s)ds)$ is given by

$$B = c \cdot 4k \int_0^1 \Sigma(s) \otimes \Sigma(s) ds + c^{-3} \cdot 4k^{(22)} \Omega \otimes \Omega + c^{-1} \cdot 4k^{(11)} \left\{ \Omega \otimes \int_0^1 \Sigma(s) ds + \int_0^1 \Sigma(s) \otimes \Omega \right\},$$

the asymptotic covariance of $vech(X^{(ij)})$ is given by

$$\int_0^1 \left[c \cdot 4k \int_0^1 \left\{ v^t \{ \Sigma^{(ij)}(s) \otimes \Sigma^{(ij)}(s) \} v \right\} ds + c^{-3} \cdot 4k^{22} \left\{ v^t \{ \Omega^{(ij)} \otimes \Omega^{(ij)} \} v \right\} + c^{-1} \cdot 8k^{11} \left\{ v^t \{ \Omega^{(ij)} \otimes \Sigma^{(ij)}(s) \} v \right\} \right] ds$$

where I used the fact that $v^t \{ \Omega^{(ij)} \otimes \int_0^1 \Sigma^{(ij)}(s) ds \} v = v^t \left( \int_0^1 \Sigma^{(ij)}(s) ds \otimes \Omega^{(ij)} \right) v$. The first term is identical to the expression derived in BNHLS (2008b, Example 1). Therefore, the novel feature of this result is given by the second and third terms. Here is the complete list
of the elements (I show only the upper diagonal parts because these matrices are symmetric):

\[
v'(\Sigma_s^{(ij)} \otimes \Sigma_s^{(ij)})v = \begin{pmatrix}
\Sigma_i^2 & \Sigma_i \Sigma_{ij} & \Sigma_{ij}^2 \\
\bullet & \frac{1}{2}(\Sigma_i \Sigma_{jj} + \Sigma_{ij}^2) & \Sigma_{jj} \Sigma_{ij} \\
\bullet & \bullet & \Sigma_{jj}^2
\end{pmatrix},
\]

\[
v'(\Omega^{(ij)} \otimes \Omega^{(ij)})v = \begin{pmatrix}
\Omega_i^2 & \Omega_i \Omega_{ij} & \Omega_{ij}^2 \\
\bullet & \frac{1}{2}(\Omega_i \Omega_{jj} + \Omega_{ij}^2) & \Omega_{jj} \Omega_{ij} \\
\bullet & \bullet & \Omega_{jj}^2
\end{pmatrix},
\]

\[
v'(\Omega^{(ij)} \otimes \Sigma_s^{(ij)})v = \begin{pmatrix}
\Omega_i \Sigma_{ii} & \frac{1}{4}(\Omega_i \Sigma_{jj} + 2\Omega_{ij} \Sigma_{ij} + \Omega_{jj} \Sigma_{ii}) & \Omega_{ij} \Sigma_{ij} \\
\bullet & \frac{1}{4}(\Omega_i \Sigma_{jj} + 2\Omega_{ij} \Sigma_{ij} + \Omega_{jj} \Sigma_{ii}) & \frac{1}{4}(\Omega_{ij} \Sigma_{jj} + \Omega_{jj} \Sigma_{ij}) \\
\bullet & \bullet & \Omega_{jj} \Sigma_{jj}
\end{pmatrix}
\]

where I suppressed the time index of \( \Sigma \) for ease of notation.

**Step 2: Asymptotic Variance of \( TSRK^+ \)-based Beta**

It is convenient to define the following variables in order to represent the asymptotic variance of \( \hat{\beta}_{ij}^{(TS)} \):

\[
Q_{gh,kl} := \frac{\int \Sigma_{gh}(v) \Sigma_{kl}(v) dv}{\int \Sigma_{gh}(v) dv \cdot \int \Sigma_{kl}(v) dv},
\]

\[
R_{gh} := \frac{\Omega_{gh}}{\int \Sigma_{gh}(v) dv}.
\]

\( Q_{gh,kl} \) is a measure of time varying co-movement of \( \Sigma_{gh} \) and \( \Sigma_{kl} \). Therefore, \( Q_{gh,kl} = 1 \) if either \( \Sigma_{gh} \) or \( \Sigma_{kl} \) is constant over time, or \( \int_0^1 \int_0^1 \Sigma(s) \otimes \Sigma(s) ds ds = \int_0^1 \Sigma(s) ds \otimes \int_0^1 \Sigma(s) ds \). \( R_{gh} \) is a measure of time varying noise-to-signal ratio between the \( g \)'s returns and the \( h \)'s returns.

High-frequency regression coefficient of the j-th asset’s returns onto the i-th asset’s returns is defined by

\[
\beta_{ij} = \left( \int_0^1 \Sigma_{ii}(s) ds \right)^{-1} \int_0^1 \Sigma_{ij}(s) ds
\]

Correspondingly, the \( TSRK^+ \)-based beta is defined as

\[
\hat{\beta}_{ij}^{(TS)} = \left( TSRK^+_n \right)^{-1} TSRK^+_n
\]

They take the form of the second element of \( TSRK^+(ij) \) divided by the first element, where \( TSRK^+(ij) \) is the submatrix of \( TSRK^+ \) constructed as previously. Therefore, we need to calculate the gradient vector of the function

\[
g(x_1, x_2, x_3) = x_1^{-1} x_2
\]
which is given by

\[
\nabla g(x) = \begin{pmatrix} -x_1^{-2}x_2 \\ x_1^{-1} \\ 0 \end{pmatrix}
\]

Apart from \( c \cdot 4k^{(0)} \), \( D_{\beta_{ij}} \) is derived as follows (we drop the last raw and column of the matrix in the center):

\[
\begin{align*}
\begin{pmatrix} -\int \frac{\Sigma_{ij}}{(f\Sigma_{ii})^2} & \frac{1}{f\Sigma_{ii}} \end{pmatrix} \left( \frac{\Omega_{ii}}{f\Sigma_{ii}} \bullet \frac{1}{4} \left( \frac{\Omega_{ii} f\Sigma_{ij} + \Omega_{ij} f\Sigma_{ii}}{f\Sigma_{ii}} \right) \right) \left( \frac{\int \frac{\Sigma_{ij}}{(f\Sigma_{ii})^2}}{f\Sigma_{ii}} \right) \\
\frac{\Omega_{ii} f\Sigma_{ii}}{(f\Sigma_{ii})^2} - 2 f\Sigma_{ii} \left( \frac{\int \frac{\Sigma_{ii} f\Sigma_{ij}}{(f\Sigma_{ii})^2} + \frac{1}{2} \int f\Sigma_{ii} f\Sigma_{jj} + f\Sigma_{ij}^2} \right)
\end{align*}
\]

An almost identical calculation delivers the expression for \( M_{\beta_{ij}} \). For \( C_{\beta_{ij}} \),

\[
\begin{align*}
\begin{pmatrix} -\int \frac{\Sigma_{ij}}{(f\Sigma_{ii})^2} & \frac{1}{f\Sigma_{ii}} \end{pmatrix} \left( \frac{\Omega_{ii}}{f\Sigma_{ii}} \bullet \frac{1}{4} \left( \frac{\Omega_{ii} f\Sigma_{ij} + \Omega_{ij} f\Sigma_{ii}}{f\Sigma_{ii}} \right) \right) \left( \frac{\int \frac{\Sigma_{ij}}{(f\Sigma_{ii})^2}}{f\Sigma_{ii}} \right) \\
\frac{\Omega_{ii} f\Sigma_{ii}}{(f\Sigma_{ii})^2} - 2 f\Sigma_{ii} \left( \frac{\int \frac{\Sigma_{ii} f\Sigma_{ij}}{(f\Sigma_{ii})^2} + \frac{1}{2} \int f\Sigma_{ii} f\Sigma_{jj} + f\Sigma_{ij}^2} \right)
\end{align*}
\]

Therefore, the asymptotic variance of \( TSRK^+ \)-based \( \beta_{ij} \) is given by

\[
V_{\beta_{ij}}(c, r) = \Phi^{(0)}_{11}(r)cD_{\beta_{ij}} + \Phi^{(2)}_{11}(r)c^{-3}M_{\beta_{ij}} + \Phi^{(1)}_{11}(r)c^{-1}C_{\beta_{ij}}
\]

with

\[
D_{\beta_{ij}} = 4k^{(0)} \left\{ \left[ Q_{ii,ii} - 2Q_{ii,ij} + 2^{-1}Q_{ij,jj} \right] \beta_{ij}^2 + 2^{-1}Q_{ii,ij} \beta_{ij} \beta_{ji}^{-1} \right\}, \\
M_{\beta_{ij}} = 4k^{(2)} \left\{ \left[ R_{ii} - 2R_{ii}R_{ij} + 2^{-1}R_{ij}^2 \right] \beta_{ij}^2 + 2^{-1}R_{ii}R_{ij} \beta_{ij} \beta_{ji}^{-1} \right\}, \\
C_{\beta_{ij}} = 8k^{(1)} \left\{ -2^{-1}R_{ij} \beta_{ij}^2 + 4^{-1}(R_{ii} + R_{ij}) \beta_{ij} \beta_{ji}^{-1} \right\}
\]

and \( \beta_{ji} = \left( \int_0^1 \Sigma_{jj}(v)dv \right)^{-1} \int_0^1 \Sigma_{ij}(v)dv. \)
Step 3: Asymptotic Variance of $TSRK^+$-based Correlation

The $TSRK^+$-based correlation takes the form of the second element divided by the square root of the product of the first and third elements in $vch(TSRK^{+(ij)})$. In other words, we need to calculate the gradient vector of the function

$$f(x_1, x_2, x_3) = (x_1 \cdot x_3)^{-1/2}x_2$$

It is given by

$$\nabla f(x) = \begin{pmatrix} -x_1^{-3/2}x_3^{-1/2}x_2/2 \\ x_1^{-1/2}x_3^{-1/2} \\ -x_1^{-1/2}x_3^{-3/2}x_2/2 \end{pmatrix}$$

A straightforward calculation gives $D_{p_{ij}}$ as follows:

$$\begin{align*}
&= \frac{1}{4}(\int \Sigma_{ii}^{-3} \left( \int \Sigma_{ij} \right)^2 \left( \int \Sigma_{jj} \right)^{-1} - \left( \int \Sigma_{ii} \right)^{-2} \left( \int \Sigma_{ij} \right) \left( \int \Sigma_{jj} \right)^{-1} - \left( \int \Sigma_{ii} \right)^{-1} \left( \int \Sigma_{ij} \right) \left( \int \Sigma_{jj} \right)^{-2} + \left( \int \Sigma_{ii} \right)^{-1} \left( \int \Sigma_{ij} \right) \left( \int \Sigma_{jj} \right)^{-1} \int \Sigma_{ij} \\
&+ \frac{1}{2} \left( \int \Sigma_{ii} \right)^{-2} \left( \int \Sigma_{ij} \right) \left( \int \Sigma_{jj} \right)^{-2} \int \Sigma_{ij}^2 \\
&+ \frac{1}{2} \left( \int \Sigma_{ij} \right)^{-1} \left( \int \Sigma_{ii} \Sigma_{jj} \right) + \left( \int \Sigma_{ij}^2 \right) - \left( \int \Sigma_{ii} \right)^{-1} \left( \int \Sigma_{ij} \right) \left( \int \Sigma_{jj} \right)^{-2} \\
&+ \frac{1}{4} \left( \int \Sigma_{ii} \right)^{-1} \left( \int \Sigma_{ij} \right) \left( \int \Sigma_{jj} \right)^{-3} \int \Sigma_{ij}^3 \\
&= \frac{1}{4} \left( \int \Sigma_{ii} \right)^{-2} \left( \int \Sigma_{ij} \right)^2 \left( \int \Sigma_{jj} \right)^{-2} + \frac{1}{2} \left( \int \Sigma_{ii} \right)^{-2} \left( \int \Sigma_{ij} \right)^2 \left( \int \Sigma_{jj} \right)^{-2} \left( \int \Sigma_{ij} \right) - \frac{1}{4} \left( \int \Sigma_{ii} \right)^{-1} \left( \int \Sigma_{ij} \right) \left( \int \Sigma_{jj} \right)^{-1} \int \Sigma_{ij} \\
&+ \frac{1}{2} \left( \int \Sigma_{ii} \right)^{-2} \left( \int \Sigma_{ij} \right)^2 \left( \int \Sigma_{jj} \right)^{-2} \left( \int \Sigma_{ij} \right) - \frac{1}{4} \left( \int \Sigma_{ii} \right)^{-1} \left( \int \Sigma_{ij} \right) \left( \int \Sigma_{jj} \right)^{-2} \int \Sigma_{ij} \\
&+ \frac{1}{4} \left( \int \Sigma_{ii} \right)^{-2} \left( \int \Sigma_{ij} \right)^2 \left( \int \Sigma_{jj} \right)^{-3} \int \Sigma_{ij}^3 \\
&= \frac{1}{2} Q_{ii,ij} + \frac{1}{4} Q_{ii,ii} - Q_{ii,ij} + \frac{1}{2} Q_{ij,ij} - Q_{jj,ij} - \frac{1}{4} Q_{jj,ij} + \frac{1}{4} Q_{jj,ji} \rho_{ij}^2 + \frac{1}{2} Q_{ij,ij} \rho_{ij}^4
\end{align*}$$

60
and similarly for $\mathcal{M}_{\rho_{ij}}$. Finally, $\mathcal{C}_{\rho_{ij}}$ is derived as follows:

$$
\begin{align*}
&\left(\begin{array}{c}
\vdots \\
\vdots \\
\vdots
\end{array}\right)^T \left(\begin{array}{c}
\Omega_{ii} f \Sigma_{ii} \\
\Omega_{ij} f \Sigma_{ij} + \Omega_{ij} f \Sigma_{ij} \\
\Omega_{ij} f \Sigma_{ij} + \Omega_{ij} f \Sigma_{ij}
\end{array}\right)
\left(\begin{array}{c}
\Omega_{ii} f \Sigma_{ii} \\
\Omega_{ij} f \Sigma_{ij} + \Omega_{ij} f \Sigma_{ij} \\
\Omega_{ij} f \Sigma_{ij} + \Omega_{ij} f \Sigma_{ij}
\end{array}\right)^{-1}
\left(\begin{array}{c}
\vdots \\
\vdots \\
\vdots
\end{array}\right) \\
&= \frac{1}{4} \left(\int \Sigma_{ii}^{-1} (\int \Sigma_{ij}^{-1}) (\int \Sigma_{ij}^{-1}) \Omega_{ii} f \Sigma_{ii}\right) - \frac{1}{2} \left(\int \Sigma_{ii}^{-1} + \int \Sigma_{ij}^{-1} \right) \left(\int \Sigma_{ij}^{-1} \right)^2 - \frac{1}{2} \int \Sigma_{ij} \left(\int \Sigma_{ij} \right)^2
\end{align*}
$$

Therefore, the asymptotic variance of $TSRK^+$-based $\rho_{ij}$ is given by

$$
V_{\rho_{ij}}(c, r) = \Phi_{11}^{(0)}(r) c \mathcal{D}_{\rho_{ij}} + \Phi_{11}^{(2)}(r) c^{-3} \mathcal{M}_{\rho_{ij}} + \Phi_{11}^{(1)}(r) c^{-1} \mathcal{C}_{\rho_{ij}}
$$

and

$$
\begin{align*}
\mathcal{D}_{\rho_{ij}} &= 4k^{(00)} \left[2^{-1} Q_{ii,ij} + \left\{4^{-1} (Q_{ii,ij} + Q_{jj,ij}) - (Q_{ii,ij} + Q_{jj,ij}) + 2^{-1} Q_{ii,ij} \right\} \rho_{ij}^2 + (2^{-1} Q_{ij,ij}) \rho_{ij}^4 \right], \\
\mathcal{M}_{\rho_{ij}} &= 4k^{(22)} \left[2^{-1} R_{ii} R_{jj} + \left\{4^{-1} (R_{ii}^2 + R_{jj}^2) - (R_{ii} + R_{jj}) R_{ij} + 2^{-1} R_{ij}^2 \right\} \rho_{ij}^2 + 2^{-1} R_{ij}^2 \rho_{ij}^4 \right], \\
\mathcal{C}_{\rho_{ij}} &= 8k^{(11)} \left[4^{-1} (R_{ii} + R_{jj}) - \left\{4^{-1} (R_{ii} + R_{jj}) + 2^{-1} R_{ij} \right\} \rho_{ij}^2 + 2^{-1} R_{ij} \rho_{ij}^4 \right].
\end{align*}
$$

$\Phi_{11}^{(i)}$ are identical to those given in Theorem 1.

**Asymptotic Normality of $TSRK^+$-based Beta and Correlation**

By the delta method for $(i, j) = (1, 2)$, the claims in Proposition 2 for $\beta_{12}$ and $\rho_{12}$ follow.

Q.E.D.
A.6 The Derivation of the Dynamic Hedge Ratio in Equation (32)

Here I derive the dynamic hedge ratio in (32) from the ex-post minimization of the portfolio variance by choosing the short position in the futures contract given the unit long position in the spot contract.

Suppose the log prices of the spot and the futures during the day-τ trading period follow the Brownian semi-martingale: for \( t \in [0, 1] \),

\[
\begin{bmatrix}
\ln P^*_s(t) \\
\ln P^*_f(t)
\end{bmatrix} = \begin{bmatrix}
\ln P^*_s(0) \\
\ln P^*_f(0)
\end{bmatrix} + \begin{bmatrix}
\mu_s \\
\mu_f
\end{bmatrix} t + \begin{bmatrix}
\Sigma_{ss}(v) & 0 \\
\rho(t) \Sigma_{sf}(v) & \sqrt{1 - \rho^2(v)} \Sigma_{ff}(v)
\end{bmatrix} \begin{bmatrix}
dW_1(v) \\
dW_2(v)
\end{bmatrix}
\]

where \( \ln P^*_s(t) \) and \( \ln P^*_f(t) \) are intra-daily efficient log prices of the spot and futures, respectively, and \( W_1 \) and \( W_2 \) are independent Brownian motions. The instantaneous correlation is given by \( \rho(v) = \Sigma_{sf}(v) / \sqrt{\Sigma_{ss}(v) \Sigma_{ff}(v)} \). The drift term incorporates the risk free rate and the dividend yield. The ex-post minimum variance hedge ratio is given by the solution to the following problem:

\[
\min_{h \in \mathbb{R}} \ Var \left( \{\ln P^*_s(1) - \ln P^*_s(0)\} - h \{\ln P^*_f(1) - \ln P^*_f(0)\} \right | \ln P^*_s)
\]

Let me assume the drift term is constant conditional on the information set at the beginning of the day. This assumption is justified by the fact that the risk free rate is approximated by the market Libor rate, which is determined in the previous day, and the empirically negligible order of the average intra-daily returns. By Ito isometry, this problem reduces to

\[
\min_{h \in \mathbb{R}} h^2 \int_0^1 \Sigma_{ff}(v) dv - 2h \int_0^1 \Sigma_{sf}(v) dv
\]

From the first order condition, I have

\[
h^* = \frac{\int_0^1 \Sigma_{sf}(v) dv}{\int_0^1 \Sigma_{ff}(v) dv}
\]

Since this is the problem specific to the day-τ, \( h^* \) depends on \( \tau \). By interpreting \([0, 1]\) as the interval from 9:35 to 16:00 at day-τ, then redefining \( h^* \) as \( \beta(\tau) \), I have the desired result.