Multivariate Two Scale Realized Kernels

Shinsuke Ikeda*

Boston University

April 24, 2009

Key Words

High Frequency Data, Multivariate Volatility, Nonparametric Estimation, Market Microstructure Noise, Realized Kernels, Bias Correction, Non-Negativity Correction, Correlation Coefficient.

JEL Classification

C14, C22

*Address: 270 Bay State Road 318. e-mail: sikeda@bu.edu. The author is grateful to ***** and ***** for advice and suggestions, and seminar participants at ***** for helpful comments.
Multivariate Two Scale Realized Kernels

Shinsuke Ikeda

[Very Preliminary: Please Do Not Cite]

Abstract

We proposed a new non-parametric bias-corrected estimator of daily covariation of multiple securities in the presence of serially dependent vector-valued microstructure noise. We proposed a simple yet general and effective finite sample correction to ensure that the estimate of that kind is positive semi-definite. We derived the consistency and asymptotic normality of both variations of vector-valued efficient log prices and market microstructure noise. The asymptotic distribution does not involve any bias. In addition, for the estimation of typical non-linear transformations of the estimated covariation matrix such as the correlation coefficient or regression coefficient, we derived the consistency, asymptotic normality and the optimal choice of the bandwidth optimized for these estimators.

Because of the absence of the asymptotic bias of our estimator, these nonlinear transformations do not have the asymptotic bias as well. This suggests that our estimators of these ratios may have better finite sample properties than those based on other estimators of covariation matrix with asymptotic bias. Through simulation studies, we confirmed that this conjecture is true for a broad specification of the noise process and the structure of cross correlation between two markets. In fact, the efficiency gain of our bias-corrected estimators in terms of the accuracy of these estimators of ratios can be very large. \( \beta \) coefficient based on our method is fairly robust to different specifications in terms of relative root mean squared errors and bias, while the other estimator is quite sensitive to the different specification.

When applied to actual data of SP500 cash, futures and 30 year T-Bond futures, we found that the estimates of financially important ratios such as the correlation coefficient or the beta coefficient based on our estimator of the covariation matrix are qualitatively different from those based on other estimators in that the linkage among those markets are estimated stronger that that based on other estimators.
1 Introduction

Previously the author proposed a way to correct the bias of a univariate realized kernel estimator of the daily quadratic variation of the efficient log price process in the presence of market microstructure noise (Ikeeda (2009)). Since it involved two realized kernel estimators with different bandwidths, the estimator was called a “Two Scale Realized Kernel” (TSRK) in analogy to the two-scale realized variance (Zhang et. al. (2005)). TSRK possesses nice statistical properties such as the asymptotic unbiasedness, the consistency at the best parametric rate of $n^{1/4}$, data-exhaustiveness, the ease of implementation and interpretation and the absence of restrictive parametric assumptions. Since TSRK does not have the asymptotic bias, we could focus on the optimization of the asymptotic variance by choosing the proportionality factor of the second bandwidth. This naturally led to smaller confidence intervals on average when applied to actual data. In addition, we proposed a simple positive semi-definite correction in finite samples, although we did not find any cases of negative estimates in simulations as well as empirical applications without such correction. Therefore, we can improve the accuracy of inference on realized measures of daily quadratic variations by TSRK without paying a cost of possible negative estimate in finite samples.

The obvious question now is if and how we can possibly generalize TSRK in a multivariate setting. This question is not trivial because the multivariate setting has several issues of its own to be addressed. First, some simple but sure positive semidefinite correction should be established because TSRK is more likely to lose positive semi-definite property in finite samples. We will propose a trick that is very intuitive and relatively free of ad hoc truncation.

Another problem is the so called non-synchronous trading. A covariation of two diffusion processes of log prices is defined as

$$\gamma_{12} = p \lim_{n \to \infty} \sum_{i=1}^{n} (\ln P_{1t_i} - \ln P_{1t_{i-1}})(\ln P_{2t_i} - \ln P_{2t_{i-1}})$$

The obvious finite-sample analogue is given by the same expression in the right hand side without $p \lim$ symbol. However, that approach implicitly assumes that (1) the number of observations for two security prices are the same, (2) the sampling times (or durations) of different security prices are coincident with one another, (3) the same news hits both markets at the same time in the same manner, and (4) we know exact times of all observations. (1), (2) and (3) are not satisfied in actual datasets, and these multiple issues are called the problem of non-synchronous trading. (4) may fail to be satisfied for some dataset. In fact, some accessible high frequency data record the observation times only in the minute level so that there are multiple observations with the same time stamps. We only know the chronological order among them, but we cannot compare the exact times of lead/lags across different security prices. We have to deal with all of these issues so as not to lose the efficiency.

In the current version of this article, we dodge these issues by using the calendar-time sparse sampling, i.e. use the observations at each grid of evenly spaced calendar-time such as 9:30, 9:31, ..., 16:00, etc with possible interpolations. The first issue is resolved by the evenly spaced sampling time and interpolation. The second issue can be mitigated by an appropriate sparse sampling interval because the activity of one market would be transmitted
to another if we take a sufficient length of time interval. However, we will need to address this issue in a later version. Another approach is the “refresh time” sampling cited in BNHLS (2008b). In essence, this is a previous-tick interpolation in the transaction time based on the transactions of the least active asset over each interval. Beyond the bivariate cases, we do not have a good method to use all data of the multiple assets in the presence of the MMN and Nonsynchronous trading.

There is another issue of the lead-lag relation among different markets. This issue is very close to the nonsynchronous trading, but it deserves separate analysis because it can happen even in the refresh time sampling. Later we will see the specific example of the lead-lag relation between the spot market and the futures market of the SP500 index. Our dataset consists of the 1min calendar-time regularly spaced sampling for both markets, but their crosscorrelation will show a clear lead-lag relation. In fact, it has been well documented that the futures market leads the corresponding spot market by a great amount. We will show that our multivariate TSRK can reveal stronger linkage between two markets measured by non-linear function of the covariation matrix such as the autocorrelation or the beta compared to that based on other estimators, thus a significant implication for the risk management.

Here is the road map of this article. In section 2, we give a basic setup of the entire model. In section 3, we propose a new non-negativity correction of any bias-corrected estimators in general. If the estimator takes the form of the original positive definite estimator with subtraction of another positive semi-definite bias estimator, we can apply our proposed method. In section 4, we state the multivariate extension of our results in the previous article, with additional non-negativity correction using the method in section 3. In section 5, we propose another two-scale realized kernel estimator of the long run variance covariance matrix of the market microstructure noise process. We included this section to complete the argument of the two scale bias correction in a financial market with noisy observations. However, it will be useful for a direct analysis of the market microstructure itself. In section 6, we propose a method of estimating the integrated quarticity consistently. [TBA] In section 7, we take up the issue of non-linear transformation of the estimated covariation matrix, and propose a method to design the bandwidth of our estimator optimized for a specific purpose of using the estimated covariation matrix. In section 8, we present simulation results to highlight the efficiency gain of our new estimator over competing data-exhaustive estimators. We will focus on the efficiency gain of our new long run variance estimator of the MMN based on two scale bias correction, and the bias/MSE properties of the estimate of beta coefficients based on our method, compared to the other method. In section 9, we will apply our method to high frequency data of SP 500 index cash, futures and 30-year Treasury bond futures and show the actual efficiency gain. In section 10 we conclude the article and suggest future courses.

2 The Basic Setup

In this section we introduce the basic setup and underlying assumptions.
2.1 Multivariate Sampling Scheme

In this section we introduce the notion of refresh time, which was proposed in Harris et al. (1995). The trading hours of a single day is normalized to [0, 1] continuous interval. Suppose we have different number and duration of observations for different assets within a trading day. Let \((t^i_k), k = 1, \ldots, N^i\) be the sequence of observation times for the asset-\(i\). We should view \((t^i_k)\) as the collection of the realizations of \(N^i\) stopping times. Although \(N^i\) is the number of observations for this asset, we can consider \(N^i\) as the realization of a counting process: \(N^i(t), t \in [0, 1]\). Suppose we can distinguish the lead-lag of all observations, i.e. we can judge if an observation of one asset is recorded before or after the all other observations. This is possible if we have the precise time stamp of every observation. In this case, we can apply the refresh time sampling as follows: Set \(\tau_1 = \max_{i=1,\ldots,d}\{t^i_1\}\), and \(\tau_j = \max_{i=1,\ldots,d}\{t^i_{N(i)+1}\}\). In other words, the refresh time is defined as the latest time of the first counting times of all assets as. This basically means that we have to wait for the first transaction of the least active assets to sample the vector of all asset prices. In general, the interval between a refresh time point and its immediate predecessor is identical to the duration of the asset with the least activity over that period of time. Therefore, there are multiple observations for the other assets in general, and the latest one of them is not guaranteed to be recorded exactly at this refresh time - strictly before it in general. In this case, we take this latest observation at/before the refresh time as the proxy of the unobserved price at the refresh time. In other words, we interpolate the missing observation at the refresh time by the previously recorded price. This is why the refresh time sampling is basically a previous-tick interpolating sampling in transaction time. However, for this sampling, we have to know the lead-lag structure of all observations. The lead-lag structure is by nature the notion in calendar time, so this is a hybrid method of transaction time sampling and calendar time sampling. As is clear now, this method discards some (and potentially large) portion of the whole data. However, we will show later that this does not disturb the asymptotic properties of our estimators. [Need qualification]

The refresh time sampling requires the knowledge of the precise lead-lag structure of the observations, and it is usually equivalent to the precise knowledge of time stamps of all observations. In reality, however, we sometimes encounter a high frequency dataset with time stamps recorded only up to the minute level, for instance. In this case, although we know the lead-lag structure within a series of prices of each asset, we do not know the structure between two assets. Suppose asset A has 5 observations with time stamp 11:05, and asset B has 4 observations with the same time stamp. Usually 5 observations for A and 4 observations for B are ordered chronologically. On the other hand, we have no information about the chronological relation between A and B. In this case, we have to give up refresh time sampling and rely instead on the classical regularly spaced calendar time sampling with previous-tick interpolation. This means that at each benchmark calendar time \(\tau_k\), we need to sample the last observation \(t^i_j\) prior to \(\tau_k\). For instance, given the regularly spaced 1 minute sampling (e.g. \(\tau_1=9:31, \tau_2=9:32, \ldots\)), we take the last observation from the observations with the time stamps strictly before \(\tau_1=9:31\).

In either case, we can define the fictitious sequence of common sampling times \((\tau_k)_{k=1,2,\ldots,n}\). The effective number of observations \(n\) satisfies \(\min_{i=1,\ldots,d} N(i) \leq n \leq \max_{i=1,\ldots,d} N(i)\).
2.2 Assumptions on DGPof Signal and Noise

Given \((\tau_k)\), we can define the synchronized vector valued dataset \(X_k \equiv X(\tau_k) \in \mathbb{R}^d, k = 1 \ldots n\).

Suppose the log of the \(d\)-dimensional vector of efficient prices \(\ln P_i^s\) follow a continuous-time Ito diffusion process of the form

\[
\ln P_i^s = \ln P_0^s + \int_0^t \mu_v dv + \int_0^t \Sigma_v^{1/2} dW_v^s
\]

where \((W_t^s)\) is \(d \times d\) multiple Brownian motion, \(\mu \in \mathbb{R}^d, \Sigma^{1/2} \in \mathbb{R}^{d \times d}\). Here we normalize one day, roughly corresponding to 6.5 hours in our datasets, as \([0, 1]\) interval. Let us assume that we take \(n\) synchronized vectors of observations, \((\ln P_k^s)_{k=0, \ldots, n}\). Using a standard argument, the covariation of \((\ln P_i^s)\) is such that

\[
QV \equiv p \lim_{n \to \infty} \sum_{i=1}^n (\ln P_{t_i}^s - \ln P_{t_{i-1}}^s)(\ln P_{t_i}^s - \ln P_{t_{i-1}}^s)' = \int_0^1 \Sigma_v^{1/2}(\Sigma_v^{1/2})' dv \equiv \int_0^1 \Sigma_v dv
\]

where \((t_i)_{i=0, \ldots, n}\) is the sequence of synchrononies observation times. In reality, however, the efficient price itself may not be observable. Instead, it is naturally contaminated by MMN defined here by \((U_i)\). This implies the following latent factor model for the observed log price \((\ln P_i)\):

\[
\ln P_i = \ln P_0^s + U_i, \quad i = 0, 1, \ldots, n
\]

For the theoretical analysis below, \(U_i\) is assumed to satisfy the following conditions:

**Assumption 1 (Market Microstructure Noise Process)**

\(U_i\) is zero-mean, \(\alpha\)-mixing, and fourth order stationary process, so that its \(j\)-th autocovariance \(\Omega_j\) and the long run variance covariance matrix \(\Omega = \sum_{j \in \mathbb{Z}} \Omega_j\) are well defined.

This assumption follows the Lemma 1 of Andrews (1991). It also substitutes the last condition in the Assumption 1 of BNHLS 2008b. Recently Kahlina and Linton (2008) raised a skeptical view against the stationarity assumption on the noise moments, especially the constancy of the short-run variance \(\Omega_0\) within a day. In this article we will stick to the above stationarity assumption, but it seems possible to extend our framework to a noise process with unconditionally time-varying moments.

3 Positivity Correction of Two Scale Estimators

Although the estimators with two scale bias correction are statistically superior to the ones without bias correction in general, the former are not assured to be positive semi-definite in finite samples. In univariate case, this may not be a serious problem. In fact, Ikeda (2009) did not find any negative estimates in simulations as well as in empirical application. However, it seems more likely that TSRK yields non-positive-definite estimates in multivariate finite samples. One way to fix this problem is to truncate the eigenvalues...
of the estimated matrix by some small positive constants. For one example in the context of the long run variance covariance estimation, see Politis (2007). The conceptual subtlety of this approach is that the truncating constant has nothing to do with the quantity to be estimated. Ikedla (2009a) instead used the RK without bias correction as the truncating value when the biased corrected version takes a negative value. This approach is a rigorous formalization of the conventional method used in the wide range of variance estimation: see Killian (****) for instance. There are three drawbacks in this approach. First, the proof of the equivalence between the version with and without bias correction may be messy. Second, the convergence in distribution of the estimator based on this procedure may not be uniform over the entire parameter space - see Ikedla (2009) and Leeb and Pötscher (2005) also. We detoured this problem by assuming explicitly that the noise to signal ratio is contained in a strict compact subset of $\mathbb{R}$, but that assumption is not entirely convincing. Third, since we use the RK without bias correction whenever TS RK is not positive semi-definite, it may create a big outlier in the trajectory of the estimates of ratios as we will see later.

In this section, we propose a different procedure to fix the problem. The procedure is more general because it can be applied to any estimators of two scale type. It is also easy to implement and easier to interpret than the truncation approach, while the proof of the asymptotic equivalence with and without that correction is surprisingly simple.

A typical two-scale estimator takes the form of

$$TS(A, B) = R_1(A) - R_2(B)$$

where $R_1$ and $R_2$ are some positive semi-definite square matrices depending on different values of a parameter $A$ and $B$. Both $A$ and $B$ depend on the sample size $n$. As $n$ tends to be larger, it is more likely that $R_1(A)$ is greater than $R_2(B)$ and it is true almost surely in the limit. However, it is not necessarily true in finite samples.

Since $R_1$ is positive semi-definite, it has the Cholesky decomposition:

$$R_1(A) = \Lambda \Lambda'$$

where $\Lambda$ is a lower triangular matrix with all diagonal element being non-negative. It also depends on the sample size through $A$. The key assumption to implement the new procedure is the following:

**Assumption 2 (Nonsingular Cholesky Decomposability)**

$R_1(A)$ is positive definite.

Although $R_1(A)$ is always positive semi-definite, we need this strict definiteness. This condition, however, is not so strong. In fact, if $R_1(A)$ is a singular matrix, we cannot use it for a standard financial risk management because we typically need its inverse.

Given this condition, $\Lambda$ is lower triangular with strictly positive diagonal elements, thus non-singular. Therefore, we can apply the following transformation:

$$TS(A, B) = \Lambda (I - \Lambda^{-1} R_2(B) \Lambda'^{-1}) \Lambda' \equiv \Lambda (I - \hat{X}) \Lambda'$$

where $\hat{X} = \Lambda^{-1} R_2(B) \Lambda'^{-1}$ is positive semi-definite because we can also apply the Cholesky decomposition to $R_2(B)$ so that $R_2(B) = VV'$ and $\hat{X} = (\Lambda^{-1} V)(\Lambda^{-1} V)'$. We put the hat on
$X$ to emphasize that it depends on the estimates and thus on the sample size. If $I - \hat{X}$ is positive semi-definite, $TS$ is positive semi-definite as well. Notice the following relation:

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

(1)

provided $\rho(\hat{X}) = \max_{i=1..K} |\lambda_i| < 1$, where $\lambda_i$ is $i$-th eigenvalue of $\hat{X}$. See Abadir and Magnus (2005), p249. Let us emphasize that this is not a statistical asymptotic relation. We fix $X$ and increase $J$ independent of $n$. That is why we still have $X$ in the limit.

This limiting equality is not justified if $\rho(\hat{X}) \geq 1$ because the power series of $\hat{X}$ is not finite. In other words, the possible non-positivity of $I - X$ is caused by using this limiting relation although $\rho(\hat{X}) < 1$ is violated. However, if we stop the summation at some finite number of terms, say $J$, then $\hat{M} = (\sum_{j=0}^{J-1} \hat{X}^j)^{-1}$ is positive definite since $\hat{X}$ is positive semi-definite and $\hat{X}^0 = I$. In other words, we can consider the finite-sample correction of the original $TS$:

$$MTS(A, B, J) = \Lambda \hat{M} \Lambda'$$

Since $\hat{M}$ is positive definite, $MTS$ is positive definite given the non-singularity of $\Lambda$. Let us emphasize that $\Lambda$ depends on the bandwidth $A$, while the matrix $\hat{M}$ depends on $A$, $B$ and $J$. To enjoy the statistical asymptotic result of $TS$, $J$ should grow at some appropriate rate as $n$ does so that $\hat{M} \xrightarrow{P} I - X$: positive semi-definite where $\hat{X} \xrightarrow{P} X$.

Now we can prove the following simple yet important result.

**Proposition 1** (Non-Negativity Correction)

Let $R_1(A)$ be a positive-definite estimator of $Q$ with a possible positive semi-definite bias, $R_2(B)$ be an positive-semi-definite estimator of that bias, $TS(A, B) = R_1(A) - R_2(B)$ and $MTS(A, B, J)$ be its finite-sample non-negativity correction defined as above. If $TS(A, B) - Q \sim O_p(n^{-a})$, $a > 0$, then

$$MTS = TS(A, B) + o_p(n^{-a}) \quad as \quad \frac{1}{n} + \frac{\ln n}{J} \to 0.$$  

(*Proof*) See the Appendix.

In practice, we should not increase $J$ so fast because of the possible $\rho(\hat{X}) \geq 1$. We recommend $J = \sqrt{n}$ that is slightly faster than $\ln n$. This does not change the numerical values of the estimates so much if they are positive. Increasing $J$ too fast may cause the near-singularity of the numerical inversion of $\sum_{j=0}^{J-1} \hat{X}^j$ adopted in any statistical package program, that is fairly consistent with the ill-definedness of $I - \hat{X}$. We can also use $\sum_{j=0}^{J-1} (G^2H^{-2})^j$ instead of $(1 - G^2H^{-2})^{-1}$ by the same reason.

4 Multivariate TSRK

In this section we extend the TSRK in Ikeda (2009) to a multivariate case. As a matter of notation, we typically use capital Greek letters to represent the multivariate analogues of
the squared Greek-letter variables. For instance, the long-run variance in a univariate case \( \omega^2 \) becomes \( \Omega \) as the long-run variance-covariance matrix, and the spot variance in a univariate case \( \sigma_t^2 \) becomes \( \Sigma_t \) as the spot variance-covariance matrix associated to the multivariate diffusion process \( \Sigma_t^{1/2} dW_t \).

Suppose the efficient log prices of \( N \) securities follow an \( N \)-dimensional Itô process:

\[
d\ln P_t^i = m_t^i dt + \Sigma_t^{1/2} dW_t
\]

where, except for minimal integrability conditions, we do not assume any structure of \( m_t \) and \( \Sigma_t \). The observed log price vector is contaminated by a vector-valued noise:

\[
\ln P_t = \ln P_t^i + U_t
\]

where \( U_t \) satisfies Assumption 1. The quantity of our interest is the \( N \times N \) matrix of multivariate QV. This is sometimes called the covariance matrix, defined by

\[
QV = \int_0^1 \Sigma_t^{1/2} (\Sigma_t^{1/2})' dt = \int_0^1 \Sigma_t dt
\]

where the integral is defined for each element. Here we use the convention that the capital Greek letter stands for the multivariate analogue of a squared Greek letter variable. For a generic bandwidth \( B \), let \( k_h = k(h/(B + 1)) \in \mathbb{R} \) be a kernel window in a class specified in BNHLS (2008b) and also employed in Ikeda (2009) with all kernel functions having non-zero second order curvature at the origin, and \( \Gamma_h = \sum_{i=h+1}^n (\ln P_i - \ln P_{i-1})(\ln P_{i-h} - \ln P_{i-h-1})' \) be \( h \)-th order multivariate realized autocovariance. Notice that \( \Gamma_{-h} = \Gamma_h' \). Multivariate realized kernel is defined as follows:

\[
RK(B) = \Gamma_0 + \sum_{h=1}^{n-1} k_h (\Gamma_h + \Gamma_{-h}) = \Gamma_0 + \sum_{h=1}^{n-1} k_h (\Gamma_h + \Gamma_h')
\]

### 4.1 Multivariate Notation

Since our object is a \( d \times d \) matrix, its asymptotic variance involves \( d^2 \times d^2 \) elements. Nguyen (1997) defined the non-symmetric matrix-valued normal distribution as follows: suppose \( X \in \mathbb{R}^{k \times n} \) is a matrix-valued random variable. Let \( X_i \) be the transpose of the \( i \)-th raw vector of \( X \) with dimension \( n \times 1 \). Given \( A \in \mathbb{R}^{k \times n} \) with the transpose of the \( i \)-th raw vector \( A_i \) and \( B \in \mathbb{R}^{n \times kn} \), \( X \sim N(A, B) \) is defined by

1. \( E[X] = A \) or \( E[\text{vec}(X)] = \text{vec}(A) \) or \( E[X_i] = A_i, \forall i = 1 \ldots k \) and
2. \( B = (B_{i,j})_{i,j=1 \ldots k} \) is a block matrix with the typical block element given by a \( n \times n \) matrix \( B_{i,j} = \text{cov}(X_i, X_j) \).

In this article, \( k = n = d \). The second condition can be restated as follows: for any four vectors \( a, b, p \) and \( q \) in \( \mathbb{R}^n \),

\[
\text{cov}(a'Xb, p'Xq) = v'_{ab}Bv_{pq}
\]

where \( v_{ab} = \text{vec}([ab' + ba']/2) \). See the appendix for the proof of this claim.
Theorem 3 (Multivariate TSRK)

Define

\[ \hat{\Omega}(G) = \left( |k'(0)|nG^{-2} \right)^{-1} RK(G), \]

and

\[ TSRK(G, H) = RK(H) - |k''(0)|nH^{-2}\hat{\Omega}(G) = RK(H) - G^2H^{-2}RK(G). \]

Then,

1. (Consistency of \( \hat{\Omega} \))
   For \( G \propto n^a, \ 0 < a < 1/2, \ as \ n \to \infty, \)
   \[ \hat{\Omega}(G) \xrightarrow{p} \Omega. \]

2. (Asymptotic Normality of \( \hat{\Omega} \))
   For \( G \propto n^a, \ 0 < a \leq 1/3, \ as \ n \to \infty, \)
   \[ \sqrt{n/G} \left( \hat{\Omega}(G) - \Omega \right) \xrightarrow{d} |k''(0)|^{-1} \cdot MN \left( QV \cdot 1_{a=1/3}, \mathcal{M} \right) \]
   where \( \mathcal{M} = 4k^{(22)} \cdot \Omega \otimes \Omega. \)

3. (Consistency of TSRK)
   For \( G \propto n^a, \ 0 < a \leq 1/3 \ and \ H \propto n^a, \ 1/3 < a < 1, \ as \ n \to \infty, \)
   \[ TSRK(G, H) \xrightarrow{p} \int_0^1 \Sigma_v dv. \]

4. (Asymptotic Normality of TSRK)
   For \( H = cn^{1/2}, \ as \ n \to \infty, \)
   \[ n^{1/4} \left( TSRK(G, H) - \int_0^1 \Sigma_v dv \right) \xrightarrow{d} MN(0, Av(c)) \]
   where
   \[ Av(c) = \int_0^1 \left[ c \cdot 4k^{00}\Sigma_s \otimes \Sigma_s + c^{-3} \cdot 4k^{22}\Omega \otimes \Omega + c^{-1} \cdot 8k^{11}\Omega \otimes \Sigma_s \right] ds. \quad (2) \]

5. (Positivity Correction of TSRK)
   Given the non-singular Cholesky decomposability of \( RK(H) \) such that \( RK(H) = \hat{\Lambda}\hat{\Lambda}' \)
   where \( \hat{\Lambda} \) is a non-singular lower triangular matrix, for \( G \propto n^a, \ a \in (0, 1/3], \ H \propto n^{1/2} \)
   and \( n^{-1} + J^{-1} \ln n \to 0, \)
   \[ MTSRK^+(G, H, J) := \hat{\mu}MTSRK(G, H, J) = TSRK(G, H) + o_p(n^{-1/4}) \]

10
where \( \hat{m} = \sum_{j=0}^{J-1} (G^2 H^{-2})^j \) and

\[
MTSRK(G, H, J) = \hat{\Lambda} \cdot \left\{ \sum_{j=0}^{J-1} \left( G^2 H^{-2} \hat{\Lambda}^{-1} RK(G)(\hat{\Lambda}')^{-1} \right)^j \right\}^{-1} \cdot \hat{\Lambda}'
\]

is positive semi-definite in finite samples.

(Proof) See the Appendix.

Now we can enjoy a list of nice asymptotic properties while we feel safer with sure positive semi-definiteness by using \(MTSRK^+\).

5 A Cleaner Estimator of the LRcov of the Noise

The literature of high frequency econometrics has been driven by the motivation to extract the variation of the efficient log return from the observed variation possibly contaminated by the noise variation. In this sense, the noise variation is nothing but a nuisance parameter. Typically it is the consistent estimation and not the precise rate of convergence nor the possibility of finite sample bias that is the sole issue of interest with respect to the noise variance. For the inference of realized measures of QV or for more direct purpose of the microstructure analysis, however, we also need to establish the result on these properties. In this section we try to estimate the noise variation and therefore distinguish between the efficient variation and the noise variation as clear as possible.

5.1 The Bias-Corrected Estimator of the LRcov of the Noise

In this section we propose the consistent estimation procedure of the long run variance covariance matrix of the market microstructure noise. There were two issues in Ikeda (2009) regarding the univariate version of the previous result on the LRV estimator \( \hat{\Omega}(G) \). First, we selected \( n^{1/3} \) so that \( \hat{\Omega}(G) \overset{D}{\to} \Omega \) at the rate of \( n^{1/3} \). The choice of \( G = n^{1/3} \) produced the bias in the asymptotic distribution, but we viewed this as another source of information about QV to be exploited in the TSRK. We can improve the rate by selecting smaller value of \( a \), but \( a \) should always be positive. If \( a = 0 \) or \( G = 1 \) and with some modification, we have

\[
(\langle |k^n(0)|/2 \rangle \hat{\Omega}(1) = (2n)^{-1} \sum_{i=1}^{n} (\ln P_i - \ln P_{i-1})^2 \overset{P}{\to} (\Omega_0 - \Omega_h)
\]

The limit is neither the long run variance \( \Omega = \sum_{h \in \mathbb{Z}} \Omega_h \) nor the short run variance \( \Omega_0 \). Since \( n^{(1-a)/2} = n^{1/2} \) only if \( a = 0 \), we cannot attain the best parametric rate of \( n^{1/2} \) derived in AMZ (2005) for the iid noise. We did not explore this issue in Ikeda (2009) because our main object was not \( \Omega \) but \( QV \). However, it may be crucial when we need a precise estimate of \( \Omega \) as mentioned earlier.

Another issue is the finite sample bias of \( \hat{\Omega}(G) \). \( \hat{\Omega}(G) \) has the bias \( (\langle |k^n(0)|nG^{-2} \rangle^{-1} QV \) in finite samples. This is asymptotically negligible if we selected \( G \propto n^a, a < 1/2 \) and that is why we needed this condition for the consistency of \( \hat{\Omega}(G) \). In finite samples, however,
this bias term cannot be ignored. In fact, several empirical results have suggested that the magnitude of $QV$ is usually greater than $\Omega$ by a digit, so the actual estimate may be severely distorted by the down-scaled version of $QV$. This issue was emphasized in Hansen and Lunde (2006) for the case of iid noise, but it is more serious for the case of serially dependent noise because the long run variance of a typical AR(1) noise with a negative root is smaller than the corresponding short run variance. Let us emphasize once again that our TSRK exploits exactly this possible severe finite sample bias (and asymptotic bias when $G \propto n^{1/3}$) so that it is advantageous for the estimation of $QV$, but the severe bias of $\hat{\Omega}$ may influence the inference because we need to estimate $\hat{\Omega}$ for the construction of the approximate confidence interval of $TSRK$.

To mitigate this finite sample bias, we apply the two-scale bias correction for the estimation of $\Omega$ as well. In particular, recall the following characterization:

$$\hat{\Omega}(G) \sim \left(\frac{G}{n} n G^{-2} \right)^{-1} QV + \Omega + O_p(n G^{-2} (Gn^{-1} + n G^{-3} + G^{-1})^{1/2})$$

The first term is the finite sample bias depending on the population value $QV$, but $QV$ can be estimated by a realized kernel with certain bandwidth, say $H$. Therefore, the bias corrected version of $\hat{\Omega}(G)$ is given by

$$TS\hat{\Omega}(G, H) \equiv \hat{\Omega}(G) - \left(\frac{G}{n} n G^{-2} \right)^{-1} RK(H) = \left(\frac{G}{n} n G^{-2} \right)^{-1} (RK(G) - RK(H))$$

where $RK(H)$ for some bandwidth $H$ is the estimate of $QV$. This is another two scale realized kernel. Let us emphasize that the two bandwidths here, $G$ and $H$, are different from $G$ and $H$ for the TSRK estimator of the $QV$. This bias corrected version is characterized by

$$TS\hat{\Omega}(G, H) \sim (1 - G^2 H^{-2})\Omega + O_p(f(n))$$

where

$$f(n) \sim n G^{-2} [(Hn^{-1} + n H^{-3} + H^{-1})^{1/2} + (Gn^{-1} + n G^{-3} + G^{-1})^{1/2}]$$

As a result, we have the following better version of the two scale estimator of the long-run variance-covariance matrix of the MMN:

$$TS\hat{\Omega}^t(G, H) = (1 - G^2 H^{-2})^{-1} TS\hat{\Omega}(G, H)$$

**Theorem 4** *(Bias Corrected Estimator of LRCOV)*

1. *(Consistency)*

   For $G \propto n^g$, $H \propto n^h$, $(g, h) \in AC(1) \cup AC(2)$, where $AC(1)$ and $AC(2)$ are indicated in the figure, as $n \to \infty$,

   $$TS\hat{\Omega}^t(G, H) \xrightarrow{p} \Omega.$$

2. *(Asymptotic Normality)*

   12
• For \((g, h) \in AC(1)\), as \(n \to \infty\),

\[
\sqrt{\frac{n}{G}} \left( T S \hat{\Omega}^+(G, H) - \Omega \right) \xrightarrow{d} MN \left( 0, |k''(0)|^{-2} \mathcal{M} \right)
\]

where \(\mathcal{M} = 4k^{(22)} \Omega \otimes \Omega\).

• For \((g, h) \in AC(2)\), as \(n \to \infty\),

\[
\sqrt{\frac{n^3}{G^4 H}} \left( T S \hat{\Omega}^+(G, H) - \Omega \right) \xrightarrow{d} MN \left( 0, |k''(0)|^{-2} \mathcal{D} \right)
\]

where \(\mathcal{D} = 4k^{(10)} \int_0^1 \Lambda_v \otimes \Lambda_v dv\) and \(\Lambda_v = \sum_v \Lambda_v'\).

3. (Positivity Correction)

Given the non-singular Cholesky decomposability of \(RK(G)\) such that \(RK(G) = \Lambda \Lambda'\) with \(\Lambda\) being non-singular, as \(n^{-1} + J^{-1} \ln n \to 0\),

\[
T S \hat{\Omega}^+(G, H, J) = T S \hat{\Omega}^t(G, H) + o_p(n^{-\alpha})
\]

where \(\hat{m} = \sum_{j=0}^{J-1} (G^2 H^{-2})^j\) and

\[
T S \hat{\Omega}^t(G, H) := \hat{m} \left( |k''(0)| n^{-2} \right)^{-1} \Lambda \left( \sum_{j=0}^{J-1} \left\{ \Lambda^{-1} RK(H)(\Lambda')^{-1} \right\}^j \right)^{-1} \Lambda'.
\]

(Proof) See the Appendix.

[Figure on AC Here]

Although \(\hat{\Omega}(G)\) proposed previously can attain the asymptotic unbiasedness if \(G \propto n^g\), \(g < 1/3\), \(T S \hat{\Omega}^+\) is likely to be better than that because of the explicit correction of the bias. We will confirm this later in the simulation study.

The limiting distribution for \((g, h) \in AC(2)\) is not useful in practice because the supremum of the rate of convergence is \(n^{1/3}\), but moreover its asymptotic variance comes from the second moment of the discretization error of the covariance matrix in the bias. On the other hand, if \((g, h) \in AC(1)\), the supremum of the rate of convergence is \(n^{1/2}\) and its asymptotic variance comes validly from the second moment of the microstructure noise. However, we still have a fairly large degree of freedom to select \((g, h)\) in this region. In other words, we still have indeterminacy of \((g, h)\) except for \(g, h \in (0, 1]\) and \(g < h < -4g + 3\). The first order asymptotic theory only tells us that we should select \(g\) as small as possible. Although we conjecture that there may be a rule according to some higher-order trade off, in this article we will resort to a simulation study to form a rule of thumb. We should emphasize that there has not been any theoretical result about the precise rate of convergence of an estimator of the long run variance covariance matrix of the MMN in the literature.\(^1\)

\(^1\) AMZ (2006) considered a similar case, but they could not pin down the rate according to some optimality criterion.

13
6 Consistent Estimator of the Integrated Quarticity

The estimation of the integrated quarticity $\int_0^1 \Sigma(s) \otimes \Sigma(s) ds$ (which is $\int_0^1 \sigma^4 s ds$ in a univariate case) is harder\(^2\). In the current version of this article, we employ the following:

**Assumption 5**

$$\int_0^1 \Sigma(s) \otimes \Sigma(s) ds = \int_0^1 \Sigma(s) ds \otimes \int_0^1 \Sigma(s) ds.$$ 

This corresponds to a conventional assumption in practice, namely $\int \sigma^4 d\tau \approx (\int \sigma^2 d\tau)^2$. [This is restrictive so we have to dispense with. TBA]

7 Nonlinear Transformation of MTSRK

In terms of the financial applications, it is probably more important to consider the precision of the estimate of nonlinear transformation of covariation matrix such as the correlation or beta.

7.1 Asymptotic Results on $\beta$ and $\rho$

Suppose we focus on the relationship between $i$-th log price and $j$-th log price. The high frequency regression coefficient and realized correlation are defined respectively by

$$\beta_{ij} = \left( \int_0^1 \Sigma_i(s) ds \right)^{-1} \int_0^1 \Sigma_{ij}(s) ds$$

$$\rho_{ij} = \left( \int_0^1 \Sigma_i(s) ds \cdot \int_0^1 \Sigma_{jj}(s) ds \right)^{-1/2} \int_0^1 \Sigma_{ij}(s) ds$$

where we regress $j$-th log prices to $i$-th log prices for $\beta_{ij}$.

To state the asymptotic result, it is convenient to define the sub-matrix of the covariance matrix under consideration. For a symmetric matrix $X \in \mathbb{R}^{d \times d}$, the $2 \times 2$ sub-matrix composed of $(i, i)$, $(i, j)$, $(j, j)$ and $(j, i)$ elements (clock-wise from the northwest element) is defined by $X^{(ii)}$. Next, since $X^{(ii)}$ is symmetric, one of its off-diagonal elements is redundant. We use the operator $vech$ to extract the $(1, 1)$, $(1, 2)$ and $(2, 2)$ elements of $X^{(ii)}$ corresponding to $(i, i)$, $(i, j)$ and $(j, j)$ elements in the original matrix $X$:

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

For $X = n^{1/4}(TSRK - \int_0^1 \Sigma(s) ds)$, we establish the asymptotic variance of $vech(X^{(ii)})$ denoted by $A_{ij}(c)$ as follows:

$$A_{ij}(c) = cD_{ij} + c^{-3}M_{ij} + c^{-1}C_{ij}$$

\(^2\)Kalnina (2008) recently proposed a bootstrapping approach to this issue.
where

\[
\begin{align*}
\mathcal{D}_{ij} &= 4k^{00} \int_0^1 \left( v'(\Sigma_s^{(ij)} \otimes \Sigma_s^{(ij)})v \right) ds, \\
\mathcal{M}_{ij} &= 4k^{22} \left( v'(\Omega^{(ij)} \otimes \Omega^{(ij)})v \right), \\
\mathcal{C}_{ij} &= 8k^{11} \int_0^1 \left( v'(\Omega_s^{(ij)} \otimes \Sigma_s^{(ij)})v \right) ds
\end{align*}
\]

where \( \Sigma_s^{(ij)} \) etc. are defined similarly as \( X^{(ij)} \) is, and

\[
v = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1/2 & 0 \\
0 & 1/2 & 0 \\
0 & 0 & 1
\end{pmatrix}
\]

Recall that \( c \) in the asymptotic variance is the proportionality factor of the main bandwidth in TSRK: \( H = cn^{1/2} \). Let us define two more vectors:

\[
\eta_{ij} := \begin{pmatrix}
-(\int_0^1 \Sigma_{ii}(s) ds)^{-1} \\
(\int_0^1 \Sigma_{ij}(s) ds)^{-1}
\end{pmatrix}, \quad \theta_{ij} := \begin{pmatrix}
-(2\int_0^1 \Sigma_{ii}(s) ds)^{-1} \\
(\int_0^1 \Sigma_{ij}(s) ds)^{-1} \\
-(2\int_0^1 \Sigma_{jj}(s) ds)^{-1}
\end{pmatrix}
\]

These are proportional to the gradient vectors associated to the transformations defining \( \beta \) and \( \rho \). Recall that \( \beta \) takes the form of the second element divided by the first element of \( \text{vech}(X^{(ij)}) \), while \( \rho \) takes the form of the second element divided by the square root of the product of the first and third elements of \( \text{vech}(X^{(ij)}) \). Then, we can state the asymptotic normality of the realized ratios as follows:

**Proposition 2 (Asymptotic Normality of the TSRK-based Realized Correlation)**

Suppose both diagonal elements of \( \int_0^1 \Sigma_s^{(ij)} ds \) are nonzero. Then, TSRK-based realized beta and realized correlation have the asymptotic distributions

\[
n^{1/4} \left( \hat{\beta}_{ij} - \beta_{ij} \right) \xrightarrow{d} MN \left( 0, \beta_{ij}^2 \eta_{ij} A_{ij}(c) \eta_{ij} \right)
\]

\[
n^{1/4} \left( \hat{\rho}_{ij} - \rho_{ij} \right) \xrightarrow{d} MN \left( 0, \rho_{ij}^2 \theta_{ij} A_{ij}(c) \theta_{ij} \right)
\]

where \( A_{ij}(c) \) is the asymptotic variance of \( \text{vech}(X^{(ij)}) \) defined as above.

(Proof) See the Appendix, though straightforward in principle by the delta method.

The notable difference between our asymptotic result and the one in BNHLS (2008b) is the absence/presence of the asymptotic bias. \( \hat{\beta}_{ij} \) based on RK without bias correction, as
proposed in BN HLS (2008b), has the asymptotic bias\textsuperscript{3}

\[
\frac{c^{-2}|k''(0)|}{\int_0^1 \Sigma^u(s)ds} (\Omega_{ij} - \Omega_{ii} \beta_{ij})
\]

Therefore, the sign of the bias is the same as that of \( \Omega_{ij} - \Omega_{ii} \beta_{ij} \), which is also the same as that of \( (\Omega_{ij}/\Omega_{ii}) - (\int \Sigma_{ij}/\int \Sigma^u) \). If the linkage of two markets is stronger in terms of the efficient log prices than in terms of the noise processes, we have a negative bias. This seems typical of the actual data. This is also natural by the following casual analysis: recall that the bias of RK is given by a positive semi-definite matrix proportional to the long run variance covariance matrix of the MMN. In the bivariate case, the positive semi-definiteness of the bias matrix can be interpreted roughly as the upward bias in the diagonal elements and the downward bias in the off-diagonal elements because the determinant of the bias matrix is positive. If we transform the estimated matrix in the form of an off-diagonal element divided by a diagonal element, that ratio has the upward bias in the denominator and the downward bias in the numerator. These two forces are working at the same direction to pull down the estimated ratio. In fact, BN HLS (2008b) reported the severe negative bias in their \( \hat{\beta} \) and \( \hat{\rho} \). Therefore, even though the magnitude of the bias may seem small in estimating the covariation matrix, it can cause a severe downward bias in the estimated ratio.

We can derive more concrete forms of these asymptotic variances. For that purpose, the key quantities are \( R_{ij} := \Omega_{ij} / \int_0^1 \Sigma_{ij}(s)ds \) and \( Q_{ij,k} := \int \Sigma_{ij} \Sigma_{kl} / \int \Sigma_{ij} \int \Sigma_{kl} \). \( R \) can be interpreted as a noise-to-signal ratio and \( Q \) as a measure of cross variability of two different covariances over time. In fact, if the covariation matrix is constant over time, \( Q_{ij,k} = 1 \). Using this notation, we can derive more concrete expression for each component in the asymptotic variance.

\textbf{Lemma 6} (Asymptotic Variance of TSRK-based \( \hat{\beta}_{ij} \))

The asymptotic variance of \( \hat{\beta}(i, j) \) is given by

\[
cD_{\hat{\beta}_{ij}} + c^{-3}M_{\hat{\beta}_{ij}} + c^{-1}C_{\hat{\beta}_{ij}}
\]

where

\[
D_{\hat{\beta}_{ij}} &= 4k^{10} \left\{ Q_{ii,i} - 2Q_{ii,ij} + \frac{1}{2}Q_{ij,ij} \right\} \hat{\beta}_{ij}^2 + \frac{1}{2}Q_{ii,jj} \hat{\beta}_{ij} \hat{\beta}_{ji}^{-1},
\]

\[
M_{\hat{\beta}_{ij}} &= 4k^{12} \left\{ R_{ii}^2 - 2R_{ii}R_{ij} + \frac{1}{2}R_{jj}^2 \right\} \hat{\beta}_{ij}^2 + \frac{1}{2}R_{ii}R_{jj} \hat{\beta}_{ij} \hat{\beta}_{ji}^{-1},
\]

\[
C_{\hat{\beta}_{ij}} &= 8k^{11} \left[ -\frac{1}{2}R_{ij} \hat{\beta}_{ij}^2 + \frac{1}{4}(R_{ii} + R_{jj}) \hat{\beta}_{ij} \hat{\beta}_{ji}^{-1} \right].
\]

It is more informative to consider a special case. Suppose the covariation matrix is constant over time so that all \( Q \)'s are one. Then, \( D_{\hat{\beta}_{ij}} \) is proportional to \( (\beta_{ij} \beta_{ji}^{-1} - \beta_{ij}^2) \). We obtain this expression from \( M \) and \( C \) as well if all \( R \)'s are one. Therefore, the asymptotic

\textsuperscript{3} \( \beta^{(\hat{\beta})} \) in BN HLS (2008b) corresponds to \( \hat{\beta}_{ji} \) in our notation.
variance of TSRK-based $\hat{\beta}_{ij}$ is proportional to $\hat{\beta}_{ij}\hat{\beta}_{ji}^{-1} - \hat{\beta}_{ij}^2$ in this case. This expression has a nice interpretation:

$$\hat{\beta}_{ij}\hat{\beta}_{ji}^{-1} - \hat{\beta}_{ij}^2 = \frac{\int \Sigma_{ij} - (\int \Sigma_{ii})^2}{\int \Sigma_{ii}} \geq 0$$

In other words, this measures the relative gain of the diversification by investing two assets (i and j) over investing a single asset. The stochastic volatility of the efficient log price and the different magnitude between the efficient variation and the noise variation causes the deviation of the asymptotic variance from $\hat{\beta}_{ij}\hat{\beta}_{ji}^{-1} - \hat{\beta}_{ij}^2$.

Similarly, we have the following result.

**Lemma 7** (Asymptotic Variance of TSRK-based $\hat{\rho}_{ij}$)

The asymptotic variance of $\hat{\rho}_{ij}$ is given by

$$cD_{\rho_{ij}} + c^{-3}M_{\rho_{ij}} + c^{-1}C_{\rho_{ij}}$$

where

$$D_{\rho_{ij}} = 4k^{00}\left[Q_{\hat{\rho}_{ij}}/2 + \left\{Q_{\hat{\rho}_{ij}} + Q_{\hat{\rho}_{jj} - \hat{\rho}_{ij}}/4 - (Q_{\hat{\rho}_{ij}} + Q_{\hat{\rho}_{jj} - \hat{\rho}_{ij}})/2\right\}^{1/2} \right]^{2} - (Q_{\hat{\rho}_{ij}}/2)^2$$

$$M_{\rho_{ij}} = 4k^{22}\left[R_{\hat{\rho}_{ij}}/2 + \left\{R_{\hat{\rho}_{ij}} + R_{\hat{\rho}_{jj}}/4 - R_{\hat{\rho}_{ij}}(R_{\hat{\rho}_{ij}} + R_{\hat{\rho}_{jj}})/(R_{\hat{\rho}_{ij}} + R_{\hat{\rho}_{jj}})/2\right\}^{1/2} \right]^{2} - (R_{\hat{\rho}_{ij}}/2)^2$$

$$C_{\rho_{ij}} = 8k^{11}\left[R_{\hat{\rho}_{ij}} + R_{\hat{\rho}_{jj}}/4 - \left\{R_{\hat{\rho}_{ij}} + R_{\hat{\rho}_{jj}}/4 - (R_{\hat{\rho}_{ij}} + R_{\hat{\rho}_{jj}})/2\right\}^{1/2} \right]^{2} - (R_{\hat{\rho}_{ij}}/2)^2$$

If the covariation matrix is constant over time, the terms in the square bracket of the first component reduces to $(1 - \rho^2)^2$, which is also the expression for the square brackets of the second and third components if all R’s are one.

So far we did not mention anything about the selection of c, the proportionality factor of $H = cn^{1/2}$ as a main bandwidth of TSRK($G, H$). In this section we focus on several choices of c according to our purpose, and see their possible consequences.

### 7.2 Optimal c in Estimating the Covariation Matrix

One obvious way to select c is by following the minimization of the asymptotic MSE. Since TSRK does not involve the asymptotic bias, this amounts to minimizing the asymptotic variance. Recall that our multivariate TSRK has three components in the asymptotic variance covariance matrix. This suggests that potentially there are numerous ways to determine a norm of the asymptotic variance covariance matrix as a criterion of the optimality. BNHLS (2008b) used a very rough approach in that they compute the asymptotic MSE-optimal $c^*$ for each security price, form a collection of d optimal bandwidth for RK, namely $F_i = c_i^* n^{3/5}$, then take their maximum, minimum or average. This is surely feasible in our case: we can compute the security-wise optimal proportionality factor by solving the
minimization problem of individual security,
\[
c_i^* = \sqrt{(k^{11}/k^{00})} \sqrt{1 + 3 k^{00} k^{22}/(k^{11})^2} \cdot \sqrt{TS \hat{\Omega}_i^2(G) / RCOV_i^{(20m)}}
\]
where we used the formula in Ikeda (2009) under the assumption of \( IQ^{1/2} \approx QV \), that is estimated by subsampling-averaging realized covariance estimator at the 20-minute sampling frequency, then take the maximum, minimum or average of \( \{ c_i^* \}_{i=1,\ldots,d} \). We took \( TS \hat{\Omega}_i^2 \) and \( RCOV_i^{(20m)} \) from the \((i,i)\) element of corresponding matrix estimates, but we could also use univariate estimators of noise variation and efficient variation for different securities. The latter approach is more advantageous to the former because we can fully exploit all observations of the security prices under consideration. In that case, we should use \( TS \hat{\omega}_i^2 \) and \( RV_i^{20m} \) in place of \( TS \hat{\Omega}_i^2 \) and \( RCOV_i^{(20m)} \).

Although the above method is feasible and easy to implement, it designs \( c \) just for minimizing the diagonal elements of the asymptotic variance covariance matrix individually. Another feasible way is to minimize the squared Frobenius norm of the asymptotic covariance matrix, \( tr(\hat{A}A') \). Given (2), we can find the minimizer of \( tr(\hat{A}A') \) from its first order condition w.r.t. \( c \) because \( tr(\hat{A}A') \) is convex in \( c \). Under Assumption 5, the equation to be solved is summarized as
\[
\alpha c^8 - \beta c^4 - \gamma c^2 - \delta = 0
\]
where
- \( \alpha = (k^{00})^2 tr \left( \int \int' \otimes \int \int' \right) \),
- \( \beta = 4 (k^{11})^2 tr \left( \Omega \Omega' \otimes \int \int' \right) + 2 k^{00} k^{11} tr \left( \int \Omega' \otimes \int \Omega \right) \),
- \( \gamma = 8 k^{22} k^{11} tr \left( \Omega \Omega' \otimes \int \int' \right) \), and
- \( \delta = 3 (k^{22})^2 tr \left( \Omega \Omega' \otimes \Omega \Omega' \right) \)

where \( \int = \int_0^t \Sigma_s ds \). In practice, we have to rely on the numerical method to solve this equation after plugging the preliminary estimates of \( \Omega \) and \( \int \Sigma_s ds \) in it. However, this method may be attractive when we try to estimate the entire covariation matrix of large dimension more accurately.

### 7.3 Optimal \( c \) in Estimating Ratios

Another possibility is to design the selection of \( c \) according to our purpose. If we try to estimate the \( TSRK \)-based regression coefficient, we can select \( c \) to minimize the asymptotic variance of \( \hat{\beta}_{ij} \). Similarly, we can select \( c \) to minimize the asymptotic variance of \( \hat{\rho}_{ij} \). Since the asymptotic variance of \( TSRK \)-based quantities take the common form of
\[
cD + c^{-3} \mathcal{M} + c^{-1} \mathcal{C},
\]
the optimal \( c \) should also have the same functional form:
Corollary 8 (Optimal Choice of the Proportionality Factor for TSRK-based Estimation)

$c^*$ minimizing the asymptotic variance of a TSRK-based estimation of covariation matrix, $\beta$ or $\rho$ is given by

$$c^* = \sqrt{\frac{C}{2\mathcal{D}}} + \sqrt{\frac{C^2}{4\mathcal{D}^2} + \frac{3\mathcal{M}}{\mathcal{D}}}$$

where $\mathcal{D}$, $\mathcal{M}$ and $C$ are given in the previous Theorem 3, Lemma 6 or Lemma 7.

(Proof) Just solve the first order conditions.

For comparison, we proved in the appendix that the asymptotic MSE-optimal proportionality factors for RK-based $\hat{\beta}_{ij}$ and $\hat{\rho}_{ij}$ under Assumption 5 are given by

$$c^*_\beta = \left(\frac{|k''(0)|^2}{k^{10}}\right)^{1/5} \cdot \left(\frac{(R_{ii} - R_{ij})^2}{((\hat{\beta}_{ij} \hat{\beta}_{ji})^{-1} - 1)/2}\right)^{1/5}$$

and

$$c^*_\rho = \left(\frac{|k''(0)|^2}{k^{10}}\right)^{1/5} \cdot \left(\frac{(R_{ii} - 2R_{ij} + R_{jj})^2}{2(1 - \rho^2)^2}\right)^{1/5}$$

What is common to all proportionality factors is that they depend on the specific choice of the kernel window as well as unknown values of $\Omega^{(ij)}$ and $\int \Sigma_s^{(ij)} ds$. To do a feasible inference, we follow a standard plug-in method as follows:

1. Estimate $\int_0^1 \Sigma^{(ij)}(s) ds$ and $\Omega^{(ij)}$ by some suboptimal methods.
2. Based on these estimates, compute $R^*$, $\hat{\beta}_{ij}$, $\hat{\beta}_{ji}$ and $\hat{\rho}_{ij}$ given Assumption 5.
3. Plug them into the above formula to obtain approximately optimal proportionality factors $c^*_\beta$ and $c^*_\rho$ and use them for the selection of the bandwidth $H^* = c^* n^H$.

However, we need a great caution here. In the context of kernel estimation of the long run variance, it is widely observed that the MSE-optimal proportionality factor seems rather small so that it tends to produce a too small bandwidth. This issue is also emphasized in BNHLS (2008c). As a practical guide, we recommend to use a biais ed estimator of the long run variance $\Omega$ so that the selected main bandwidth $H = cn^{1/2}$ is not so small.

8 Simulation Studies

There are three goals in this section. First, since we obtained the consistent estimator of the long-run variance of the MMN, it is worthwhile to check its finite sample properties. This is important because we will use this estimator for the subsequent simulation of TSRK. Second, given the list of nice asymptotic properties of TSRK such as the best rate of convergence, asymptotic normality and asymptotic unbiasedness, the obvious question is if it is good in finite samples. Third, we have to make sure if the non-linear transformation of the estimated covariation matrix such as $\hat{\beta}$ or $\hat{\rho}$ possesses good finite sample properties as well. BNHLS (2008b) reported that their $\hat{\beta}$ and $\hat{\rho}$ did not work well in finite samples.

We mainly follow BNHLS (2008a,b) and Ikeya (2009) for the design of the simulation. The 6.5 trading hours are normalized as [0, 1], so that 1 second corresponds to 1/23400 unit.
We generated a multivariate 2-factor stochastic volatility model for efficient log prices: For three Brownian motion processes \( B^{(1)}, B^{(2)} \) and \( W \) such that \( B^{(1)} \perp B^{(2)} \) and \( B^{(i)} \perp W \),

\[
\begin{align*}
d \ln P^{(i)}_t &= \mu_t dt + \sigma^{(i)}_t dZ^{(i)}_t \\
\sigma^{(i)}_t &= \exp(\beta_0 + \beta_1 \tau^{(i)}_t) \\
d\tau^{(i)}_t &= \alpha \tau^{(i)}_t dt + dB^{(i)}_t \\
dZ^{(i)}_t &= \phi^{(i)} dB^{(i)}_t + \sqrt{1 - (\phi^{(i)})^2} dW_t
\end{align*}
\]

with \( \mu = .03, \beta_1 = .125, \alpha = -.025 \) and \( \beta_0 = \beta^2/(2\alpha) \) and \( \phi = -.3 \) which are common to both processes. Notice that in this setting \( \phi^{(i)} \) controls both the leverage effect (the negative correlation between the spot return and the spot volatility) and the correlation between two efficient log prices given by \( \Pi_{i=1,2}(1 - (\phi^{(i)})^2)^{1/2} \), which is always positive. This is restrictive and we will introduce another parameter later to consider the possibility of the negative correlation between two efficient log prices.

The simulation of the vector-valued MMN process has not been well explored in the literature. One thing we should bear in our minds is that we will use some sparse sampling framework such as the calender-time equi-distanced sampling or the refresh-time sampling in the current version of this article. This implies that the general remark on the correlation structure of the MMN given by Hansen and Lunde (2006) and Corradi, Distaso and *** (2007) such that IID assumption for the MMN seems reasonable unless we use a sampling frequency higher than 1 minute. This remark may apply to our setting as well. Therefore, as a benchmark, we will use the same specification as BNHLS (2008b) by assuming that two MMN processes are independent with each other and individually IID. Later we will try other specifications for the robustness check of the result obtained from the IID specification. The noise variance \( \sigma^2_z \) is specified through the noise-to-signal ratio \( \xi^2 \), i.e. \( \sigma^2_z = \xi^2 \sqrt{\sum_i \sigma^2_i n^{-1}} \) where we try three configurations \( \xi^2 = 10^{-2}, 10^{-3} \) and \( 10^{-4} \). In the current draft, we will not pursue the issue of the non-synchronous trading. We will consider the equi-distanced and synchronized observations of vector processes.

We will compare the RK without bias correction and our TSRK. They are designed as follows:

- **RK** \((F)\) with \( F = \frac{1}{2} \sum_{i=1,2} F_i \) where
  \[
  F_i := (\hat{\omega}^2_i / RV^{(20m)}_i)^{1/5} (k^{11} / k^{00})^{1/5} n^{3/5},
  \]
  \( \hat{\omega}^2_i \) is the subsampling-averaging estimator of the \( i \)-th short run variance based on the skip-2-minute sampling recommended in BNHLS 2008c, and \( RV^{(20m)}_i \) is the estimator of \( \sqrt{\hat{\Omega}_i} \) based on subsampling-averaging based on the skip-20-minute sampling.

- **TSRK** \((G, H)\) with \( G = n^{1/3} \) and \( H = \frac{1}{2} \sum_{i=1,2} H_i \) where
  \[
  H_i = \sqrt{\hat{\Omega}_i (G) / RV^{(20m)}_i} \cdot \sqrt{\left( k^{11} / k^{00} \right) \sqrt{1 + 3k^{00} k^{22} / k^{11} \cdot n^{1/2}}}
  \]

We use \( \hat{\Omega} \) in Ikeda (2009) as a biased estimator to select a sufficiently large bandwidth.
Besides, we also compare the estimators of $\beta_{12}$, $\beta_{21}$ and $\rho$ based on these four different estimators. For RK and TSRK, we select the proportionality factor to minimize the asymptotic MSE as explained in Corollary 8. For the estimation of $\Omega$, we use the collection of individual estimators of the short-run variances for RK-based $\beta$ and $\rho$, while we use $TS\hat{\Omega}^+$ for TSRK-based $\beta$ and $\rho$. For the estimation of $\int \Sigma_s ds$,

The Parzen window was used in all cases:

$$k(x) = (1 - 6|x|^2 + 6|x|^3) \cdot 1_{[0,1/2]}(|x|) + 2(1 - |x|)^3 \cdot 1_{[1/2,1]}(|x|)$$

so that $|k^6(0)| = 12$, $k^{00} = .269$, $k^{11} = 1.5$, $k^{22} = 24$. We replicated 25000 days for two sample sizes: $n = 3400$ (comparable to actual dataset) and $n = 400$ (roughly comparable to the 1 minute sampling). In the following table, $rBias$ means the relative bias and $rRMSE$ means the relative root mean squared errors of estimating a parameter (or some integral of time-varying quantity) $\theta$ defined as

$$rBias = \frac{\text{mean}\{\theta_{est} - \theta_{true}\}}{\theta_{true}}$$

$$rRMSE = \sqrt{\text{mean}\{(\theta_{est} - \theta_{true})^2\}}$$

For $MTS\hat{\Omega}$, $IIDb$ means this relative bias for IID noise, and $IIDm$ means this relative root mean squared errors for IID noise. Similarly for other cases.

### 8.1 MTS$\hat{\Omega}$

No other estimates on the long run variance-covariance matrix has been proposed in the literature, so we will give a comprehensive simulation in this subsection. The area of consistency of $(g, h)$ was established, but no specific rule to select the particular pair was proposed. From the asymptotic characterization of $MT\hat{\Omega}$, we know that the pair $(g, h)$ should be selected from the south-west boundary of the area of consistency defined by $h = -4g + 2$ to make the rate of convergence as close to $n^{1/2}$ as possible. We tried a univariate case with three different noise configurations (IID, AR(1) and MA(1) used in Ikeda (2009)).
8.1.1 Simulation Results on $TS\hat{\Omega}(G, H, J)$: Univariate Case

<table>
<thead>
<tr>
<th>$\xi^2$</th>
<th>$n$ : 3400</th>
<th>IIDb</th>
<th>IIDm</th>
<th>ARb</th>
<th>ARm</th>
<th>MAb</th>
<th>MAm</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>$\hat{\Omega}$</td>
<td>.13</td>
<td>.94</td>
<td>1.02</td>
<td>2.37</td>
<td>-.05</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>$\hat{\omega}^2$</td>
<td>.25</td>
<td>.25</td>
<td>5.40</td>
<td>5.41</td>
<td>-.33</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td>$TS\hat{\Omega}^+$</td>
<td>-.13</td>
<td>.23</td>
<td>.03</td>
<td>.44</td>
<td>-.16</td>
<td>.20</td>
</tr>
<tr>
<td>.001</td>
<td>$\hat{\Sigma}$</td>
<td>4.36</td>
<td>7.78</td>
<td>13.54</td>
<td>22.52</td>
<td>1.67</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>$\hat{\omega}^2$</td>
<td>2.49</td>
<td>2.49</td>
<td>11.87</td>
<td>11.88</td>
<td>.66</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td>$TS\hat{\Sigma}^+$</td>
<td>.21</td>
<td>1.06</td>
<td>1.49</td>
<td>2.95</td>
<td>-.10</td>
<td>.58</td>
</tr>
<tr>
<td>.0001</td>
<td>$\hat{\Sigma}$</td>
<td>47.84</td>
<td>77.77</td>
<td>139.28</td>
<td>224.87</td>
<td>20.99</td>
<td>34.54</td>
</tr>
<tr>
<td></td>
<td>$\hat{\omega}^2$</td>
<td>25.49</td>
<td>25.58</td>
<td>78.35</td>
<td>78.58</td>
<td>10.89</td>
<td>10.93</td>
</tr>
<tr>
<td></td>
<td>$TS\hat{\Sigma}^+$</td>
<td>6.05</td>
<td>9.97</td>
<td>18.53</td>
<td>29.08</td>
<td>2.39</td>
<td>4.39</td>
</tr>
</tbody>
</table>

Table 2. Relative Bias and Relative RMSE, $n = 3400$.

<table>
<thead>
<tr>
<th>$\xi^2$</th>
<th>$n$ : 23400</th>
<th>IIDb</th>
<th>IIDm</th>
<th>ARb</th>
<th>ARm</th>
<th>MAb</th>
<th>MAm</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>$\hat{\Omega}$</td>
<td>.05</td>
<td>.77</td>
<td>.60</td>
<td>1.82</td>
<td>-.03</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>$\hat{\omega}^2$</td>
<td>.26</td>
<td>.26</td>
<td>5.40</td>
<td>5.41</td>
<td>-.33</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td>$TS\hat{\Omega}^+$</td>
<td>-.06</td>
<td>.11</td>
<td>.01</td>
<td>.22</td>
<td>-.07</td>
<td>.10</td>
</tr>
<tr>
<td>.001</td>
<td>$\hat{\Sigma}$</td>
<td>2.95</td>
<td>5.85</td>
<td>9.38</td>
<td>16.79</td>
<td>1.07</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td>$\hat{\omega}^2$</td>
<td>2.55</td>
<td>2.56</td>
<td>12.03</td>
<td>12.05</td>
<td>.69</td>
<td>.70</td>
</tr>
<tr>
<td></td>
<td>$TS\hat{\Sigma}^+$</td>
<td>-.02</td>
<td>.62</td>
<td>.42</td>
<td>1.46</td>
<td>-.07</td>
<td>.32</td>
</tr>
<tr>
<td>.0001</td>
<td>$\hat{\Sigma}$</td>
<td>33.57</td>
<td>57.97</td>
<td>97.92</td>
<td>167.50</td>
<td>14.64</td>
<td>25.76</td>
</tr>
<tr>
<td></td>
<td>$\hat{\omega}^2$</td>
<td>24.86</td>
<td>24.94</td>
<td>76.53</td>
<td>76.76</td>
<td>10.61</td>
<td>10.64</td>
</tr>
<tr>
<td></td>
<td>$TS\hat{\Sigma}^+$</td>
<td>2.29</td>
<td>4.64</td>
<td>7.55</td>
<td>13.41</td>
<td>.77</td>
<td>2.11</td>
</tr>
</tbody>
</table>

Table 3. Relative Bias and Relative RMSE, $n = 23400$.

<table>
<thead>
<tr>
<th>$\xi^2$</th>
<th>IID$_1$</th>
<th>AR$_1$</th>
<th>MA$_1$</th>
<th>IID$_2$</th>
<th>AR$_2$</th>
<th>MA$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>.176</td>
<td>.639</td>
<td>.196</td>
<td>.970</td>
<td>.995</td>
<td>.760</td>
</tr>
<tr>
<td>.001</td>
<td>.179</td>
<td>.430</td>
<td>.099</td>
<td>.974</td>
<td>.975</td>
<td>.979</td>
</tr>
<tr>
<td>.0001</td>
<td>.369</td>
<td>.363</td>
<td>.382</td>
<td>.969</td>
<td>.968</td>
<td>.969</td>
</tr>
</tbody>
</table>

Table: Coverage Rates of Confidence Interval, $n = 3400$.

<table>
<thead>
<tr>
<th>$\xi^2$</th>
<th>IID$_1$</th>
<th>AR$_1$</th>
<th>MA$_1$</th>
<th>IID$_2$</th>
<th>AR$_2$</th>
<th>MA$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>.146</td>
<td>.292</td>
<td>.168</td>
<td>.945</td>
<td>.992</td>
<td>.703</td>
</tr>
<tr>
<td>.001</td>
<td>.068</td>
<td>.498</td>
<td>.051</td>
<td>.970</td>
<td>.973</td>
<td>.959</td>
</tr>
<tr>
<td>.0001</td>
<td>.465</td>
<td>.462</td>
<td>.474</td>
<td>.968</td>
<td>.968</td>
<td>.969</td>
</tr>
</tbody>
</table>
Table: Coverage Rates of Confidence Interval, \( n = 23400 \).

The major findings for the univariate case (i.e. \( TS\hat{\omega}^2 \)) are summarized as follow:

- In general, the short-run variance and the long-run variance are not so different unless the noise process is non-stationary. This is confirmed in our simulation as well: their difference is smaller than the standard deviation of any estimators of the LRV. Therefore, a good estimate of the SRV is also a good estimate of the LRV in finite sample, although the former will never be a consistent estimator of the latter.

- We have three estimators to be compared: subsampling-averaging estimator of the short-run variance \( \hat{\omega}^2 \) proposed by BNHLS (2008c), the modified kernel estimator of the long-run variance \( \hat{\Omega}(G) \) proposed by Ikeda (2009) and its two-scale correction \( TS\hat{\Omega}^+(G, H, J) \). In the simulation with 3240 transactions a day, \( \hat{\omega}^2 \) beats \( \hat{\Omega}(G) \), while with 23400 transactions a day, they are very similar. This is not surprising because the subsampling-averaging estimator has an implicit structure of the realized kernel. In either case, however, \( TS\hat{\Omega}^+(G, H, J) \) outperforms both in terms of the relative bias and relative RMSE.

- The smaller the magnitude of the noise, the harder to estimate accurately. This is reflected by the fact that all estimators under consideration perform worse as \( \xi^2 \) decreases, and the positive bias relative to the true value of the long run variance is no less than 2. In all of the cases, however, our \( TS\hat{\Omega} \) outperforms two other estimators by a digit.

- The coverage rate of the approximate confidence interval of \( TS\hat{\Omega} \) for \( (g, h) = (1/2, 3/5) \) in \( AC_1 \) is poorly small, while the one for \( (g, h) = (2/5, 1/2) \) in \( AC_2 \) is reasonably good. As emphasized in Ikeda (2009), the combination of MA(1) noise and \( \xi^2 = .01 \) is unrealistic, so for a reasonable range of the specification, we should use the combination of \( (g, h) \) taken from \( AC_2 \).

8.2 TSRK-based \( \beta \)

We now turn to the finite sample performance of TSRK-based \( \hat{\beta} \) compared to those based on \( RK \) without bias correction or \( RCOV \). We fix the number of observations within a day at 3400. The subscript 1 in the first row (such as \( RK_1 \)) means the result for the bivariate IID processes, while the subscript 2 means the bivariate AR and MA case. We generated 25000 days.
<table>
<thead>
<tr>
<th></th>
<th>$RK_1$</th>
<th>$TSRK_1$</th>
<th>$RK_2$</th>
<th>$TSRK_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n : 400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rBias</td>
<td>-2.51</td>
<td>-1.02</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>rRMSE</td>
<td>5.64</td>
<td>5.59</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>$n : 1400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rBias</td>
<td>-2.13</td>
<td>-0.99</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>rRMSE</td>
<td>4.37</td>
<td>3.99</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>$n : 3400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rBias</td>
<td>-1.91</td>
<td>-1.02</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>rRMSE</td>
<td>3.69</td>
<td>3.24</td>
<td>TBA</td>
<td>TBA</td>
</tr>
</tbody>
</table>

Table: Relative RMSE and Bias (%), $\xi^2 = .001$.

<table>
<thead>
<tr>
<th></th>
<th>$RK_1$</th>
<th>$TSRK_1$</th>
<th>$RK_2$</th>
<th>$TSRK_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n : 400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rBias</td>
<td>-11.57</td>
<td>-8.09</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>rRMSE</td>
<td>14.03</td>
<td>13.38</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>$n : 1400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rBias</td>
<td>-9.96</td>
<td>-7.27</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>rRMSE</td>
<td>11.67</td>
<td>11.24</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>$n : 3400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rBias</td>
<td>-8.96</td>
<td>-6.76</td>
<td>TBA</td>
<td>TBA</td>
</tr>
<tr>
<td>rRMSE</td>
<td>10.32</td>
<td>9.99</td>
<td>TBA</td>
<td>TBA</td>
</tr>
</tbody>
</table>

Table: Relative RMSE and Bias (%), $\xi^2 = .01$.

Several features emerge from these tables.

- RK-based $\beta$ is negatively biased, but its variance is small. On the other hand, the bias of TSRK-based $\beta$ is modest while its variance is not small as RK-based $\beta$ is. We should have expected this by the asymptotic characterization of these estimators: RK has the asymptotic bias with only a single component in the asymptotic variance. TSRK has no asymptotic bias with two additional components in the asymptotic variance.

- Statistically the gains shown in the table are not small, but it may be arguable if they are economically not large. Later we will consider this issue in terms of the economic gain of the dynamic hedging strategy.

- Our TSRK is still subject to the negative bias. This means that the negative bias in the $\beta$ estimator may be caused by something which may not be captured by the first order asymptotic theory.
9 Application-I: Bivariate Covariations

In this section we present two bivariate applications of the modified multivariate TSRK approach. More concretely, we try to estimate

1. the hedge ratio of a futures position against the associated spot position, and
2. the degree of flight-to-quality from the stock market to the bond market.

The estimation of the degree of linkage between two asset prices are crucial for any financial risk management. To estimate possibly time-varying covariances, the standard method is to set up latent conditional dynamics of the covariance in the class of multivariate GARCH model. However, it involves so many parameters to be estimated and cause easily the issue of identification. Once we can access to the intra-daily data, however, we can estimate this time-varying covariances without imposing any restrictive assumptions on the latent dynamics of the variance-covariance matrix of the returns. Since the observed prices are likely to be contaminated by the MMN, we have to disentangle the efficient covariation matrix from the noisy bivariate return processes. One of the goal in this section is to estimate beta coefficients and correlation coefficients as ratios of the covariation between efficient spot returns and efficient futures returns to the measures of variations such as quadratic variations of the efficient futures returns or geometric average of quadratic variations of two return processes by using the multivariate TSRK framework. To the best of our knowledge, this is the first result of applying RK framework to the estimation of dynamic hedge ratio between spot and futures positions. The in-sample fitting of the MTSRK compared to the competitors such as traditional realized covariance or RK without bias correction will produce remarkably different estimates of $\beta_t$.

The second goal of this section is to analyze the out-of-sample forecasting performance of the various hedge ratio estimates. We can use a particular estimate of $\beta_t$ to form a portfolio to assess its risk in the next period by keeping the ratio fixed. By “risk” of the portfolio, the traditional measure is given by the variance of the portfolio returns. Another one that seems appropriate is the coherent risk measure. We will give the hedging performances give by these two measures and show that the methodology based on MTSRK outperforms any other methodology. [TBA]

The final goal of this section is to convert the statistical gain obtained so far to the economic gain, and compare it to the standard transaction cost such as brokerage or clearing fees. We employ the method of Bandi, Russell and Lin (2008) to assess the forecasting performance of the volatility estimators by utility gain of mean-variance hedgers through the certainty equivalence. Since the futures trading are highly stylized and professionalized, most of the investors have to pay these transaction costs to re-balance their futures positions. [TBA]

9.1 Analysis of Linkage between SP500 Spot and Futures

9.1.1 Intra-Daily Patterns between Spot and Futures

The next figure shows the evolution of spot prices and associated futures prices of SP500 index on 27-Dec-2006. First, we can observe the opening drift of the spot prices. Second,
initially the spot prices are higher than the futures prices. Given the no-arbitrage relation between spot and futures markets, the latter should not be the case. Therefore, the opening drift is nothing but the adjustment of the spot market prices to the no-arbitrage justified level. In another words, over the first 5 minutes since the opening, the spot prices were not efficient: they were on the tatonnement process.

This opening drift is widely observed in the spot market relative to the associated futures market. If the opening price of the spot market are not justified by the rational futures-spot spread, it is the spot market which is adjusted to the level given by the futures prices. It is an interesting research question in the future to ask why this is the case.

The analysis of linkage between one market and another has been conducted repeatedly in the literature\(^4\), but there are surprisingly few results are available for the intra-daily linkage. Our leading example is the linkage between SP500 index spot market and associated futures market. The frequency of sampling is one-minute calendar-time because we can only access to the data on spot market at that frequency. It has been well documented that a futures market leads spot market in terms of dissemination and processing of information. If this is also true within a day, we have non-trivial lagged cross correlation of spot returns and futures returns. The following picture shows that this is indeed the case.

Interestingly, the cross correlations have changed over time in a clear fashion. In the first 1/3 of the sample corresponding to the pre-2001 period, futures leads the spot by a great margin. The largest values of the cross correlations appear in the 1 minute lead of the futures prices. Even the 5 minute futures-lead cross correlations seem informative in this period. Since the beginning of the year 2001, the lead of the futures market becomes smaller. In fact, the cross correlations tend to take the greatest values for \( k = 0 \) since then. This suggests that the linkage between these two markets has been deepening. However, even in this later period, the 1 minute futures-lead cross correlation is not negligible. On the other hand, we do not observe such patterns for the spot-lead cross correlations at all. This suggests that the high frequency estimators of the covariation between the spot and futures should be robust to this particular lead-lag structure. By the previous simulation studies, we know that our MTSRK seems appropriate for this purpose.

\([TBA]\)

In the next section we will compare the performance of MTSRK with other candidates.

### 9.2 Spot-Futures Dynamic Hedge Ratio

In this subsection we apply the multivariate TSRK to the estimation of the time-varying variance-optimal hedge ratio of SP500 futures position relative to the associated cash position. Futures contract of some underlying security (the underlying/spot/cash position) is a traditional tool of hedging the risk in the spot position. The theoretical justification of this conventional hedging method is given by the following no-arbitrage relation:

\[
\ln F_{t,T} = \ln S_t + (r - \delta)(T - t)
\]

where \( F \) and \( S \) are futures and spot prices, \( r \) is the constant risk-free rate, \( \delta \) is the dividend yield of the spot contract, and \( T \) is the maturity date (or \( T - t \) is the time-to-maturity).

\(^4\)To name a few, ****
Figure 1: Crosscorrelations b/w $r'^S_t$ and $r'^F_{t-k}$, $k = 5, \ldots, -2$ if futures leads spot ($k \geq 0$) or spot leads futures ($k < 0$)
Traditionally, \( r - \delta \) is assumed to be constant so that log futures price and log spot price have the deterministic spread called the basis. Given the absence of the basis risk, two prices are positively correlated perfectly. In this case, we can completely hedge the fluctuation of the spot price by exactly the opposite position (i.e. the same magnitude but the opposite sign) of the futures contract. In reality, the basis evolves randomly over time, so this naive hedging strategy may not be successful. Moreover, it seems likely that the moments of the returns of these contracts vary over time. To take these into account, we have to estimate the dynamic hedge ratio denoted by \( \beta_t \).

Suppose we try to take the short position of the index futures as a fraction of the cash position to minimize the \( \mathcal{F}_t \)-conditional variance of the portfolio return of these two positions. If we can only access to the daily closing prices, this is formalized as

\[
\min_h \quad \text{var}_t \left( (\ln S_{t+1} - \ln S_t) - h(\ln F_{t+1} - \ln F_t) \right)
\]

The negative sign in front of \( h \) indicates that we are going to take a short position of the futures, but \( h \) can take negative values as well. This is nothing but a population version of the OLS regression of \( \Delta \ln S \) onto \( \Delta \ln F \), and therefore the conditional-variance-optimal hedge ratio \( h \) is given by

\[
h^*_t = \frac{\text{cov}_t(\ln F_{t+1} - \ln F_t, \ln S_{t+1} - \ln S_t)}{\text{var}_t(\ln F_{t+1} - \ln F_t)}
\]

If we can access to intra-daily data, we can modify this result as follows: suppose the index \( t \) represents the end of the \( t \)-th trading day, say 4:00 PM EST. If we disregard the information content of the overnight trading, the information set at \( t \) also represents the one just prior to the opening time of the \( t + 1 \)-st trading day. The intra-daily joint dynamics of the log prices of spot and futures is given by

\[
\left( \frac{d \ln S_t}{d \ln F_t} \right) = \left( \begin{array}{c} \mu_t^s \\ \mu_t^f \end{array} \right) dt + \sqrt{\sigma_t^s \sigma_t^f} \sqrt{1 - \rho_t^2} \left( \begin{array}{c} 0 \\ \rho_t \sigma_t^f \end{array} \right) \left( \begin{array}{c} dW_t^s \\ dW_t^f \end{array} \right)
\]

where \( \sigma_t^s \) is the spot volatility of the log stock price, \( \sigma_t^f \) the spot volatility of the log futures price, and \( \rho_t \) the spot correlation between \( W_t^s \) and \( W_t^f \). We will not specify further what \( \mu \)'s, \( \sigma \)'s and \( \rho \) are all about except very weak regularity conditions such as certain integrability or the uniformly positive lower bound to \( \Sigma_t \). Let us represent

\[
\Sigma(t) = \Sigma_t^{1/2} (\Sigma_t^{-1/2})' = (\Sigma_{ij}(t))_{i,j=1,2} = \left( \begin{array}{cc} (\sigma_t^s)^2 & \rho_t \sigma_t^s \sigma_t^f \\ \rho_t \sigma_t^s \sigma_t^f & (\sigma_t^f)^2 \end{array} \right)
\]

Using this notation, we can represent the optimal hedge ratio as

\[
h^*_t = \frac{\text{cov}_t(\ln F_{t+1} - \ln F_t, \ln S_{t+1} - \ln S_t)}{\text{var}_t(\ln F_{t+1} - \ln F_t)} = \frac{\int_t^{t+1} \Sigma_{12}(v)dv}{\int_t^{t+1} \Sigma_{22}(v)dv}
\]

28
As is clear now, we can estimate the matrix $\int_{t}^{t+1} \Sigma_v dv$ and thus $h_{t:t+1}^\ast$ ex-post non-parametrically by RCOV, RK or TSRK. We should emphasize that we did not assume any inter-daily dynamics of specification (3), so $\int_{t}^{t+1} \Sigma_v dv$ can change for different $t$ corresponding to different trading days.

The next figure plots the trajectories of $\hat{\beta}_t$.

![High Frequency Estimates of Dynamic Hedge Ratio (1 minute sampling)](image)

**Figure 2: Dynamic Hedge Ratio**

The lower trajectory is the one based on the realized covariance (green dotted). It is systematically below any other trajectories based on RK (magenta), TSRK (blue dotted) and 22-day rolling-window OLE based estimator. TSRK-based beta is also above the RK-based beta, which is consistent with our earlier prediction that the positive semi-definite biases of RCOV and RK may mask the reality in the non-linear transformation of the estimated covariation matrix. Around the blue dotted trajectory is the thick black trajectory based on 22-day rolling-window estimate of beta based on daily closing prices. It seems that our MTSRK-based beta using only intra-daily information seems consistent with the black trajectory which use the additional information across days. RK and RCOV-based betas are systematically below it. All three high-frequency estimates are systematically below the black trajectory before the late 1999. We should recall that in this initial period the futures market led the spot market by a wide margin. This suggests that even our MTSRK may
not be able to capture the entire covariation due to the clear lead-lag structure. In their simulation works, BN HLS (2008b) reported that RK-based covariance is systematically below the low-frequency estimator based on open-to-close variation. Open-to-close variation or the close-to-close variation is relatively free of this possible downward bias due to the intra-daily lead lag structure. The issue of possibly non-trivial lead-lag structure is, though very close to, different from the non-synchronous trading because even if we have completely synchronized data we may still face the lead-lag structure. This issue is regarding the deep linkage between two different markets, and it is quite challenging for the high frequency volatility estimation. However, our MTSRK is surely one of the competitive methods to estimate the dynamic hedge ratio. The difference between MTSRK and RK is more clear in the next figure.

![Graph of difference between TSRK-based beta and RK-based beta](image)

The difference is positive in general, that is fairly consistent with our theoretical prediction. Two big downward spikes correspond to the 17-Sep-2001 and 22-Jul-2004. The former is the first trading day since 11-Sep-2001. The following is the plot of intra-daily evolution of the prices within those days.
Figure 3: Big Downward Spikes: Intra-Daily Analysis. (Red: Futures, Blue: Spot)

It is now clear that staggered trajectories may cause the trouble for our MTSRK approach. Therefore, our MTSRK requires more careful data cleaning procedure. In the next figure is the plot of $\hat{\rho}$ based on MTSRK (top) and the $MTSRK^+$ (middle), and their difference (bottom). As is clear, it is very hard to distinguish between the one based on $TSRK$ and another based on $MTSRK^+$ except for few outliers. In other words, the effect of non-negativity correction is minor if $TSRK$ produces a positive estimate. In the bottom panels, we put circles on the largest spikes in the differences, and corresponding values based on TSRK are also indicated in the upper and middle panels. It is now clear that these big spikes are corresponding to the excellent performance of MTSRK to keep the covariation matrix positive semi-definite so that the correlation does not exceed one.

The estimated trajectories of dynamic hedge ratio may not be so useful directly for actual hedging strategy because usually the re-balancing of the futures position incurs the transaction cost such as the brokerage fee or the clear-housing fee. In other words, such a high-frequent re-balancing may not be optimal in terms of the total economic gain net of the transaction cost. We can assess this statistically by constructing a confidence band for $MTSRK^+$-based beta.

From this figure, we can infer that statistically the gain of dynamic hedging strategy over the simplest strategy corresponding to $\beta = 1$ is small. In fact, the number of days in which
Figure 4: Correlation based on TSRK and MTSRK and their difference
Figure 5: MTSRK-based $\beta$ and 95% Confidence Band by Limiting Approximation
the confidence band including one is 1386, the band above one is 3 and below one is 943, all out of 2332. This means that in more than half of the sample the dynamic hedging strategy is indistinguishable from the simplest strategy.

9.3 Economic Gains of Dynamic Hedging Strategy

[Bandi, Russell and Liu 2008]
9.4 Flight-to-Quality from SP500 to 30 year T-bond

In this subsection we apply our method to the issue of flight-to-quality from SP500 index futures market to the 30-year T-bond futures market. The next figure shows this.

![Correlation between SP and TB](image)

**Figure 6: Time-Varying Correlation b/w TB futures and SP500 futures based on Realized Covariance (Top), RK (Middle) and TSRK (Bottom)**

Now we can see another example that the trajectory of estimated ratio based on TSRK has a larger absolute magnitude than those based on other estimators. In other words, the linkage between SP500 market and TB market may be stronger than we might consider based on other estimators. Moreover, its relation is far from constant. In particular, we can observe the drastic change of the regime from positive correlation to the negative correlation in the middle of the sample. This is the change from Oct-22-1997 (1887th, pre-shift) to Oct-23-1997 (1888th, post-shift). Since then, the correlation between these two markets stays in the negative region in general. Moreover, the fluctuation of the trajectory becomes wilder than it used to be. We have to be careful of this latter claim because the trajectory based on TSRK seems wilder than others in general. To draw a shaper conclusion about the possibility of the change in the volatility-of-correlation, we need to improve the efficiency of our estimator. Regardless of a type of estimator we used, however, we equally observed the level shift of the correlation from Oct-22-1997 to Oct-23-1997.
In the next figure we plot the intra-daily evolutions of prices on Oct-22-1997 and Oct-23-1997. We can see immediately that these two security prices exhibit totally different behavior on 22nd and 23rd.

![Graph](image)

*Figure 7: Intra-Daily Evolution of SP (blue) and TB (green).*

Finally, the next figure compares our result with the estimates based on daily returns using a 22-day rolling window OLS method as employed in Bansal et. al. (2008). We can see immediately that the trajectory based on TSRK seems closer to the one based on the inter-daily procedure. All high frequency estimates can detect the acute change such as the one explained above earlier than the inter-daily estimates, of course. Another property is the locally systematic difference of the signs of estimated correlations. For example, the rolling-window correlation around May 01 shows a positive sign systematically, but all high frequency estimates show negative estimates in general. This discrepancy of the estimated signs may cause a great impact on the issue of the flight-to-quality. It is surprising that the trajectory of the rolling-window estimates are sometimes sticking out of the other high frequency estimates. We do not know if this is the evidence of the poor performance of the daily estimates or the evidence of the possibly downward bias in high frequency estimates. [Need Backup by Simulation]
Figure 8: Correlation Estimates: Intra-Daily (high-frequency) vs. Inter-Daily (22-day rolling window)
10 Conclusion

In this article we proposed a bias correction of multivariate realized kernel estimator of the covariation matrix of multiple security prices in the presence of market microstructure noise with unknown serial dependence. The bias correction is indispensable in the multivariate setting because in most of the cases we want to apply a non-linear transformation of the estimated covariation matrix such as the correlation coefficient or the beta coefficient. The estimates of these ratios based on other approaches in the literature confront a serious downward bias in magnitude while those based on our approach are more accurate and relatively smaller bias in finite samples according to simulation studies. When applied to actual intra-daily dataset of SP500 index, its futures and 30-year T-bond futures, we found that those theoretical prediction also applies in practice as well: the estimates of those ratios based on our method are greater in magnitude than those based on other biased methods, and consequently the degree of linkage among those markets seems much tighter.

There are several issues to be addressed in the future version of this article. In particular, we have to deal with the fact that our estimator seems to produce wilder trajectory than those based on other approaches. Because our estimator has two more components in the asymptotic variance compared to the one without bias correction, this should be theoretically conceivable. However, the actual trajectory seems subject to more variability. We suspect this as a consequence of employing the equi-distance calendar-time sampling because it discards a large portion of the data available. In the future version, we should inquire theoretically the possibility or impossibility of efficiency gain by subsampling, resampling or some imputation method popular in the context of semi-parametric missing data analysis. We will update the current version of this article in the near future by incorporating the analysis of these factors.
References


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.


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


http://personal.lse.ac.uk/KALNINA/jmp.pdf


Appendix

Proof of 4

\( \hat{\Omega}(G) \) is characterized as follows:

\[
\hat{\Omega}(G) \sim (|k''(0)| h^{-2})^{-1} QV + \Omega + |k''(0)|^{-1} O_p(n^{-1}G^2 [n^{-1}G + nG^{-3} + G^{-1}]^{1/2})
\]

If we estimate \( QV \) by the realized kernel \( RK(H) \) and subtract the bias term from \( \hat{\Omega}(G) \),

\[
TS\hat{\Omega}(G, H) \sim -(|k''(0)| h^{-2})^{-1} (RK(H) - QV) + \Omega + |k''(0)|^{-1} O_p(n^{-1}G^2 [n^{-1}G + nG^{-3} + G^{-1}]^{1/2})
\]

Recall that \( RK(H) - QV \) is characterized as follows:

\[
RK(H) - QV \sim |k''(0)| h^{-2} \Omega + O_p(n^{-1}H + nH^{-3} + H^{-1})^{1/2})
\]

Substituting this in the first term of the \( TS\hat{\Omega}(G, H) \), we have

\[
TS\hat{\Omega}(G, H) \sim (1 - G^2H^{-2})^{-1} \Omega + O_p \left( n^{-1}G^2 \left\{ (n^{-1}H + nH^{-3} + H^{-1})^{1/2} + (n^{-1}G + nG^{-3} + G^{-1})^{1/2} \right\} \right)
\]

where we should remember that \( O_p \) terms are divided by \( |k''(0)| \). As a result, we can define the finite sample correction of \( TS\hat{\Omega}(G, H) \):

\[
TS\hat{\Omega}^*(G, H) = (1 - G^2H^{-2})^{-1} TS\hat{\Omega}(G, H)
\]

This implicitly assumes that \( G \neq H \). Let \( G \propto n^g \) and \( H \propto n^h \) so that \( g \neq h \). For consistency, we need to select \( g, h > 0 \). Since we cannot increase the bandwidth faster than the total sample size, \( g, h \leq 1 \). Moreover, for \( G^2H^{-2} \) to be asymptotically negligible, we need \( g < h \). In other words, we should limit our attention to the intersection of \( (0, 1] \times (0, 1] \) and \( h > g \) in the \( (g, h) \) space.

Let us focus on the \( O_p \) terms of the asymptotic standard deviation:

\[
TS\hat{\Omega}^*(G, H) - \Omega \sim O_p \left( n^{-1}G^2 \left\{ (n^{-1}H + nH^{-3} + H^{-1})^{1/2} + (n^{-1}G + nG^{-3} + G^{-1})^{1/2} \right\} \right)
\]

\[
\sim O_p \left( \left\{ n^{4g+h-3} + n^{4g-3h-1} + n^{4g-h-2} \right\}^{1/2} + \left\{ n^{5g-3} + n^{g-1} + n^{3g-2} \right\}^{1/2} \right)
\]

- Consistency. \( TS\hat{\Omega}^*(G, H) \) is consistent for \( \Omega \) if all exponents are strictly negative. Given \( g, h > 0 \) and \( h > g \), these conditions reduce to intersection of \( (0, 1] \times (0, 1] \), \( h > g \) and \( h < -4g + 3 \). These define the semi-open area with four vertices \( (0, 0), (0, 1), (1/2, 1) \) and \( (3/5, 3/5) \). In particular, \( g \) can take a value greater than \( 1/2 \) as long as it is smaller than \( 3/5 \). Recall that \( \Omega(G) \) in Ikeda (2009) was consistent for \( \Omega \) only if \( g < 1/2 \). Therefore, the bias correction allows us to select the rate of increase of the main bandwidth \( G \) for LIV more flexibly.

- Asymptotic Normality. We should notice immediately that the asymptotic variance is increasing w.r.t. \( g \), so in the end we will see that it is better to select smaller \( g \).

For the first block in the asymptotic variance

\[
O_p \left( \left\{ n^{4g+h-3} + n^{g-1} + n^{3g-2} \right\}^{1/2} \right),
\]

it is straightforward to show that

\[
h > 1/2 \iff 1st > 3rd > 2nd,
\]

42
the opposite relation if \( h < 1/2 \), and all terms are balanced if \( h = 1/2 \). Similarly, for the second block

\[
O_p\left(\left\{n^{5g-3} + n^{g-1} + n^{2g-3}\right\}^{1/2}\right).
\]

it is straightforward to show that

\[
g > 1/2 \iff 1st > 3rd > 2nd,
\]

the opposite relation if \( g < 1/2 \), and all terms are balanced if \( g = 1/2 \). Since \( h > g \), we only have to consider the following five cases:

1. \( g > 1/2, h > 1/2 \): in this case, we should compare \( 4g + h - 3 \) and \( 5g - 3 \). Given \( h > g \), \( 4g + h - 3 \) is larger. Therefore, in this region, the leading component of the asymptotic variance is \( n^{4g+h-3} \) due to the discretization error of \( \int \Sigma ds \) in the bias.

2. \( g = 1/2, h > 1/2 \): in this case, three components in the second block are balanced, with common order \( n^{-1/4} \). \( n^{(4g+h-3)/2} \) becomes \( n^{(h-1)/2} \), which dominates \( n^{-1/4} \) because \( h > 1/2 \). However, for faster rate of convergence, we need smaller \( h \). This means that \( h \) should be as close as possible to \( 1/2 \). Since \( h > g = 1/2 \), there is indeterminancy of the rate of convergence.

3. \( g < 1/2, h > 1/2 \): in this case, we have to compare \( 4g + h - 3 \) and \( g - 1 \). It turns out that \( 4g + h - 3 \) is dominant for \( h > -3g + 2 \), \( g - 1 \) dominant for \( h < -3g + 2 \) and they are balanced for \( h = -3g + 2 \). In either case, however, we should select smaller \( g \).

4. \( g < 1/2, h = 1/2 \): in this case, three components in the first block are balanced with the common order \( n^{4g-5/2} \). Therefore, we have to compare \( 4g - 5/2 \) and \( g - 1 \). Given \( g < 1/2 \), \( g - 1 \) dominates. Again, we should select smaller \( g \).

5. \( g < 1/2, h < 1/2 \): in this case, we have to compare \( 4g - 3h - 1 \) and \( g - 1 \). Given \( h > g \), \( g - 1 \) is dominant. Again, we should select smaller \( g \).

Although there are rates of convergence depending on the combination of \( (g, h) \), we should always select smaller \( g \). This is obvious if we recall the asymptotic variance: it is increasing in \( g \). In reality, we are going to estimate the LONG-RUN variance covariance matrix so we need a \( g > 0 \) so that \( G = n^{g} \) increases as the sample size increases.

Q.E.D.

**Proof of Proposition 1**

It is sufficient to prove \( P = n^{a}\left(MTS - TS\right) \sim o_p(1), a > 0 \).

\[
P = n^{a}\Lambda\left(I - A\right) - \sum_{j=0}^{J-1} \bar{A}^{j}\left(I - A\right)^{-1}\Lambda'^{'}
\]

\[
= n^{a}\Lambda\sum_{j=0}^{J-1} \bar{A}^{j}\left(I - A\right)^{-1}\left(I - A\right)\Lambda'^{'}
\]

\[
= \Lambda\sum_{j=0}^{J-1} \bar{A}^{j}\left(a^{1/4} \bar{A}^{j}\right)\Lambda'^{'}
\]

If we recall that \( T S - Q \sim O_p(n^{-a}) \), \( T S \overset{D}{\to} Q > 0 \). This means that \( \Lambda(I - A)\Lambda' > 0 \) asymptotically. Since \( \Lambda \) is non-singular, this implies \( I - A > 0 \), \( \rho(A) < 1 \) and \( \sum_{j=0}^{J-1} \bar{A}^{j} \) well defined asymptotically. In particular, if \( A = \lim_{n \to \infty} \bar{A}, \sum_{j=0}^{\infty} \bar{A}^{j} = (I - A)^{-1} \) is bounded and positive semi-definite. By the dominated convergence
theorem and continuous mapping theorem,
\[
(\sum_{j=0}^{J-1} A^j)^{-1} \to P I - A > 0
\]
so that \(\sum_{j=0}^{J-1} A^j \sim O_p(1)\). In conjunction with \(\mathbf{A} \xrightarrow{P} \mathbf{A}'\) such that \(\mathbf{A}\mathbf{A}' = p\lim R_1(\mathbf{A})\), the order of \(\mathbf{P}\) is determined by \(n^n A^j\), which in turn is determined by \(n^n \rho(\mathbf{A})^j\) because the eigenvalue of \(A^j\) is given by \(\lambda^j\) where \(\lambda\) is the eigenvalue of \(A\). By exponentiating,
\[
n^n A^j = \exp\left(a \ln n + \ln \rho(\mathbf{A}) \cdot J\right)
\]
Since \(\rho(\mathbf{A}) < 1\) asymptotically, the coefficient of \(J\) in the exponent becomes negative asymptotically. This means that by letting \(J\) grow at the rate faster than \(\ln n\), \(n^n A^j \sim o_p(1)\).

Q.E.D.

10.0.1 Proof of Proposition 2

We first need to state the definition and the properties of a matrix-valued normal distribution. We focus on the square and symmetric case, but the definition applies to the non-square matrix as well. The following definition follows Dinh and Nguyen (1994).

Definition 9 An \(n \times n\) symmetric matrix-valued random variable \(X\) follows an “matrix-valued random variable with mean \(A\) and variance-covariance matrix \(B^T\) (and denoted as \(X \sim N(A, B)\)) iff

1. \(E[X] = A\) element-by-element, and
2. \(B = (B_{i,j})_{i,j=1,...,n}\) is block matrix such that the typical \((i, j)\) block is given by the \(n \times n\) matrix \(B_{i,j} = \text{cov}(X_i, X_j) = E[X_i X_j^T] - E[X_i]E[X_j]\) where \(X_i\) is the transpose of the \(i\)-th raw vector of \(X\).

The second property about \(B\) has the following operational characterization: for any vectors \(a, b, c, d \in \mathbb{R}^n\),
\[
\text{Cov}(a' X b, c' X d) = \text{vec}(\mathbf{B})' \cdot \text{vec}(c d' + d c')/2,
\]
that is adopted in BNHS (2008b). The equivalence comes from the following observation: by the property of Kronecker product, we can show that (Abadir and Magnus (2005) p282)
\[
a' X b = (b' \otimes a') \text{vec}(X) = (b \otimes a)' \text{vec}(X')
\]
where the second equality follows by the relation between Kronecker product and the transpose operator and the fact that \(X\) is assumed to be symmetric. However, since \(a' X b\) is a scalar, we also have
\[
a' X b = b' X' a = (a' \otimes b') \text{vec}(X') = (a \otimes b)' \text{vec}(X')
\]
From this, we can rewrite the above quadratic form as
\[
a' X b = ([a \otimes b + b \otimes a]/2)' \text{vec}(X') = v_{ab}' \text{vec}(X')
\]
where \(v_{ab}\) is as defined in BNHS (2008b). Since \(X' = [X_1, \ldots, X_n]\) where \(X_i\) is the transpose of \(i\)-th raw vector of \(X\),
\[
\text{vec}(X') = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}
\]
Therefore,
\[
\text{Cov}(a' X b, c' X d) = v_{ab}' B v_{cd}
\]

44
Using this property, we can state the asymptotic result on the vecb-representation of two assets. We can represent the diagonal and upper (or lower) off-diagonal elements of the sub-matrix of $X$ corresponding to $i, j$ rows and columns (denoted by $X^{(ij)} \in \mathbb{R}^{2 \times 2}$) by

$$vecb(X^{(ij)}) = \begin{pmatrix} X_{ii} & X_{ij} \\ X_{ji} & X_{jj} \end{pmatrix} = \begin{pmatrix} e_i^T X e_i \\ e_i^T X e_j \\ e_j^T X e_i \\ e_j^T X e_j \end{pmatrix}$$

where $e_i \in \mathbb{R}^d$ is the $i$-th unit vector ($1$ in $i$-th element and zero otherwise). Given $X = n^{1/4}(TSRK - \int \Sigma_d dv)$, the asymptotic variance of $vecb(X^{(ij)})$ is represented in terms of the asymptotic variance covariance block matrix $B$ as follows:

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

where $v = (v_{ii}, v_{ij}, v_{jj}) \in \mathbb{R}^{d \times 3}$ and $v_{ij} = vec(e_i e_j^T + e_j e_i^T) / 2 \in \mathbb{R}^d$. If we focus on particular pair of two assets $(i, j)$, we can derive more operational rule to calculate this asymptotic variance covariance matrix for $X_{ii}$, $X_{ij}$ and $X_{jj}$ as follows. Given two unit vectors $e_1, e_2 \in \mathbb{R}^2$,

$$vecb(X^{(ij)}) = \begin{pmatrix} e_1^T X^{(ij)} e_1 \\ e_2^T X^{(ij)} e_2 \end{pmatrix}$$

Therefore, its asymptotic variance is

$$\begin{pmatrix} v_{11} \\ v_{12} \\ v_{22} \end{pmatrix} Avar(X^{(ij)})(v_{11}, v_{12}, v_{22})$$

where

$$v \equiv (v_{11}, v_{12}, v_{22}) = \begin{pmatrix} vec(e_1 e_1^T), vec(e_1 e_2^T + e_2 e_1^T) / 2, vec(e_2 e_2^T) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 1 \end{pmatrix}$$

and $Avar(X^{(ij)})$ is the block matrix $(B_{ij}^{pq})_{p,q=1,2}$ where typical element $B_{ij}^{pq}$ is $2 \times 2$ matrix of variance covariance matrix between $p$-th raw vector and $q$-th raw vector in $X^{(ij)}$. Since $X^{(ij)} = \begin{pmatrix} X_{ii} & X_{ij} \\ X_{ji} & X_{jj} \end{pmatrix}$ and we take $X = n^{1/4}(TSRK - \int \Sigma_d dv)$, we have

$$Avar(X^{(ij)}) = c \cdot 4k^{00} \int_0^1 \Lambda_v^{(ij)} \otimes \Lambda_v^{(ij)} dv + c^{-3} \cdot 4k^{22} \Omega^{(ij)} \otimes \Omega^{(ij)} + c^{-1} \cdot 8k^{11} \Omega^{(ij)} \otimes \int_0^1 \Lambda_v^{(ij)} dv$$

Combine everything together, the asymptotic variance of $vecb(X^{(ij)})$ is given by

$$\int_0^1 \left[ c \cdot 4k^{00} \left( v^T (\Lambda_v^{(ij)} \otimes \Lambda_v^{(ij)}) v \right) ds + c^{-3} \cdot 4k^{22} \left( v^T (\Omega^{(ij)} \otimes \Omega^{(ij)}) v \right) + c^{-1} \cdot 8k^{11} \left( v^T (\Omega^{(ij)} \otimes \Lambda_v^{(ij)}) v \right) \right] ds \quad (4)$$

45
The first term is identical to the one derived in BN HLS (2008b). Therefore, the novel feature of our multivariate TSRK is given by the second and third terms. Here is the list of complete elements:

\[
\begin{align*}
u' \left( \Sigma^{(ij)}_s \otimes \Sigma^{(ij)}_s \right)^v &= \left( \begin{array}{ccc}
\Sigma^2_i & \Sigma_i \Sigma_{ij} & \Sigma^2_j \\
\Sigma_i \Sigma_{ij} & \frac{1}{2}(\Sigma_i \Sigma_{jj} + \Sigma^2_j) & \Sigma_{jj} \Sigma_{ij} \\
\Sigma_{jj} \Sigma_{ij} & \frac{1}{2}(\Sigma_{jj} \Sigma_{ii} + \Sigma^2_i) & \Sigma_{ii} \Sigma_{ij} \\
\end{array} \right), \\
u' \left( \Omega^{(ij)}_s \otimes \Omega^{(ij)}_s \right)^v &= \left( \begin{array}{ccc}
\Omega^2_i & \Omega_i \Omega_{ij} & \Omega^2_j \\
\Omega_i \Omega_{ij} & \frac{1}{2}(\Omega_i \Omega_{jj} + \Omega^2_j) & \Omega_{jj} \Omega_{ij} \\
\Omega_{jj} \Omega_{ij} & \frac{1}{2}(\Omega_{jj} \Omega_{ii} + \Omega^2_i) & \Omega_{ii} \Omega_{ij} \\
\end{array} \right), \\
u' \left( \Omega^{(ij)}_s \otimes \Sigma^{(ij)}_s \right)^v &= \left( \begin{array}{ccc}
\Omega_{ii} \Sigma_{ij} & \frac{1}{2}(\Omega_{ii} \Sigma_{jj} + \Omega_{ij} \Sigma_{ii}) & \frac{1}{2}(\Omega_{jj} \Sigma_{ii} + \Omega_{ij} \Sigma_{jj}) \\
\frac{1}{2}(\Omega_{ii} \Sigma_{jj} + \Omega_{ij} \Sigma_{ii}) & \Omega_{ii} \Sigma_{ij} + 2\Omega_{ij} \Sigma_{jj} & \Omega_{jj} \Sigma_{ij} + \Omega_{ij} \Sigma_{jj} \\
\frac{1}{2}(\Omega_{jj} \Sigma_{ii} + \Omega_{ij} \Sigma_{jj}) & \Omega_{jj} \Sigma_{ij} + \Omega_{ij} \Sigma_{jj} & \Omega_{ii} \Sigma_{ij} + \Omega_{ij} \Sigma_{jj} \\
\end{array} \right),
\end{align*}
\]

where we suppressed the time index \( s \) of \( \Sigma \) for ease of notation.²

Now it is straightforward to derive the asymptotic variance of the limiting distribution of ratio such as the TSRK-based realized correlation or the realized regression. High-frequency correlation between \( i \)-th and \( j \)-th assets is defined by

\[
\rho(i, j) = \left( \int_0^1 \Sigma_{ij}(s) ds \cdot \int_0^1 \Sigma_{jj}(s) ds \right)^{-1/2} \int_0^1 \Sigma_{ij}(s) ds
\]

Correspondingly, the (TSRK-based) realized correlation between \( i \) and \( j \) is defined as

\[
\hat{\rho}(i, j) = \left( T \text{SRK}_{ii} \cdot T \text{SRK}_{jj} \right)^{-1/2} T \text{SRK}_{ij}
\]

They take the form of the second element divided by the square root of the product of the first and third elements in \( vech(X^{(ij)}) \). In other words, we need to calculate the gradient vector of the function

\[
f : \mathbb{R}^3 \ni x = (x_1, x_2, x_3) \mapsto f(x) = (x_1 \cdot x_3)^{-1/2} x_2 \in \mathbb{R}
\]

which is given by

\[
\nabla f(x) = \begin{pmatrix}
-x_1^{-3/2} x_2^{-1/2} / 2 \\
-x_1^{-1/2} x_3^{-3/2} / 2 \\
-x_1^{-1/2} x_2^{-1/2} / 2
\end{pmatrix} = \begin{pmatrix}
-f(x) / (2 x_1) \\
-f(x) / (2 x_2) \\
-f(x) / (2 x_3)
\end{pmatrix} = \rho(i, j) \begin{pmatrix}
-(2 \int_0^1 \Sigma_{ii}(s) ds)^{-1} \\
-2 \int_0^1 \Sigma_{ij}(s) ds^{-1} \\
-(2 \int_0^1 \Sigma_{jj}(s) ds)^{-1}
\end{pmatrix} = \rho(i, j) \theta
\]

The last equality follows by evaluating \( x \) at the population value, i.e. \( vech(\int_0^1 \Sigma^{(ij)}(s) ds) \).

Similarly, we can consider the regression coefficient (or “beta” in financial context) as another ratio depending on the variance covariance block matrix. High-frequency beta between \( i \)-th and \( j \)-th assets (regression of \( j \)-th log prices to \( i \)-th log prices) is defined by

\[
\beta(i, j) = \left( \int_0^1 \Sigma_{ii}(s) ds \cdot \int_0^1 \Sigma_{jj}(s) ds \right)^{-1} \int_0^1 \Sigma_{ij}(s) ds
\]

Correspondingly, the (TSRK-based) realized beta of regression of \( j \)-th to \( i \)-th log prices is defined as

\[
\hat{\beta}(i, j) = \left( T \text{SRK}_{ii} \right)^{-1} T \text{SRK}_{ij}
\]

²Notice that BN HLS (2008b; July 25, 2008 version) has a small typo in the asymptotic variance B of the \( \beta^{(ij)} \) in p12. The (1,2) element of the matrix inside the integral should be \( 2 \Sigma_{jj} \Sigma_{ij} \), not \( 2 \Sigma_{ii} \Sigma_{ij} \).
They take the form of the second element divided by the first element in $\text{vech}(X^{(i,j)})$. In other words, we need to calculate the gradient vector of the function

$$g : \mathbb{R}^3 \ni x = (x_1, x_2, x_3)^t \mapsto g(x) = (x_1)^{-1}x_2 \in \mathbb{R}$$

which is given by

$$\nabla g(x) = \left(\begin{array}{c}
-x_1^{-2}x_2 \\
x_1^{-1} \\
0
\end{array}\right) = \left(\begin{array}{c}
-g(x)/x_1 \\
g(x)/x_2 \\
0
\end{array}\right) = \beta(i, j) \left(\begin{array}{c}
-(\int_0^t \Sigma_i(s)ds)^{-1} \\
(\int_0^t \Sigma_{ij}(s)ds)^{-1} \\
0
\end{array}\right) \equiv \beta(i, j)\nu$$

Using the results so far, we can show the following result:

**Proposition 3** (Asymptotic Normality of the TSRK-based Realized Correlation)

The TSRK-based realized correlation and realized beta have the asymptotic distributions

$$n^{1/4} \left(\hat{\rho}(i, j) - \rho(i, j)\right) \xrightarrow{d} MN\left(0, \rho(i, j)^2 \theta^2 \text{Av}(\text{vech}(X^{(i,j)}, c)\theta)\right)$$

$$n^{1/4} \left(\hat{\beta}_{ij} - \beta_{ij}\right) \xrightarrow{d} MN\left(0, \beta_{ij}^2 \theta^2 \text{Av}(\text{vech}(X^{(i,j)}, c)\nu)\right)$$

where $\text{Av}(\text{vech}(X^{(i,j)}), c)$ is the asymptotic variance of $\text{vech}(X^{(i,j)})$ defined by (4).

(Proof) Straightforward by the delta method.

The Asymptotic MSE of RK-based ratios and optimal $c$

Using our notation and the result above, the asymptotic normality of the $i,j$-submatrix of the multivariate RK estimator is expressed as follows:

$$n^{1/5} \left(\text{vech}(\text{RK}^{(i,j)}) - \text{vech}\left(\int \Sigma^{(i,j)}(s)ds\right)\right) \xrightarrow{d} MN\left(e^{-2}[\mu'(0)\text{vech}(\Omega^{(i,j)}), c \cdot 4k^{00} \int_0^1 [\mu'(\Sigma_s^{(i,j)} \otimes \Sigma_s^{(j,i)}\nu)]ds\right)$$

where more concrete expression of $\nu'((\Sigma_s^{(i,j)} \otimes \Sigma_s^{(j,i)}))\nu$ was given previously.

Now we can derive the asymptotic distribution of the ratios. First, the gradient vector of the transformation defining the realized beta $\beta_{ij} := \int \Sigma_{ij}/\int \Sigma_{ii}$ is given by

$$\nabla g(x) = \left(\begin{array}{c}
-x_1^{-2}x_2 \\
x_1^{-1} \\
0
\end{array}\right) = \left[\int \Sigma_{ii}\right]^{-1} \left(\begin{array}{c}
-\beta_{ij} \\
1 \\
0
\end{array}\right)$$

where the last equality follows by evaluating $x = \int \Sigma^{(i,j)}$. The delta method then gives the following:

$$n^{1/5}(\hat{\beta}_{ij} - \beta_{ij}) \xrightarrow{d} MN(e^{-2}B_{\beta}, cV_{\beta})$$

The asymptotic bias and variance are given by

$$B_{\beta} = -[\mu'(0)](R_{ii} - R_{ij})\beta_{ij},$$

$$V_{\beta} = 4k^{00} \left\{(Q_{ii,ii} - 2Q_{ii,ij})\beta_{ij}^2 + (Q_{ii,ij}\beta_{ij}\beta_{ij}^{-1} + Q_{ij,ij}\beta_{ij}^2/2 \right\}.$$  

where $R_{ij} = \Omega_{ij}/\int \Sigma_{ij}$ and $Q_{ij,kl} = \int \Sigma_{ij} \Sigma_{kl}/(\int \Sigma_{ij} \int \Sigma_{kl})$. Although there is no restriction on the sign of $R_{ii} - R_{ij}$, it is likely to be positive. If the covariance matrix is constant, all $Q$’s are one so that the asymptotic variance reduces to $(\beta_{ij}\beta_{ij}^{-1} - \beta_{ij}^2)/2$. Similarly, we can deduce the asymptotic normality and its asymptotic bias and variance of the realized regression of $j$-th security to the $i$-th security (this is stated in BNHLS (2008b)). It has the identical expression except that $i$ and $j$ are interchanged.
Finally, the bias and variance of the asymptotic distribution of the RK-based $\hat{\rho}_{ij}$ are

$$n^{1/5}(\hat{\rho}_{ij} - \rho_{ij}) \overset{d}{\rightarrow} MN(c^{-2}B_{\rho}eV_{\rho})$$

where

$$B_{\rho} = -|k(0)|(|R_{ii} - 2R_{ij} + R_{jj}|)/2,$$

$$V_{\rho} = 4k^0 \left[ Q_{ii,ij}/2 + \left\{ (Q_{ii,ij} + Q_{jj,ij})/4 - (Q_{ii,ij} + Q_{jj,ij} + Q_{ij,ij}/2)\right\}\rho^2 + Q_{ij,ij}\rho^4/2 \right]$$

Therefore, the asymptotic MSE of RK-based ratio $\hat{\beta}_{ij}$ is given by

$$MSE_{\beta} = c^{-2}B_{\beta}^2 + eV_{\beta}$$

and similarly for $MSE_{\rho}$. This is clearly convex in $c$ so we can find the global minimizer from the first order condition, which is

$$c_0^* = (4B_{\rho}^2/V_{\rho})^{1/5} = \left( \frac{|k(0)|^2}{k^0} \right)^{1/5} \cdot \frac{(R_{ii} - R_{ij})^2}{2Q_{ii,ij} + \left\{ (Q_{ii,ij} + Q_{jj,ij}) - 4(Q_{ii,ij} + Q_{jj,ij} + Q_{ij,ij})\rho^2 + 2Q_{ij,ij}\rho^4/2 \right\}^{1/5}}$$

Similarly,

$$c_0^* = (4B_{\rho}^2/V_{\rho})^{1/5} = \left( \frac{|k(0)|^2}{k^0} \right)^{1/5} \cdot \frac{(R_{ii} - 2R_{ij} + R_{jj})^2}{2Q_{ii,ij} + \left\{ (Q_{ii,ij} + Q_{jj,ij}) - 4(Q_{ii,ij} + Q_{jj,ij} + Q_{ij,ij})\rho^2 + 2Q_{ij,ij}\rho^4/2 \right\}^{1/5}}$$

If we impose Assumption 5, all $Q$'s are one. In this special case which we exploited in the simulation, we have

$$c_0^* \approx \left( \frac{|k(0)|^2}{k^0} \right)^{1/5} \cdot \frac{(R_{ii} - R_{ij})^2}{((\hat{\beta}_{ij})^2 - 1)/2}^{1/5}$$

and

$$c_0^* \approx \left( \frac{|k(0)|^2}{k^0} \right)^{1/5} \cdot \frac{(R_{ii} - 2R_{ij} + 2R_{jj})^2}{2(1 - \rho^2)^2}^{1/5}$$