CONSISTENT ESTIMATOR OF EX-POST COVARIATION OF DISCRETELY OBSERVED DIFFUSION PROCESSES AND ITS APPLICATION TO HIGH FREQUENCY FINANCIAL TIME SERIES

Sujin Park
Department of Statistics
London School of Economics

A thesis submitted for the degree of
Doctor of Philosophy
October 2011

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Acknowledgement

I would like to express gratitude to Professor Oliver Linton for continuous academic and financial support since the start of PhD. He has been a resourceful supervisor and his belief from very early on in my ability to carry out the technically and computationally challenging research has been a motivational factor for me to press on. I would like to thank Professor Qiwei Yao for his thoughtful comments in various stages of my research and giving me an opportunity to learn how to carry out a rigorous statistical research. I thank the Department of Statistics and Department of Economics at London School of Economics for providing the intellectually stimulating environment.

I dedicate my thesis to my father Seuk-kun Park and mother Seun-hye Kim who supported me with unconditional love.

Abstract

First chapter of my thesis reviews recent developments in the theory and practice of volatility measurement. We review the basic theoretical framework and describe the main approaches to volatility measurement in continuous time. In this literature the central parameter of interest is the integrated variance and its multivariate counterpart. We describe the measurement of these parameters under ideal circumstances and when the data are subject to measurement error, microstructure issues. We also describe some common applications of this literature.

In the second chapter, we propose a new estimator of multivariate ex-post volatility that is robust to microstructure noise and asynchronous data timing. The method is based on Fourier domain techniques. The advantage of this method is that it does not require an explicit time alignment, unlike existing methods in the literature. We derive the large sample properties of our estimator under general assumptions allowing for the number of sample points for different assets to be of different order of magnitude. We show in extensive simulations that our method outperforms the time domain estimator especially when two assets are traded very asynchronously and with different liquidity.

In the third chapter, we propose to model high frequency price series by a time-deformed Lévy process. The deformation function is modeled by a piecewise linear function of a physical time with a slope depending on the marks associated with intra-day transaction data. The performance of a quasi-MLE and an estimator based on a permutation-like statistic is examined in extensive simulations. We also consider estimating the deformation function nonparametrically by pulling together many time series. We show that financial returns spaced by equal elapse of estimated deformed

time are homogenous. We propose an order execution strategy using the fitted deformation time.

Contents

List of Tables

List of Figures

Chapter 1

Realized Volatility: theory and application

1.1 Introduction

This chapter reviews some recent developments in the theory and practice of volatility measurement. Volatility is a fundamental quantity for investment decisions. Its measurement is necessary for the implementation of most economic or financial theories that guide such investment. Volatility is also important for assessing the quality of performance of financial markets, with very volatile markets being perceived as not functioning effectively as a way of channeling saving into investment. Despite its importance, volatility is not an easy quantity to measure, and there are many approaches to do that. From the point of view of the investor facing investment opportunities with returns r_t at time t and information $\mathcal{F}_s$ at time $s < t$, one might be interested in the matrix $\text{var}(r_t|\mathcal{F}_s)$. This represents a challenge because the time horizon $t - s$ might be unknown or be stochastic, the information set $\mathcal{F}_s$ might be extremely large containing current and past values of many variables, and the probability distribution $f(r_t|\mathcal{F}_s)$ may be unknown. The recent emphasis on continuous time methods of volatility measurement in some way addresses all of these concerns.

We review the basic theoretical framework and describe the main approaches to volatility measurement in continuous time. In this literature the central parameter of interest is the integrated variance and its multivariate counterpart. We describe the measurement of these parameters under ideal circumstances and when the data are subject to measurement error. We discuss the main types of measurement error models that apply and how they may arise from the way the market operates at the fine grain, i.e., microstructure issues. We also describe some common applications of this literature. Our review is necessarily selective and there are many topics and papers that we do not cover.

1.2 Modeling Framework

1.2.1 Efficient price

We start by setting the modeling framework. Under the standard assumptions that the return process does not allow for arbitrage and has a finite instantaneous mean, the asset price process, as well as smooth transformations thereof, belong to the class of special semi-martingale processes, as detailed by Back (1991). If, in addition, it is assumed that the sample paths are continuous, we have the Martingale Representation Theorem (e.g. Protter (1990)). Specifically, there exists a representation for the log price Y_t, such that for all $t \in [0, T]$,

$$Y_t = \int_0^t \mu_u du + \int_0^t \sigma_u dW_u, \tag{1.1}$$

where μ_u is a predictable locally bounded drift, σ_u is a cádlág volatility process and W_u is an independent Brownian motion, and the integral is of the Itô form. Let Y_{t_j} denote an observed log prices on the time grid, $0 = t_0 < \cdots < t_n = T$, where we take $T = 1$ for simplicity. Note that $\{t_i\}$ is usually assumed to be a non-decreasing deterministic sequence. Crucial to semimartingales, and to the economics of financial risk, is the Quadratic Variation (QV) process. Let Γ_t be a set of points that partition the interval $[0, t]$ with $\Gamma = \Gamma_1$. The quadratic variation of Y over the time interval

$[0, t]$ is given by

$$[Y, Y]_t = \operatorname*{plim}_{\sup_{t_j \in \Gamma_t} |t_j - t_{j-1}| \to 0} \sum_{0 \leq t_j \leq t} (Y_{t_j} - Y_{t_{j-1}})^2, \tag{1.2}$$

with $[Y, Y] = [Y, Y]_1$. This quantity is a measure of ex-post volatility. Under (1.1), the following holds almost surely

$$[Y, Y] = \int_0^1 \sigma_u^2 du. \tag{1.3}$$

The quadratic variation is also called the integrated variance, for obvious reasons. It is the key parameter of interest that this survey will focus on. It is an integral over the sample path of the stochastic process σ_u^2, and hence itself is a random variable. The specification of the process σ_u^2 is very general and nonparametric, i.e., it may depend on the entire past of Y_t and additional sources of randomness. The averaging inherent in (1.3) suggests gains in terms of estimability.

We now relate the parameter of interest to other concepts of volatility. A natural theoretical notion of ex-post return variability in this setting is notional volatility, Anderson, Bollerslev, Diebold, and Labys (2000). Under the maintained assumption of continuous sample path, the notional volatility equals the integrated volatility. The notional volatility over an interval $[t-h, t]$, is

$$\upsilon^2(t, h) \equiv [Y, Y]_t - [Y, Y]_{t-h} = \int_{t-h}^t \sigma_u^2 du.$$

Let $\mathcal{F}_t$ denote information on Y up to and including time t. Now, in the above setting, the conditional volatility, or expected volatility, over $[t-h, t]$, is defined by

$$
\begin{aligned}
\operatorname{var}(Y_t|\mathcal{F}_{t-h}) &\equiv E\left[\{Y_t - E(Y_t|\mathcal{F}_{t-h})\}^2 |\mathcal{F}_{t-h}\right] \\
&= E\left[\left\{\int_{t-h}^{t} \mu_u du - E\left(\int_{t-h}^{t} \mu_u du|\mathcal{F}_{t-h}\right) + \int_{t-h}^{t} \sigma_u dW_u\right\}^2 |\mathcal{F}_{t-h}\right] \\
&= E\left[\left\{\int_{t-h}^{t} \{\mu_u - E(\mu_u|\mathcal{F}_{t-h})\} du\right\}^2 |\mathcal{F}_{t-h}\right] \qquad (1.4)\\
&\quad + E\left[\left\{\int_{t-h}^{t} \sigma_u dW_u\right\}^2 |\mathcal{F}_{t-h}\right] \qquad (1.5)\\
&\quad + 2E\left[\int_{t-h}^{t} \{\mu_u - E(\mu_u|\mathcal{F}_{t-h})\} du \int_{t-h}^{t} \sigma_u dW_u |\mathcal{F}_{t-h}\right]. \qquad (1.6)
\end{aligned}
$$

Denote $A_h = O_{a.s.}(B_h)$ when A_h/B_h converges almost surely to a finite constant as $h \to 0$. We have that (1.4)$= O_{a.s.}(h^2)$, (1.5)$= \int_{t-h}^{t} \sigma_u^2 du = O_{a.s.}(h)$, and (1.6)$= O_{a.s.}(h^{3/2})$, so that (1.5) is the dominant term. Therefore, we have

$$
\operatorname{var}(Y_t|\mathcal{F}_{t-h}) \simeq E[\upsilon^2(t,h)|\mathcal{F}_{t-h}].
$$

In other words, the conditional variance of returns volatility is well approximated by the expected notional volatility, i.e., it is an approximately unbiased proxy. The above approximation is exact if the mean process, $\mu_u = 0$, or if μ_u is measurable with respect to $\mathcal{F}_{t-h}$. However, the result remains approximately valid for a stochastically evolving mean return process over relevant horizons, as long as the returns are sampled at sufficiently high frequencies. This gives further justification for $[Y, Y]$ as a parameter of interest.

Notional volatility or integrated volatility is latent. However, it can be estimated consistently using the so-called Realized Volatility. The Realized Variance (RV) for the time interval $[0, 1]$ is the discrete sum in (1.2);

$$
[Y, Y]^n = \sum_{j=1}^{n} (Y_{t_j} - Y_{t_{j-1}})^2, \qquad (1.7)
$$

where $t = 1$.

Barndorff-Nielsen and Shephard (2002) showed that the RV is a $\sqrt{n}$ consistent estimator of the QV and is asymptotically mixed Gaussian under infill asymptotics. We can also generalize the above specification for the process driven by Lévy process. In this case the Realized Variance converges in probability to the quadratic variation of the process, which includes contributions from the jumps. We discuss estimation further below.

1.2.2 Measurement error

Empirical evidence suggests that the price process deviates from the semimartingale assumption in (1.1). The "volatility signature plot" (which shows (1.7) against sampling frequency) in Figure 1.5 suggests a component in observed price that has an infinite quadratic variation. Previous authors have identified this component as microstructure noise, meaning that it is due to the fine grain structure of how observed prices are determined in financial markets. A common way of modeling this is as follows. Let X_{t_j} be an observed log price and Y_{t_j} be discretely sampled from the process in (1.1). Then suppose that

$$X_{t_j} = Y_{t_j} + \varepsilon_{t_j}, \tag{1.8}$$

where ε_{t_j} is a random error term. The simplest case is where the microstructure noise ε_{t_j} is i.i.d. with zero mean, independent of the process Y. This model was first considered in Zhou (1996). In this case, Zhang, Mykland, and Aït-Sahalia (2005) showed that $RV = 2nE(\varepsilon^2) + O_p(n^{1/2})$, which implies that RV is inconsistent and that divided by $2n$ it is an asymptotically unbiased estimator of the variance of the microstructure noise. The noise can also be assumed to be serially correlated, and there are some theoretical results for this case, which we discuss below. One may want to allow for heteroscedasticity in the noise (1.8), which has been taken up by Kalnina and Linton (2008). This is motivated by the stylized fact in market microstructure literature that intra-daily spreads and intra-daily stock price volatility are described typically by a U-shape (or reverse J-shape).

Also to closely mimic the high frequency transaction data authors considered rounding error noise or non-additive noise that is generated from specific model of order book dynamics. Li and Mykland (2007) discuss the rounding model,

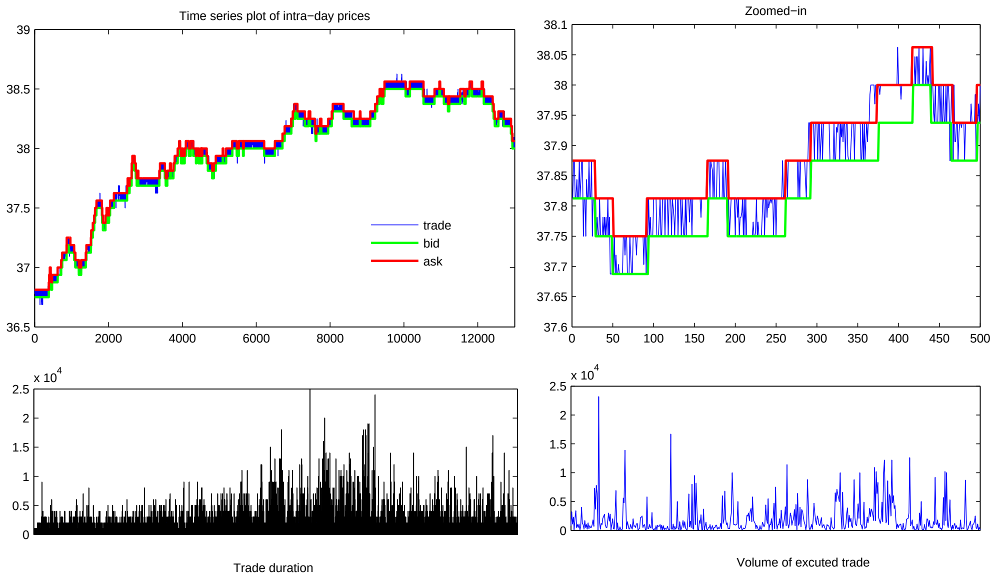
$$X_{t_j} = \log(\delta[\exp(Y_{t_j} + \varepsilon_{t_j})/\delta]) \vee \log \delta, \tag{1.9}$$

where $\delta[s/\delta]$ denotes the value of s rounded to the nearest multiples of δ which is a small positive number. This is consistent with the market that has a minimum price change, tick sizes for stocks and futures and pips for foreign exchange. The rounding model (1.9) is much more complex to work with than (1.8), due to the nonlinear way in which the efficient price enters. For example, even assuming no microstructure noise the quadratic variation of X_t is given by $[f(Y), f(Y)]_t$ where $f(Y) = E(X|Y)$ is a complicated nonlinear function, although we are interested in estimating $[Y, Y]_t$. Li et al. (2007) showed that when var(ε) is large, we have $f(Y_t) \simeq Y_t$, whereas for a small noise variance, the divergence of two quadratic variations can be large. In any case, under the presence of such microstructure noise the Realized Variance is no longer a consistent estimator of the integrated variance. We explore the impact of different microstructure noise assumptions on RV and the class of consistent estimators under (1.1) and (1.8) in Section 1.4.

1.3 Issues in Handling Intra-day Transaction Database

Before examining volatility estimators based on high frequency data, it is important to understand the basic statistical features of such dataset. In this section we provide a brief summary of the stylized features of intra-day transaction data. Goodhart and O'Hara (1997) and Guillaume et al. (1997) provide early reviews. The distributional properties of high frequency returns varies with sampling frequency. At higher frequency, there is a stronger evidence of return distribution being non-Gaussian. The empirical evidence suggests that high frequency returns are approximately symmetric

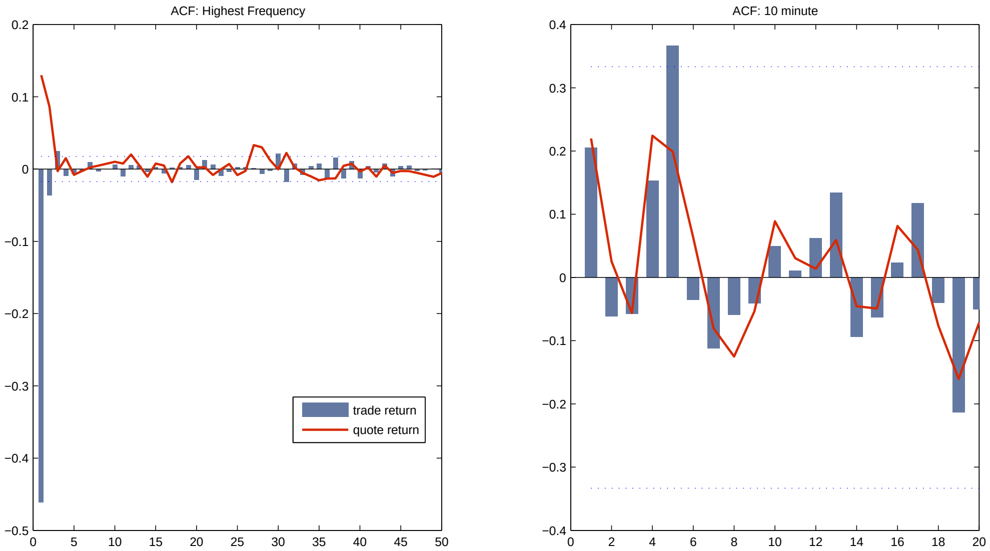
Figure 1.1: Time series of intra-day price, trade duration and volume over a day



with finite second moment but with large fourth moment and the tail of the distribution declines according to a power law. In fact, prices are discrete, taking values that are integer multiples of tick sizes, which vary according to assets and time period (in the US stock market tick size changed from being 1/8 of a dollar to 1/100 of a dollar during a few years at the beginning of the last decade), see Figure 1.1 which plots intra-day price of the Dell stock over a single day. The data we use in this paper is a National Best Bid and Offer (NBBO) trade and quote consolidated dataset from TAQ. This puts together the best available quotes from multiple venues and matches the trades to NBBO quotes. Therefore, trade and quote price dynamics should be indicative of that from the single order book. However, returns, whether defined logarithmically or exactly, are less discrete, since the normalization changes over time, so this comment mostly just affects the study of prices within a single day.

The returns of executed trade prices (trade returns) are negatively serially correlated. This is due to bid-ask bounce: at the tick level, buy orders are likely to be followed by sell orders and vice versa. Absolute returns and trade activity variables such as volume, spread and trade duration exhibit strong serial correlation. Andersen

Figure 1.2: ACF of trade and quote returns at different sampling frequency

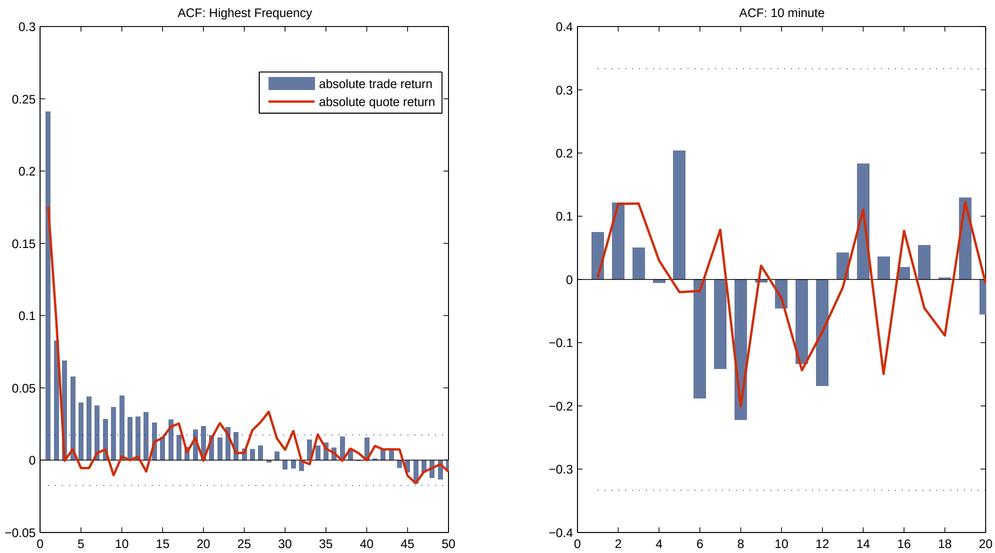


and Bollerslev (1997) showed that the absolute trade returns, after eliminating the short term periodic component, have an hyperbolic decaying autocorrelation function. This can affect the construction of the standard errors and forecasting. Variables associated with transaction activity show periodic patterns due to trading convention. Activities are high at the start and at the end of the trading session and this induces a particular pattern in activity variables. See bottom left of Figure 1.1. Periodicity can be modeled by introducing periodic dummies, frequency domain filtering and analysis at the activity time scale. The intra-day periodicity and long memory structure can be explained by the presence of the information arrival process that drives the price formation process.

1.3.1 Which price to use?

In intra-day we typically have different types of prices. We briefly describe workings of stock market order book. Order book is a collection of sell and buy orders at any point of time, recording a price, time stamp and volume associated with each order. The bid is the maximum buy price and the ask is the minimum selling price. The

Figure 1.3: ACF of absolute trade and quote returns at different sampling frequency



spread is defined by ask minus bid. Depending on the type of the order sent, it either adds to the order book, gets canceled or generates a trade. For example a buy limit order with price above current bid but below current ask tighten the spread. Same order with price below current bid joins the que. Buy order with price specified at the current ask, assuming that the order is filled, takes off the liquidity. This is always the case for a market order. Orders that are stored in the order book are referred as quotes. The quote return is defined by the change in the mid quote, which is an average of bid and ask. The trade returns is the return associated with the price of the executed trade. In terms of time series behavior, trade returns show significant negative first order autocorrelation due to bid-ask bounce. In comparison, quote returns show positive first order autocorrelation in a short interval. See Figure 1.2. If the data is based on the higher frequency sampling, for example at a tick time or one second, the quote and trade price have distinctively different features in returns and in absolute values. The difference disappears in lower frequency sampling. See Figure 1.3, which also shows that absolute returns are quite persistent. In certain cases, we may want to construct a price series that reflects the information at the deeper level of order book and also the volumes of these orders. We may construct such price as a weighted sum of quote prices at different levels of order book where weight is given by the associated volume. Such price construction has the advantage that it uses more information available regarding investor's anticipation of price movement and discreteness is less severe by construction. Related but not the same VWAP (volume weighted average price) can be also used. Over specified period, it is constructed by taking sum of executed trade price weighted by its volume. This quantity is used in a common strategy for execution of large transactions, see Almgren and Chriss (2001) for example.

One of the important conclusions we can draw from the analysis is that in ultra-high frequency, the choice of quote or trade price will sometimes affect the results of empirical modeling. For example, in calculating the naive Realized Variance measure of integrated variance based on low frequency returns, at 10-20 minutes which is a popular choice, the choice of quote or trade returns will not have a discernible impact on the final quantity. However for more recently proposed methods that use un-sampled tick data, we should compare the results using quote and trade returns. See Barndorff-Nielsen, Hansen, Lunde and Shephard (2008b) for such studies.

1.3.2 High frequency data pre-processing

Prior to analysis, the tick data has to be pre-processed to remove non sensible prices and duplicated transaction data points. Barndorff-Nielsen et al. (2008b) provide a guideline to do this for equity intra-day data. Brownlees and Gallo (2006) summarize the structure of the TAQ high frequency dataset and address various issues in high frequency data management including: outlier detection and how to treat non-simultaneous observations, irregular spacing, issues of bid-ask bounce, and methods for identifying exact opening and closing prices. The authors also present the effect of data handling on the result of empirical analysis.

For the market where there is a centralized exchange and trading is electronic the intra-day transaction data should be available easily. The example of such market is equities and commodity futures market. Most empirical work has so far concentrated on NYSE traded stocks and major currencies. Empirical application in other markets - geographically and also other fixed income markets will be of interest.

1.3.3 How to and how often to sample?

Intra-day prices are observed on the discrete and irregular intervals. For volatility estimation one can ask what is the effect of using all the data versus using sparsely sampled data, for example at 10-20 minutes. For covariance estimation, the problem is more substantial. Naturally the estimation of covariance involves the cross product of returns. How should we align the data points observed at a different times and what is the statistical impact of the synchronization method on the estimators? This section discusses two data sampling/alignment method: *fixed clock time* and *refresh time* method. We will present the synchronization method for d number of assets. The sampling method for univariate series is a special case for $d = 1$. In a given interval (for simplicity one day) $[0, 1]$, we observe intra-day transaction prices of the i-th asset, X_i at discrete time points $\{t_{i,j}; j = 0, \dots, n_i\}$ where n_i is a total number of observations on that interval. The set of

$$\{X_{i,t_{i,j}}, t_{i,j}; i = 1, \cdots, d, j = 1, \dots, n_i\},$$

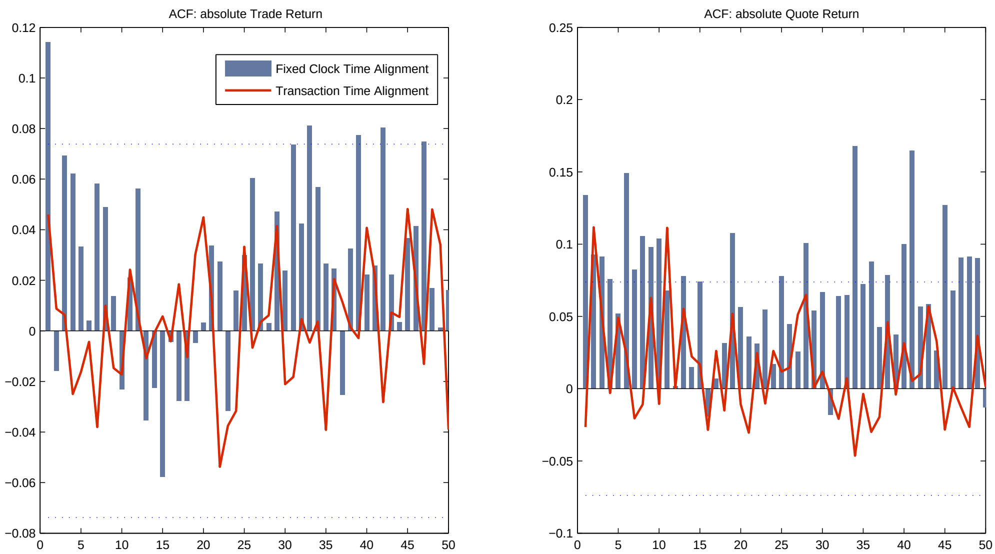
gives us the tick database of prices for d numbers of assets. We can associate the counting process to $\{t_{i,j}\}$

$$N_i(t) := \sum_{j=1}^{n_i} 1(t_{i,j} \leq t),$$

recording the number of transactions that occurred for the i-th asset up to and including the time t. Let $0 = \tau_0 < \cdots < \tau_n = 1$ be an artificially created time grid and let $\{s_{i,j}\}$ be the actual time points of the data for i-th asset to be aligned on the $\{\tau_j\}$'s grid. Regardless of how τ is defined we take the data that is closest to this artificial grid,

$$s_{i,j} = \max_{0 \leq l \leq n_i} \{t_{i,l} \leq \tau_j\}.$$

Figure 1.4: ACF of absolute trade and quote returns sampled by fixed clock time and transaction time



We will denote an aligned dataset as

$$\{X_{i,\tau_j}, \tau_j; i = 1, \cdots, d, j = 1, \cdots, n\},$$

with $X_{i,\tau_j} := X_{i,s_{i,j}}$. If no observation is available during the given interval we repeat the previous data point.

First, consider the problem of sampling scheme for univariate time series of intraday prices. One can use the raw tick data of prices observed at $\{t_{i,j}\}$ or work with instead sparser sampling. One method of sparse sampling is called *fixed clock time*. For example, we might want to create one minute returns from the irregularly spaced tick data

$$\tau_j = jh \ , \ h = 1/60, \tag{1.10}$$

so that $\tau_j - \tau_{j-1} = h$, for all i. Empirical work shows that the effect of microstructure noise become attenuated when return are sparsely sampled. Aït-Sahalia, Mykland and Zhang (2005) derived the optimal sampling rate h minimizing the mean square of the Realized Variance under the presence of i.i.d microstructure noise. When market microstructure noise is present but unaccounted for, they showed that the optimal sampling frequency is finite and derived its closed-form expression. The optimal sampling frequency is often found to be between one and five minutes. See Bandi and Russell (2008) and reference therein for further discussion of the optimal sampling rate in estimating integrated variance. However, modeling the noise and using all the data should yield a better solution, see Section 1.4.1. on the noise robust estimators. The second method for sparse sampling is to sample the price per given number of transactions. For example, data sampled per h number of transactions is

$$\tau_{j+1} = t_{i,N_i(\tau_j)+h}. \tag{1.11}$$

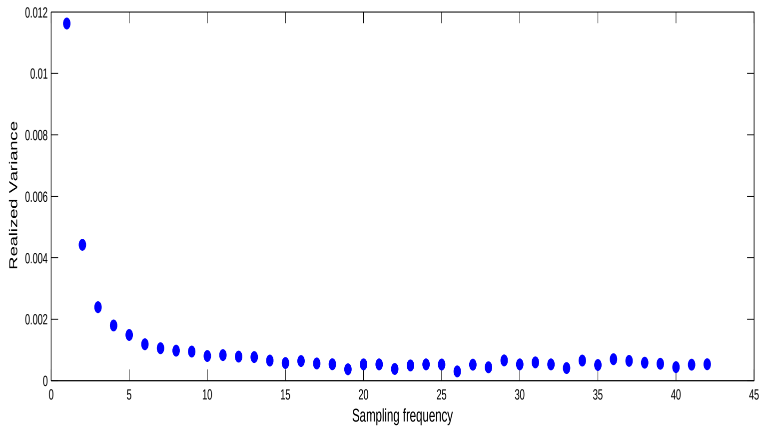
Griffin and Oomen (2008) argued that under the transaction time sampling, returns are less serially correlated and microstructure noise is closer to i.i.d. They note the bias correction procedures that rely on the noise being independent are better implemented in transaction time. Figure 1.4 shows that the ACF of absolute returns at a different sampling scheme - verifying that the transaction time sampling scheme reduces the serial correlation and the process is closer to i.i.d.

For the multivariate case, the additional issue of synchronicity arises, whereby trading for different assets occurs at different times. It is necessary to align the returns of asynchronously traded assets to calculate the covariance estimator that involves the cross product of returns. One method is to use the *fixed clock time* as given in (1.10). Another method, called the *Refresh time*, proposed by Barndorff-Nielsen, Hansen, Lunde and Shephard (2011) can be thought as the multivariate version of the transaction time alignment given in (1.11). It is constructed by

$$\tau_{j+1} = \max_{1 \leq i \leq d} \{t_{i,N_i(\tau_j)+1}\}. \tag{1.12}$$

As we sample the returns at higher frequency, zero returns (stale price) induce the downward bias in covariance estimators. This is known as the Epps effect. Hayashi and Yoshida (2005) showed analytically the bias induced by the *fixed clock time* assuming independent homogenous Poisson process for $N_i(t)$. The *refresh time* also induces synchronization bias and the problem is more severe for a high dimension

Figure 1.5: Realized variance calculated at different calendar time frequencies



covariance estimation since the method effectively collects the transaction time of the most illiquid asset. See Zhang (2010) for further studies on the refresh time bias and its effect on the time domain based estimator of integrated covariance matrix. See Section 1.4.2 for a discussion of covariance estimator robust to the synchronization bias.

1.4 Realized variance and covariance

1.4.1 Univariate volatility estimators

We first present the results for realized volatility in the perfect world where there is no measurement error. The case of no noise is dealt with by Andersen, Bollerslev, Diebold and Labys (2001), Barndorff-Nielsen and Shephard (2002), and Mykland and Zhang (2006). Barndorff-Nielsen et al. (2002) showed that the error using the RV to estimate the QV is asymptotically normal with rate $\sqrt{n}$, i.e.,

$$\sqrt{n}\frac{\sum_j y_{t_j}^2 - \int_0^1 \sigma_u^2 du}{\sqrt{2\int_0^1 \sigma_u^4 du}} \Longrightarrow N(0,1),$$

where $y_{t_j} = Y_{t_j} - Y_{t_{j-1}}$ is the observed return and $\Longrightarrow$ denotes convergence in distribution. We remark that their proof does not require that $Ey_{t_j}^4 < \infty$ or even $Ey_{t_j}^2 < \infty$ as would generally be the case for a central limit theorem to hold. The reason is that the data generating process assumes a different type of structure, namely that locally the process is even Gaussian, and it is this feature that permits the arrival of the normal distribution in the limit. Note that this CLT is statistically infeasible since it involves a random unknown quantity called integrated quarticity (IQ), $\int_0^1 \sigma_u^4 du$. However we can consistently estimate this by the following sample quantity

$$\widehat{IQ} = \frac{n}{3} \sum_j y_{t_j}^4 \to_p IQ.$$

Therefore, the feasible CLT is given by

$$\sqrt{n} \frac{\sum_j y_{t_j}^2 - \int_0^1 \sigma_u^2 du}{\sqrt{2\widehat{IQ}}} \Longrightarrow N(0,1).$$

This implies that $\sum_j y_{t_j}^2 \pm z_{\alpha/2} \sqrt{\frac{2}{3} \sum_j y_{t_j}^4}$ gives a valid α-level confidence interval for $\int_0^1 \sigma_u^2 du$.

Measurement Error

Motivated by some of the issues observed in the intra-day financial time series largely to do with the presence of microstructure noise, authors have proposed competing estimators of the QV. The assumption on a microstructure noise has been generalized from a white noise to a noise process with some of following characteristics: autocorrelation, heteroscedasticity, rounding models. McAleer and Medeiros (2008) provide a summary of the theoretical properties of different estimators of QV under different assumptions of microstructure noise.

Suppose that the efficient prices process is given by (1.1) and we observe (1.8). In this case, the Realized Variance is inconsistent. The first consistent estimator under this scheme was the two time scale estimator (TSRV) of Zhang, Mykland and Aït-Sahalia (2005). Split the sample of size n into K subsamples, with the i^{th} subsample containing n_i observations. Let $[X,X]^{n_i}$ denote the i^{th} subsample estimator based on a K-spaced subsample of size n_i, and let $[X,X]^{avg}$ denote the averaged estimator:

$$
\begin{aligned}
[X,X]^{n_i} &= \sum_{j=1}^{n_i-1} \left(X_{t_{jK+i}} - X_{t_{(j-1)K+i}} \right)^2, \ i = 1, \ldots, K, \\
[X,X]^{avg} &= \frac{1}{K} \sum_{i=1}^{K} [X,X]^{n_i}.
\end{aligned}
$$

To simplify the notation, we assume that n is divisible by K and hence the number of data points is the same across subsamples, $n_1 = n_2 = ... = n_K = n/K$. Let $\overline{n} = n/K$. Define the adjusted TSRV estimator as

$$
\widehat{[X,X]} = [X,X]^{avg} - \left(\frac{\overline{n}}{n} \right) [X,X]^n. \tag{1.13}
$$

Zhang et al. (2005) show that this estimator is consistent and show that

$$
\frac{n^{1/6} \left(\widehat{[X,X]} - [X,X] \right)}{\sqrt{8c^{-2}E^2\varepsilon^2 + \frac{4}{3}cIQ}} \Longrightarrow N(0,1),
$$

provided that $K = cn^{2/3}$ for any $c \in (0, \infty)$. Zhang (2006) extended this work to the multiscale estimator (MSRV). She shows that this estimator is more efficient than the two time scale estimator and achieves the best convergence rate of $O_p(n^{1/4})$, (i.e., the same as the MLE with complete specification of the observed process).

Kalnina and Linton (2008) proposed a modification of the TSRV estimator that is consistent under heteroscedasticity and endogenous noise. Aït-Sahalia, Mykland and Zhang (2010b) modified TSRV and MSRV estimators and achieve consistency in the presence of serially correlated microstructure noise.

An alternative class of estimators is given by the so-called, the Realized Kernel estimators. The motivation for this class of estimators is to recognize the connection between the problem of estimating the long run variance of a discrete time process, Bartlett (1946). Define the symmetric realized autocovariance sequence

$$\gamma_h(X) := \sum_{j=h+1}^{n} x_{t_j} x_{t_{j-h}}, \tag{1.14}$$

for $h \in \mathbb{Z}^+$ and $\gamma_{-h}(X) = \gamma_h(X)$. At a zero lag, $\gamma_0(X)$ gives us the usual sum of squared high frequency returns, i.e., RV. The kernel estimators smooth the realized autocovariances with the weight function given by $k(\cdot)$, where $k(0) = 1$, $k(s) \to 0$ as $s \to \infty$ and the bandwidth H controls the bias-variance trade-off. Specifically, consider

$$\widehat{[X,X]} = \sum_{|h|<n} k\left(\frac{h}{H+1}\right) \gamma_h(X). \tag{1.15}$$

Zhou (1996) was the first to consider the use of the kernel method to deal with the problem of microstructure noise. Hansen and Lunde (2006) examined the properties of Zhou's estimator and showed that, although unbiased under the presence of i.i.d microstructure noise, the estimator is not consistent. However, they advocated that, while inconsistent, Zhou's kernel method is able to uncover several properties of the microstructure noise.

Barndorff-Nielsen et al. (2011) proposed an estimator of the form in (1.15) with a second order kernel $k(\cdot)$. Their important contribution is to show that it is consistent under the presence of second order stationary noise, and that furthermore, it is asymptotically normal with rate $O_p(n^{1/5})$ and

$$\frac{n^{1/5}\left(\widehat{[X,X]} - [X,X] - c^{-2}|k''(0)|w^2\right)}{\sqrt{4c||k||^2 IQ}} \Longrightarrow N(0,1), \tag{1.16}$$

provided that $H = cn^{3/5}$ for $c \in (0,\infty)$, where $||k||^2 := \int_{-\infty}^{\infty} k(s)^2 ds$ and $w^2 = \sum_h E(\varepsilon_t \varepsilon_{t-h})$, a long run variance of the noise process. The estimator is guaranteed to be positive definite and note that the limiting distribution has an asymptotic bias component. For inference, Zhang et al. (2005) showed that $\frac{[X,X]^n}{2n}$ is consistent estimator of $E\varepsilon^2$. The integrated quarticity can be estimated by the bipower type estimator of Barndorff-Nielsen and Shephard (2004b) which is guaranteed to be positive definite but rate inefficient at $O_p(n^{1/5})$. In Barndorff-Nielsen et al. (2008a), they had a realized kernel estimator with a flat-top kernel i.e. $k(0) = k(|1|/H) = 1$ and the realized autocovariance γ_h was defined such that the sum runs from 1 not $h+1$. Their flat top realized kernel is unbiased under the presence of i.i.d microstructure noise and achieves the optimal convergence rate, $O_p(n^{1/4})$. The drawback of the earlier version, however is that the resulting estimator is not guaranteed to be p.s.d.

We should briefly mention the promising pre-averaging method analyzed for example in Jacod, Li, Mykland, Podolskij and Vetter (2009), which involves averaging observed prices over a moderate number of time points to reduce the measurement error. Consider

$$\overline{X}_t = \frac{1}{n_t} \sum_{|t-t_j|<\epsilon_T} X_{t_j} \quad ; \quad \overline{x}_t = \frac{1}{n_t} \sum_{|t-t_j|<\epsilon_T} x_{t_j},$$

where n_t is the number of time points with $|t - t_j| < \epsilon_T$ for some small $\epsilon_T \to 0$. Then $\overline{X}_t = n_t^{-1} \sum_{|t-t_j|<\epsilon_T} Y_{t_j} + O_p(n_t^{-1})$ and $\overline{x}_t = n_t^{-1} \sum_{|t-t_j|<\epsilon_T} y_{t_j} + O_p(n_t^{-1})$, so that now the noise is small provided n_t is large. The preaveraged data can then be used in a variety of the above procedures.

The final method involves a little departure. Parkinson (1980) and Alizadeh, Brandt and Diebold (2002) proposed a range-based volatility proxy defined by the extreme prices over the pre-determined interval. Specifically, let

$$\mathcal{R} = \sup_{0 \le t \le 1} X_t - \inf_{0 \le t \le 1} X_t.$$

This is an alternative measure of volatility to QV. In some special cases it has a known positive relationship with QV. Specifically, if $X_t = \sigma W_t$, then $\mathcal{R}$ is a stochastic variable, while the quadratic variation is the constant σ^2. In fact, $\mathcal{R} = \sigma \left[\sup_{0 \le u \le 1} W(u) - \inf_{0 \le u \le 1} W(u) \right]$, from which one can compute $E\mathcal{R}^\kappa = \lambda_\kappa \sigma^{\kappa/2}$ for $\kappa \ge 1$, where λ_κ are known constants. More generally the relationship between $\mathcal{R}$ and QV is likely to be rather complex. In practice, one may compute

$$\mathcal{R}_n = \max_{1 \le j \le n} X_{t_j} - \min_{1 \le j \le n} X_{t_j},$$

from a given sample of data observed at times $t_1, \ldots, t_n$. One can expect that $\mathcal{R}_n \to \mathcal{R}$ with probability one under quite general conditions. The most rigorous analysis of the realized range has been in Christensen and Podolskij (2007), except that they only compute $\mathcal{R}$ over small subintervals, which is like assuming that locally $X_t = \sigma W_t$ for some σ, and then average the resulting values of $\mathcal{R}_n$ over these subintervals. Alizadeh et al. (2002) recommend using the log of the sample range, as it is closer to a normally distributed random variable.

The realized range has the significant advantage that one can find the daily value in the newspapers for a variety of financial instruments, and so one has a readily available volatility measure without recourse to analysis of the intra-day price path. Alizadeh et al. (2002) also argue that the method is relatively robust to a measurement error of a bid-ask bounce variety, since the intra-day maximum is likely to be at the ask price and the daily minimum at the bid price of a single quote and so one expects a bias corresponding only to an average spread. By contrast, in computing the realized variance one can be cumulating these biases over many small periods, thereby greatly expanding the total effect.

A number of authors have carried out empirical studies to rank the performance of competing estimators of QV. One way to do this is by simulating the process given in (1.1). To test for the robustness of the estimator, we may introduce jumps in the price or in the volatility, assume different settings for microstructure noise or sampling scheme. Gatheral and Oomen (2010) took a different approach to this and simulated the order book directly. They compared QV estimators under the realistic microstructure setting and compared if the theoretical prediction matches well with actual small sample properties. They found that subsampling estimator, realized kernel, and maximum likelihood estimator deliver superior performance in terms of efficiency and robustness to different parameterizations of microstructure noise.

The actual data may deviate from the assumed model. Then to directly test the competency of the estimators when population quantity is unknown, a popular method is to look at the volatility signature plot which plots the $\widehat{[X,X]}$ against the sampling frequencies. The estimator prone to a microstructure bias will show upward sloping pattern as data is sampled increasingly frequently. See Barndorff-Neilsen et al. (2008b) for example. Authors also compared forecasts of QV estimators under different scenarios of underlying stochastic volatility process and the distribution of microstructure noise. Aït-Sahalia and Mancini (2008) found that TSRV in (1.13) outperforms the RV under varying degree of assumptions. Bandi, Russel and Yang (2008) considered comparison in the context of option pricing and Voev (2009) in the context of an unconditional measure of portfolio performance.

1.4.2 Multivariate volatility estimators

In this section we discuss estimators of integrated covariance matrix. We present a framework for the bivariate case, as this allows treatment of the main issues. We suppose that the efficient price process follows a Brownian semimartingale. For the i-th asset, $i = 1, 2$, we have

$$Y_{i,t} = \int_0^t \mu_{i,u} du + \int_0^t \sigma_{i,u} dW_{i,u}, \tag{1.17}$$

where $\mu_{i,u}$ is a predictable locally bounded drift, $\sigma_{i,u}$ is a cádlág volatility process, and $W_{i,u}$ is a Brownian motions with $E[dW_{1t}dW_{2t}] = \rho_t dt$. The time span we consider is fixed and scaled to vary between $[0, 1]$. We observe a (log) price at discrete time points, $0 = t_{i,0} < \cdots < t_{i,n_i} = 1$. Let Υ be a set of points that partition the interval $[0, 1]$. Define $m_i(n) := \sup_{j:t_{i,j}\in\Upsilon} |t_{i,j} - t_{i,j-1}|$ and assume that as $n \to \infty$, $m(n) := m_1(n) \vee m_2(n) \to 0$, so that the observation grid is becoming finer and finer. Denote by $Y_{i,t_{i,j}}$ the discretely sampled log prices. Suppose that the two prices series are observed on the synchronous time points $\{\tau_j, j = 1, \ldots, n\}$. The quadratic co-variation of Y_1 and Y_2 over a time interval $[0, 1]$ is defined by

$$[Y_1, Y_2] = \lim_{m(n)\to 0} \sum_{j=1}^{n} (Y_{1,\tau_j} - Y_{1,\tau_{j-1}})(Y_{2,\tau_j} - Y_{2,\tau_{j-1}}) = \int_0^1 \sigma_{1,u}\sigma_{2,u}\rho_u du, \tag{1.18}$$

where the last equality holds with probability one. We may denote the quadratic variation of general $d \times 1$ vector of Y as $[Y, Y]_t = \int_0^t \Sigma(u)du$ where $\Sigma_{i,j}(t)$ denotes the instantaneous covariation between i-th and j-th element of Y. The natural estimator of quadratic co-variation is the discrete sum in (1.18), called the Realized Covariance

$$[Y_1, Y_2]^n = \sum_{j=1}^{n} (Y_{1,\tau_j} - Y_{1,\tau_{j-1}})(Y_{2,\tau_j} - Y_{2,\tau_{j-1}}). \tag{1.19}$$

Under perfect synchronization, Barndorff-Nielsen and Shephard (2004a) showed that the Realized Covariance is a $\sqrt{n}$ consistent estimator of the integrated covariance and is asymptotically mixed normal under (1.17). Let us denote the returns for the i-th asset by $y_{i,\tau_j} := Y_{i,\tau_j} - Y_{i,\tau_{j-1}}$. Then, we have

$$\sqrt{n} \frac{\sum_{j=1}^{n} y_{1,\tau_j} y_{2,\tau_j} - \int_0^1 \Sigma_{1,2}(u)du}{\sqrt{\int_0^1 \Sigma_{1,1}(u)\Sigma_{2,2}(u) + (\Sigma_{1,2}(u))^2 \, du}} \Longrightarrow N(0,1).$$

The corresponding feasible CLT is given by

$$\frac{\sum_{j=1}^{n} y_{1,\tau_j} y_{2,\tau_j} - \int_0^1 \Sigma_{1,2}(u)du}{\sqrt{\sum_j y_{1,\tau_j}^2 y_{2,\tau_j}^2 - \sum_j y_{1,\tau_j} y_{1,\tau_{j+1}} y_{2,\tau_j} y_{2,\tau_{j+1}}}} \Longrightarrow N(0,1).$$

Compare this with the univariate case in the previous section. A similar asymptotic argument can be carried out for the realized regression coefficient or the realized betas in the capital asset pricing model (CAPM).

The time stamp for transactions of two different securities rarely matches, and so some data synchronization method is typically employed. This will have an impact on the finite sample as well as on the asymptotic behavior of the resulting covariance estimate. The well known Epps effect refers to the phenomenon that the sample correlation tends to have a strong bias towards zero as the sampling interval progressively shrinks. Hayashi and Yoshida (2005) showed that the realized covariance calculated from the aligned data using the *fixed clock time* alignment method described in the Section 1.3.3 is biased. They proposed a modified covariance estimator

$$\widehat{[Y_1, Y_2]} = \sum_{i=1}^{n} \sum_{j=1}^{n} y_{1,t_{1,i}} y_{2,t_{2,j}} 1_{\{\Delta t_{1,i} \cap \Delta t_{2,j} \neq \emptyset\}},$$

which they show is unbiased and $\sqrt{n}$ consistent. Under presence of asynchronicity but with no microstructure noise this estimator is theoretically the best one. Essentially their estimator takes the cross product of returns only if the portion of transaction time intervals of two assets overlaps.

Malliavin and Mancino (2009) proposed an estimator of the integrated covariance that does not require synchronization. They establish the relationship between the Fourier transform of returns and the Fourier transform of spot volatility. Under (1.17), their estimator is consistent and asymptotically normal. Their estimator is defined by

$$\widehat{[Y_1, Y_2]} = \frac{1}{2m+1} \sum_{|k| \leq m} \mathcal{F}_n(Y_1)(k) \mathcal{F}_n(Y_2)(-k),$$

where $\mathcal{F}_n(Y_i)(\cdot)$ denotes the discretized Fourier transform of i-th asset returns. For $k \in \mathbb{Z}$ and assuming that the time interval is re-scaled to vary $[0, 2\pi]$,

$$\mathcal{F}_n(Y_i)(k) := \sum_{j=1}^{n} e^{ikt_{i,j}} (Y_{i,t_{i,j}} - Y_{i,t_{i,j-1}}) \to_p \int_0^{2\pi} e^{ikt} dY_t.$$

In fact, they have a stronger result where the Fejer Fourier inversion of the above estimator gives a consistent estimator of the instantaneous (co)volatility.

Finally, we should mention some work on the multivariate range based estimation. Brandt and Diebold (2006) extended the work on the realized range to the multivariate case. It is not immediately obvious how to extend such notion to the multivariate case, and indeed their cunning idea relies on the specific structures that arise in a number of settings, notably exchange rates. Suppose we observe the exchange rates between three currencies: A, B, and C, denoted $X_{A:B}$, $X_{B:C}$, and $X_{A:C}$, then we know that in the absence of arbitrage $X_{A:C} = X_{A:B} X_{B:C}$. Taking logs and differencing, we obtain

$$\text{cov}(\Delta \ln X_{A:B}, \Delta \ln X_{A:C}) = \frac{1}{2} \left[\text{var}(\Delta \ln X_{A:C}) + \text{var}(\Delta \ln X_{A:B}) - \text{var}(\Delta \ln X_{B:C}) \right].$$

Therefore, using the relationship between the variance and the range, they obtain an estimate of the covariance between the two exchange rates. The advantage of this method as before is that it does not require high frequency data so that the effect of measurement error is minimized.

Measurement Error

So far we have considered the case where the only source of error is observation error, i.e., discretization error of the continuous semimartingale and the non-synchronicity of the observed price. We next consider the presence of an infinite quadratic variation component in the observed prices due to a further measurement issue, microstructure noise. There has not been a uniform approach to modeling multivariate microstructure noise, perhaps due to the confounding effects of asynchronicity. Furthermore, it is not clear if the microstructure noise between two assets should be correlated and if so how to parameterize this quantity. Let us assume an additive noise for each asset

$$X_{i,t_{i,j}} = Y_{i,t_{i,j}} + \epsilon_{i,t_{i,j}} \text{ for } i = 1, \ldots, d, 0 = t_{i,0} < t_{i,1} < \cdots < t_{i,n_i} = 1.$$

Zhang (2010) assumed that $\{\epsilon_{1,t_{1,j}}, \epsilon_{2,t_{2,j}}\}$ are stationary and exponentially alpha mixing. She proposed a Two Scales Realized Covariance estimator (TSCV), which is defined as a bivariate version of (1.13) applied to an aligned data,

$$\widehat{[Y_1, Y_2]} = [Y_1, Y_2]^K - \left(\frac{\overline{n_K}}{\overline{n_J}}\right) [Y_1, Y_2]^J ,$$

where the average lag K realized covariance is defined by

$$[Y_1, Y_2]^K = \frac{1}{K} \sum_{j=K}^{n} (Y_{1,\tau_j} - Y_{1,\tau_{j-K}})(Y_{2,\tau_j} - Y_{2,\tau_{j-K}}),$$

for $1 \leq J \ll K$. Let summation of sample sizes of two assets as $N = n_1 + n_2$ and recall that the number of points for the aligned time stamp τ is n. Define $\overline{n_K} = n - K + 1)/K$ and similarly for $\overline{n_J}$. Then the above estimator is $O_p(n^{1/6})$ consistent and asymptotically normal under the presence of noise and asynchronous trading, provided that $K = O(N^{2/3})$.

Barndorff-Nielsen et al. (2011) proposed to synchronize the high frequency prices using *refresh Time* explained in Section 1.3.3. They assumed that the microstructure noise $\{\epsilon_{i,\tau_j}, i = 1, \ldots, d\}$ is a second order stationary process with respect to refresh time $\{\tau_j\}$. Their Multivariate Realized Kernels (MRK) is given in (1.15) with realized autocovariance defined by

$$\gamma_h(X) = \sum_j x_{\tau_j} x_{\tau_{j-h}}^T, \quad h = 0, \pm 1, \pm 2, \cdots$$

where $\sum_j = \sum_{h<j\leq n}$ for $h \geq 0$, and $\sum_j = \sum_{1\leq j\leq n+h}$ for $h < 0$ and $x = [x_1 : \cdots : x_d]$ is a matrix of *refresh time* aligned returns for d number of assets. The MRK is $O_p(n^{1/5})-$consistent and asymptotically normal and its asymptotic distribution is given in (1.16) modified with relevant multivariate quantities, under the second order kernel. It is also guaranteed to be positive semi-definite at the cost of asymptotic bias. Note that the asymptotic rate is based on the sample size of the aligned time stamp. Aït-Sahalia, Fan and Xiu (2010a) proposed an $O_p(n^{1/4})$ consistent estimator based on the quasi-MLE and a generalized time synchronization method. An advantage of their estimator over TSCV and MRK is that it does not involve choosing or estimating tuning parameters such as bandwidth. However they adopt a somewhat restrictive assumption on the microstructure noise - it is a white noise that is mutually independent across assets.

Christensen, Kinnebrock and Podolskij (2010) proposed a multivariate pre-averaging estimator. Voev and Lunde (2007) proposed a modified Hayashi and Yoshida estimator to bias-correct for the microstructure noise. Park and Linton (2011a) proposed a covariation estimator that is robust to both microstructure noise and asynchronicity based on the Fourier analysis of returns, extending Mallianvin and Mancino (2009). Griffin and Oomen (2011) ranked the performance in terms of efficiency of the three estimators: realized covariance, realized covariance plus lead- and lag-adjustments, and the Hayashi and Yoshida estimator. They found that the performance of competing estimators depends on the level of microstructure noise as well as on the magnitude of correlation.

1.5 Modeling and Forecasting

We will designate the class of estimators of quadratic (co) variation based on the high frequency data as "realized measures". In this section, we review how realized measures can be used to model and forecast the (co)variances. We will summarize the studies that compare these competing models in terms of forecasting power where the forecasting variable is a general function of volatility such as Value at Risk and portfolio performance. We also consider extensions to a dynamic model of the realized covariance matrix.

1.5.1 Time series models of (co) volatility

There is a large literature on time series models of volatilities. In the well-known GARCH and Stochastic volatility family of models, volatility is treated as a latent variable. The method we discuss here takes a different stance. We treat the Realized Variance as ex-post observed variance. Given the sequence of RVs (or the robust estimator discussed in Section 1.4.1), we use traditional time series techniques such as ARMA to fit a model and carry out forecasts. The key feature of the time series of the Realized Variance is that it is highly persistent. To account for this, Andersen, Bollerslev, Diebold, and Labys (2003) proposed an autoregressive fractionally integrated moving average (ARFIMA) to model the time series of the Realized Variance. Let h_t denote an estimator of integrated variance for t-th day, $t = 1, \ldots, T$. The ARFIMA model for h_t is given by

$$\Phi(L)(1-L)^{\nu}(h_t - \mu) = \Theta(L)\epsilon_t,\ \epsilon_t \sim \mathrm{WN}(0,1), \tag{1.20}$$

where $\Theta(L)$ is a polynomial of lag operators and ν is a real-valued parameter that measures the degree of fractional integration. The model can be estimated by maximum likelihood method. Lanne (2006) modified (1.20) by making parameters in $\Theta(\cdot)$ time varying and letting ϵ_t be a non-Gaussian. In practice these methods can be problematic as estimation of ν is non-trivial and influential on other features of the model. A simpler model that seems to capture lag dependencies well is the Heterogenous Autoregressive model of Corsi (2009);

$$h_{t+1} = \theta_0 + \theta_D h_t + \theta_W h_t^{(W)} + \theta_M h_t^{(M)} + \epsilon_{t+1}, \tag{1.21}$$

where $h_t^{(W)} := \frac{1}{5}(h_t + \cdots + h_{t-4})$ is a Realized Variance over a week and similarly defined $h_t^{(M)}$ denotes a Realized Variance over a month. Shephard and Sheppard (2010) who proposed a model that is a hybrid of a GARCH augmented with a realized measure and a reduced form time series model for the Realized Variance. See similar approach in Hansen, Huang and Shek (2011) who jointly modeled returns and realized measures of volatility. Liu and Maheu (2009) carried out Bayesian averaging over both different measures of integrated variance and different time series models.

In the multivariate setting, a key issue is that the fitted model should produce a positive definite covariance matrix. Also, if we were to model a high dimensional covariance matrix, we need to address the dimension issue, which grows rapidly with the number of assets considered. Voev (2007) proposed a method to combine volatility and bivariate co-volatility forecasts to produce a positive definite matrix. The problem with this method is that interaction between elements of covariance matrix is not taken into account. The full joint modeling of covariance matrix is an important issue. For example, the variance of one asset and covariance with another asset have significant dependencies, especially during episodes of market crashes and large economic events. Compared with the univariate volatility modeling literature, such multivariate models have been sparsely researched mainly due to the fact that consistently estimating a general $d \times d$ covariance matrix for $d > 2$ has been difficult, plagued by bias induced by synchronization as well as microstructure noise. However with recent work in Section 1.4.2 this area of research can progress further.

Let $H_t, t = 1, \ldots, T$, be a time series of such estimates of the integrated covariance matrix. A natural way to model the persistency and lead-lag dependencies in the elements of matrix H_t is to fit a multivariate version of model given in (1.20), called Vector ARFIMA model. We fit a model for $h_t = vech(g(H_t))$ where $vech(\cdot)$ operation stacks the lower triangular matrix of an argument. The dimension of h_t is given by

$m = d(d+1)/2$. A range of transformation function $g(\cdot)$ is considered for the purpose of dimension reduction and to guarantee a p.s.d. matrix forecast. We will discuss this in a moment. First consider the vector ARFIMA model

$$\Phi(L)D(L)(h_t - BZ_t) = \Theta(L)\epsilon_t,\ \epsilon_t \sim \mathrm{WN}(0, I_m), \tag{1.22}$$

where $\Theta(L) = I_m - \Theta_1 L \cdots - \Theta_q L^q$ is a matrix lag polynomial of degree $q \in \mathbb{Z}$ for the MA component, $\Phi(L)$ is defined similarly for AR component. $D(L) = diag\{(1 - L)^{\nu_1}, \ldots, (1 - L)^{\nu_m}\}$ is a matrix fractional difference operator with $\nu_1, \ldots, \nu_m$ the degrees of fractional integration for each element of h_t. Z_t are exogenous variables that affect the dynamics of volatility; candidate variable are trading activity variables and macroeconomic state variables. B is a restriction matrix. We can estimate such a model by maximum likelihood. The one step ahead prediction is then $\widehat{h}_{t+s} = E(h_t|h_s, s \leq t)$. We obtain a covariance matrix forecast by $\widehat{H}_{t+s} = vech^{-1}(\widehat{h}_{t+s}))$ where the $vech^{-1}(\cdot)$ re-stacks the vector into a symmetric matrix.

Bauer and Vorkink (2011) fitted the vech of $\log(H_t)$ (rather like a matrix E-GARCH model) to an AR model where the right hand side lagged variables are dimension reduced by principal component analysis. Chiriac and Voev (2011) carried out a Cholesky decomposition of the covariance matrix and model the lower dimensional factors by a vector ARFIMA model. They showed this method outperforms in terms of root mean square error, a number of models including: the Heterogeneous Autoregressive model, a multivariate version of (1.21), Wishart Autoregressive (WAR) model of Gourieroux, Jasiak, and Sufana (2009) and the Dynamic Conditional Correlation model. We may use the Realized Variance to proxy true H_{t+s} and compare the Frobenius norm of the bias $\|\widehat{H}_{t+s} - H_{t+s}\|$, across different models and different horizons s. Authors also compare minimum variance portfolio efficient frontiers using different covariance matrix forecast.

1.5.2 Forecast comparison

Since volatility itself is unobservable, the comparison of volatility forecasts relies on an observable proxy for the latent volatility process. See Patton (2011) on the method robust to the measurement error in the volatility proxy. In the previous section, we presented how we can compare root-MSE of covariance forecasts. We might be interested in economically meaningful loss functions. Brownlees and Gallo (2010) compared the Value at Risk forecasts from different time series models of RV. Bandi et al. (2008) considered the forecast comparison in the context of option pricing.

An important research question is whether there is a gain in using the high frequency data over traditional daily volatility models. We can compare the dynamic model of estimators of ex-post variation calculated from the high frequency data against the latent volatility models such as GARCH and Stochastic Volatility. Koopman, Jungbacker and Hol (2005) found that the ARFIMA model of RV delivers the best out-of-sample forecast compared with the GARCH or the SV model fitted to a daily S&P500 index. Shephard and Sheppard (2010) showed their hybrid model using the realized measures outperforms the daily GARCH model in terms of various criteria. Siu and Okunev (2009) compared historical, realized and implied volatility measures for predicting over multiple horizons.

We are also interested in ranking the competing realized measures in Section 1.4. Ghysels, Santa-Clara and Valkanov (2006) proposed a framework to do this, called the mixed data sampling (MIDAS) regression, comparing measures of ex-post variation in terms of their forecasting ability at various horizons. Ghysels and Sinko (2011) found that the microstructure robust realized measures deliver better forecasts. Likewise, Aït-Sahalia and Mancini (2008) reach similar conclusion where the TSRV estimator in (1.13) outperforms the RV under diverse setting of volatility process and assumptions on the noise.

1.6 Asset Pricing

1.6.1 Distribution of returns conditional on the volatility measure

Authors found the evidence that the de-volatized returns by the class of RV estimators are Gaussian or approximately so. Andersen, Bollerslev, Diebold and Labys

(2001) found that daily returns standardized by the realized volatility approximate the Gaussian distribution. Thomakos and Wang (2003) also found such evidence for a futures market.

Peters and de Vilder (2006) studied the volatility and return dependence by sampling the returns in financial time. They tested if the return series are a realization of a local martingale using the theorem by Dubins and Schwarz (1965) who stated that any continuous local martingale $Y_t \in \mathcal{F}_t$ is a time-changed Brownian motion. Formally stated,

$$B_s = Y_{T_s}\ ,\ T_s = \inf\{t|[Y]_t \geq s\}, \tag{1.23}$$

where $B_s \in \mathcal{F}_{T_s}$ is an independent Brownian motion and T_s is a stopping times. It is the first time the quadratic variation reaches a specified level. Equivalently, the theorem implies that

$$Y_t = B_{[Y]_t}, \tag{1.24}$$

which states that every continuous martingale is a time-changed Brownian motion where the time change is given by the quadratic variation. In empirical analysis, (1.23) is more useful, since it states that between the unit interval of the transformed time, $[T_{(j-1)a}, T_{ja}]$, Y has a constant QV at a. Given an interval of physical time, Y is sampled more frequently when QV is large. More precisely, the (discretized) transformed time is constructed by: $T_0 = 0$, $T_{(j+1)a} = T_{ja} + \Delta T_{(j+1)a}$,

$$\Delta T_{(j+1)a} = \inf\{t|[Y]_{[T_{ja}, T_{ja}+t)} \geq a\}, \tag{1.25}$$

where $[Y]_{[T_{ja}, T_{ja}+t)}$ denotes the quadratic variation in the interval $[T_{ja}, T_{ja}+t)$. The standardized increment in financial time

$$\xi = \frac{Y_{T_{ja}} - Y_{T_{(j-1)a}}}{\sqrt{a}}, \tag{1.26}$$

is i.i.d standard normal. Observe the trade-off between having large and small a. We need to have a large a to have many data points to consistently estimate QV by a realized measure but large a means sparse sampling of Y. Note also that we can explicitly derive the distributional features of the stopping time T when the Y

process is completely specified. Testing for the hypothesis that Y_t is a local martingale is then equivalent to testing for i.i.d standard normality of the return series that is spaced by T_s. Peters and de Vilders (2006) tested if the S&P500 intra-day return is a local martingale where they constructed the stopping time T_s based on the Realized Variance. They concluded that we cannot reject the null hypothesis that returns are the realization of a martingale process at various time scales (> 1 day) based on the tests for Gaussianity, independence and serial correlation.

1.6.2 Application to factor pricing model

We next discuss applications to asset pricing models for cross-sections of returns. Denote a stock return for i-th firm at time t by $y_{i,t}$, with $i = 1, \ldots, d$ and $t = 1, \ldots, T$. The K factor pricing model for stock returns is given by

$$y_{i,t} = \beta_i^\top f_t + \varepsilon_{i,t}, \tag{1.27}$$

where the factor loadings $\beta_i = (\beta_{i,1}, \ldots, \beta_{i,K})^\top$ are unrestricted. The sampling unit t is typically a low frequency such as monthly. In some cases f_t are unobserved statistical factors, while in others they are the returns on carefully constructed portfolios. In the latter case, $\beta_{i,k}$ can be given the interpretation of the covariance between return on portfolio k and asset i divided by the variance of the return on portfolio k. The continuous time framework allows us to measure the time varying beta between two assets using the high frequency data. The realized beta between asset i and k in period $[t-1, t]$ calculated from high frequency returns $\{y_{\cdot,t}\}$ is given by,

$$\hat{\beta}_{i,k}(t) = \frac{\sum_j y_{i,j} y_{k,j}}{\sum_j y_{k,j}^2} \to_p \frac{\int_{t-1}^t \Sigma_{i,k}(s)ds}{\int_{t-1}^t \Sigma_{k,k}(s)ds} := \beta_{i,k}(t),$$

where the convergence in probability holds under (1.17) and as mesh goes to zero.For studies on the relationship between returns and volatility, see Ghosh and Linton (2007), Bollerslev, Litvinova, and Tauchen (2006) and Bali and Peng (2006). Ghosh and Linton (2007) showed that the estimating the risk-return trade-off parameters can be posed as a GMM estimation problem. They used the Realized Variance as a conditional volatility proxy and showed that there is a significant time-variation in the risk-return slope coefficient. Bali et al. (2006) found a positive and statistically significant relation between the conditional mean and conditional volatility of market returns at a daily level where volatility is proxied by RV. Bollerslev et al. (2006) made use of the time aggregation formula between lower and high frequency covariance. They found that the correlations between absolute high-frequency returns and current and past high-frequency returns are significantly negative for several days.

Andersen, Bollerslev, Diebold and Wu (2005) and Bandi and Russell (2005) estimated the beta in CAPM by a realized covariation. Bandi et al. (2005) provided the MSE-based optimal sampling frequency for calculating the realized beta designed to reduce the effect of market microstructure noise. Bollerslev and Zhang (2003) estimated the factor loadings in the three-factor Fama-French model using the high frequency data adopting a simple adjustment procedure to account for non-synchronous trading effects. Bannouh, Martens, Oomen and van Dijk (2009) and Kyj, Ostdiek and Ensor (2009) used a mixed frequency framework, using the high-frequency data to obtain an estimate of the factor covariance matrix and using the daily data to estimate the factor loadings. This method avoids the non-synchronicity between a individual stock and usually more liquid factor prices.

The economic value of using the realized covariance in portfolio management is discussed by Fleming, Kirby Ostdiek (2003) and Liu (2009). Fleming et al. (2003) found that a risk-averse investor is willing to pay between 50 and 200 basis points per annum to switch from a covariance measurement based on the daily data to the one based on intra-day data whereas Liu (2009) found that the benefits depend upon the re-balancing frequency and estimation horizon of portfolio optimization decision. See Fan, Li and Yu (2010) for estimating high dimensional covariance matrix using high-frequency data and its benefit in portfolio selection.

1.6.3 Effects of algorithmic trading

Recently, the effects of high frequency or algorithmic trading have been the focus of policy discussions, arising part from the flash crash of May 2010, where the US market suffered rapid price decreases followed ultimately by a recovery. Chaboud, Chiquoine, Hjalmarsson and Vega (2009) investigated the effects of algorithmic trading on volatility in the foreign exchange market. They considered the following regression equation

$$RV_{it} = \alpha_i + \beta_i AT_{it} + \gamma_i^{\top} \tau_{it} + \sum_{k=1}^{22} \delta_{ik} RV_{i,t-k} + \varepsilon_{it},$$

where RV_{it} is the log of realized volatility of currency i during day t computed using one minute returns, AT_{it} is the fraction of algorithm trading in that day and currency, which was recorded by the trade matching engine, and τ_{it} are dummy and time trend variables. The latter are included because the AT series has a pronounced upward trend, while volatility appears to be stationary. They recognized AT is endogenous variables since high frequency automated trading algorithms may trade more in volatile times. They therefore instrument it with a variable that measures the capacity for computer trading in a given currency/period combination. The estimation strategy matters here, so that using OLS yields a positive effect, $\beta_i > 0$, but the instrumental variable estimator finds $\beta_i < 0$ but not statistically significant. They conclude that intra-day algorithmic trading does not by itself lead to higher daily volatility. For other studies that use realized measure of volatility to determine the effects of high frequency trading, see Hendershott, Jones and Menkveld (2009) and Hendershott and Riordan (2009).

1.6.4 Application to option pricing

In recent years, volatility has been thought of as an asset class in its own right. One can trade volatility through a position in puts and calls but this has an additional exposure to a price movement. Swaps and options on quadratic variation have been developed for a pure exposure on the volatility. For a discussion on the volatility as an asset class, see Demeterfi, Derman, Kamal, and Zou (1999). An investor of volatility swap is swapping a fixed volatility $SW_{t,T}$ for a floating (actual) volatility $[Y]_{t,T}$, denoting quadratic variation accumulated over $[t,T]$. The floating leg is usually given by a sum of squared daily log returns over the relevant time interval. Given N notional amount in dollar terms per annualized volatility point, its payoff at expiration is equal to

$$([Y]_{t,T} - SW_{t,T})N.$$

Denote r a risk-free discount rate corresponding to an expiration date T. The value of such forward contract is given by the expected present value of the future payoff under a risk neutral measure $\mathbb{Q}$, a probability measure such that the discounted price of traded asset is a martingale,

$$E^{\mathbb{Q}}[e^{rT}([Y]_{t,T} - SW_{t,T})].$$

Then the strike for which the contract has zero present value is

$$SW^*_{t,T} = E^{\mathbb{Q}}([Y]_{t,T}).$$

Carr, Geman, Madan and Yor (2005) proposed a method of pricing options on quadratic variation via Laplace transform when returns follow pure jump Lévy process. Itkin and Carr (2010) considered a pricing problem when returns are time changed Lévy processes. Britten-Jones and Neuberger (2000) proposed a method to estimate $E^{\mathbb{Q}}([Y]_T)$, an option-implied (i.e. risk-neutral) integrated variance over the life of the option contract, assuming price follows stochastic volatility diffusion process. $SW^*_{t,T}$ can be labeled as a model-free implied variance as well as being a no-arbitrage variance swap rate. Carr and Wu (2009) showed that the variance swap rate is well approximated by the value of a particular portfolio of options. They established that the difference between the Realized Variance and this synthetic variance swap rate, given by

$$[Y]_{t,T} - SW^*_{t,T},$$

quantifies the variance risk premium. They have analyzed the variance swaps for stocks and found it to be significantly negative. This means that investors are willing to pay a premium to hedge away upward movement in the return variance.

Bollerslev, Gibson and Zhou (2010) proposed a method for constructing a volatility risk premium relying on sample moments of the Realized Variance and a option-implied volatility estimator. Wu (2010) studied the variance risk premium using both variance swap rates constructed from the option prices and the quadratic variance estimates using the high frequency data and found a strong evidence for negative variance risk premium in the equity market.

1.7 Estimating continuous time models

In this section we review how realized measures can be used to estimate the parameters of a continuous time model. Consider a diffusion model for financial prices X_t,

$$dX_t = \mu(X_t, \theta)dt + \sigma(X_t, \theta)dB_t, \tag{1.28}$$

where B_t is an independent Brownian motion, $\mu(X_t, \theta)$ is a drift function and $\sigma(X_t, \theta)$ is a given diffusion coefficient function. We are interested in estimating vector of parameters θ. X_t is non-homogenous in a sense that the diffusion coefficient is not constant. This specification includes geometric Brownian motion, Ornstein–Uhlenbeck process, and Cox–Ingersoll–Ross process as special cases. Since X_t in (1.28) is markov we can write down a log likelihood in terms of transition density if a closed form for this exists. For discretely observed data $\{X_{t_i}\}_{0 \leq i \leq n}$ on the equally spaced grid, $\Delta t_i = 1/n$, the transition density is given by $\mathbb{P}[X_{\frac{i}{n}} \mid X_{\frac{i-1}{n}}; \theta]$. Such exact maximum likelihood method yields a consistent and efficient estimator under usual regularity conditions.

When transition density does not have a closed form expression, we may use Euler scheme and its higher order refinement to approximate the process or use a closed-form approximation to the transition density itself. See Phillips and Yu (2009b) for a survey on maximum likelihood estimation of a model in (1.28). If X_t can be observed continuously, the likelihood function for the continuous record can be obtained via the Girsanov theorem.

However in practice we observe the data at discrete time points and even for densely sampled high frequency data, it deviates from the model in (1.28) due to a presence of microstructure noise. Phillips and Yu (2009a) proposed a two stage estimation method based on the realized variance to estimate parameters in diffusion coefficient $\sigma(X_t, \theta)$ and using the infill likelihood to estimate the drift parameters, $\mu(X_t, \theta)$. Yu and Phillips (2001) also showed that the time changed Brownian motion given in (1.23) can be used to construct an exact Gaussian maximum likelihood for a non-homogeneous Itô-processes.

Once the model departs from the Markovian property, we cannot decompose the likelihood into a transition density involving just observable quantities. There are large literature on computationally intensive estimation method, however the availability of high frequency data gives us alternative route to estimate such model. Consider the stochastic volatility specified by the OU process and assume that there is an additive measurement error in the Realized Variance. Then the Realized Volatility has an ARMA representation and the parameter can be estimated by the quasi-maximum likelihood constructed using the output of the Kalman filter. Barndorff-Nielsen and Shephard (2002) showed that the method yields quite precise estimates even for non-Gaussian driven volatility processes. See also Barndorff-Nielsen and Shephard (2006) for related approach for estimating a time deformed Lévy processes. In this case the source of stochastic volatility is through a deformation of time and we are interested in estimating the parameter for the autocovariance function of a deformed time process.

Bollerslev and Zhou (2002) proposed a Generalized Method of Moment type estimator for parameters of a Brownian motion driven stochastic volatility model under no microstructure noise. Their method is by matching the sample moments of the realized volatility to the population moments of the integrated volatility implied by a assumed continuous-time model. Todorov, Tauchen and Grynkiv (2010) proposed a method, first integrating intra-day data into the Realized Laplace Transform (Todorov and Tauchen (2010)) of volatility and matching moments of the integrated joint Laplace transform with those implied by the assume stochastic volatility model. This method is robust to the presence of jumps in the price.

Chapter 2

Estimating the Quadratic Covariation Matrix for an Asynchronously Observed Continuous Time Signal Masked by Additive Noise

2.1 Introduction

There have been many advances in the theory and application of volatility measurement from high frequency data. The ex-post measure of volatility called the quadratic variation has been the focus of much attention. The theory has been developed in a series of papers including: Andersen, Bollerslev, Diebold and Labys (2001), Barndorff-Nielsen and Shephard (2002,2004a) and Mykland and Zhang (2006). This work has been extended to take account of what is called microstructure noise when an underlying efficient price diffusion is distorted by measurement error in papers by Zhang, Mykland, and Aït-Sahalia (2005). Their two time scale estimator is the first consistent estimator of the quadratic variation under the presence of the additive noise. Zhang (2006) extended this work to the multiscale estimator which converges to the target faster. Kalnina and Linton (2008) proposed a modification of the two time scale estimator that is consistent under heteroscedastic and endogenous noise. Aït-Sahalia, Mykland and Zhang (2010b) modified their earlier estimator so that it achieves consistency in the presence of serially correlated microstructure noise. Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a) generalized this idea on a kernel smoothing technique for the problem of estimating the integrated variance : their estimator using the flat-top kernel achieves the fastest possible convergence rate (the same as an infeasible MLE in a special case) although it is not guaranteed to be positive definite. Jacod, Li, Mykland, Podolskij and Vetter (2009) introduced the pre-averaging method, which involves first averaging the observed prices over a moderate number of time points to reduce the measurement error.

In the multivariate case an additional issue arises, namely that the observations are asynchronous, i.e., transactions occur at different time points for different assets. Hayashi and Yoshida (2005) proposed estimators of the integrated covariance that does not require synchronization. However their estimator is inconsistent under the presence of microstructure noise. Malliavin and Mancino (2009) proposed a fourier domain approach that does not require data alignment but they have not work out the theoretical results when noise is present. Estimators addressing both the non-synchronicity and the microstructure noise were proposed by Zhang (2010), Barndorff-Nielsen, Hansen, Lunde and Shephard (2011) and Aït-Sahalia, Fan and Xiu (2010a). The estimators are consistent and convergence rates are respectively $O_p(n^{1/6})$,$O_p(n^{1/5})$ and $O_p(n^{1/4})$. First two papers assume microstructure noise is stationary and exponentially alpha mixing with respect to transaction time and estimators still require aligning the data although the consistency is robust to the alignment. However the hidden cost of data alignment and non-synchronicity for these estimators are that the sample size n that appears in the convergence rate is the sample size of aligned data. Also the drawback of Zhang (2010) and Aït-Sahalia et al. (2010a) is that the estimator cannot be generalized to dimensions higher than two unless the covariance matrix is estimated element-wise which does not guarantee the positive definite estimator. See Park and Linton (2011) for a more detailed survey.

The goal of this paper is to propose a new estimator of the general multivariate volatility measure that is robust to microstructure noise and to asynchronous data timing. The method is based on Fourier domain techniques, which have been widely used in discrete time series. The advantage of this method is that it does not require an explicit time alignment. This class of techniques was first proposed in Malliavin and Mancino (2009), who analyze the case with no microstructure noise. The by-product of our Fourier domain based estimator is that we have a consistent estimator of the instantaneous co-volatility even under the presence of microstructure noise. We apply these results for multivariate regression estimation in continuous time and show that we can consistently estimate the regression coefficients for variables that are non-synchronously observed.

In Section 2.2 we give a set up of the model and assumptions regarding the sampling scheme. In Section 2.3, we propose a Fourier domain based estimator of integrated covariance. The Fourier domain estimator is closely related to a time domain estimator and we show their relationship and what it implies for conditions on the smoothing windows. Section 2.4 presents the asymptotic properties of the proposed estimator without and with the presence of microstructure noise. We devote a subsection giving an intuitive explanation for the source of the bias in the time domain estimator using a simple example. In Section 2.5 the Fourier method is further extended to estimate the instantaneous covariance matrix of diffusion process and to estimate the autocovariance function of the microstructure noise. Section 2.6 discuss the estimation of some economically interesting scalar functions of the integrated covariance matrix. We carried out extensive simulations in Section 2.7.

Some notation. For scalars a and b, $a \wedge b$ and $a \vee b$ denote the minimum and maximum value. For a series $t_{i,j}$, denote $\Delta t_{i,j} = t_{i,j} - t_{i-1,j}$, and for any function g, let $\Delta g(t_{i,j}) = g(t_{i,j}) - g(t_{i-1,j})$. We use $\longrightarrow_p$ to denote convergence in probability, and $\Longrightarrow$ to mean stable convergence described in the Appendix. For real sequences a_n and b_n, $a_n \simeq b_n$ means $a_n = b_n + o_p(b_n)$. For a matrix A, $\|A\|_2 = \mathrm{tr}(A^\intercal A)^{1/2}$. Let L denote the discrete time lag operator, so that $LX_t = X_{t-1}$.

2.2 The Model and assumptions

2.2.1 Efficient Price and Parameter of Interest

The following assumption describes the general setting used throughout the paper.

ASSUMPTION 1. *The efficient price process follows a Brownian semimartingale. For a $d \times 1$ vector of logarithmic prices $P(t) = [P_1(t), \ldots, P_d(t)]^\intercal$ defined on the filtered probability space $(\Sigma, \mathcal{F}, \mathcal{F}_{t\geq 0}, \mathbb{P})$, we have*

$$P(t) = \int_0^t \mu(u)du + \int_0^t \sigma(u)dW(u),$$

where $\mu(u) = [\mu_1(u), \ldots, \mu_d(u)]^\intercal$ is a vector of predictable locally bounded drifts and $\sigma(u)$ is a symmetric $d \times d$ matrix of locally bounded cádlág processes with $\int_0^t \sigma(u)\sigma(u)^\intercal \otimes \sigma(u)\sigma(u)^\intercal du < \infty$ a.s. $W(u)$ is a $d \times 1$ vector of independent Brownian motion and is independent from the volatility process.

The matrix $\int_0^t \sigma(u)\sigma(u)^\intercal \otimes \sigma(u)\sigma(u)^\intercal du$, which we call integrated quarticity, appears in the asymptotic variance of the estimator below. The assumption of locally bounded drift and diffusion coefficient are required to apply Girsanov's theorem to remove the drift term in the theoretical derivation. Consider the discrete time grid $0 = t_0 < \cdots < t_n = T$, where T is fixed, and let $P(t_i)$ denote the (log) price at those points. The quadratic covariation matrix of P over a time interval $[0, t], t \leq T$ is defined by

$$[P, P]_t = \operatorname*{plim}_{n\to\infty} \sum_{i;t_i\leq t} \{P(t_i) - P(t_{i-1})\}\{P(t_i) - P(t_{i-1})\}^\intercal, \tag{2.1}$$

where the limit is finite and well defined with probability one. Under Assumption 1, this is almost surely equal to the integrated covariance matrix

$$[P, P]_t = \int_0^t \sigma(u)\sigma(u)^\intercal du. \tag{2.2}$$

A natural estimator of (2.2) is the finite sum given in the definition of quadratic variation, which is called the Realized Covariance. Barndorff-Nielsen and Shephard (2002) showed that the Realized Covariance is unbiased and is a $\sqrt{n}$ consistent estimator of the integrated covariance under Assumption 1 and assuming synchronous trading. Throughout this paper we will reserve the square bracket to denote the quadratic variation, following the convention in the stochastic processes literature. The objective of this paper is to consistently estimate the integrated covariation matrix. The integrated covariance is related to the covariance matrix of prices by

$$\mathrm{cov}\{P(t)\} = E\{\int_0^t \sigma(u)dW(u)(\int_0^t \sigma(u)dW(u))^\intercal\} = \int_0^t E\{\sigma(u)\sigma(u)^\intercal\}du = E[P,P]_t,$$

where the second equality follows from Itô's formula. Let $[P,P] := [P,P]_T$. We will denote the (i,j)- th element of an instantaneous covariance matrix by $\Sigma_{i,j}(u) = \{\sigma(u)\sigma(u)^\intercal\}_{i,j}$. The j-th diagonal element gives an integrated variance $[P_j, P_j] = \int_0^T \Sigma_{j,j}(u)du$.

Two problems are present in estimating (2.2). First, prices of different assets are observed at different times. Second, observed prices are distorted by noise and do not satisfy Assumption 1. We propose below an estimator that is robust to these two problems. We will examine in detail the two problems in the following sections.

2.2.2 Sampling scheme

In this section we describe the main assumptions we make on the observation times. We allow for unequally spaced and asynchronous observation times.

Assumption 2. *The time span is fixed and scaled to vary between* $[0, 2\pi]$. *We observe log prices at discrete time points:* $0 = t_{0,\ell} < \cdots < t_{n_\ell,\ell} = 2\pi$ *for* $\ell = 1,\ldots,d$, *where* n_ℓ *is the total number of observations for the* ℓ*-th asset. The discrete time points are allowed to be stochastic and assumed to be independent of price and volatility process. The total number of observation points* n_ℓ *is large and* $n := \min_\ell(n_\ell) \to \infty$. *Unless otherwise stated, all convergence below holds with probability one. For all* $a, b, \ell \in \{1,\ldots,d\}$:

1. *The discrete time points satisfy* $\sup_{0\le i<n_\ell}(t_{i,\ell} - t_{i-1,\ell}) = O(\frac{1}{n_\ell})$.

2. *Denote the interval* $I_{i,a} = [t_{i-1,a}, t_{i,a})$ *and* $I_{j,b} := [t_{j-1,b}, t_{j,b})$. *Define the empirical quadratic covariation of time by*

$$\mathcal{Q}^{(n)}_{aabb}(t) = (n_a \wedge n_b) \sum_{i,j:t_{i,a},t_{j,b}<t} \Delta t_{i,a} \Delta t_{j,b} 1_{\{I_{i,a}\cap I_{j,b}\neq\emptyset\}}$$
$$\mathcal{Q}^{(n)}_{abab}(t) = (n_a \wedge n_b) \sum_{i,j,k,l:t_{i,a},t_{j,b},t_{k,a},t_{l,b}<t} (t_{i,a} \wedge t_{j,b} - t_{i-1,a} \vee t_{j-1,b})$$
$$\times (t_{k,a} \wedge t_{l,b} - t_{k-1,a} \vee t_{l-1,b}) 1_{\{I_{i,a}\cap I_{j,b}\neq\emptyset\}} 1_{\{I_{k,a}\cap I_{l,b}\neq\emptyset\}} 1_{\{i=k\}}, \text{ for } n_a < n_b,$$

where the last indicator function is replaced by $1_{\{j=l\}}$, for $n_b \leq n_a$. *The empirical quadratic covariation satisfies* $\mathcal{Q}^{(n)}_{\cdot}(t) \longrightarrow \mathcal{Q}_{\cdot}(t)$ as $n_a \wedge n_b \to \infty$, *where* $\mathcal{Q}_{\cdot}(t)$ *is continuously differentiable.*

3. *The degree of non-synchronicity satisfies* $\sup_{i,j} |t_{i,a} - t_{j,b}| 1_{\{I_{i,a}\cap I_{j,b}\neq\emptyset\}} = O(\frac{1}{n_a \wedge n_b})$. *Given any set of* $\{t_{i,a}, t_{j,b}\}$ *such that* $n_a < n_b$, *we assume that*

$$\sup_{0\leq j\leq n_b} \#\{t_{j,b} \in [t_{i-1,a}, t_{i,a}) | 1_{\{I_{i,a}\cap I_{j,b}\neq\emptyset\}}\} = O(\frac{n_a \vee n_b}{n_a \wedge n_b}).$$

In Assumption 2.2, the expression specializes to $\mathcal{Q}^{(n)}_{aa}(t) = n_a \sum_{i,:t_{i,a}<t} (\Delta t_{i,a})^2$ in univariate case which will appear in the asymptotic variance of the integrated variance estimator. Assumption 2 does not restrict the ratio of sample sizes of different assets to be bounded away from zero or infinity. One asset can be allowed to be much more liquid than the other. This allows for quite a lot of generality. Define

$$\{T^{(ab)}_l\}_{1\leq l\leq N^{(ab)}_T} := \{t_{i,a} \cup t_{l,b}, i = 1, \dots, n_a, l = 1, \dots, n_b\},$$

where $N^{(ab)}_T$ is a total number of data points for union of time stamps. If Assumption 2.1 is further restricted to $\inf_i \Delta t_{i,\ell} = O(\frac{1}{n_\ell})$ and $\sup_i \Delta t_{i,\ell} = O(\frac{1}{n_\ell})$, then Assumption 2.3 is implied. One way of showing this is as follows. Let a be the less liquid asset such that $n_a < n_b$, then it holds that $\frac{\Delta t_{j,b}}{\Delta t_{i,a}} \leq \sup_{i,j} \frac{\Delta t_{j,b}}{\Delta t_{i,a}} \leq \frac{\sup_j \Delta t_{j,b}}{\inf_i \Delta t_{i,a}} = O(\frac{n_a \vee n_b}{n_a \wedge n_b})$. The sample size of the union of time stamps, $N^{(ab)}_T$ is of order $O(n_a \vee n_b)$. We will use the fact that $\{1_{\{I_{i,a}\cap I_{j,b}\neq\emptyset\}} = 1\}$ if and only if $\{u_{ij} := t_{i,a} \wedge t_{j,b} > t_{i-1,a} \vee t_{j-1,b} := l_{i,j}\}$.

We introduce here some notation we will use in the sequel. Denote the average interval size for asset ℓ by $\overline{\Delta t}_\ell := 2\pi/n_\ell$. When comparing asset a and asset b, denote for convenience the average interval size of the more liquid asset by $\widetilde{\Delta t}_{ab} = 2\pi/(n_a \vee n_b)$. We may drop the asset index whenever it is obvious.

The set of $\{(P_\ell(t_{i,\ell}), t_{i,\ell})\,; i = 1, \ldots, n_\ell, \ell = 1, \ldots, d\}$, gives us a tick database of prices for d number of assets. Let $0 = \tau_0 < \cdots < \tau_N = 2\pi$ be a (subjectively) specified time grid with a total of $N + 1$ points. We can align the observed times to such a common grid by the previous tick-time method among many methods available. Define for asset $\ell = 1, \ldots, d$ and time index $k = 1, \ldots, N$, the closest previous observation time and the corresponding price

$$\tau_{k,\ell} = \max_i \{t_{i,\ell} : t_{i,\ell} \leq \tau_k\}, \quad P_\ell(\tau_k) = P_\ell(\tau_{k,\ell}).$$

The $\{\tau_{k,\ell}\}$ associated with each ℓ-th asset is a time stamp sampled to be aligned on the $\{\tau_k\}$'s grid. We assume that $N \to \infty$ and $\max_k(\tau_k - \tau_{k-1}) \to 0$. To create the $\{\tau_k\}$, two schemes are often employed: *fixed clock times* and *refresh times*. Let $\tau_0 = 0,\ \tau_{0,\ell} = 0$ and $\tau_{n,\ell} = 2\pi$ for $\ell = 1, \ldots, d$.

Fixed Clock Times: For $k \geq 1$, let

$$\tau_k = kh, \quad h := 2\pi/N. \tag{2.3}$$

Refresh Times: The refresh time was proposed by Barndorff-Nielsen et al. (2011) . First define a counting process associated with occurrences of transactions $N_\ell(t) := \sum_{i=1}^{n_\ell} 1\{t_{i,\ell} \leq t\}$. The refresh time grid is defined for $k \geq 1$ by

$$\tau_{k+1} = \max_\ell \{t_{N_\ell(\tau_k)+1,\ell}, \ell = 1, \ldots, d\}. \tag{2.4}$$

In words, the refresh time is the time at which all the assets are traded at least once since the last refresh time. In practice, if d is large this may lead to quite a small number of sample size. In this case, we may define refresh times pairwise.

The data alignment technique is not without problems. For example, if we use pairwise refresh times we obtain estimated covariance matrices that are not guaranteed to be positive definite. More seriously the single summation estimators such as Realized Covariance computed using aligned data are biased, i.e. in general $E\sum_{i=1}^{N}\Delta P_1(\tau_i)\Delta P_2(\tau_i) \neq E[P_1, P_2]$. Hayashi and Yoshida (2005) show the bias of the Realized Covariance calculated from *fixed clock time* aligned data, while Zhang (2010) shows the bias when the data is aligned by the *refresh times*. In practical example give in Section 2.3.2 and Section 2.4.1, we in detail analyze the bias induced by the data synchronization on time domain estimators of integrated covariance.

2.3 Estimation

2.3.1 Our Estimator

We propose to use the Fourier domain approach, which does not require data alignment at all. The nonparametric method based on Fourier analysis of returns was first introduced by Malliavin and Mancino (2009). Frequency domain techniques are widely used in estimating the long run variance of time series in traditional discrete time framework. Considerable attention has been paid to estimating the covariance matrix in the presence of autocorrelation of unknown form [see, inter alia: Bartlett (1946), Newey and West (1987), Andrews (1991), Hansen (1992).] An important application is the estimation of the long-run variance of nonstationary time series analysis. This is the special case of spectral density estimation at frequency zero. We draw a natural link of such traditional method to the estimating the quadratic covariation of continuous time processes.

The Fourier basis given by $\{g_t(q) := e^{iqt}, q \in \mathbb{Z}\}$ where $i = \sqrt{-1}$ and $\overline{g_t(q)}$ denoting its complex conjugate, constitutes an orthonormal basis on the interval $t \in [0, 2\pi]$,

$$\frac{1}{2\pi}\int_0^{2\pi} g_t(k)\overline{g_t(j)}dt = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{otherwise.} \end{cases}$$

We can express the continuous time signal $\{\Sigma(t)\}_{t\in[0,2\pi]}$ as a linear combination of

Fourier basis with coefficient denoted by $\mathcal{F}(\Sigma)(q)$ for $q \in \mathbb{Z}$

$$\Sigma(t) = \frac{1}{2\pi} \sum_{q=-\infty}^{\infty} \mathcal{F}(\Sigma)(q) e^{iqt}, \tag{2.5}$$

and its Fourier pair by

$$\mathcal{F}(\Sigma)(q) := \int_0^{2\pi} e^{-iqt} \Sigma(t) dt, \qquad q = 0, \pm 1, \pm 2, \ldots. \tag{2.6}$$

This is the continuous time Fourier transform of an instantaneous covariation matrix and at $q = 0$ we have the integrated covariance. We will propose an estimator for the above general form in (2.6). The above Fourier pair suggests that once we estimate the Fourier coefficient by $\widehat{\mathcal{F}}(\Sigma)(q)$, we may reconstruct the signal by replacing the infinite sum by the finite sum

$$\widehat{\Sigma}(t) = \frac{1}{2\pi} \sum_{q=-n}^{n} \widehat{\mathcal{F}}(\Sigma)(q) e^{iqt}.$$

By Assumption 1, we have $\{\Sigma(t)\} \in L^2([0, 2\pi])$ which guarantees that (2.5) is finite and $\|\widehat{\Sigma}(t) - \Sigma(t)\|_2 \to 0$. We next show how we can estimate (2.6) from the Fourier transform of the return process. We define the continuous time Fourier transform of return $dP_\ell(t)\,, \ell = 1, \ldots, d$ satisfying Assumption 1

$$\mathcal{F}(P_\ell)(\alpha) = \int_0^{2\pi} e^{-i\alpha t} dP_\ell(t), \qquad \alpha = 0, \pm 1, \pm 2, \ldots \tag{2.7}$$

where the integral is a stochastic integral. The discrete Fourier transform of the ℓ-th asset is

$$\mathcal{F}_n(P_\ell)(\alpha) = \sum_{j=1}^{n_\ell} e^{-i\alpha t_{j,\ell}} \Delta P_\ell(t_{j,\ell}). \tag{2.8}$$

Let $\mathcal{F}_n(P)(\alpha) = \{\mathcal{F}_n(P_1)(\alpha), \ldots, \mathcal{F}_n(P_d)(\alpha)\}^\intercal$ for $\alpha \in \mathbb{Z}$ denote the vector of such Fourier transforms. Denote a weight function, called the amplitude window, by $K_H(\cdot) : [-\pi, \pi] \to \mathbb{R}$. It suffices to note for now that the function is symmetric, centered at zero and integrates to a finite number over its support. The construction and the properties of the weight function are given in the next section. Our proposed estimator of (2.6) is given by

$$\widehat{\mathcal{F}}(\Sigma)(q) = \sum_{|\alpha| \leq m/2} K_H(\lambda_\alpha) \mathcal{F}_n(P)(\alpha) \mathcal{F}_n(P)(q-\alpha)^\intercal, \tag{2.9}$$

where for $\rho(n) := \max_{\ell=1,\cdots,d} n_\ell$, we define $\lambda_\alpha = 2\pi\alpha/\rho(n)$, for $\alpha \in \mathbb{Z}$. We let $m = \rho(n)/H$ where the bandwidth $H \to \infty$ and $\rho(n), m \to \infty$ as $n \to \infty$. We are smoothing λ_α over the interval $[-\pi/H, \pi/H]$ where H controls the width of the smoothing window. The main focus is on the case $q = 0$. We name our estimator, *Fourier Realized Kernel*. For $q = 0$, we may define the realized cross periodogram between assets 1 and 2 by $I_{12}(\alpha) := \mathcal{F}_n(P_1)(\alpha)\mathcal{F}_n(P_2)(-\alpha)$. Then $(1,2)$-th element of $\widehat{\mathcal{F}}(\Sigma)(0)$ is given by kernel smoothing the realized cross periodogram around the zero frequency

$$\widehat{\mathcal{F}}(\Sigma_{12})(0) = \sum_{|\alpha| \leq m/2} K_H(\lambda_\alpha) I_{12}(\alpha). \tag{2.10}$$

What is hidden in the frequency domain formulated estimator is that we can conveniently express our estimator as weighted double summation estimator given in (2.20).

2.3.2 Comparison with some Time domain estimators

For data that is synchronized at $\{\tau_i\}$, we may define a realized autocovariance function

$$\gamma_{12}(h) = \sum_i \Delta P_1(\tau_i) \Delta P_2(\tau_{i-h}), \quad h = 0, \pm 1, \pm 2, \cdots, \tag{2.11}$$

where $\sum_i = \sum_{h<i\leq n}$ for $h \geq 0$, and $\sum_i = \sum_{1\leq i \leq n+h}$ for $h < 0$. The realized periodogram is closely related to the realized autocovariance in the aligned case. In the case that τ_i are equally spaced and synchronous, i.e. $\tau_i = \tau_j + (i-j)2\pi/n$, we can conveniently write down the realized cross periodogram as a Fourier transform of the realized autocovariance, i.e., $I_{12}(\alpha) = \sum_{|h|<n} e^{-i\alpha h 2\pi/n} \gamma_{12}(h)$. We next make a comparison with the covariation estimator of Hayashi and Yoshida (2005). Their estimator is a realized cross periodogram at zero frequency over the interval that

overlaps, i.e.,

$$HY = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Delta P_1(t_{i,1}) \Delta P_2(t_{j,2}) 1_{\{I_{i,1} \cap J_{j,2} \neq \emptyset\}}.$$

The realized cross periodogram at zero frequency is given by

$$I_{12}(0) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Delta P_1(t_{i,1}) \Delta P_2(t_{j,2}). \tag{2.12}$$

Then the centered realized cross periodogram (2.12) can be decomposed into $I_{12}(0) - \int_0^{2\pi} \Sigma_{12}(t)dt = M_1 + M_2$, where

$$M_1 = HY - \int_0^{2\pi} \Sigma_{12}(t)dt, \quad M_2 = \sum_{i,j} \Delta P_1(t_i) \Delta P_2(s_j) 1_{\{I_{i,1} \cap J_{j,2} = \emptyset\}}.$$

Hayashi and Yoshida (2008) showed that $\sqrt{n} M_1$ is asymptotically zero mean Gaussian. M_2 has a zero mean and is a leading order term of $O_p(1)$ since $I_{12}(0) = \{P_1(2\pi) - P_1(0)\}\{P_2(2\pi) - P_2(0)\}$. In summary, if no microstructure noise is present, the Hayashi and Yoshida estimator has a zero bias and achieves $\sqrt{n}$ consistency. The realized periodogram is unbiased but inconsistent due to the extra term in M_2.

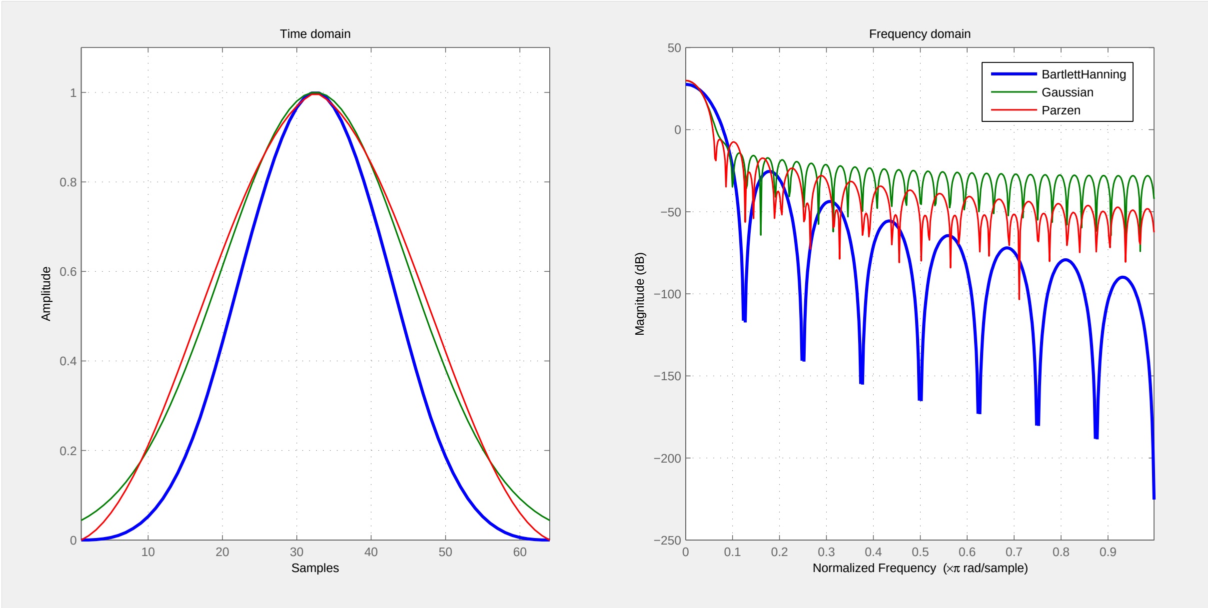
We next compare our estimator (2.10) to an estimator given by smoothing the realized autocovariances of the aligned data. Given a smoothing window in time domain $k(\cdot)$, define

$$\widetilde{\Sigma}_{12} = \sum_{|h|<n} k\left(\frac{h}{H}\right) \gamma_{12}(h). \tag{2.13}$$

This was first proposed by Barndorff-Nielsen et al. (2008a). To establish the relation between the time domain (2.13) and the frequency domain estimator (2.10) we now discuss the construction and properties of smoothing windows. We assume that the lag window satisfies the following conditions given by Barndorff-Nielsen et al. (2010). We will first work with a lag window for continuous time denoted by $k(x), x \in \mathbb{R}$ and a spectral kernel for continuous and bandlimited frequency denoted by $K(\lambda), \lambda \in [-\pi, \pi]$.

Assumption 3. *The lag window $k(\cdot)$ satisfies the following conditions:* (i) $k(0) =$

Figure 2.1: Lag and spectral window satisfying Assumption 3



$1, k'(0) = 0$; (ii) *k is twice continuously differentiable; (iii)* $\|k\|^2 := \int_{-\infty}^{\infty} |k(x)|^2 dx <$ $, \infty,$ $\|k^2\|^2 := \int_{-\infty}^{\infty} |k(x)|^4 dx < \infty,$ $\|k'\|^2 := \int_{-\infty}^{\infty} |k'(x)|^2 dx < \infty,$ $\|k''\|^2 := \int_{-\infty}^{\infty} |k''(x)|^2 dx < \infty,$ *where prime denotes the derivatives of kernel function. And* $k(x) \to 0$ *as* $x \to \infty$; (iv) $\int_{-\infty}^{\infty} k(x) \exp(-i\lambda x) dx \geq 0, \forall \lambda \in [-\pi, \pi]$.

The spectral window is defined by the Fourier transform of the lag window and vice versa

$$K(\lambda) = \int_{-\infty}^{\infty} k(t) e^{-i\lambda t} dt \qquad k(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} K(\lambda) e^{i\lambda t} d\lambda, \tag{2.14}$$

where λ denotes the angular frequency. Given this relation, Assumption 3 on the lag window k is equivalent to the following conditions on the spectral window K.

Assumption 3' *The spectral window K satisfies the following conditions:* (i) $\int_{-\pi}^{\pi} K(\lambda) d\lambda = 1$, $\int_{-\pi}^{\pi} \lambda K(\lambda) d\lambda = 0$; (ii) $\|K\|^2 := \int_{-\pi}^{\pi} |K(\lambda)|^2 d\lambda < \infty$, $\mu_1^2(K) := \int_{-\pi}^{\pi} |\lambda K(\lambda)|^2 d\lambda < \infty$ *and* $\mu_2^2(K) := \int_{-\pi}^{\pi} |\lambda^2 K(\lambda)|^2 d\lambda < \infty$; (iii) $K(\lambda) \geq 0$, $\forall \lambda \in [-\pi, \pi]$.

The condition (iv) in Assumption 3 and equivalently (iii) in Assumption 3' are needed to guarantee that the estimators defined in (2.13) and (2.10) are p.s.d. The realized periodogram is Hermitian and positive semi definite as long as the spectral window is non negative, i.e., $K(\lambda) \geq 0, \forall \lambda$. To avoid the aliasing problem we assume that the signal is zero for frequencies that falls outside of the Nyquist critical frequency, $f_c = n/2$. The results can be summarized as follows. The discrete time and discrete frequency Fourier pair are given by

$$K_H(\lambda_\alpha) = \tfrac{1}{n} \textstyle\sum_{|h| \leq H} k\left(\tfrac{h}{H}\right) e^{-i\lambda_\alpha h} \qquad k\left(\tfrac{h}{H}\right) = \textstyle\sum_{\alpha=-m/2}^{m/2-1} K_H(\lambda_\alpha) e^{i\lambda_\alpha h}. \tag{2.15}$$

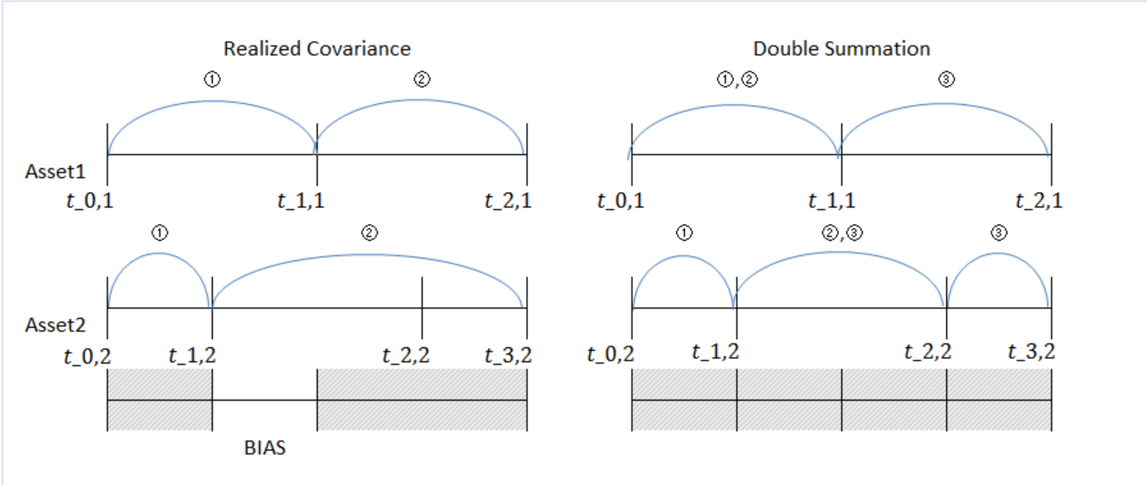
Figure 2.1 shows weighting functions that satisfy the Assumptions 3.

Proposition 1. *When trading times are synchronized and equally spaced the two estimators in* (2.10) *and* (2.13) *are identical when* (2.15) *holds.*

It is of interest how our estimator is related to other time domain estimators such as the multivariate two time scale estimator of Zhang (2010), and the Modulated Realized Covariance (multivariate pre-averaging estimator) of Christensen, Kinnebrock and Podolskij (2010). In the univariate setting, Jacod et al. (2009) showed that their pre-averaging estimator, the univariate two time scale estimator of Zhang et al. (2005) and the flat-top Realized Kernel of Barndorff-Nielsen et al. (2008a) can be written as a smoothed realized autocovariances where the difference between the estimators comes from the contribution of the end points. This result holds also for the multivariate versions of the three estimators when observation points are synchronized. Our estimator can be expressed as a Realized Kernel only when sampling points are equally spaced and aligned. The relation between the smoothed periodogram to estimate the spectrum and data tapering (i.e. Fourier transforming the weighted return) is analogous to the relation between our estimator and the pre-averaging estimator.

Toy Example We now consider an example to clarify the source of the bias in estimating integrated covariance due to aligning the non-synchronous observation points. Suppose that $P_1(t) = P_2(t) = B(t)$, an independent Brownian motion. Then $[P_1, P_2](1) = \int_0^1 dt = 1$. Assume that the observed price is given by $P_1(t_{i,1})$ with $\{t_{0,1} = 0, t_{1,1} = 1/2, t_{2,1} = 1\}$ and $P_2(t_{i,2})$ with $\{t_{0,2} = 0, t_{1,2} = 1/4, t_{2,2} = 3/4, t_{3,2} = 1\}$. Denote the union of time grid by $T_l := \{t_{i,1} \cup t_{j,2}, i = 0, 1, 2, j = 0, \ldots, 3\}$. The union of time grid is then simply $\{T_l = l/4, l = 0, \ldots, 4\}$. The refresh time grid is the same as the time stamp of the first asset $\{\tau_i = t_{i,1}, i = 0, 1, 2\}$. The previous

Figure 2.2: Toy Example



tick time for the first asset is obviously $\{\tau_{i,1} = t_{i,1}, i = 0, 1, 2\}$ and the second asset is $\{\tau_{0,2} = 0, \tau_{1,2} = t_{1,2} = 1/4, \tau_{2,2} = t_{2,3} = 1\}$. The expectation of the Realized Covariance is then given by

$$
\begin{aligned}
&E\{P_1(\tau_1) - P_1(\tau_0)\}\{P_2(\tau_1) - P_2(\tau_0)\} + E\{P_1(\tau_2) - P_1(\tau_1)\}\{P_2(\tau_2) - P_2(\tau_1)\} \\
&= E\{B(1/2) - B(0)\}\{B(1/4) - B(0)\} + E\{B(1) - B(1/2)\}\{B(1) - B(1/4)\} \\
&= 1/4 + 1/2 = 3/4.
\end{aligned}
$$

So the bias of Realized Covariance due to non-synchronicity is given by $1-3/4$. If two assets are positively correlated then the Realized Covariance will have a downward bias according to this derivation. However if we consider a double sum estimation of the form

$$
\begin{aligned}
&E\sum_{i,j}\{P_1(t_{i,1}) - P_1(t_{i-1,1})\}\{P_2(t_{j,2}) - P_2(t_{j-1,2})\} \\
=&E\sum_{0\le l\le 4}\{P_1(T_l) - P_1(T_{l-1})\}\{P_2(T_l) - P_2(T_{l-1})\} \\
=&E\left(\{B(1/4) - B(0)\}^2 + \{B(1/2) - B(1/4)\}^2 + \{B(3/4) - B(1/2)\}^2 + \{B(1) - B(3/4)\}^2\right) \\
=&4 \times 1/4 = 1.
\end{aligned}
$$

Therefore there is no bias induced by aligning the non-synchronously observed data. This is described graphically in Figure 2.2.

2.4 Asymptotic Properties

We first consider the asymptotic bias of the time domain estimator for subsequent comparison with our own. We then derive the asymptotic properties of our estimator in two cases: with and without microstructure noise.

2.4.1 Asymptotic Bias of the time domain estimator

This section focuses on showing the effect of the data synchronization on the covariance estimators. First, consider the Realized Covariance applied to the refresh time $\{\tau_i\}_{i=1}^N$ aligned data

$$\sum_{i=1}^{N} \Delta P_1(\tau_i)\Delta P_2(\tau_i) - \mathcal{F}(\Sigma_{12})(0). \tag{2.16}$$

Let $u_i = \tau_{i,1} \wedge \tau_{i,2}$ and $l_i = \tau_{i-1,1} \vee \tau_{i-1,2}$, then (*2.16*) can be expressed by

$$\sum_i \int_{l_i}^{u_i} \{P_1(t) - P_1(l_i)\} dP_2(t) + \int_{l_i}^{u_i} \{P_2(t) - P_2(l_i)\} dP_1(t) \tag{2.17}$$

$$+ \sum_i \int_{l_i}^{u_i} \Sigma_{12}(t)dt - \int_0^{2\pi} \Sigma_{12}(t)dt. \tag{2.18}$$

The order of (2.17) is $O_p(N^{-1/2})$ with zero expectation, while (2.18) contributes to a stochastic bias term, which is an analytical form for the so-called Epps effect. Theorem 1 of Zhang (2010) shows the order of magnitude for (2.18) is

$$\left(\sum_{i=1}^{n} \int_{l_i}^{u_i} - \int_0^{2\pi} \right) \Sigma_{12}(t)dt = -\int_0^{2\pi} \Sigma_{12}(u)dF(u) + O_p(1/n),$$

where $F(t) = \sum_{\{i:\tau_{i,1}\wedge\tau_{i,2}\le t\}} |\tau_{i,1} - \tau_{i,2}|$. She shows that $F(t) = O_p(\frac{N}{n_1+n_2})$. See Zhang (2010) Corollary 4 for the analytical form of the bias when arrival times are random with stochastic intensity. Consider now a Realized Kernel applied to the data aligned on $\{\tau\}_{i=1}^{N}$,

$$\sum_{|h|<n} k\left(\frac{h}{H}\right) r_{12}(h) = \sum_{i,j=1}^{N} \Delta P_1(\tau_i)\Delta P_2(\tau_j) k\left(\frac{i-j}{H}\right). \tag{2.19}$$

We also recognize that the estimator in (2.10) can be expressed as similar form,

$$\sum_{|\alpha|\leq m/2} K_H(\lambda_\alpha) I_{12}(\alpha) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \Delta P_1(t_i)\Delta P_2(s_j) k_H(t_i - s_j). \tag{2.20}$$

where we defined $k_H(t_i - s_j) := k(\frac{(t_i-s_j)/\widetilde{\Delta} t}{H}) = \sum_{|\alpha|\leq m/2} K_H(\lambda_\alpha) e^{-i(t_i-s_j)\alpha}$, i.e. we are scaling the difference of time stamps by the average interval size for more liquid asset. We first show for asymptotic bias of (2.20). Define $u_{ij} = t_i \wedge s_j$ and $l_{ij} = t_{i-1} \vee s_{j-1}$. What matters for the bias term, as shown in the proof for Theorem 1 is, conditionally on $1_{\{i,j|u_{ij}>l_{ij}\}}$,

$$\begin{aligned}
&\sum_{i,j} \Delta P_1(t_i)\Delta P_2(s_j) k_H(t_i - s_j) - \int_0^{2\pi} \Sigma_{12}(t) e^{-iqt} dt \\
&= \sum_{i,j} \int_{l_{i,j}}^{u_{i,j}} \{P_1(t) - P_1(l_{i,j})\} dP_2(t) + \int_{l_{i,j}}^{u_{i,j}} \{P_2(t) - P_2(l_{i,j})\} dP_1(t) && (2.21) \\
&+ \sum_{i,j} \int_{l_{i,j}}^{u_{i,j}} dP_1(t) \int_{l_{i,j}}^{u_{i,j}} dP_2(s)\{1 - k_H(t_i - s_j)\} && (2.22) \\
&+ \sum_{i,j} \int_{l_{i,j}}^{u_{i,j}} \Sigma_{12}(t) dt - \int_0^{2\pi} \Sigma_{12}(t) dt && (2.23) \\
&= O_p(\{n_1 \vee n_2\}^{-1/2}) + O_p(\{\frac{n_1 \vee n_2}{H(n_1 \wedge n_2)}\}^2) + O_p(\{n_1 \vee n_2\}^{-1}).
\end{aligned}$$

The asymptotic bias term (2.21) and (2.23) are due to the discretization error of the continuous time signal, which depends inversely on the number of union of time stamps for two assets. (2.22) is due to smoothing, which can be controlled as it depends on the bandwidth. See Appendix. We have an asymptotic bias that vanishes in large sample and we do not have a synchronization error of form (2.18). The asymptotic bias term of (2.19) can be derived similarly, by replacing the transaction time stamp t_i by refresh time $\tau_{i,1}$ and s_j by $\tau_{j,2}$. The order of union of refresh time aligned grid, $\{\tau_{i,1} \cup \tau_{j,2}\}$ is $O_p(N)$ which usually in practice is $o_p(n_1+n_2)$.Whereas the union of all time stamps $\{t_i \cup s_j\}$ is larger order at $O_p(n_1 \vee n_2)$. For the estimator of form (2.19) and (2.20), we effectively discretize the signal by the union of two times. Using all the data realized at transaction times is a much finer approximation for the real line $[0, 2\pi]$ than the coarser refresh time. The asymptotic bias term of the Realized Kernel is given by $O_p(N^{-1/2}) + O_p(H^{-2}) + O_p(N^{-1})$, where bandwidth H is chosen for the Realized Kernel. If we let $N = (n_1 \wedge n_2)^R$, then under the optimal bandwidth, our estimator converges faster at $(n_1 \wedge n_2)^{\frac{4-2\beta}{5}}$ than the Realized Kernel at $(n_1 \wedge n_2)^{\frac{2R}{5}}$, when $R < 2- \beta$ i.e. $\frac{N}{n_1 \wedge n_2} = o(1)$. In 2 dimensional case, the condition will hold when two assets are traded very asynchronously and it will likely hold when we are estimating the large dimensional covariance matrix.

Another conceptual problem of the refresh time alignment method is that it necessitates the return of the illiquid asset leads the return of the liquid asset, which is undesirable. It is also more natural to formulate the assumption on the microstructure noise in terms of the actual transaction time rather than the refresh time.

2.4.2 Asymptotic Distribution of our Estimator without Microstructure Noise

We consider the case where the sample sizes of different assets may not be of the same order of magnitude. This situation arises often in practice, since some assets are traded much more frequently than others. To apply the We need the following rate condition.

Assumption 4. *H is a bandwidth satisfying $H \propto n^{\alpha}$ with $\alpha \in (0,1)$ so that we have as $n \to \infty$, $H \to \infty$ and $m := n/H \to \infty$. Also assume that $\frac{n_a \vee n_b}{n_a \wedge n_b} = o(H)$ for all $a, b \in \{1, \dots, d\}$.*

Remark Let β be a degree of liquidity parameter so that $n_a \vee n_b = O((n_a \wedge n_b)^{\beta})$, $1 \leq \beta$. Then Assumption 4 implies that $1 \leq \beta < 2$.

By balancing the squared bias and the variance given in Proposition 1, the optimal bandwidth is given by $H = C_0 n^{\alpha^*}$, $\alpha^* = \frac{4\beta-3}{5}$, where $C_0 \in (0, \infty)$. Then the convergence rate of the estimator under the optimal bandwidth is given by $(n_1 \wedge n_2)^{\vartheta}$,

$0 < \vartheta := -\frac{2}{5}\beta + \frac{4}{5} \leq \frac{2}{5}$. The result makes intuitive sense that for unbalanced sample sizes, the estimator converges at slower rate than the balanced case, $n^{2/5}$. As the discrepancy between the liquidity of asset increases (higher β), the estimator becomes less efficient. Define for each $a = 1, \ldots, d$, $\mathcal{B}_{aa} = 0$ and

$$\mathcal{B}_{ab} = C_0^{-2}|k''(0)|\frac{\mathcal{A}^2}{2}\int_0^{2\pi} e^{-itq}|\Sigma_{ab}(t)|dt, \quad \mathcal{A} := \lim_{n_1\wedge n_2\to\infty} \sup_{i,j} \frac{n_1 \wedge n_2}{2\pi}|t_i - s_j|1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}},$$

where $0 \leq \mathcal{A} < \infty$ under Assumption 2. $\mathcal{A}$ could be thought as a measure of the degree of non-synchronicity. When the two series are perfectly synchronized and balanced then $\mathcal{A} = 0$; otherwise it is $O(1)$ under the Assumption 2.3. Define the asymptotic variance for the typical diagonal and off diagonal element:

$$\begin{aligned}
\mathcal{V}_{aa} &= 2C_0||k||^2 \int_0^{2\pi} e^{-i2qt}\Sigma_{aa}^2(t)d\mathcal{Q}_{aa}(t) \\
\mathcal{V}_{ab} &= C_0||k||^2 \int_0^{2\pi} e^{-i2qt}\left\{\Sigma_{aa}(t)\Sigma_{bb}(t)d\mathcal{Q}_{aabb}(t) + \Sigma_{ab}^2(t)d\mathcal{Q}_{abab}(t)\right\}.
\end{aligned}$$

The covariation between the integrated covariance estimator of asset a and b with the estimator of c and d is given by

$$\mathcal{V}_{ab,cd} = C_0\|k\|^2 \int_0^{2\pi} e^{-i2tq}\left\{\Sigma_{ac}(t)\Sigma_{bd}(t)d\mathcal{Q}_{acbd}(t) + \Sigma_{ad}\Sigma_{bc}(t)d\mathcal{Q}_{adbc}(t)\right\},$$

and let $\mathcal{B}$ and $\mathcal{V}$ be a *vech* of bias and covariance matrix of our estimator. Define $\mathcal{D}_n^*$ be the matrix of convergence rates,

$$\begin{aligned}
\mathcal{D}_n = \text{diag}\left\{\text{vech}(\mathcal{D}_n^*)\right\} \quad ; \quad & \{\mathcal{D}_n^*\}_{a,a} = \sqrt{n_a} \\
& \{\mathcal{D}_n^*\}_{a,b} = (n_a \wedge n_b)^{\vartheta}, \quad \vartheta = \frac{4-2\beta}{5}, 1 \leq \beta < 2,
\end{aligned}$$

where the upper bound $\vartheta = 2/5$ is obtained when the sample sizes are of the same order.

THEOREM 1 *Suppose that Assumptions 1 - 4 hold. Then for each* $q \in \mathbb{Z}$,

$$\mathcal{D}_n \text{vech}\left\{\widehat{\mathcal{F}}(\Sigma)(q) - \mathcal{F}(\Sigma)(q)\right\} \Longrightarrow N\left(\mathcal{B}, \mathcal{V}\right).$$

REMARK When data is synchronized and balanced we have $\mathcal{B}_{ab} = 0$ and the covariation estimator achieves the same rate of convergence as the variance estimator. Our result is comparable with Malliavin and Mancino (2009) whose results were under sub-optimal bandwidth.

REMARK ON EFFICIENCY If our goal is to achieve the most efficient estimator, we can estimate the asymptotic bias term and subtract it from our estimator. In that case we can get $\sqrt{n}$ convergence rate at the cost of sacrificing the positive definiteness of the estimator. We also may estimate each element of the covariance matrix in most efficient way and use the clipping method to achieve p.s.d i.e. we can project a $d \times d$ symmetric covariance matrix estimate which has singular value decomposition, $U^T diag[\lambda_1, \cdots, \lambda_p]U$ as $U^T diag[\lambda_1^+, \cdots, \lambda_p^+]U$ where $\lambda_j^+ = \max\{\lambda_j, 0\}$. Whether we should emphasize on the efficiency of an estimator or on the covariance estimator that is guaranteed to be positive definite depends on problem at hand and we leave this choice to the practitioner.

2.4.3 Asymptotic Distribution of our Estimator with Microstructure Noise

Assumption on the microstructure noise

The empirical evidence from the volatility signature plot suggests that the observed price deviates from the semimartingale assumption. More precisely various studies document that the observed high frequency returns have infinite quadratic variation. To model this phenomena, we make the following assumption.

ASSUMPTION 5. *Let* $X_j(t_{i,j})$ *is an observed log price of* j*-th asset which has two additive components. One is a discretely observed continuous signal* $P_j(t_{i,j})$ *that satisfies the semimartingale Assumption 1 and another component is a noise process with respect to the realization of transaction time* $U_j(t_{i,j})$ *that has an infinite quadratic* *variation*

$$X_j(t_{i,j}) = P_j(t_{i,j}) + U_j(t_{i,j}). \tag{2.24}$$

In univariate studies, it is usually assumed that $U_j(t_{i,j})$ is a stationary time series, which has been supported by empirical studies. There has not been a lot of empirical work studying the cross autocorrelation of the microstructure noise for the multiple asset case. In the limited theoretical work in this area, Aït-Sahalia et al. (2010a) assumed i.i.d noise that is uncorrelated across different assets. Barndorff-Nielsen et al. (2011) assumed that the noise is covariance stationary with respect to a refresh time. Their assumption on the diagonal makes sense - there is an evidence that under financial clock i.e., transaction time, the process is homogeneous and less serially correlated. The off-diagonal assumption needs verification. Zhang (2010) assumed the alpha mixing condition with respect to an observation time. We think it is realistic to assume the following for the microstructure noise.

Assumption 6. *Let $U_j(.), j = 1, \ldots, d$ be a n dimensional stationary process, independent of the efficient price process with $E(U_j(.)) = 0$ and covariance function defined by $EU_a(t_{i,a})U_b(t_{j,b}) = \gamma(|t_{i,a} - t_{j,b}|/\widetilde{\Delta t}_{ab})$ that satisfies*

$$\frac{1}{n_a \wedge n_b} \sum_{i=1}^{n_a-1} \sum_{j=1}^{n_b-1} \gamma(|t_{i,a} - t_{j,b}|/\widetilde{\Delta t}_{ab}) \to \Gamma_{ab},$$

where Γ is a $d \times d$ p.s.d. covariance matrix with (a,b)-th element denoted by Γ_{ab}. We also assume that $|\mathrm{E}(U_a(t_{i,a})U_b(t_{j,b}), U_c(t_{r,c})U_d(t_{l,d}))| \le \rho(M)$, where $M := \sup_{\{p,s\}\in\{\{a,c\},\{b,d\}\}}^{\{u,v\}} \{(t_{u,p} - t_{v,s})/\widetilde{\Delta t}_{ps}\}$ and $\sum_{\nu}^{\infty} \rho(\nu)(1+\epsilon)^{\nu} < \infty$ for some $\epsilon > 0$.

This assumption is consistent with the usual univariate microstructure noise model. For the equally spaced balanced case the assumption simplifies to

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma(|i-j|) \to \Gamma_{ab} = O(1) \quad ; \quad M := \max\{|i-j|, |h-l|\}.$$

This allows cross-sectional correlation in the measurement error process.

Distribution Theory

In this section we will show that our estimator in (2.9) is a consistent estimator of the Fourier transform of the covariance matrix even under the presence of microstructure noise. We have the following decomposition for (2.9) at zero frequency $q = 0$:

$$
\begin{aligned}
\hat{\mathcal{F}}(\Sigma)(0) - \mathcal{F}(\Sigma)(0) &= \sum_{\alpha} K_H(\lambda_\alpha) I(\alpha) - \mathcal{F}(\Sigma)(0) \\
&= \sum_{\alpha} K_H(\lambda_\alpha)[\mathcal{F}_n(P)(\alpha)\mathcal{F}_n(P)(-\alpha)^\intercal - \mathcal{F}(P)(\alpha)\mathcal{F}(P)(-\alpha)^\intercal] \\
&+ \sum_{\alpha} K_H(\lambda_\alpha)\mathcal{F}(P)(\alpha)\mathcal{F}(P)(-\alpha)^\intercal - \mathcal{F}(\Sigma)(0) \\
&+ \sum_{\alpha} K_H(\lambda_\alpha)\mathcal{F}_n(dU)(\alpha)\mathcal{F}_n(dU)(-\alpha)^\intercal \\
&+ \sum_{\alpha} K_H(\lambda_\alpha)[\mathcal{F}_n(dU)(\alpha)\mathcal{F}_n(P)(-\alpha)^\intercal + \mathcal{F}_n(P)(\alpha)\mathcal{F}_n(dU)(-\alpha)^\intercal] \\
&= (i) + (ii) + (iii) + (iv).
\end{aligned}
$$

The term (i) is the error due to sampling the continuous time signal at discrete points. (ii) is the error due to smoothing. (iii) is a contribution from the smoothed realized periodogram applied to a microstructure noise and (iv) is due to the cross term between the efficient price and the noise. We will show that (ii) is a leading order term with $O_p(\sqrt{\frac{H}{n}})$ and is asymptotically normal. The bias term is given by (i)+(iii)+(iv), where the leading term is (iii) with $O_p(n/H^2)$. The estimator is asymptotically unbiased when $n/H^2 \to 0$ as $n, H \to \infty$. We add one further assumption on the end points.

Assumption 7. *The two end points, $X_j(t_{0,j})$ and $X_j(t_{n,j})$ are respectively an average of m_0 number of distinct observations on the interval $[t_{-1,j}, t_{0,j})$ and $[t_{n,j}, t_{n+1,j})$.*

This assumption turns about to be crucial for our estimator to achieve consistency. The time domain estimator by Barndorff-Nielsen et al. (2011) also assumes this condition. We derive the rate of convergence of our estimator by balancing the asymptotic variance of order $O_p(\frac{H}{n_1 \wedge n_2})$ and the asymptotic bias of order $O_p(\frac{n_1 \vee n_2}{H^2})$. See Appendix. Let the reference sample size, $n = \min(n_1, n_2) = n_1$ and let $n_2 = O(n^\beta)$ and $H = O(n^\alpha)$. For the asymptotic variance to vanish we require $n^{\alpha-1} \to 0$ and for the asymptotic bias to vanish we require $n^{\beta-2\alpha} \to 0$. Assumption 4 guarantees this. At the optimum, we balance the order of the squared bias and variance; $n^{2\beta-4\alpha} = n^{\alpha-1}$. Solving this for the bandwidth rate, the optimal bandwidth is given by

$$H = C_0 n^{\alpha^*}, \ \alpha^* = \frac{2\beta + 1}{5},$$

where $C_0 \in (0, \infty)$. When the two sample sizes are the same order i.e.($\beta = 1$), then $\alpha^* = 3/5$. In general, liquidity parameter $1 \leq \beta < 2$ implies that $\frac{3}{5} \leq \alpha^* < 1$. The rate of convergence is then $(n_1 \wedge n_2)^\vartheta$, $0 < \vartheta := \frac{2-\beta}{5} \leq \frac{1}{5}$, where the upper bound is achieved when the sample sizes are of the same order. We define a finite tuning parameter η in a following way. There exists $C^* \in (0, \infty)$ such that $n_\ell \sup_{0 \leq i \leq n_\ell} \Delta t_{i,\ell} \leq C^*$ for $\forall \ell = 1, \cdots, d$ under Assumption 2.1. We define $\eta = (\frac{C^*}{2\pi})^2$. Let denote

$$\mathcal{B} = \text{vech}(C_0^{-2} \eta |k''(0)| \Gamma),$$

and let $\mathcal{V}$ be as defined in Theorem 1. Let $\mathcal{D}_n^*$ be the matrix of convergence rates

$$\mathcal{D}_n = \text{diag}\left\{\text{vech}(\mathcal{D}_n^*)\right\} \quad \{\mathcal{D}_n^*\}_{a,b} = (n_a \wedge n_b)^\vartheta \quad \vartheta = \frac{2-\beta}{5}, 1 \leq \beta < 2,$$

where the degree of liquidity parameter β is defined in Theorem 1. The the upper bound for ϑ is 1/5 which is obtained when $n_a/n_b = O(1)$.

Theorem 2. *Suppose that Assumptions 1-7 hold. Then for each $q \in \mathbb{Z}$*

$$\mathcal{D}_n \text{vech}\left\{\widehat{\mathcal{F}}(\Sigma)(q) - \mathcal{F}(\Sigma)(q)\right\} \Longrightarrow N(\mathcal{B}, \mathcal{V}).$$

2.5 Extension

In this section, we further extend the Fourier method discussed above to estimate the instantaneous covariance matrix of diffusion process. We also appeal to Fourier analysis to estimate the autocovariance function of the microstructure noise.

2.5.1 Estimation of the Instantaneous covariance matrix

The *instantaneous* covariance matrix is also a parameter of interest, see Kristensen (2010). We can construct an estimator of instantaneous covariation matrix by Fourier inverting the estimator given in (2.9)

$$\widehat{\Sigma}(t) = \frac{1}{2\pi} \sum_{|q|\leq m/2} K_H(\lambda_q) \exp(iqt) \widehat{\mathcal{F}}(\Sigma)(q). \tag{2.25}$$

Suppose that the modulus of continuity of $\Sigma(t)$ denoted by $\mathcal{C}(h)$ is given by

$$\mathcal{C}(h) := \sup_{|t-s|\leq h} \|\Sigma(t) - \Sigma(s)\|_2. \tag{2.26}$$

The continuity assumption is met when each element of $\Sigma(t)$ in Assumption 1 does not contain jumps, for example $\Sigma(t)$ is a Brownian semimartingale.

THEOREM 3. *(Consistency of the instantaneous covariance matrix estimator under the presence of noise). Suppose that the assumptions of Theorem 2 hold and that* (2.26) *holds. Then, there exists a sequence* $\delta(n) \to 0$, *such that*

$$\lim_{n\to\infty} \sup_{\delta(n)\leq t\leq 2\pi-\delta(n)} \|\widehat{\Sigma}(t) - \Sigma(t)\|_2 = 0.$$

2.5.2 Estimation of ACFs of Microstructure noise

Under Assumption 6, we may appeal to the benefit of frequency domain analysis to estimate the cross autocorrelation structure of the microstructure noise of high frequency prices for multiple assets. The idea is that the conventional spectral density estimation applied to the high frequency returns and Fourier inversion of it will reveal the ACF structure of the microstructure noise. This is recognizing that the observed return has an $O_p(1)$ component that comes from the microstructure noise and a smaller vanishing term $O_p(n^{-1/2})$ that is coming from the semimartingale component.

See Matsuda and Yajima (2009) for studies of periodogram applied to multidimensional processes observed on asynchronous points in the space. For asynchronously observed data we can take following steps. In terms of notation, the double subscript f_{xx} and I_{xx} is to emphasize that we are referring to a second order spectral density and periodogram that we know conventionally in discrete time setting, for example as in Brockwell and Davis (1991):

1. Estimate the spectral density of the observed returns $\{\Delta X_j(t_{i,j})\}_{i,j}$ using the conventional method by smoothing the periodogram

$$\widehat{f}_{xx}(q) = \sum_{|q-\alpha|\leq m/2} K_H(\lambda_q - \lambda_\alpha) I_{xx}(\alpha), \tag{2.27}$$

   where $\lambda_q = q\widetilde{\Delta t}$ and the second order periodogram is given by

$$I_{xx}(\alpha) := \frac{1}{n} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Delta X_1(t_i) \Delta X_2(s_j) e^{-i(t_i - s_j)\alpha}.$$

2. Fourier invert the estimated spectral density to obtain the estimate of autocovariance.

$$\widehat{\gamma}_{uu}(\tau) = \frac{1}{2\pi} \sum_{|q|\leq m/2} K_H(\lambda_q) \exp(i\lambda_q \tau) \widehat{f}_{xx}(q).$$

3. Reconstruct the ACF of the un-differenced noise

$$\widehat{\gamma}_{UU}(\tau) = \frac{\widehat{\gamma}_{uu}(\tau)}{(1-L)^2}.$$

THEOREM 4. *Suppose that Assumptions 2-4 and 6 hold. Then there exists a sequence $\tau(n) \to \infty$ such that*

$$\lim_{n\to\infty} \sup_{|h|\leq \tau(n)} \|\widehat{\gamma}_{uu}(h) - \gamma_{uu}(h)\|_2 = 0.$$

2.6 Application - Multivariate Regression

In this section we provide a framework for continuous time multivariate regression with non-synchronously observed data and show how to consistently estimate the regression coefficient. Often, practitioners encounter a problem of running a regression between variables that are asynchronously observed - for example we might be interested in the effect of returns and order book information of one asset on another asset. Hannan (1975) and Robinson (1975) are the earlier literature on using frequency domain to solve such problems. Mykland and Zhang (2006) discussed a general the set up of analysis of variance for continuous time regression.

Let $S(t)$ be a dependent variable and $P_j(t)$, $j = 1, \ldots, d$ be d regressors. We assume that $\{S(t), P_j(t), j = 1, \ldots, d\}$ satisfies Assumption 1. We define a residual process $Z(t)$ by

$$dZ(t) = dS(t) - \sum_{j=1}^{d} \beta_j dP_j(t).$$

The regression coefficients are estimated by minimizing the quadratic variation of the residual process

$$\min_{\beta_j, j=1,\ldots,d} [Z, Z]_t.$$

The regressors are correlated in a sense of Assumption 1. When we have two regressors, the solution to the optimization problem is

$$\beta_1(t) = \frac{d[P_1, S](t)}{d[P_1, P_1](t)} - \frac{d[P_2, P_1](t)}{d[P_1, P_1](t)}\beta_2(t).$$

Plugging in the solution for $\beta_2(t)$, we have

$$\beta_1(t) = \frac{d[\widetilde{P_1}, S](t)}{d[\widetilde{P_1}, \widetilde{P_1}](t)},$$

where $d\widetilde{P_1}$ is an orthogonal projection of dP_1 on dP_2

$$d\widetilde{P_1}(t) = dP_1(t) - dP_2(t)\frac{d[P_2, P_1](t)}{d[P_2, P_2](t)}.$$

In particular, for non-time varying coefficient $\beta_1 = \widetilde{[P_1, S]}(t)/[P_1, P_1](t)$. It is easy to see the analogy with a discrete time linear regression problem. However we do not observe the continuous time process. Using the method described in the previous sections we can consistently estimate quadratic (co) variations based on discretely observed data and can estimate the regression coefficient even when the variables are observed with error (under the presence of microstructure noise).

In general we are often interested in a scalar function of $\theta = \mathrm{vech}(\Sigma)$. We will denote such function by $\upsilon(\theta)$. The simplest example is the selection operation that picks out one element of θ. The theory for this is given in Theorem 1 and 2. The linear regression problem discussed above also can be thought in this framework, where $\upsilon(\cdot)$ is a non-linear function in the elements of θ. Other examples of $\upsilon(\cdot)$ are eigenvalues, trace and determinant of the covariance matrix. For portfolio management, we are interested in $\upsilon(\theta) = w^\intercal \theta$ for a vector of weights w. We may study the asymptotic distribution for a scalar function of the integrated covariance matrix. Under smoothness conditions, we have $\delta_n\{\upsilon(\hat{\theta}) - \upsilon(\theta)\} = O_p(1)$ where δ_n is such that $\delta_n \frac{d\upsilon(\theta)}{d\theta^\intercal}\mathcal{D}_n^{-1} \to 1$ and $\mathcal{D}_n$ is a convergence rate matrix given in Theorem 1 and 2.

2.7 Numerical Study

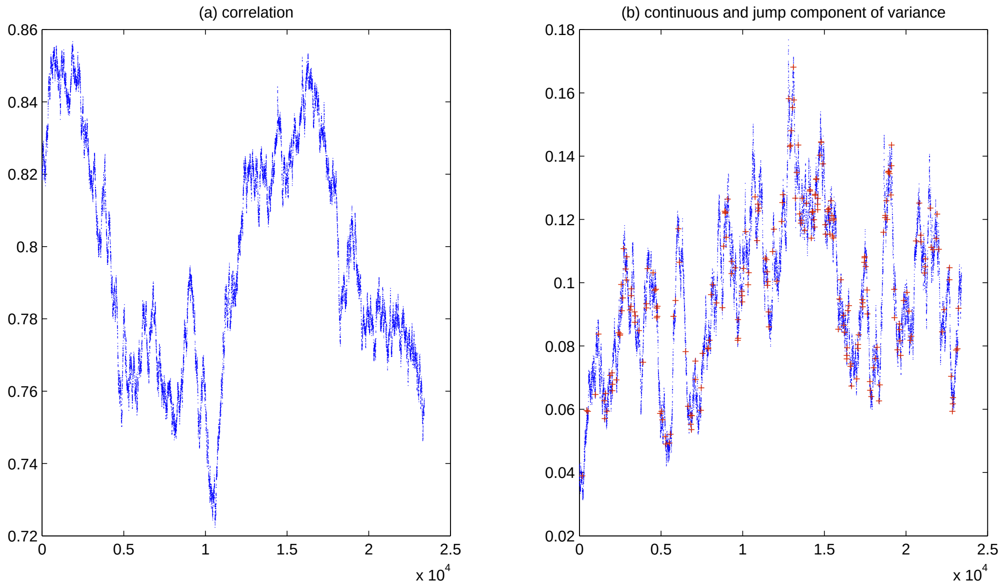
2.7.1 Estimator of co-volatility comparison

We have the following versions of our estimator:

$$\begin{aligned}
\sum_{\alpha} K_H(\lambda_\alpha) I(\lambda_\alpha) &:= \sum_{i,j=1}^{n} \Delta P_1(t_i)\Delta P_2(s_j) \sum_{\alpha} K_H(\lambda_\alpha) e^{-i(t_i - s_j)\alpha} \\
&=_{(1)} \sum_{i,j=1}^{n} \Delta P_1(t_i)\Delta P_2(s_j) k_H(t_i - s_j) \\
&=_{(2)} \sum_{|h|<n} k\left(\frac{h}{H}\right)\gamma_h,
\end{aligned}$$

where (2) holds only when the discretization points are synchronous and equally spaced. The form of estimator we will implement is (1). In the theoretical work, we

Figure 2.3: Simulated intraday instantaneous co-volatility and variance



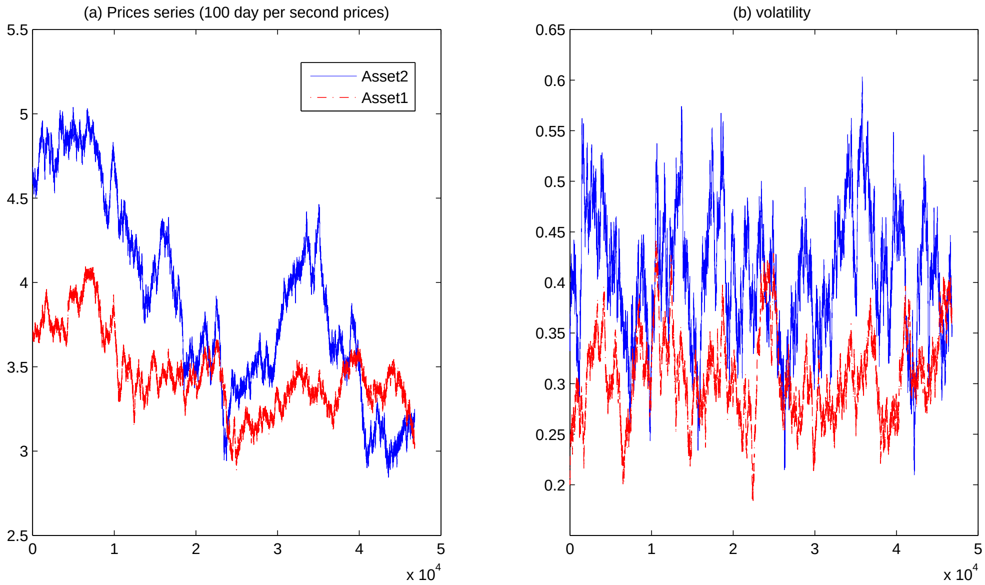
assumed no leverage between the volatility and the return process. In the simulation studies, we relax this assumption and see if our estimator is robust to a presence of the leverage. We consider two data generating processes for asset returns. For first simulation, we consider the stochastic volatility model with a perfect leverage given in Barndorff-Nielsen et al. (2011). The volatility process is continuous and the instantaneous co-volatility is constant. For $j = 1, 2$-th asset;

$$
\begin{aligned}
dP_j(t) &= 0.03dt - 0.3\sigma_j(t)dB_j(t) + \sqrt{1-(0.3)^2}\sigma_j(t)dW(t) \qquad (2.28)\\
\sigma_j(t) &= \exp\{-5/16 + 1/8\varrho_j(t)\} \quad ; \quad d\varrho_j(t) = -1/40\varrho_j(t)dt + dB_j(t).
\end{aligned}
$$

$\varrho_j(t)$ is initialized by $\varrho_j(0) \sim N(0, 20)$. The model implies that the covariance between the returns are $EdP_1(t)dP_2(t) = 0.91\sigma_1(t)\sigma_2(t)dt$. There is a perfect statistical leverage since a single Brownian motion $B_j(t)$ which is present in the return equation, drives the volatility process.

For a second simulation, the stochastic volatility is specified as a jump diffusion process and the instantaneous co-volatility coefficient follows CIR process. This is

Figure 2.4: Simulated price and variance - per second observation



modification of DGP considered in Aït-Sahalia et al. (2010a) and Barndorff-Nielsen et al. (2004a). For $j = 1, 2$-th asset

$$dP_j(t) = \sigma_j(t)dW_j(t) \tag{2.29}$$
$$d\sigma_j^2(t) = \kappa_j\{\bar{\sigma}_j^2 - \sigma_j^2(t)\} + a_j\sigma_j(t)dB_j(t) + \sigma_j(t_-)J_j(t)dN_j(t),$$

The jump size follows $J_j(t) = \exp\{z_j(t)\}$ with $z_j(t) \sim N(\mu_j, s_j)$ and $N_j(t)$ is a poisson process with intensity λ_j. Let the leverage effect be given by $EdW_j(t)dB_j(t) = \delta_j dt$. We use parameter values given in Aït-Sahalia et al. (2010a). The covariance between the Brownian motions that are present in the price equation is given by $EdB_1(t)dB_2(t) = \rho_t dt$. We let $\rho(t) = \frac{e^{2x(t)}-1}{e^{2x(t)}+1}$ and $x(t)$ follows CIR process

$$dx(t) = 0.03(0.64 - x(t))dt + 0.118x(t)dB_{xt}.$$

The Figure 2.3(a) shows the time series plot of ρ_t and (b) shows $\sigma_1^2(t)$ decomposed into continuous and discontinuous component. The Figure 2.4(a) shows the time series plot of $P_j(t), j = 1, 2$ and (b) shows $\sigma_j^2(t), j = 1, 2$. The DGP of Microstructure noise is formed with respect to transaction time. We consider correlated AR(1) noise processes with smooth decaying cross autocovariances. This can be implemented by

$$U_j(t_{i,j}) = \bar{U}_j(t_{i,j}) + \varepsilon(t_{i,j}) \quad : \quad \bar{U}_j(t_{i,j}) = \alpha\bar{U}_j(t_{i-1,j}) + \epsilon_j(t_{i,j}), \tag{2.30}$$

where idiosyncratic errors are independent Gaussian; $\epsilon_j(t_{i,j}) \sim NID(0, 1)$. The common disturbance that will drive the correlation between the two microstructure noise is simulated by

$$\varepsilon_l = 0.5\varepsilon_{l-1}+\xi_l, \text{ for } \{T_l\}_{1\leq l\leq N_T} = \{t_{i,1}\cup t_{j,2}, i = 1, \ldots, n_1, j = 1, \ldots, n_2\}, \ \xi \sim NID(0, 1).$$

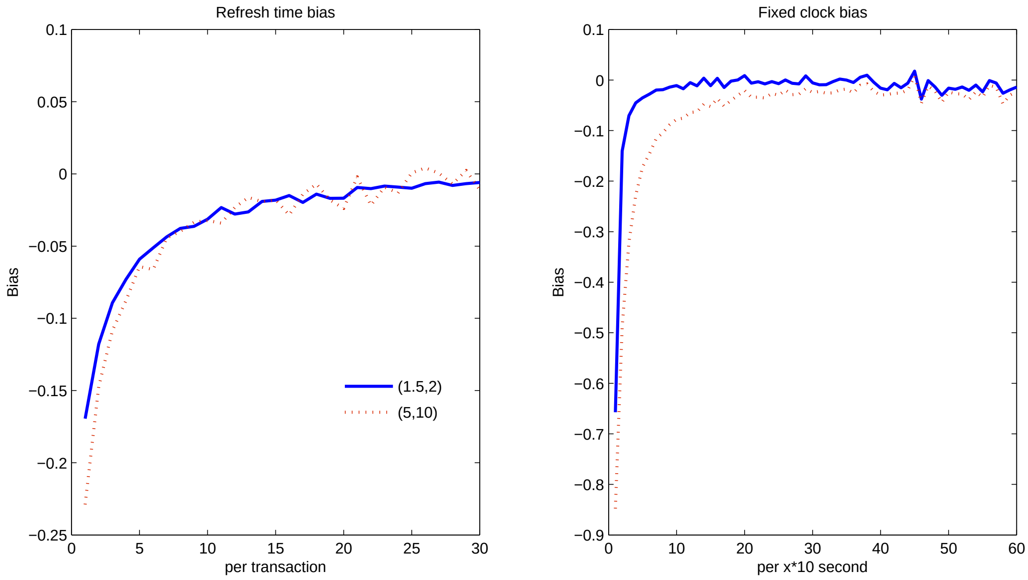
Then we define $\{\varepsilon(t_{i,1})\}_{1\leq i\leq n_1}$ as $\{\varepsilon_l\}_{1\leq l\leq N_T}$ sampled at $\{T_l\cap t_{i,1}\}$ points. $\{\varepsilon(t_{j,2})\}_{1\leq j\leq n_2}$ is similarly defined. The variance of the noise is set to be proportionate to the sample integrated quarticity; $\zeta^2\sqrt{n_j{}^{-1}\sum_{i=1}^{n_j}\sigma_j^4(t_{i,j})}$, where $\zeta = \{0, 0.0.001, 0.01\}$ is a noise to signal ratio. We simulated the one second data assuming 6.5 hour daily trading, which give us 23,400 daily data points over 100 monte carlo sample. We designed the simulation to assess the impact of the asynchronicity on the estimator.

Finally, we examine properties of estimators in higher dimension. We consider a simple setting where log prices are given by $P(t) = AB(t)$ where $P(t)$ is 10×1 vector of prices, $B(t)$ is 3×1 independent Brownian motion and A is a factor loading matrix. This is poisson sampled at rate $\{2, 2, 4, 4, 8, 8, 10, 10, 30, 30\}$ and masked by i.i.d gaussian noise. Table 2.2 and 2.3 reports the results for estimating the 2 dimensional covariation matrix, where first asset is more often traded then the second asset. Table 2.4 reports the results for higher dimension.

Realized Covariation: bias induced by data synchronization

Table 2.1 reports the finite sample properties of the Realized Covariance. The efficient price follows Brownian semimartingale, given in (2.28). The transaction time follows a homogenous poisson process and the microstructure noise are correlated AR(1) processes given in (2.30). Asynchronous data is aligned using the 5 minute fixed clock time and the refresh time. The negative bias when no noise is present is consistent

Figure 2.5: Covariation signature plot



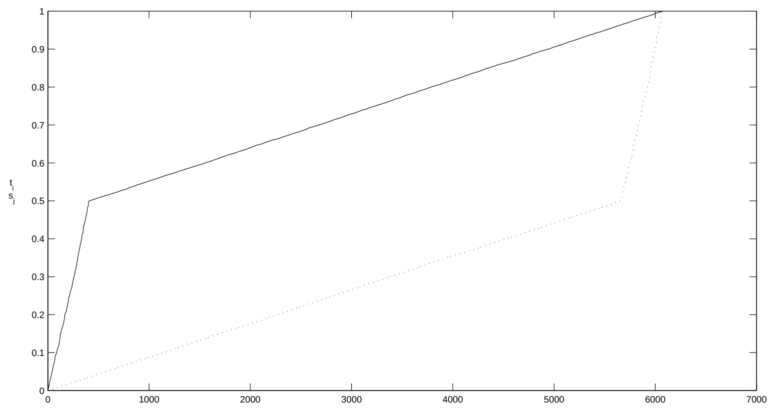
with the result in Section 2.4.1.

When microstructure noise is present, the variance estimate has a large positive bias. The sparse sampling (5 minute aligned data) is able to reduce such bias. However the covariance estimate has a negative bias induced by the Epps effect which dominates the positive bias induced by the microstructure noise. The degree at which Epps effect dominates the noise effect depends on the degree of non-synchronicity. The Figure 2.5 shows the covariation signature plot for the simulated series when the price is observed without the noise. It shows that given varying degrees of non-synchronicity (rate at which assets is traded), the higher frequency we align the data (moving leftwards in x axis) the more bias it induces in estimating the integrated covariance.

Balanced Sample example

We designed the simulation to assess the impact of the asynchronicity on the estimator. We created the non-synchronously observed prices by poisson sampling at the rate $(3/2, 30)$ in a following way. We first simulate the equally spaced data per one

Figure 2.6: Time stamps of two assets traded at opposite liquidity



second for two assets. For the first asset, we sample on average at 1.5 second for the first half of the sample and at 30 second for the last half of the sample. For the second asset, we do this in reverse order - sample at 30 second for the first half and 1.5 second for later part. See Figure 2.6. Then we have two assets that have the same number of transactions each day but traded very asynchronously. This is like a case where two assets have opposite liquidity profile over a day.

The sample size is $607,774$ over one hundred days and the refresh time aligned data reduces to a size around 750 per day. The large reduction in the sample size of the aligned data is due to severe non-synchronicity by simulation design. We compare the Realized Kernel and the proposed method over the range of bandwidths, $H = \{1, 5, 10, 20, 50, 100, \ldots, 750, 800\}$. The Figure 2.7 shows that the proposed estimator is less sensitive to the choice of bandwidth - especially for large H. With large H, we can reduce the bias for the off-diagonal element more than we can for the Realized Kernel. Our estimator is less sensitive to the choice of bandwidth for large values of H.

Un-Balanced Sample example

We carried out the same exercise as above but with the unbalanced sample sizes. We poisson sampled the data at rate $\{(3/2, 30), (3/2, 2), (20, 30)\}$. For example, sampling rate $(3/2, 2)$ means that we sample the first asset on average per 1.5 second and the second asset per 2 second. First rate is to examine the effect of different liquidity and different sample sizes. Second and third rates are to examine the effect of sparse and intense sampling of asset prices of similar liquidity.

Figure 2.8 shows the results for sampling rate $(3/2, 30)$. The proposed estimator has a less bias and is less sensitive to a choice of the bandwidth for large values of H. When the noise is not present, the proposed method estimates the variance of more liquid asset more precisely. When the noise is present, the bandwidth should be large for the proposed estimator to perform better. The conclusion is similar for sampling rate $(20, 30)$ as shown in Figure 2.9. The improvement of using the Fourier domain estimator is most evident when estimating the variance of more liquid asset when two sample sizes are very different. The proposed covariation estimator performs better under large bandwidth. For sampling rate $(3/2, 2)$ in Figure 2.10, the difference of two estimator is less pronounced.

Each of these figures also show the accuracy of estimating the scalar function of the covariation matrix. We examined the maximum eigenvalues and the variance of portfolio with weight $[0.5, \sqrt{0.75}]$. Under the realistic noise to signal ratio and when two assets are of different liquidity, the proposed method delivers superior estimates. Regardless of sampling scheme, the proposed method does better in estimating these quantities when effect of microstructure noise is not too dominant.

Overall Comparison and Higher Dimension Case

Table 2.2 and 2.3 shows that the proposed estimator has the best bias profile. With carefully chosen bandwidth we can achieve the best root MSE under the presence of noise. When no noise is present, the Hayashi and Yoshida estimator performs well. The refresh time aligned method often performs better in estimating the integrated variance of the less traded asset; (2,2) element. This is since it effectively aligns on the time stamp of less traded asset. As shown in the analysis of asymptotic bias, when no noise is present and number of refresh time sample is size smaller then realized kernel under performs in terms of bias. The proposed estimator overall estimates the off-diagonal elements better. We observe also that the realized covariance estimator aligned on sparesly sampled data often performs well - this is because there is two opposing effect in terms of bias, negative bias from epps effect and positive bias from microstructure noise. The advantage of our estimator is most clear in estimating higher dimension covariance matrix as shown in Table 2.4. We estimate 10 dimensional integrated covariance matrix and compare maximum of eigenvalues and variance of the equally weighted portfolio. Under no presence of noise the refresh time based method has large bias. We calculate the optimal bandwidth as given in Theorem 1 and 2 for each element of covariance matrix and take the minimum, maximum of these and average of the two. We note that our estimator seems to have large variance, however it performs best under the optimal bandwidth.

2.7.2 Empirical Application

In this section we apply the Fourier Realized Kernel to a high frequency data. We analyzed five stocks of different liquidity - Microsoft (MSFT), Dell (Dell), J P Morgan (JPM) and less frequently traded Caterpillar Inc (CAT) and Banco de Chile (BCH) from WRDS TAQ database. The period of analysis is for 20 days during 05-30 March 2007. The liquidity of stocks is in order of least liquid BCH, CAT, DELL, JPM, MSFT with average daily sample sizes, $\{48, 7526, 8337, 10337, 11451\}$. We may calculate the optimal bandwidth for individual asset by equalizing the squared bias and the variance term given in Theorem 2. Let $n_\ell = n^{\beta_\ell}$ where n is a minimum of all sample sizes, then it is given by $H_\ell = \{\frac{\eta|k''(0)|}{\|k\|}\}^{2/5}\zeta_\ell^{4/5}n^{\frac{1+2\beta_\ell}{5}}$. The ζ_ℓ^2 is a squared noise to signal ratio for each asset given by $\Gamma_{\ell,\ell}/\sqrt{IQ_{\ell,\ell}}$. We may estimate the variance of the microstructure noise by the Realized Variance applied on the tick data divided by $2n$, i.e. $E(U^2) \simeq RV/2n$. See Zhang et al. (2005) The square root of the integrated quarticity is estimated by the Realized Variance applied on the sparsely sampled data e.g. 10 minutes. We applied maximum, minimum and average of the above individual bandwidths. The Figure 2.11 and Figure 2.12 compare the Realized Covariance and the proposed method in estimating the daily covariation matrix. Since the first asset is least traded, the all refresh time is effectively aligned on the trading time of the first asset. In estimating the integrated variance, the proposed method lies between the RV using pairwise refresh time (which will be dominated by the microstructure noise) and the RV using all refresh time (which is more sparsely sampled, therefore less affected by the noise). Most interesting case is the performance in estimating covariation for assets of different liquidity - i.e. $(1,4)$ and $(1,5)$-th element of the estimator in our case. The daily Realized Covariance take values closer to zero due to Epps effect whereas proposed estimator clearly gives us non trivial estimates.

2.8 Appendix

2.8.1 Remark on Assumption 3

The condition (ii) needs a verification and the rest is straightforward to derive given (2.14). The Parseval's identity between the Fourier transform pair given by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|K(\lambda)|^2 d\lambda = \int_{-\infty}^{\infty}|k(x)|^2 dx,$$

which can be easily derived by

$$\begin{aligned}\int_{-\infty}^{\infty}|k(x)|^2 dx &= \frac{1}{2\pi}\int_{-\infty}^{\infty}k(x)^*\int_{-\pi}^{\pi}K(\lambda)e^{ix\lambda}d\lambda dx = \frac{1}{2\pi}\int_{-\pi}^{\pi}K(\lambda)\int_{-\infty}^{\infty}k(x)^* e^{ix\lambda}dxd\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}K(\lambda)K(\lambda)^* d\lambda = \frac{1}{2\pi}\int_{-\pi}^{\pi}|K(\lambda)|^2 d\lambda.\end{aligned}$$

This gives us condition $\|k\|^2 = \frac{1}{2\pi}\|K\|^2 < \infty$ and

$$\begin{aligned}\int_{-\infty}^{\infty}|k'(x)|^2 dx &= \frac{1}{2\pi}\int_{-\infty}^{\infty}k'(x)^*\int_{-\pi}^{\pi}i\lambda K(\lambda)e^{ix\lambda}d\lambda dx = \frac{1}{2\pi}\int_{-\pi}^{\pi}i\lambda K(\lambda)\int_{-\infty}^{\infty}k'(x)^* e^{ix\lambda}dxd\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}i\lambda K(\lambda)\left(k(x)e^{-ix\lambda}|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}-i\lambda k(x)e^{-ix\lambda}dx\right)^* d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}|i\lambda|^2 K(\lambda)\left(\int_{-\infty}^{\infty}k(x)e^{-ix\lambda}dx\right)^* = \frac{1}{2\pi}\int_{-\pi}^{\pi}|\lambda K(\lambda)|^2 d\lambda.\end{aligned}$$

This gives us the condition $\|k'\|^2 = \frac{1}{2\pi}\|\lambda K(\lambda)\|^2 < \infty$. With similar argument rest of condition can be verified.

2.8.2 Lemmas

We will prove the theorems for the general version of our estimator given in (2.9). We derive the results conditionally on the volatility matrix and the discretization time points therefore we regard these variables deterministic in the proofs. Throughout the proof we denote $C, C_1, C_2, \cdots$ finite constants.

LEMMA 1. *Let $P(t)$ defined on the filtered probability space* $(\Sigma, \mathcal{F}, \mathcal{F}_{t\geq 0}, \mathbb{P})$ *satisfies Assumption 1 and* $f(t,s;q)$ be *a bounded and measurable function. Define square bracket operation to denote a quadratic covariation process defined in* (2.1). *Then,*

$$E\left[\int_0^{2\pi}\int_0^{2\pi} f(t,s;q)dP_a(s)dP_b(t), \int_0^{2\pi}\int_0^{2\pi} f(t,s;q')dP_c(s)dP_d(t)\right] \tag{2.31}$$
$$= \int_0^{2\pi}\int_0^{2\pi} f(t,s;q)f(t,s;q')d[P_a,P_c](s)d[P_b,P_d](t)$$
$$+ \int_0^{2\pi}\int_0^{2\pi} f(t,s;q)f(s,t;q')d[P_a,P_d](s)d[P_b,P_c](t).$$

where double stochastic integral is Wiener-Itô sense.

PROOF. The double Wiener-Itô integral can be written as

$$\int_0^{2\pi}\int_0^{2\pi} f(t,s;q)dP_a(s)dP_b(t)$$
$$= \int_0^{2\pi}\int_0^{t} f(t,s;q)dP_a(s)dP_b(t) + \int_0^{2\pi}\int_0^{t} f(s,t;q)dP_b(s)dP_a(t),$$

so that the integrand is measurable with respect to $\mathcal{F}_t$ and the stochastic integration is well defined. Two terms above are martingale. Therefore (2.31) can be expressed as

$$\left[\begin{array}{c} \int_0^{2\pi}\int_0^t f(t,s;q)dP_a(s)dP_b(t) + \int_0^{2\pi}\int_0^t f(s,t;q)dP_b(s)dP_a(t), \\ \int_0^{2\pi}\int_0^t f(t,s;q')dP_c(s)dP_d(t) + \int_0^{2\pi}\int_0^t f(s,t;q')dP_d(s)dP_c(t) \end{array}\right].$$

Lets consider one of the cross product terms among four possible terms from above. By Itô's isometry,

$$
\begin{aligned}
&E\Big[\int_0^{2\pi}\int_0^t f(t,s;q)dP_a(s)dP_b(t), \int_0^{2\pi}\int_0^t f(s,t;q')dP_d(s)dP_c(t)\Big] \qquad (2.32)\\
&= E\int_0^{2\pi}\left(\int_0^t f(t,s;q)dP_a(s)\right)\left(\int_0^t f(s,t;q')dP_d(s)\right)d[P_b,P_c](t),
\end{aligned}
$$

where $d[P_b, P_c](t)$ means $[P_b, P_c]'(t)dt$, where the prime denotes the time derivative. By Fubini's theorem,

$$
\begin{aligned}
&\int_0^{2\pi} E\left(\int_0^t f(t,s;q)dP_a(s)\right)\left(\int_0^t f(s,t;q')dP_d(s)\right)d[P_b,P_c](t).\\
&= \int_0^{2\pi}\int_0^t f(t,s;q)f(s,t;q')d[P_a,P_d](s)d[P_b,P_c](t).
\end{aligned}
$$

Together with the expected quadratic covariation of following terms,

$$
\begin{aligned}
&E\Big[\int_0^{2\pi}\int_0^t f(s,t;q)dP_b(s)dP_a(t), \int_0^{2\pi}\int_0^t f(t,s;q')dP_c(s)dP_d(t)\Big]\\
&= \int_0^{2\pi}\int_0^t f(s,t;q)f(t,s;q')d[P_b,P_c](s)d[P_a,P_d](t),
\end{aligned}
$$

we have

$$
\int_0^{2\pi}\int_0^{2\pi} f(t,s;q)f(s,t;q')d[P_a,P_d](s)d[P_b,P_c](t).
$$

For example, when $a = b = c = d$ and simplifying the integrand $H_t = \int_0^t f(t,s;q)dP_s \in \mathcal{F}_t$, (2.32) is given by $E\int_0^{2\pi} H_t^2 d[P,P]_t$. By interchange the expectation and the integration,

$$
\begin{aligned}
&\int_0^{2\pi} E\left\{\int_0^t H_s dH_s + d[H,H]_s\right\} d[P,P]_t\\
=\quad &\int_0^{2\pi}\int_0^t f_n^2(t,s;q)d[P,P]_s d[P,P]_t.
\end{aligned}
$$

■

LEMMA 2. *Define a off-diagonal step function*

$$\begin{aligned} f_n(t,s;q) &= \sum_{i,j} e^{-is_jq} e^{-i(t_i-s_j)\alpha} 1_{[t_{i-1},t_i[}(t) 1_{[s_{j-1},s_j[}(s) 1_{\{I_{i,1}\cap I_{j,2}=\emptyset\}}(t,s) \\ g_n(t,s;q) &= \sum_{i,j} e^{-is_jq} e^{-i(t_i-s_j)\alpha} 1_{[t_{i-1},t_i[}(t) 1_{[s_{j-1},s_j[}(s). \end{aligned}$$

where discretization points $\{t_i, s_j\}$ satisfy Assumption 2.1. Then

$$\int_0^{2\pi}\int_0^{2\pi} f_n(t,s;q)dsdt = \int_0^{2\pi}\int_0^{2\pi} g_n(t,s;q)dsdt + O(\frac{1}{n}).$$

PROOF. When we use a single discretizing point $\{t_i\}$,

$$\begin{aligned} &\int_0^{2\pi}\int_0^{2\pi} \{\sum_{i,j} - \sum_{i\neq j}\} e^{-it_jq} e^{-i(t_i-t_j)\alpha} 1_{[t_{i-1},t_i[}(t) 1_{[t_{j-1},t_j[}(s) dsdt \\ &= \int_0^{2\pi}\int_0^{2\pi} \sum_i e^{-it_iq} 1_{[t_{i-1},t_i[}(t) 1_{[t_{i-1},t_i[}(s) dsdt = \sum_i e^{-it_iq} \Delta t_i^2 \\ &\leq C \sup_i \Delta t_i = O(\frac{1}{n}), \end{aligned}$$

under Assumption 2.1. Likewise, using two discretizing point $\{t_i, s_j\}$,

$$\begin{aligned} &\int_0^{2\pi}\int_0^{2\pi} \sum_{i,j} \{1 - 1_{\{I_{i,1}\cap I_{j,2}=\emptyset\}}(t,s)\} e^{-it_jq} e^{-i(t_i-s_j)\alpha} 1_{[t_{i-1},t_i[}(t) 1_{[s_{j-1},s_j[}(s) dsdt \\ &= \int_0^{2\pi}\int_0^{2\pi} \sum_{i,j} e^{-is_jq} e^{-i(t_i-s_j)\alpha} 1_{[t_{i-1},t_i[}(t) 1_{[s_{j-1},s_j[}(s) 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}}(t,s) dsdt \\ &\leq \sum_{i,j} \Delta t_i \Delta s_j 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}} \leq \sup_i(\Delta t_i) \sup_j(\Delta s_j) \sum_{i,j} 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}} \\ &= \frac{C}{n_1 n_2} \sharp\{t_i \cup s_j, 0 \leq t_i, s_j \leq 2\pi\} = O(\frac{1}{n_1 \wedge n_2}), \end{aligned}$$

where the penultimate equality is using Assumption 2.1 $\sharp\{t_i \cup s_j, 0 \leq t_i, s_j \leq 2\pi\}$ means the number of union of points for two discretization grid $\{t_i, s_j\}$ on the interval $[0, 2\pi]$. Its order is bounded by $O(n_1 \vee n_2)$. ■

LEMMA 3. *Define a off-diagonal step function weighted by kernel by*

$$f_n(t,s;q) = \sum_{i\neq j} e^{-it_jq} k_H(t_i - t_j) 1_{[t_{i-1},t_i[}(t) 1_{[t_{j-1},t_j[}(s) \tag{2.33}$$
$$g_n(t,s;q) = \sum_{i,j} e^{-it_jq} k_H(t_i - t_j) 1_{[t_{i-1},t_i[}(t) 1_{[t_{j-1},t_j[}(s) \ ,$$

where discretization points $\{t_i\}$ *satisfy Assumption 2. Let* $k_H(t_i - t_j) = k\left(\frac{(t_i-t_j)/\overline{\Delta t}}{H}\right)$ *where* $k(\cdot)$ *is a lag window Assumption 3. Then*

$$\frac{n}{H}\int_0^{2\pi}\int_0^{2\pi}\left\{f_n^2(t,s;q) + f_n(t,s;q)f_n(s,t;q)\right\} d[P,P](t)d[P,P](s) \tag{2.34}$$
$$\to 2\|k^2\| \int_0^{2\pi} e^{-i2tq}([P,P]'(t))^2 d\mathcal{Q}_{11}(t),$$

where $\mathcal{Q}_{11}(t)$ *is defined in Assumption 2.*

PROOF. First note that for any function $d(\cdot,\cdot)$, it holds that $\sum_{i,j=1}^n d(i,j) = \sum_{h=0}^{n-1}\sum_{j=1}^{n-h} d(j,j+h) + \sum_{h=1}^{n-1}\sum_{j=1+h}^{n} d(j,j-h)$, which we will denote by $\sum_{i,j=1}^n d(i,j) = \sum_{h=0}^{n-1}\sum_{j=1}^{n-h} d(j,j+h)[2]$. By Lemma 1, we can replace f_n by g_n with error $O(n^{-1})$. Then (2.34) is approximated by

$$\frac{n}{H}\int_0^{2\pi}\int_0^{2\pi}\left\{g_n^2(t,s;q) + g_n(t,s;q)g_n(s,t;q)\right\} d[P,P](t)d[P,P](s)$$
$$= \frac{n}{H}\sum_{h=0}^{n-1}\sum_{j=1}^{n-h}(e^{-it_j2q} + e^{-it_jq}e^{-it_{j+h}q})k_H^2(t_{j+h} - t_j)$$
$$\times [P,P]'(t_{j+h})[P,P]'(t_j)\Delta t_{j+h}\Delta t_j[2] + O(\frac{1}{nH}) \tag{2.35}$$
$$\simeq \frac{n}{H}\sum_{h=0}^{n-1} k^2\left(\frac{t_h/\overline{\Delta t}}{H}\right)\sum_{j=1}^{n-h}(e^{-it_j2q} + e^{-it_jq}e^{-it_{j+h}q})$$
$$\times [P,P]'(t_{j+h})[P,P]'(t_j)\Delta t_{j+h}\Delta t_j[2] \tag{2.36}$$
$$\to 2\|k\|^2 \int_0^{2\pi} e^{-it2q}([P,P]')^2(t)d\mathcal{Q}_{11}(t).$$

In (2.35) the error is from approximating the integral by the discrete sum. (2.36) holds since $t_i - t_{i-h} \simeq t_h$ under Assumption 2. The last convergence holds from the fact that $\frac{n}{H}\sum_{h=0}^{n-1} k^2\left(\frac{h}{H}\right) \sum_{j=1}^{n-h} \Delta t_{j+h}\Delta t_j[2]$ and $\frac{n}{H}\sum_{|h|<n} k^2\left(\frac{h}{H}\right) \sum_{j=1}^{n} \Delta t_j \Delta t_j$ approaches the same limit due to presence of the kernel weights and the convergence of the Riemann approximation of an integral $\int_{-\infty}^{\infty} k^2(x)dx$. ■

LEMMA 4. *Define a off-diagonal step function weighted by kernel by*

$$
\begin{aligned}
f_n(t,s;q) &= \sum_{i,j} e^{-is_jq} k_H(t_i - s_j) 1_{[t_{i-1},t_i[}(t) 1_{[s_{j-1},s_j[}(s) 1_{\{I_{i,1}\cap I_{j,2}=\emptyset\}}(t,s) \quad (2.37)\\
g_n(t,s;q) &= \sum_{i,j} e^{-is_jq} k_H(t_i - s_j) 1_{[t_{i-1},t_i[}(t) 1_{[s_{j-1},s_j[}(s),
\end{aligned}
$$

where discretization points $\{t_i, s_j\}$ *satisfy Assumption 2. Let* $k_H(t_i - s_j) = k\left(\frac{(t_i-s_j)/\widetilde{\Delta t}}{H}\right)$ *where* $k(\cdot)$ *is a lag window Assumption 3. Then*

$$
\begin{aligned}
&\frac{n_1 \wedge n_2}{H} \int_0^{2\pi}\int_0^{2\pi} f_n^2(t,s;q) d[P_1,P_1](t) d[P_2,P_2](s)\\
&\qquad \to \|k^2\| \int_0^{2\pi} e^{-i2tq} [P_1,P_1]'(t)[P_2,P_2]'(t) d\mathcal{Q}_{1122}(t) \quad (2.38)\\
&\frac{n_1 \wedge n_2}{H} \int_0^{2\pi}\int_0^{2\pi} f_n(t,s;q) f_n(s,t;q) d[P_1,P_2](t) d[P_2,P_1](s)\\
&\qquad \to \|k\|^2 \int_0^{2\pi} e^{-i2tq} ([P_1,P_2]')^2(t) d\mathcal{Q}_{1212}(t). \quad (2.39)
\end{aligned}
$$

In general define a step function by

$$
f_n(t,s;q,a,b) = \sum_{i,j} e^{-it_{j,b}q} k_H(t_{i,a} - t_{j,b}) 1_{[t_{i-1,a},t_{i,a}[}(t) 1_{[t_{j-1,b},t_{j,b}[}(s) 1_{\{I_{i,a}\cap I_{j,b}=\emptyset\}}(t,s), \quad (2.40)
$$

then it holds that

$$
\begin{aligned}
&\frac{n_a \wedge n_b \wedge n_c \wedge n_d}{H} \int_0^{2\pi}\int_0^{2\pi} f_n(t,s;q,a,b) f_n(t,s;q,c,d) d[P_a,P_c](s) d[P_b,P_d](t) \quad (2.41)\\
&\to \|k\|^2 \int_0^{2\pi} e^{-i2tq} [P_a,P_c]'(t)[P_b,P_d]'(t) d\mathcal{Q}_{acbd}(t),
\end{aligned}
$$

where $\mathcal{Q}.(t)$ is defined in Assumption 2.

PROOF. We first note that

$$f_n^2(t,s;q) = \sum_{i,j} e^{-i2s_jq} k_H^2(t_i - s_j) 1_{[t_{i-1},t_i[}(t) 1_{[s_{j-1},s_j[}(s) 1_{\{I_{i,1}\cap I_{j,2}=\emptyset\}}(t,s).$$

Then using Lemma 2, it holds that

$$\begin{aligned}
&\int_0^{2\pi}\int_0^{2\pi} f_n^2(t,s;q)[P_1,P_1]'(t)[P_2,P_2]'(s)dtds \\
=&\sum_{i,j} e^{-i2s_jq} k_H^2(t_i-s_j)\int_{t_{i-1}}^{t_i}[P_1,P_1]'(t)dt\int_{s_{j-1}}^{s_j}[P_2,P_2]'(s)ds + O(\frac{1}{n_1\wedge n_2}).
\end{aligned}$$

Note the following inequality

$$\begin{aligned}
\sum_{i,j}\Delta t_i\Delta s_j &\geq \sum_{i,j}\Delta t_i \Delta s_j k_H^2(t_i-s_j) \\
&\geq \sum_{i,j}\Delta t_i\Delta s_j 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}} \geq \sum_{i,j}(t_i\wedge s_j - t_i\vee s_j)^2,
\end{aligned} \tag{2.42}$$

since the first three quantities are of order $O(1)$, $O(\frac{H}{n_1\wedge n_2})$ and $O(\frac{1}{n_1\wedge n_2})$ respectively by Assumption 2.1. Recalling that $\{T_l\}_{1\leq l\leq N_T}$ are union of time stamps,

$$\sum_{i,j}(t_i\wedge s_j - t_i\vee s_j)^2 \leq \sup_l |T_l - T_{l-1}| \sum_{1\leq l\leq N_T}|T_l - T_{l-1}| = O(\frac{1}{n_1\vee n_2}).$$

Then it holds that

$$\begin{aligned}
&\frac{n_1\wedge n_2}{H}\sum_{i,j} e^{-i2s_jq}k_H^2(t_i-s_j)\int_{t_{i-1}}^{t_i} d[P_1,P_1](t)\int_{s_{j-1}}^{s_j} d[P_2,P_2](s) \\
\simeq&\frac{n_1\wedge n_2}{H}\sum_{i,j} e^{-i2s_jq}k_H^2(t_i-s_j)[P_1,P_1]'(t_i)[P_2,P_2]'(s_j)\Delta t_i\Delta s_j,
\end{aligned}$$

where error is in approximating a continuous integral by discrete sum. Given (2.42), the above has the same order of magnitude as

$$
\begin{aligned}
&\|k\|^2(n_1 \wedge n_2)\sum_{i,j} e^{-i2s_jq}[P_1,P_1]'(t_i)[P_2,P_2]'(s_j)\Delta t_i \Delta s_j 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}}\\
\to&\|k\|^2\int_0^{2\pi} e^{-i2tq}[P_1,P_1]'(t)[P_2,P_2]'(t)d\mathcal{Q}_{1122}(t),
\end{aligned}
$$

by the Riemann approximation of a continuous integral under Assumption 2.2. We now turn to expression (2.39). By Lemma 2, we may replace the function $f_n(t,s;q)$ by $g_n(t,s;q)$. Then the cross product term simplifies to

$$
\begin{aligned}
g_n(t,s;q)g_n(s,t;q) = &\sum_{i,j,k,l} e^{-is_jq}e^{-is_lq}k_H(t_i-s_j)k_H(t_k-s_l)\\
&\qquad 1_{[t_{i-1}\vee s_{l-1},t_i\wedge s_l[}(t)1_{[t_{k-1}\vee s_{j-1},t_k\wedge s_j[}(s).
\end{aligned}
$$

Then (2.39) is given by

$$
\begin{aligned}
&\frac{n_1\wedge n_2}{H}\int_0^{2\pi}\int_0^{2\pi} f_n(t,s;q)f_n(s,t;q)d[P_1,P_2](t)d[P_2,P_1](s)\\
\simeq&\frac{n_1\wedge n_2}{H}\sum_{i,j,k,l} e^{-is_jq}e^{-is_lq}k_H(t_i-s_j)k_H(t_k-s_l)[P_1,P_2]'(t_i\wedge s_l)[P_2,P_1]'(t_k\wedge s_j)\\
&\times(t_i\wedge s_l - t_{i-1}\vee s_{l-1})(t_k\wedge s_j - t_{k-1}\vee s_{j-1})1_{\{I_{i,1}\cap I_{l,2}\neq\emptyset\}}1_{\{I_{k,1}\cap I_{j,2}\neq\emptyset\}}\\
\simeq&\|k\|^2(n_1\wedge n_2)\sum_{i,j,l} e^{-is_jq}e^{-is_lq}[P_1,P_2]'(t_i\wedge s_l)[P_2,P_1]'(t_i\wedge s_j)\\
&\times(t_i\wedge s_l - t_{i-1}\vee s_{l-1})(t_i\wedge s_j - t_{i-1}\vee s_{j-1})1_{\{I_{i,1}\cap I_{l,2}\neq\emptyset\}}1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}}\\
\to&\|k\|^2\int_0^{2\pi} e^{-i2tq}([P_1,P_2]')^2(t)d\mathcal{Q}_{1212}(t).
\end{aligned}
$$

The cross product term involving the diagonal step function in (2.41) is given by

$$
\begin{aligned}
&g_n(t,s;q,a,b)g_n(t,s;q,c,d)\\
&=\sum_{i,j,k,l} e^{-it_{j,b}q}e^{-it_{l,d}q}k_H(t_{i,a}-t_{j,b})k_H(t_{k,c}-t_{l,d})1_{[t_{i-1,a},t_{i,a}[}(t)1_{[t_{j-1,b},t_{j,b}[}(s)1_{[t_{k-1,c},t_{k,c}[}(t)1_{[t_{l-1,d},t_{l,d}[}(s)\\
&=\sum_{i,j,k,l} e^{-it_{j,b}q}e^{-it_{l,d}q}k_H(t_{i,a}-t_{j,b})k_H(t_{k,c}-t_{l,d})1_{[t_{i-1,a}\vee t_{k-1,c},t_{i,a}\wedge t_{k,c}[}(t)1_{[t_{j-1,b}\vee t_{l-1,d},t_{j,b}\wedge t_{l,d}[}(s).
\end{aligned}
$$

We note following inequality. (2) $\geq$ (1) $\geq$ (3) $\geq$ (4)

$$
(1)\sum_{i,j,k,l} k_H(t_{i,a}-t_{j,b})k_H(t_{k,c}-t_{l,d})(t_{i,a}\wedge t_{k,c}-t_{i-1,a}\vee t_{k-1,c})(t_{j,b}\wedge t_{l,d}-t_{j-1,b}\vee t_{l-1,d})\leq(2),
$$

where upper bound (2) is given by

$$
(2)\sum_{i,j,k,l}(t_{i,a}\wedge t_{k,c}-t_{i-1,a}\vee t_{k-1,c})(t_{j,b}\wedge t_{l,d}-t_{j-1,b}\vee t_{l-1,d})1_{\{I_{i,a}\cap I_{k,c}\neq\emptyset\}}1_{\{I_{j,b}\cap I_{l,d}\neq\emptyset\}}=O(1).
$$

Under Assumption 2.1, the lower bound is given by

$$
\begin{aligned}
&(3)\sum_{i,j,k,l}(t_{i,a}\wedge t_{k,c}-t_{i-1,a}\vee t_{k-1,c})(t_{j,b}\wedge t_{l,d}-t_{j-1,b}\vee t_{l-1,d})1_{\{I_{i,a}\cap I_{j,b}\cap I_{k,c}\cap I_{l,d}\neq\emptyset\}}\\
&\leq C\frac{1}{n_a\vee n_c}\frac{1}{n_b\vee n_d}\sum_{i,j,k,l}\sharp 1_{\{I_{i,a}\cap I_{j,b}\cap I_{k,c}\cap I_{l,d}\neq\emptyset\}}\\
&=C\frac{1}{n_a\vee n_c}\frac{1}{n_b\vee n_d}\sharp\{t_{i,a}\cup t_{j,b}\cup t_{k,c}\cup t_{l,d},0\leq t_{i,a},t_{j,b},t_{k,c},t_{l,d}\leq 2\pi\}\\
&=O(\frac{n_a\vee n_b\vee n_c\vee n_d}{(n_a\vee n_c)(n_b\vee n_d)}),
\end{aligned}
$$

which will be order of inverse of second or third largest sample size. (3) is bigger or equal to (4),

$$
\begin{aligned}
&(4)\sum_{i,j,k,l}(t_{i,a}\wedge t_{j,b}\wedge t_{k,c}\wedge t_{l,d}-t_{i-1,a}\vee t_{j-1,b}\vee t_{k-1,c}\vee t_{l-1,d})^2 1_{\{I_{i,a}\cap I_{j,b}\cap I_{k,c}\cap I_{l,d}\neq\emptyset\}}\\
&=\sum_{\ell=1}^{N_T}(T_\ell-T_{\ell-1})^2\leq 2\pi\sup_\ell|T_\ell-T_{\ell-1}|=O(\frac{1}{n_a\vee n_b\vee n_c\vee n_d}),
\end{aligned}
$$

where union of four time stamps is constructed by

$$\{T_\ell\}_{1\leq \ell\leq N_T} := \{t_{i,a}\cup t_{j,b}\cup t_{k,c}\cup t_{l,d}, 0\leq t_{i,a}, t_{j,b}, t_{k,c}, t_{l,d}\leq 2\pi\}.$$

N_T is a total number of member points in the union of time stamps. The equality (3) = (4) holds when time stamps are synchronous. Let $I_{min\{(i,a),(j,b)\}}$ denote $I_{i,a}$ if $n_a < n_b$ and $I_{j,b}$ otherwise. $I_{min\{(k,c),(l,d)\}}$ is equivalently defined. For simplicity assume that $n_a < n_b < n_c < n_d$, then

$$\begin{aligned}
&\sum_{i,j,k,l: t_{i,a},t_{j,b},t_{k,c},t_{l,d}<t} (t_{i,a}\wedge t_{j,b} - t_{i-1,a}\vee t_{j-1,b})(t_{k,c}\wedge t_{l,d} - t_{k-1,c}\vee t_{l-1,d}) \\
&\times 1_{\{I_{i,a}\cap I_{j,b}\neq\emptyset\}} 1_{\{I_{k,c}\cap I_{l,d}\neq\emptyset\}} 1_{\{I_{min\{(i,a),(j,b)\}}\cap I_{min\{(k,c),(l,d)\}}\neq\emptyset\}} \\
&\leq C\frac{1}{n_b}\frac{1}{n_d}\sum_{i=1}^{n_a}\sum_{k=1}^{n_c}\sharp\{t_{k,c}\in[t_{i-1,a},t_{i,a}[\}\sum_{j=1}^{n_b}\sharp\{t_{j,b}\in[t_{i-1,a},t_{i,a}[\}\sum_{l=1}^{n_d}\sharp\{t_{l,d}\in[t_{k-1,c},t_{k,c}[\},
\end{aligned}$$

which is order of $C\frac{1}{n_b}\frac{1}{n_d}n_a\frac{n_c}{n_a}\frac{n_b}{n_a}\frac{n_d}{n_c} = O(\frac{1}{n_a})$ under Assumption 2.3. Then (2.41) is given by

$$\begin{aligned}
&\frac{n_a\wedge n_b\wedge n_c\wedge n_d}{H}\int_0^{2\pi}\int_0^{2\pi} f_n(t,s;q,a,b)f_n(t,s;q,c,d)d[P_a,P_c](s)d[P_b,P_d](t) \\
\simeq&\frac{n_a\wedge n_b\wedge n_c\wedge n_d}{H}\sum_{i,j,k,l} e^{-i2t_{j,b}q}e^{-i2t_{l,d}q}k_H(t_{i,a}-t_{j,b})k_H(t_{k,c}-t_{l,d}) \\
&\times [P_a,P_c]'(t_{i,a}\wedge t_{k,c})[P_b,P_d]'(t_{j,b}\wedge t_{l,d}) \\
&\times (t_{i,a}\wedge t_{k,c} - t_{i-1,a}\vee t_{k-1,c})(t_{j,b}\wedge t_{l,d} - t_{j-1,b}\vee t_{l-1,d}) \\
&\times 1_{\{I_{i,a}\cap I_{k,c}\neq\emptyset\}}1_{\{I_{j,b}\cap I_{l,d}\neq\emptyset\}}1_{\{I_{min\{(i,a),(k,c)\}}\cap I_{min\{(j,b),(l,d)\}}\neq\emptyset\}} \\
\simeq&\|k\|^2(n_a\wedge n_b\wedge n_c\wedge n_d)\sum_{i,j,k,l} e^{-i2t_{j,b}q}e^{-i2t_{l,d}q}[P_a,P_c]'(t_{i,a}\wedge t_{k,c})[P_b,P_d]'(t_{j,b}\wedge t_{l,d}) \\
&\times (t_{i,a}\wedge t_{k,c} - t_{i-1,a}\vee t_{k-1,c})(t_{j,b}\wedge t_{l,d} - t_{j-1,b}\vee t_{l-1,d}) \\
&\times 1_{\{I_{i,a}\cap I_{k,c}\neq\emptyset\}}1_{\{I_{j,b}\cap I_{l,d}\neq\emptyset\}}1_{\{I_{min\{(i,a),(k,c)\}}\cap I_{min\{(j,b),(l,d)\}}\neq\emptyset\}} \\
\to&\|k\|^2\int_0^{2\pi} e^{-i2tq}[P_a,P_c]'(t)[P_b,P_d]'(t)d\mathcal{Q}_{acbd}(t),
\end{aligned}$$

where $\mathcal{Q}_{acbd}(t)$ is limit of

$$\mathcal{Q}^{(n)}_{abcd}(t) = (n_a \wedge n_b \wedge n_c \wedge n_d)\times$$
$$\sum_{i,j,k,l:t_{i,a},t_{j,b},t_{k,c},t_{l,d}<t} (t_{i,a} \wedge t_{j,b} - t_{i-1,a} \vee t_{j-1,b})(t_{k,c} \wedge t_{l,d} - t_{k-1,c} \vee t_{l-1,d}) \tag{2.43}$$
$$\times\, 1_{\{I_{i,a}\cap I_{j,b}\neq\emptyset\}} 1_{\{I_{k,c}\cap I_{l,d}\neq\emptyset\}} 1_{\{I_{min\{(i,a),(j,b)\}}\cap I_{min\{(k,c),(l,d)\}}\neq\emptyset\}}.$$

We make a concrete example here to show an asymptotic variance. Consider when time stamps are synchronous but sample sizes are unbalanced. For simplicity we let volatility to be constant. The time stamps are nested and equally spaced over the fixed interval $[0, T]$ with $T = 2\pi$ and let $n_d > n_c > n_b > n_a$. This is example of time stamps satisfying Assumption 2. The discretization points can be expressed as

$$t_{i,a} = \frac{T}{n_a} i, i = 0, \ldots, n_a; \quad t_{j,b} = \frac{T}{n_b} j, j = 0, \ldots, n_b,$$

and so on. We can express the time stamp of more liquid asset$\{t_{j,b}\}$ in terms of less liquid asset by

$$t_{[\frac{n_b}{n_a}](j-1)+u} = \frac{T(j-1)}{n_a} + \frac{Tu}{n_b}; j = 1, \ldots, n_a, u = 0, \ldots, \frac{n_b}{n_a} - 1.$$

Likewise, we may express the time stamp of most liquid asset by

$$t_{\ell,d} = T(\frac{(\ell-1)}{n_a} + \frac{\beta}{n_b} + \frac{\lambda}{n_c} + \frac{\gamma}{n_d}); \ell = 1, \ldots, n_a, \beta = 0, \ldots, \frac{n_b}{n_a} - 1,$$
$$\lambda = 0, \ldots, \frac{n_c}{n_b} - 1, \gamma = 0, \ldots, \frac{n_d}{n_c} - 1.$$

The quantities associated with the quadratic covariation of time is given by

$$\begin{aligned}\mathcal{Q}^{(n)}_{abcd}(T) =& n_a \sum_{i,j,k,l=1}^{n_a} \sum_{u,\beta,\tau=0}^{\frac{n_b}{n_a}-1} \sum_{\alpha,\lambda=0}^{\frac{n_c}{n_b}-1} \sum_{\gamma=0}^{\frac{n_d}{n_c}-1} T^4 \frac{1}{n_b n_d} 1_{\{i=j=k=l\}} 1_{\{\beta=\tau\}} 1_{\{\alpha=\lambda\}} \\ =& T^4 \frac{1}{n_b n_d} n_a^2 \left(\frac{n_b}{n_a}\right)^2 \frac{n_c}{n_b} \frac{n_d}{n_c} = O(1).\end{aligned}$$

Also note that by Taylor expansion,

$$k^2\left(\frac{\frac{n_2}{n_1}h-u}{H}\right)=k^2\left(\frac{\frac{n_2}{n_1}h}{H}\right)+\frac{u}{H}2k\left(\frac{\frac{n_2}{n_1}h}{H}\right)k'\left(\frac{\frac{n_2}{n_1}h}{H}\right)+O(H^{-2}).$$

Then (2.41) is given by

$$\begin{aligned}
&\sum_{i,j,k,l}k_H(t_{i,a}-t_{j,b})k_H(t_{k,c}-t_{l,d})(t_{i,a}\wedge t_{k,c}-t_{i-1,a}\vee t_{k-1,c})(t_{j,b}\wedge t_{l,d}-t_{j-1,b}\vee t_{l-1,d})\\
&=\sum_{i,j,k,l=1}^{n_a}\sum_{u,\beta=0}^{\frac{n_b}{n_a}-1}\sum_{\alpha=0}^{\frac{n_c}{n_a}-1}\sum_{\gamma=0}^{\frac{n_d}{n_b}-1}k\left(\frac{\frac{n_b}{n_a}(i-j+1)-u}{H}\right)k\left(\frac{\frac{n_d}{n_a}(k-l)-\frac{n_d}{n_b}\beta+\frac{n_d}{n_c}\alpha-\gamma}{H}\right)\\
&\quad\times\frac{T}{n_a\vee n_c}1_{\{i=k\}}\frac{T}{n_b\vee n_d}1_{\{j=l,u=\beta\}}\\
&=\frac{T^2}{(n_a\vee n_c)(n_b\vee n_d)}\sum_{i,j=1}^{n_a}\sum_{u=0}^{\frac{n_b}{n_a}-1}\sum_{\alpha=0}^{\frac{n_c}{n_a}-1}\sum_{\gamma=0}^{\frac{n_d}{n_b}-1}k\left(\frac{\frac{n_b}{n_a}(i-j+1)-u}{H}\right)k\left(\frac{\frac{n_d}{n_a}(i-j)-\frac{n_d}{n_b}u+\frac{n_d}{n_c}\alpha-\gamma}{H}\right)\\
&\simeq\frac{T^2}{(n_a\vee n_c)(n_b\vee n_d)}\sum_{|h|<n_a}(n_a-|h|)\sum_{u=0}^{\frac{n_b}{n_a}-1}\sum_{\alpha=0}^{\frac{n_c}{n_a}-1}\sum_{\gamma=0}^{\frac{n_d}{n_b}-1}k\left(\frac{\frac{n_b}{n_a}|h|-u}{H}\right)k\left(\frac{\frac{n_d}{n_a}|h|-\frac{n_d}{n_b}u+\frac{n_d}{n_c}\alpha-\gamma}{H}\right)\\
&\simeq T^2\frac{n_a\wedge n_b\wedge n_c\wedge n_d}{(n_a\vee n_c)(n_b\vee n_d)}\sum_{|h|<n_a}\sum_{u=0}^{\frac{n_b}{n_a}-1}\sum_{\alpha=0}^{\frac{n_c}{n_a}-1}\sum_{\gamma=0}^{\frac{n_d}{n_b}-1}\left\{k\left(\frac{\frac{n_b}{n_a}|h|}{H}\right)-\frac{u}{H}k'\left(\frac{\frac{n_b}{n_a}|h|}{H}\right)\right\} \qquad (2.44)\\
&\quad\times\left\{k(\frac{\frac{n_d}{n_a}|h|}{H})+\frac{1}{H}(-\frac{n_d}{n_b}u+\frac{n_d}{n_c}\alpha-\gamma)k'\left(\frac{\frac{n_d}{n_a}|h|}{H}\right)\right\}+O(H^{-2}).
\end{aligned}$$

Lets look at the cross product terms in (2.44). The first of cross product terms is given by,

$$\begin{aligned}
&T^2\frac{n_a}{n_cn_d}\sum_{|h|<n_a}\sum_{u=0}^{\frac{n_b}{n_a}-1}\sum_{\alpha=0}^{\frac{n_c}{n_a}-1}\sum_{\gamma=0}^{\frac{n_d}{n_b}-1}k\left(\frac{\frac{n_b}{n_a}|h|}{H}\right)k\left(\frac{\frac{n_d}{n_a}|h|}{H}\right)\\
&=T^2\frac{n_a}{n_cn_d}\frac{n_b}{n_a}\frac{n_c}{n_a}\frac{n_d}{n_b}\sum_{|h|<n_a}k\left(\frac{\frac{n_b}{n_a}|h|}{H}\right)k\left(\frac{\frac{n_d}{n_a}|h|}{H}\right)=O(\frac{H}{n_a}),
\end{aligned}$$

where the last approximation holds when $\frac{n_b}{Hn_a}=o(1)$ and $\frac{n_d}{Hn_a}=o(1)$, which is satisfied under Assumption 4. The remainder term is given by

$$T^2 \frac{n_a}{n_c n_d} \sum_{|h|<n_a-1} \sum_{u=0}^{\frac{n_b}{n_a}-1} \sum_{\alpha=0}^{\frac{n_c}{n_a}-1} \sum_{\gamma=0}^{\frac{n_d}{n_b}-1} \left\{ -\frac{u}{H} k' \left(\frac{\frac{n_b}{n_a}|h|}{H} \right) \right\} \left\{ \frac{1}{H} (-\frac{n_d}{n_b} u + \frac{n_d}{n_c} \alpha - \gamma) k' \left(\frac{\frac{n_d}{n_a}|h|}{H} \right) \right\}$$

$$= T^2 \frac{n_a}{n_c n_d} \frac{1}{H} \sum_{u=0}^{\frac{n_b}{n_a}-1} \sum_{\alpha=0}^{\frac{n_c}{n_a}-1} \sum_{\gamma=0}^{\frac{n_d}{n_b}-1} u(\frac{n_d}{n_b} u - \frac{n_d}{n_c} \alpha + \gamma) \left\{ \frac{1}{H} \sum_{|h|<n_a-1} k' \left(\frac{\frac{n_b}{n_a}|h|}{H} \right) k' \left(\frac{\frac{n_d}{n_a}|h|}{H} \right) \right\}$$

$$\simeq T^2 \frac{n_a}{n_c n_d} \frac{1}{H} \sum_{u=0}^{\frac{n_b}{n_a}-1} u(\frac{n_d}{n_b} u - \frac{n_d}{n_c} \sum_{\alpha=0}^{\frac{n_c}{n_a}-1} \alpha + \sum_{\gamma=0}^{\frac{n_d}{n_b}-1} \gamma) ||k'||^2$$

$$= O(\frac{n_b}{H n_a^2} + \frac{n_d}{H n_c n_b}).$$

Since the kernel function is symmetric, other cross terms in (2.44) are zero. Then (2.41) is given by

$$\sum_{i,,`j,k,l} k_H(t_{i,a} - t_{j,b}) k_H(t_{k,c} - t_{l,d})(t_{i,a} \wedge t_{k,c} - t_{i-1,a} \vee t_{k-1,c})(t_{j,b} \wedge t_{l,d} - t_{j-1,b} \vee t_{l-1,d})$$

$$= O(\frac{H}{n_a}) + O(\frac{n_b}{H n_a^2} + \frac{n_d}{H n_c n_b}).$$

When sample size is balanced, then above simplifies to $O(\frac{H}{n}) + O(\frac{1}{Hn})$. The first term is of leading order under Assumption 4.

■

LEMMA 5. *Let $P(t)$ defined on probability space $(\Sigma, \mathcal{F}, \mathcal{F}_{t\geq 0}, \mathbb{P})$ satisfying the Assumption 1 and let sub-σ-*field of $\mathcal{F}$ by $\mathcal{G} = \sigma(P)$. *The Z is a standard normal variable on the suitable extension of probability space and $\mathcal{V}$ is a $\mathcal{G}$-measurable stochastic variance. Then it holds that for $f_n(\cdot)$ given in Lemma 2,*

$$\sqrt{\frac{n}{H}} \int_0^{2\pi} \int_0^{2\pi} f_n(t, s; q) dP_1(s) dP_2(t) \Longrightarrow \sqrt{\mathcal{V}} Z.$$

where convergence is $\mathcal{G}$-stably in law.

PROOF. Stable convergence is notion of joint convergence and stronger than the convergence in law. See Aldous and Eagleson (1978) Proposition 1 for the definition of a stable convergence. Let the discretized filtration by $\mathcal{F}_i$, $i = \max_j\{t_j \leq t\}$. For the discretized sequence

$$\chi_i^n = \sqrt{\frac{n}{H}}\Delta P_1(t_i) \sum_{j:s_j<t_i} \Delta P_2(s_j)k_H(t_i - s_j)e^{-is_jq},$$

which is adopted to $\mathcal{F}_i$,we show the stable convergence of $Z_t^n := \sum_{\max_i\{t_i \leq t\}} \chi_i^n$ to $Z_t = \int_0^t v_s dW_s$, a $\mathcal{F}_t$-conditional Gaussian martingale. Under the following conditions:

$$(1) \sum_i E(|\chi_i^n|^2|\mathcal{F}_{i-1}) \to_p [Z, Z]_t; \quad (2) \sum_i E(|\chi_i|^2 1_{\{|\chi_i^n|>\epsilon\}}|\mathcal{F}_{i-1}) \to_p 0 \quad \forall\epsilon,$$

we have $Z^n \implies Z$ stably. See the proof for Theorem 3.2 in Jacod (1997). The sufficient condition for the conditional Lindberg condition in (2) is the Liapanov condition $\sum_i E(|\chi_i^n|^{2+\varepsilon}|\mathcal{F}_{i-1}) \to_p 0$,for $\varepsilon > 0$. We will show for $\varepsilon = 2$ in the proofs for Theorem 1 and Theorem 2.

2.8.3 Proof of Theorem 1

We first prove for the diagonal element. Consider the first element of the centered estimator

$$\mathcal{E} = \sum_{|\alpha|\leq m/2} K_H(\lambda_\alpha)\mathcal{F}_n(P_1)(\alpha)\mathcal{F}_n(P_1)(q - \alpha) - \mathcal{F}(\Sigma_{11})(q).$$

We drop the subscript denoting asset for now. We can decompose the centered estimator into two terms

$$\mathcal{E} = M_1 + M_2,$$

$$M_1 = \sum_{i=1}^n \Delta P^2(t_i)e^{-it_iq} - \int_0^{2\pi} e^{-iqt}d[P, P](t) \quad ; \quad M_2 = \sum_{i\neq j} \Delta P(t_i)\Delta P(t_j)k_H(t_i - t_j)e^{-it_jq}.$$

We will show that $\sqrt{\frac{n}{H}}M_1 = o_p(1)$ and $\sqrt{\frac{n}{H}}M_2$ stably converges to a zero mean Gaussian variable. By Itô's formula

$$\begin{aligned}\Delta P^2(t_i) &= P^2(t_i) - P^2(t_{i-1}) - 2P(t_{i-1})\{P(t_i) - P(t_{i-1})\} \\ &= 2\int_{t_{i-1}}^{t_i} \{P(t) - P(t_{i-1})\}dP(t) + \int_{t_{i-1}}^{t_i} d[P,P](t).\end{aligned}$$

Then M_1 can be further decomposed into a martingale M_{11} and a predictable finite variation component A :

$$\begin{aligned}M_{11} &= 2\sum_{i=0}^{n}\int_{t_{i-1}}^{t_i} \{P(t) - P(t_{i-1})\}e^{-ikt_i}dP(t) = O_p(n^{-1/2}) \\ A &= \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} (e^{-itk} - e^{-it_ik})d[P,P](t) = O(n^{-1}).\end{aligned}$$

This is the Euler approximation error and its distribution is given by the Theorem 5.5 of Jacod and Protter (1998). Therefore, $\sqrt{\frac{n}{H}}M_1 = o_p(1)$. The expectation of M_2 is zero. Given the step function $g_n(t,s;q)$ on $[0,2\pi]^2$ defined in (2.33), it can be expressed as

$$M_2 = \int_0^{2\pi}\int_0^{2\pi} f_n(t,s;q)dP(s)dP(t),$$

where double integration is Wiener-Itô sense. Then by Lemma 1,

$$E[M_2, M_2] = 2E\int_0^{2\pi}\int_0^{2\pi} f_n^2(t,s;q)d[P,P](s)d[P,P](t).$$

This is equal to (2.34) in Lemma 3. To verify the condition (2) in Lemma 5, let

$$\chi_i^n = \{\sum_{j<i}\sqrt{\frac{n}{H}}\Delta P(t_i)\Delta P(t_j)k_H(t_i - t_j)\left(e^{-it_jq} + e^{-it_iq}\right)\}.$$

Then, $E|\chi_i^n|^4$ for $i = n$ is bounded by $2^4\times$

$$
\begin{aligned}
&(\frac{n}{H})^2 E\{\sum_{h=1}^{n} \Delta P(t_i)\Delta P(t_{i-h}) k\left(\frac{t_h}{\overline{\Delta t}H}\right)\}^4 \qquad (2.45)\\
&= (\frac{n}{H})^2 \sum_{h=1}^{n} E\{\Delta P^4(t_i)\}E\{\Delta P^4(t_{i-h})\}k^4\left(\frac{t_h}{\overline{\Delta t}H}\right)\\
&+ 6(\frac{n}{H})^2 \sum_{h,l=1}^{n} E\{\Delta P^4(t_i)\}E\{\Delta P^2(t_{i-h})\}E\{\Delta P^2(t_{i-l})\}k^2\left(\frac{t_h}{\overline{\Delta t}H}\right)k^2\left(\frac{t_l}{\overline{\Delta t}H}\right).
\end{aligned}
$$

The fourth moment of the return is given by

$$
\begin{aligned}
&E\Delta P^4(t_i) = E\left(2\int_{t_{i-1}}^{t_i}\{P(t) - P(t_{i-1})\}dP(t) + d[P,P](t)\right)^2\\
&= 4E\left(\int_{t_{i-1}}^{t_i}\{P(t) - P(t_{i-1})\}dP(t)\right)^2 + E\left(\int_{t_{i-1}}^{t_i} d[P,P](t)\right)^2\\
&= 2E\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} d[P,P](s)d[P,P](t) + E\left(\int_{t_{i-1}}^{t_i} d[P,P](t)\right)^2 = 3E\left(\int_{t_{i-1}}^{t_i} d[P,P]_t\right)^2.
\end{aligned}
$$

Denote $[P,P]'(t)dt = d[P,P](t)$. In univariate case this simplifies to $[P,P]'(t) = \sigma^2(t)$. Then (2.45) equals to

$$
\begin{aligned}
9(\frac{n}{H})^2 \sum_{h=1}^{n} &E\left(\int_{t_{i-1}}^{t_i}[P,P]'(t)dt\right)^2\left(\int_{t_{i-h-1}}^{t_{i-h}}[P,P]'(t)dt\right)^2 k^4\left(\frac{t_h}{\overline{\Delta t}H}\right)\\
&+ 18(\frac{n}{H})^2 \sum_{h,l=1}^{n} E\left(\int_{t_{i-1}}^{t_i}[P,P]'(t)dt\right)^2\\
&\times \int_{t_{i-h-1}}^{t_{i-h}}[P,P]'(t)dt \int_{t_{i-l-1}}^{t_{i-l}}[P,P]'(t)dt k^2\left(\frac{t_h}{\overline{\Delta t}H}\right)k^2\left(\frac{t_l}{\overline{\Delta t}H}\right),
\end{aligned}
$$

which is bounded by

$$\begin{aligned}
&\sup_t \ ([P,P]'(t))^4 \{9(\frac{n}{H})^2 \sum_{h=1}^{n} \Delta t_i{}^2 \Delta t_{i-h}{}^2 k^4 \left(\frac{t_h}{\overline{\Delta t} H}\right) \\
&+ 18(\frac{n}{H})^2 \sum_{h,l=1}^{n} \Delta t_i{}^2 \Delta t_{i-h} \Delta t_{i-l} k^2 \left(\frac{t_h}{\overline{\Delta t} H}\right) k^2 \left(\frac{t_l}{\overline{\Delta t} H}\right)\} \\
&\leq 9n^2 H^{-1} \sup_t \ ([P,P]'(t))^4 \sup_i (\Delta t_i{}^4) \\
&\times \{\frac{1}{H} \sum_{h=1}^{n} k^4 \left(\frac{t_h}{\overline{\Delta t} H}\right) + \frac{2}{H} \sum_{h,l=1}^{n} k^2 \left(\frac{t_h}{\overline{\Delta t} H}\right) k^2 \left(\frac{t_l}{\overline{\Delta t} H}\right)\} \\
&= n^{-2} H^{-1} C_1 \int_0^\infty k^4(x) dx + n^{-2} C_2 (\int_0^\infty k^2(x) dx)^2 = O(n^{-2}).
\end{aligned}$$

The pen-ultimate equality is using Assumption 2.1. Therefore the condition (2) in Lemma 5 is satisfied. ■

We now give a result for the off-diagonal element of the estimator. When time stamps are synchronous and sample sizes are balanced, the proof is same as the univariate case. We will give a proof for the most general case, when time stamps are asynchronous and sample sizes are unbalanced. We first show for the bivariate case and will extend the result to general $d \times d$ dimension. Denote the transaction time of the first asset $t_{i,1} = t_i$ and the second asset $t_{j,2} = s_j$. The centered estimator in (2.9) is given by

$$\mathcal{E} = \sum_{\alpha} K_H(\lambda_\alpha) \mathcal{F}_n(P_1)(\alpha) \mathcal{F}_n(P_2)(q-\alpha) - \mathcal{F}(\Sigma_{12})(q).$$

It can be decomposed into $\mathcal{E} = M_1 + M_2$,

$$\begin{aligned}
M_1 &= \sum_{i,j} e^{-is_j q} k_H(t_i - s_j) \Delta P_1(t_i) \Delta P_2(s_j) 1_{\{I_{i,1} \cap I_{j,2} \neq \emptyset\}} - \int_0^{2\pi} e^{-iqt} d[P_1, P_2](t), \\
M_2 &= \sum_{i,j} e^{-is_j q} k_H(t_i - s_j) \Delta P_1(t_i) \Delta P_2(s_j) 1_{\{I_{i,1} \cap I_{j,2} = \emptyset\}}.
\end{aligned}$$

We prove the following proposition.

PROPOSITION 1 *Suppose Assumptions 1-4 hold. Define $\mathcal{B}_{12}$ a bias and $\mathcal{V}_{12}$ variance of $\widehat{\mathcal{F}}(\Sigma_{12})(q)$. Then,*

$$\mathcal{B}_{12} = \left(\frac{n_1 \vee n_2}{(n_1 \wedge n_2)H}\right)^2 \frac{1}{2}\mathcal{A}^2 |k''(0)| \int_0^{2\pi} e^{-itq} d\,|[P_1, P_2]|\,(t)$$

$$\mathcal{V}_{12} = \frac{H}{n_1 \wedge n_2}\|k\|^2 \int_0^{2\pi} e^{-i2tq}\{[P_1, P_1]'(t)[P_2, P_2]'(t)d\mathcal{Q}_{1122}(t) + ([P_1, P_2]')^2(t)d\mathcal{Q}_{1212}(t)\},$$

where $\mathcal{A}$ is defined in Theorem 1.

Let $u_{ij} = t_i \wedge s_j$ and $l_{ij} = t_{i-1} \vee s_{j-1}$. Then,

$$\begin{aligned} E(M_1) &= E(\sum_{i,j} e^{-is_jq} \int_{l_{i,j}}^{u_{i,j}} dP_1(t) \int_{l_{i,j}}^{u_{i,j}} dP_2(s) 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}} - \int_0^{2\pi} e^{-iqt} d[P_1, P_2](t)) \\ &\quad - E(\sum_{i,j} e^{-is_jq} \int_{l_{i,j}}^{u_{i,j}} dP_1(t) \int_{l_{i,j}}^{u_{i,j}} dP_2(s)\{1 - k_H(t_i - s_j)\} 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}}). \end{aligned}$$

By multivariate Itô's calculus,

$$\begin{aligned} &\{P_1(u_{ij}) - P_1(l_{ij})\}\{P_2(u_{ij}) - P_2(l_{ij})\} \\ &= \int_{l_{ij}}^{u_{ij}} \{P_1(t) - P_1(l_{ij})\}dP_2(t) + \{P_2(t) - P_2(l_{ij})\}dP_1(t) + d[P_1, P_2](t). \end{aligned}$$

Conditionally on $1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}}$, $E(M_1)$ is given by the expectation of following terms

$$\sum_{i,j} e^{-is_jq} \int_{l_{i,j}}^{u_{i,j}} \{P_1(t) - P_1(l_{i,j})\}dP_2(t) + e^{-is_jq} \int_{l_{i,j}}^{u_{i,j}} \{P_2(t) - P_2(l_{i,j})\}dP_1(t) \quad (2.46)$$

$$+ \sum_{i,j} \int_{l_{i,j}}^{u_{i,j}} (e^{-is_jq} - e^{-itq})d[P_1, P_2](t) \quad (2.47)$$

$$- \sum_{i,j} e^{-is_jq} \int_{l_{i,j}}^{u_{i,j}} dP_1(t) \int_{l_{i,j}}^{u_{i,j}} dP_2(s)\{1 - k_H(t_i - s_j)\}. \quad (2.48)$$

Recalling the definition of the union of time stamps in Assumption 2, the order of magnitude of the first term in (2.46) is given by

$$\begin{aligned}
& \sum_{1\le i\le n_1, 1\le j\le n_2} e^{-is_j q} \int_{l_{i,j}}^{u_{i,j}} \{P_1(t) - P_1(l_{i,j})\} dP_2(t) \\
= & \sum_{l=1}^{N_T} \int_{T_{l-1}}^{T_l} \{P_1(t) - P_1(T_{l-1})\} dP_2(t) - \sum_{i,j} (1 - e^{-is_j q}) \int_{l_{i,j}}^{u_{i,j}} \{P_1(t) - P_1(l_{i,j})\} dP_2(t) \\
= & \ O_p(N_T^{-1/2}) + O_p(n_2^{-1} N_T^{-1/2}).
\end{aligned}$$

The order of the magnitude for the second term in (2.46) is derived in a similar way. The change of discretization points to the union of the time points are without error and holds analytically. In (2.47), we are discretizing the deterministic function e^{-itq} over the time stamp of s_j. Therefore we can express (2.47)

$$\sum_{i,j} \int_{l_{i,j}}^{u_{i,j}} (e^{-is_j q} - e^{-itq}) d[P_1, P_2](t) = \sum_{1\le j\le n_2} \int_{s_{j-1}}^{s_j} (e^{-is_j q} - e^{-itq}) d[P_1, P_2](t) = O(n_2^{-1}).$$

This term is zero for an integrated (co)variance estimator, $q = 0$. For (2.48), observe that

$$\begin{aligned}
& \left|1 - k_H(t_i - s_j) 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}}\right| = \left| k(0) - k\left(\frac{(t_i - s_j)/\widetilde{\Delta t}}{H}\right) 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}} \right| \\
\simeq & \frac{1}{2} |k''(0)| \left(\frac{(t_i - s_j)/\widetilde{\Delta t}}{H}\right)^2 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}} \\
\le & \frac{1}{2} |k''(0)| \left\{ \frac{n_1 \wedge n_2}{2\pi} \sup_{i,j} |t_i - s_j| 1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}} \right\}^2 \left(\frac{n_1 \vee n_2}{(n_1 \wedge n_2) H} \right)^2 \\
= & \frac{1}{2} |k''(0)| \mathcal{A}^2 \left(\frac{n_1 \vee n_2}{(n_1 \wedge n_2) H} \right)^2.
\end{aligned}$$

by Assumption 2.3. By Assumption 3, $k'(0) = 0$. The explicit asymptotic bias term conditional on the volatility path is given by

$$
\begin{aligned}
& E\sum_{i,j} e^{-is_j q}\int_{l_{i,j}}^{u_{i,j}} dP_1(t)\int_{l_{i,j}}^{u_{i,j}} dP_2(s)\{1-k_H(t_i-s_j)\}1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}} \\
\simeq & \sum_{i,j} e^{-is_j q}\int_{l_{i,j}}^{u_{i,j}} d[P_1,P_2](t)\{-\frac{k''(0)}{2}\left(\frac{(t_i-s_j)}{\widetilde{\Delta} tH}\right)^2\}1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}} \\
\leq & \left(\frac{n_1\vee n_2}{(n_1\wedge n_2)H}\right)^2\left\{\frac{n_1\wedge n_2}{2\pi}\sup_{i,j}|t_i-s_j|1_{\{I_{i,1}\cap I_{j,2}\neq\emptyset\}}\right\}^2\frac{|k''(0)|}{2}\sum_{i,j} e^{-is_j q}\int_{l_{i,j}}^{u_{i,j}} d\left|[P_1,P_2]\right|(t) \\
= & \left(\frac{n_1\vee n_2}{(n_1\wedge n_2)H}\right)^2\mathcal{A}^2\frac{|k''(0)|}{2}\int_0^{2\pi} e^{-itq} d\left|[P_1,P_2]\right|(t),
\end{aligned}
$$

since

$$
\lim_{n\to\infty}\sum_{\{i,j;u_{i,j}>l_{i,j}\}} e^{-is_j q}\int_{l_{i,j}}^{u_{i,j}} d[P_1,P_2](t)=\int_0^{2\pi} e^{-itq} d[P_1,P_2](t).
$$

Then the order of the stochastic bias M_1 is given by $O_p(N_T^{-1/2})+O_p(n_2^{-1})+O_p(\{\frac{n_1\vee n_2}{(n_1\wedge n_2)H}\}^2)$ for estimator at non-zero frequency and $O_p(N_T^{-1/2})+O_p(\{\frac{n_1\vee n_2}{(n_1\wedge n_2)H}\}^2)$ for integrated (co)variance estimator. In both cases, the leading order term for the bias is the last term under the optimal bandwidth. We next analyze M_2 which can be expressed as

$$
M_2=\int_0^{2\pi}\int_{s<t} f_n(t,s;q)dP_2(s)dP_1(t)+\int_0^{2\pi}\int_{s<t} f_n(s,t;q)dP_1(s)dP_2(t),
$$

where $f_n(t,s;q)$ is given in (2.37) in Lemma 4. It has a zero expectation and by Lemma 1, the expectation of quadratic variation is given by

$$
\begin{aligned}
& E\left[M_2,M_2\right] \\
&= E\{\int_0^{2\pi}\int_{s<t} f_n^2(t,s;q)d[P_2,P_2](s)d[P_1,P_1](t)+\int_0^{2\pi}\int_{s<t} f_n^2(t,s;q)d[P_1,P_1](s)d[P_2,P_2](t) \qquad (2.49)\\
&\quad +2\int_0^{2\pi}\int_{s<t} f_n(t,s;q)f_n(s,t;q)d[P_2,P_1](s)d[P_1,P_2](t)\}. \qquad (2.50)
\end{aligned}
$$

By Lemma 4, above expression multiplied by the rate of convergence is equal to

(2.38)+(2.39). To complete the proof for the stable convergence, define

$$\chi_i^n = \{\sum_{j:s_j<t_i} \sqrt{\frac{n}{H}} \Delta P_1(t_i)\Delta P_2(s_j) k_H(t_i - s_j) e^{-is_j q} 1_{\{I_{i,1}\cap I_{j,2}=\emptyset\}}\}.$$

Then $\sup_i E|\chi_i^n|^4 = O((n_1 \wedge n_2)^{-2})$ which can be proved similarly as the univariate case. Therefore the condition (2) in Lemma 5 is met. Denote the reference sample size, $n = \min(n_1, n_2) = n_1$ and let $n_2 = O(n^\beta)$ and $H = O(n^\alpha)$. Under Assumption 4, the estimator is asymptotically unbiased, $\frac{n_1 \vee n_2}{n_1 \wedge n_2}\frac{1}{H} = o(1)$ and consistent, $\frac{H}{n_1 \wedge n_2} = o(1)$. By balancing the squared bias and the variance, the optimal bandwidth is given by $H \propto (n_1 \wedge n_2)^{\alpha^*}$, $\alpha^* = \frac{4\beta-3}{5}$. Then the convergence rate of the estimator under the optimal bandwidth is given by $(n_1 \wedge n_2)^{\vartheta}$, $0 < \vartheta := -\frac{2}{5}\beta^* + \frac{4}{5} \leq \frac{2}{5}$.

To show a convergence of covariation matrix estimator to a multivariate Gaussian distribution by a Cramer-Wold device, it is sufficient and necessary to show that the linear combination of the elements of the matrix estimator converges to a univariate Gaussian random variable. Let denote $\mathcal{R}(q) := \widehat{\mathcal{F}}(\Sigma)(q) - \mathcal{F}(\Sigma)(q)$ and consider the linear combination of the element $a^\top \mathcal{R}(q) b$ and $c^\top \mathcal{R}(q) d$. Note that

$$a^\top \mathcal{R}(q) c b^\top \mathcal{R}(q) d = \mathrm{tr}(\mathcal{R}(q) ab^\top \mathcal{R}(q) dc^\top) = \mathrm{vech}(ab^\top)^\top (\mathcal{R}(q) \otimes \mathcal{R}(q)) \mathrm{vech}(dc^\top).$$

The expectation of the above expression depends on $E\{\mathcal{R}(q) \otimes \mathcal{R}(q)\}$. Each element of this is given in Lemma 3 and Lemma 4. For example, the covariation between the integrated covariance estimator for asset a and b with the estimator for asset c and d,

$$\begin{aligned}
&\left[\widehat{\mathcal{F}}(\Sigma_{ab})(q) - \mathcal{F}(\Sigma_{ab})(q), \widehat{\mathcal{F}}(\Sigma_{cd})(q) - \mathcal{F}(\Sigma_{cd})(q)\right] \\
&= \begin{bmatrix} \sum_{i,j} e^{-it_{j,b}q} k_H(t_{i,a} - t_{j,b}) \Delta P_a(t_{i,a}) \Delta P_b(t_{j,b}) 1_{\{I_{i,a}\cap I_{j,b}=\emptyset\}}, \\ \sum_{k,l} e^{-it_{l,d}q} k_H(t_{k,c} - t_{l,d}) \Delta P_c(t_{k,c}) \Delta P_d(t_{l,d}) 1_{\{I_{k,c}\cap I_{l,d}=\emptyset\}} \end{bmatrix}.
\end{aligned}$$

is given by Lemma 4 (2.41).

■

2.8.4 Proof of Theorem 2

We analyze following quantity,

$$\sum_{|\alpha|\le m/2} K_H(\lambda_\alpha)\mathcal{F}(dU_1)(\alpha)\mathcal{F}(dU_2)(q-\alpha)$$
$$= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \Delta U_1(t_i)\Delta U_2(s_j)e^{-is_jq}\sum_{|\alpha|\le m/2} K_H(\lambda_\alpha)e^{-i(t_i-s_j)\alpha}. \quad (2.51)$$

Proposition 2. *Suppose Assumptions 2-7 holds. Then*

$$E\left\{\sum_{|\alpha|\le m/2} K_H(\lambda_\alpha)\mathcal{F}(dU_1)(\alpha)\mathcal{F}(dU_2)(q-\alpha)\right\} \simeq \eta\frac{n_1\vee n_2}{H^2}|k''(0)|\Gamma_{12}$$
$$E\left\{\left|\sum_{|\alpha|\le m/2} K_H(\lambda_\alpha)\mathcal{F}(dU_1)(\alpha)\mathcal{F}(dU_2)(q-\alpha)\right|^2\right\} = O\Big(\frac{(n_1\vee n_2)^2}{n_1\wedge n_2}H^{2\mu-3}\Big),$$

for $0<\mu<1$ and η is defined in Theorem 2.

Proof. We first derive the expression for (2.51) in terms of Us not ΔUs, separating the end terms and the rest. The end term is defined by either U_0 and U_n. The terms not affected by the end points are given by

$$\sum_{i=1}^{n_1-1}\sum_{j=1}^{n_2-1} U_1(t_i)U_2(s_j)\{-k_H(t_{i+1}-s_j)e^{-is_j\alpha} + k_H(t_i-s_j)e^{-is_j\alpha} \quad (2.52)$$
$$+ k_H(t_{i+1}-s_{j+1})e^{-is_{j+1}\alpha} - k_H(t_i-s_{j+1})e^{-is_{j+1}\alpha}\}.$$

Let denote $\overline{H} := H\widetilde{\Delta t}$ Observe that

$$\begin{aligned}
&-e^{-is_j\alpha}\{k_H(t_{i+1}-s_j)-k_H(t_i-s_j)\} + e^{-is_{j+1}\alpha}\{k_H(t_{i+1}-s_{j+1})-k_H(t_i-s_{j+1})\}\\
=&-e^{-is_j\alpha}\{k_H(t_i-s_j+\Delta t_{i+1})-k_H(t_i-s_j)\}\\
&+e^{-is_{j+1}\alpha}\{k_H(t_i-s_j-\Delta s_{j+1}+\Delta t_{i+1})-k_H(t_i-s_j-\Delta s_{j+1})\}\\
\simeq&-e^{-is_j\alpha}\Delta t_{i+1}\Delta s_{j+1}\frac{1}{\overline{H}^2}k''\left(\frac{t_i-s_j}{\overline{H}}\right) + e^{-is_j\alpha}(e^{-i\Delta s_{j+1}\alpha}-1)\Delta t_{i+1}\frac{1}{\overline{H}}k'\left(\frac{t_i-s_j}{\overline{H}}\right).
\end{aligned}$$

Under the Assumption 2 and 4, we have $\Delta t_{i+1}\Delta s_{j+1}\frac{1}{\overline{H}^2} = o(1)$ and $\Delta t_{i+1}\frac{1}{\overline{H}} = O(\frac{1}{H})$. For equally spaced and balanced sample, the above expression collapses to

$$-e^{-is_j\alpha}\frac{1}{H^2}k^{''}(\frac{i-j}{H}) + e^{-is_j\alpha}(e^{-i\Delta s_{j+1}\alpha}-1)\frac{1}{H}k^{'}(\frac{i-j}{H}).$$

Then (2.52) can be simplified into two terms,

$$\frac{1}{\overline{H}^2}\sum_{i=1}^{n_1-1}\sum_{j=1}^{n_2-1}U_1(t_i)U_2(s_j)e^{-is_j\alpha}k^{''}(\frac{t_i-s_j}{\overline{H}})\Delta t_{i+1}\Delta s_{j+1} \tag{2.53}$$

$$+\frac{1}{\overline{H}}\sum_{i=1}^{n_1-1}\sum_{j=1}^{n_2-1}U_1(t_i)U_2(s_j)e^{-is_j\alpha}(e^{-i\Delta s_{j+1}\alpha}-1)k^{'}\left(\frac{t_i-s_j}{\overline{H}}\right)\Delta t_{i+1}. \tag{2.54}$$

We show that (2.53) is a leading order term. The upper bound for expectations of (2.54) is given by.

$$\begin{aligned}
&\frac{1}{\overline{H}}\sup_j\left|1-e^{-i\Delta s_{j+1}\alpha}\right|\sup_i(\Delta t_{i+1})\left|\sum_{i=1}^{n_1-1}\sum_{j=1}^{n_2-1}E\{U_1(t_i)U_2(s_j)\}k^{'}\left(\frac{t_i-s_j}{\overline{H}}\right)\right| \\
&= C_1\frac{n_1\vee n_2}{H}\frac{1}{n_1n_2}\left|\{\sum_{|t_i-s_j|/\widetilde{\Delta t}\le\sqrt{H}}+\sum_{|t_i-s_j|/\widetilde{\Delta t}>\sqrt{H}}\}E\{U_1(t_i)U_2(s_j)\}k^{'}\left(\frac{t_i-s_j}{\overline{H}}\right)\right| \\
&\le C_1(n_1\wedge n_2)^{-1}H^{-1}\{\sup_{|t_i-s_j|/\widetilde{\Delta t}\le\sqrt{H}}|k^{'}(\frac{t_i-s_j}{\overline{H}})|\left|\sum_{|t_i-s_j|/\widetilde{\Delta t}\le\sqrt{H}}\gamma(\{t_i-s_j\}/\widetilde{\Delta t})\right| \\
&+\sup_{|t_i-s_j|/\widetilde{\Delta t}>\sqrt{H}}|\gamma(\{t_i-s_j\}/\widetilde{\Delta t})|\left|\sum_{|t_i-s_j|/\widetilde{\Delta t}>\sqrt{H}}k^{'}\left(\frac{t_i-s_j}{\overline{H}}\right)\right|\} = o(H^{-1}),
\end{aligned}$$

since $\sup_{|t_i-s_j|/\widetilde{\Delta t}\le\sqrt{H}}|k^{'}(\frac{t_i-s_j}{\overline{H}})| \to k'(0) = 0$ under the Assumption 2. By Assumption 6, $\sum_{|t_i-s_j|/\widetilde{\Delta t}\le\sqrt{H}}\gamma(\{t_i-s_j\}/\widetilde{\Delta t}) = O(n_1\wedge n_2)$. The last supremum term vanishes at the exponential rate by Assumption 6. The expectation of squares of (2.54) is bounded by

$$C_2(n_1 \wedge n_2)^{-2}H^{-2}\left\{\sum_{i,j,r,l} EU_1(t_i)U_2(s_j)U_1(t_r)U_2(s_l)k^{'}(\frac{t_i - s_j}{\overline{H}})k^{'}(\frac{t_r - s_l}{\overline{H}})\right\} \tag{2.55}$$
$$= O((n_1 \wedge n_2)^{-1}H^{2\mu-1}).$$

Denote a set $\mathcal{S} := \{i, j, r, l; (t_i - t_r)/\overline{\Delta t} < H^{\mu}, (s_j - s_l)/\overline{\Delta s} < H^{\mu}\}$ where $0 < \mu < 1$. Then the terms in the curly bracket in (2.55) is given by

$$\begin{aligned}
&\left\{\sum_{i,j,r,l\in\mathcal{S}} + \sum_{i,j,r,l\in\mathcal{S}^c}\right\} EU_1(t_i)U_2(s_j)U_1(t_r)U_2(s_l)k'(\frac{t_i - s_j}{\overline{H}})k'(\frac{t_r - s_l}{\bar{H}}) \\
\leq\ & \sup_{i,j,r,l\in\mathcal{S}} |EU_1(t_i)U_2(s_j)U_1(t_r)U_2(s_l)| \left|\sum_{i,j}\ \sum_{|h|,|v|<H^{\mu}} k'(\frac{t_i - s_j}{\overline{H}})k'(\frac{t_{i-h} - s_{j-v}}{\overline{H}})\right| \\
& + C_3 n_1^2 n_2^2 \sup_{i,j,r,l\in\mathcal{S}^c} |EU_1(t_i)U_2(s_j)U_1(t_r)U_2(s_l)| \\
=\ & (i) + (ii).
\end{aligned}$$

For balanced and equally spaced case, (i) simplifies to

$$\begin{aligned}
&\sup_{|i-r|<H^{\mu},|h-v|<H^{\mu}} |EU_1(t_i)U_2(s_{i-h})U_1(t_r)U_2(s_{r-v})| \\
&\qquad \times \left|\sum_{|i-r|<H^{\mu},|h-v|<H^{\mu}} k^{'}\left(\frac{t_i - s_{i-h}}{\overline{H}}\right) k^{'}\left(\frac{t_r - s_{r-v}}{\overline{H}}\right)\right|.
\end{aligned} \tag{2.56}$$

When the sample size is balanced, it holds that $\frac{t_i - s_{i-h}}{\overline{H}} \simeq \frac{h}{H}$ under Assumption 2.

Then (2.56) is given by

$$
\begin{aligned}
&\sum_{|i-r|<H^\mu}\sum_{|h-v|<H^\mu} k^{'}(\frac{h}{H})k^{'}(\frac{v}{H}) = 2H^\mu n \sum_{|h-v|<H^\mu} k^{'}(\frac{h}{H})k^{'}(\frac{v}{H}) \\
&= 2H^\mu n \left\{ \sum_{0\le l<H^\mu}\sum_{h=1+l}^{n} k^{'}(\frac{h}{H})k^{'}(\frac{h-l}{H}) + \sum_{0<l<H^\mu}\sum_{h=1}^{n-l} k^{'}(\frac{h}{H})k^{'}(\frac{h+l}{H}) \right\} \\
&\le 4H^{2\mu} n \sum_{h=1}^{n} \{k^{'}(\frac{h}{H})\}^2.
\end{aligned}
$$

For unbalanced case, we use the fact that $\sum_{i,j}\{k^{'}\left(\frac{t_i-s_j}{\overline{H}}\right)\}^2 \simeq (n_1 \wedge n_2) H \int_{-\infty}^{\infty}\{k'(x)\}^2 dx$ and that the order of $\#\{0 \le i, r \le n_1; \frac{t_i - t_r}{\widetilde{\Delta t}} < H^\mu\}$ is same as when the data is equally spaced under Assumption 2. Then

$$
(i) = \rho(0) 4 (n_1 \wedge n_2) H^{2\mu+1} \int_{-\infty}^{\infty} \{k'(x)\}^2 dx.
$$

We have $(ii) = C_3 n_1^2 n_2^2 \sup_{|\tau|>H^\mu} \rho(\tau)$ which is exponentially vanishing by Assumption 6. The expectation of (2.53) is given by

$$
\begin{aligned}
& \frac{1}{\overline{H}^2} \left\{ \sum_{|t_i - s_j|/\widetilde{\Delta t} \le \sqrt{H}} + \sum_{|t_i - s_j|/\widetilde{\Delta t} > \sqrt{H}} \right\} E\{U_1(t_i)U_2(s_j)\} e^{-is_j\alpha} k^{''}(\frac{t_i - s_j}{\overline{H}}) \Delta t_{i+1} \Delta s_{j+1} \\
= \;& (i) + (ii).
\end{aligned}
$$

(ii) is bounded by

$$
\begin{aligned}
& \frac{1}{\overline{H}^2} \sup_i(\Delta t_{i+1}) \sup_j(\Delta s_{j+1}) \sup_{|t_i - s_j|/\widetilde{\Delta t} > \sqrt{H}} |EU_1(t_i)U_2(s_j)| \sum_{|t_i - s_j|/\widetilde{\Delta t} > \sqrt{H}} |k^{''}(\frac{t_i - s_j}{\overline{H}})| \\
\le \;& C_4 \frac{n_1 \vee n_2}{H} \sup_{|t_i - s_j|/\widetilde{\Delta t} > \sqrt{H}} \left|\gamma(|t_i - s_j|/\widetilde{\Delta t})\right| \int_{-\infty}^{\infty} |k''(x)| dx,
\end{aligned}
$$

which vanishes at the exponential rate by the Assumption 6. $\int_{-\infty}^{\infty} |k''(x)|dx$ is well defined by the Assumption 3. Given definition of η in Theorem 2, (i) is bounded by

$$\begin{aligned}
&\eta \frac{n_1 \vee n_2}{H^2(n_1 \wedge n_2)} \sum_{|t_i - s_j|/\widetilde{\Delta} t \leq \sqrt{H}} E\{U_1(t_i)U_2(s_j)\}k^{''}(\frac{t_i - s_j}{\overline{H}})e^{-is_j\alpha} \\
&\simeq \eta \frac{n_1 \vee n_2}{H^2(n_1 \wedge n_2)} |k''(0)| \sum_{|t_i - s_j|/\widetilde{\Delta} t \leq \sqrt{H}} EU_1(t_i)U_2(s_j)e^{-is_j\alpha} \simeq \eta \frac{n_1 \vee n_2}{H^2} |k''(0)| \Gamma_{12},
\end{aligned}$$

by Assumption 6. The order of (2.53) is derived similarly as (2.55). The expectation of squares of (2.53) is bounded by

$$\begin{aligned}
&C_5 \left(\frac{n_1 \vee n_2}{H^2(n_1 \wedge n_2)} \right)^2 E \left\{ \sum_{i=1}^{n_1 - 1} \sum_{j=1}^{n_2 - 1} U_1(t_i)U_2(s_j)k^{''}(\frac{t_i - s_j}{\overline{H}}) \right\}^2 \\
&\simeq C_5 \left(\frac{n_1 \vee n_2}{H^2(n_1 \wedge n_2)} \right)^2 \rho(0)4(n_1 \wedge n_2)H^{2\mu+1} \int_{-\infty}^{\infty} \{k''(x)\}^2 dx \\
&= O(\frac{(n_1 \vee n_2)^2}{n_1 \wedge n_2} H^{2\mu-3}). \qquad (2.57)
\end{aligned}$$

Then by Markov inequality (2.53)$=O_p(\frac{n_1 \vee n_2}{(n_1 \wedge n_2)^{1/2}} H^{(2\mu-3)/2})$. With some algebra it can be shown that under the optimal bandwidth given in Theorem 2, $H \propto n^{\frac{2\beta+1}{5}}$, the square root of (2.57) multiplied by the rate of convergence of the distribution $n^{\vartheta}, \vartheta = \frac{2-\beta}{5}$ is $o(1)$. All other terms that involve the end terms are of smaller order by similar argument given in Lemma A.4 and Lemma A.5, Barndorff-Nielsen et al. (2011) Therefore the microstructure noise only contributes to the asymptotic bias. This results coincides with Lemma A.5 of Barndorff-Nielsen et al. (2011) when data is synchronous and sample size is balanced. ■

2.8.5 Proof of Theorem 3

By the triangular inequality

$$
\begin{aligned}
&\|\Sigma(t) - \frac{1}{2\pi} \sum_{|q| \le m/2} K_H(\lambda_q) \exp(iqt) \widehat{\mathcal{F}}(\Sigma)(q)\|_2 \\
&\le \|\Sigma(t) - \frac{1}{2\pi} \sum_{|q| \le m/2} K_H(\lambda_q) \exp(iqt) \mathcal{F}(\Sigma)(q)\|_2 \\
&+ \|\frac{1}{2\pi} \sum_{|q| \le m/2} K_H(\lambda_q) \exp(iqt) \{\mathcal{F}(\Sigma)(q) - \widehat{\mathcal{F}}(\Sigma)(q)\}\|_2.
\end{aligned}
$$

Theorem 2 implies that $\sup_{|q| \le m/2} \|\widehat{\mathcal{F}}(\Sigma)(q) - \mathcal{F}(\Sigma)(q)\|_2 \to_p 0$. If we assume the modulus of continuity of $\Sigma(t)$ is available and given by (2.26) then there exists sequence $\delta(n) \to 0$ such that

$$
\sup_{\delta(n) \le t \le 2\pi - \delta(n)} \|\Sigma(t) - \frac{1}{2\pi} \sum_{|q| \le m/2} K_H(\lambda_q) \exp(iqt) \mathcal{F}(\Sigma)(q)\|_2 \le \mathcal{C}(\frac{4}{m}).
$$

■

Remark To shed some light on the result, consider an instantaneous volatility estimator for asset a. We may use the same amplitude window for smoothing and Fourier inversion.

$$
\begin{aligned}
\widehat{\Sigma}_{aa}(t) &= \frac{1}{2\pi} \sum_{|q| \le m/2} K_H(\lambda_q) \exp(iqt) \sum_{|\alpha| \le m/2} K_H(\lambda_\alpha) \mathcal{F}_n(P_a)(\alpha) \mathcal{F}_n(P_a)(q - \alpha) \\
&= \frac{1}{2\pi} \sum_{i,j} \Delta P_a(t_i) \Delta P_a(t_j) k_H(t_i - t_j) k_H(t - t_j),
\end{aligned}
$$

where $k_H(t-t_j) = \sum_{|q|\leq m/2} K_H(\lambda_q)\exp(-iq(t-t_j))$. Then conditionally on the volatility path,

$$\begin{aligned}
E[\widehat{\Sigma}_{aa}(t) - \Sigma_{aa}(t)] &= E\frac{1}{2\pi}\sum_{i=1}^{n}[2\int_{t_{i-1}}^{t_i}\{P_a(t) - P_a(t_{i-1})\}dP_a(t) \\
&\quad + \int_{t_{i-1}}^{t_i}\Sigma_{aa}(t)dt]k_H(t-t_i) - \Sigma_{aa}(t) \\
&\simeq \frac{1}{2\pi}\sum_i \Delta t_i \Sigma_{aa}(t_{i-1})k_H(t-t_i) - \Sigma_{aa}(t) \to 0.
\end{aligned}$$

2.8.6 Proof of Theorem 4

The spectral density estimator (2.27) can be expressed as

$$\widehat{f}_{xx}(q) = \frac{1}{n}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\Delta X_1(t_i)\Delta X_2(s_j)k_H(t_i - s_j)e^{-i(t_i-s_j)q}. \tag{2.58}$$

The estimator can be decomposed into

$$\widehat{f}_{xx}(q) = \widehat{f}_{uu}(q) + \widehat{f}_{pp}(q) + \widehat{f}_{up}(q) + \widehat{f}_{pu}(q) = \widehat{f}_{uu}(q) + o_p(1),$$

where $\widehat{f}_{up}(q)$ denotes the estimator (2.58) applied to $\Delta U \Delta P$. The leading term of (2.58) comes from the spectral density estimate of the first differenced noise. This makes intuitive sense since $\Delta P = O_p(n^{-1/2})$ and $\Delta U = O_p(1)$ and when the conventional spectral density estimator is applied to these two terms, ΔU will drive the order of magnitude. We will show that for each q ,

$$\widehat{f}_{xx}(q) \to_p f_{uu}(q) := \lim_{n\to\infty}\sum_{|h|<n_1\vee n_2}\gamma_{uu}(h)e^{-i\widetilde{\Delta t}hq}. \tag{2.59}$$

We still assume that the spectral density is symmetric We show the results for simplified case when two time stamps are nested. The time stamps of more liquid asset denoted by s_j can be expressed in terms of the time stamps of less liquid asset, t_i by

$$s_{m_{l-1}(l-1)+u} = t_{l-1} + s^*_{u,l-1}; l = 1,\ldots,n_1, u = 0,\ldots,m_{l-1},$$

where $m_l = \#\{s_j \in [t_{l-1}, t_l)\}$ and $s^*_{u,l-1} := s_{m_{l-1}(l-1)+u} - t_{l-1}$ with $s^*_{0,0} = 0$. For example with $l = 1$, we have $\{s_u = s^*_{u,0}; u = 0, \ldots, m_0\}$. We mean t_{-h} by $-t_h$ and s_{-j} by $-s_j$. Then it holds that under Assumption 2.1 and 2.3,

$$
\begin{aligned}
& \frac{1}{n_1 \wedge n_2} \sum_{i,l=1}^{n_1} \sum_{u=0}^{m_{l-1}} \gamma_{uu}(\{t_i - t_{l-1} - s^*_{u,l}\}/\widetilde{\Delta t}) k\left(\frac{t_i - t_{l-1} - s^*_{u,l}}{\overline{H}}\right) \cos(\{t_i - t_{l-1} - s^*_{u,l}\}q) \\
\simeq\ & \frac{1}{n_1 \wedge n_2} \sum_{|h|<n_1} \sum_{u=0}^{m_{|h|-1}} (n_1 - |h|)\gamma_{uu}(\{t_h + s^*_{u,|h|}\}/\widetilde{\Delta t}) k\left(\frac{t_h + s^*_{u,|h|}}{\overline{H}}\right) \cos(\{t_h + s^*_{u,|h|}\}q) \\
\simeq\ & \sum_{|j|<n_2} \gamma_{uu}(s_j/\widetilde{\Delta t}) k\left(\frac{s_j}{\widetilde{\Delta t} H}\right) \cos(s_j q) \to f_{uu}(q).
\end{aligned}
$$

Using similar argument given by Brockwell and Davis (1991), it can be shown that $E\widehat{f}_{uu}(q) - f_{uu}(q) = O(H^{-2})$ and $\mathrm{var}\{\widehat{f}_{uu}(q) - f_{uu}(q)\} = O(\frac{H}{n})$. If we assume that $\gamma_{uu}(h)$ is continuous in h and the modulus of continuity is given by

$$
\mathcal{C}(\tau) := \sup_{|h-s|\leq\tau} |\gamma_{uu}(h) - \gamma_{uu}(s)|,
$$

then there exists a sequence $\tau(n) \to \infty$ such that

$$
\sup_{|h|\leq\tau(n)} |\gamma_{uu}(h) - \sum_{|q|\leq m/2} K_H(\lambda_q) \exp(i\widetilde{\Delta t} hq) \widehat{f}_{xx}(q)| \leq \mathcal{C}(\frac{4}{m}).
$$

■

Table 2.1: Realized Covariance

		Realized Covariance 5 min aligned					
		BIAS			rMSE		
Sampling	NoiseSignal	(1,1)	(2,2)	**(1,2)**	(1,1)	(2,2)	**(1,2)**
Equal	0	(0.01)	0.00	**(0.00)**	0.17	0.17	**0.16**
	0.001	0.02	0.03	**0.01**	0.17	0.18	**0.16**
	0.01	0.24	0.26	**0.13**	0.31	0.34	**0.23**
(3/2,30)	0	(0.00)	0.01	**(0.08)**	0.17	0.17	**0.17**
	0.001	0.02	0.03	**(0.08)**	0.17	0.18	**0.17**
	0.01	0.25	0.26	**(0.07)**	0.32	0.33	**0.18**
(3/2,2)	0	(0.00)	0.01	**(0.00)**	0.17	0.17	**0.16**
	0.001	0.03	0.03	**0.01**	0.18	0.18	**0.16**
	0.01	0.27	0.26	**0.09**	0.34	0.35	**0.21**
(20,30)	0	(0.00)	0.01	**(0.07)**	0.17	0.17	**0.17**
	0.001	0.02	0.02	**(0.07)**	0.17	0.17	**0.18**
	0.01	0.24	0.22	**(0.04)**	0.31	0.30	**0.19**

		Realized Covariance Refresh Time aligned					
		BIAS			rMSE		
Sampling	NoiseSignal	(1,1)	(2,2)	**(1,2)**	(1,1)	(2,2)	**(1,2)**
Equal	0	(0.00)	0.00	**(0.00)**	0.01	0.01	**0.01**
	0.001	4.52	4.54	**2.10**	4.52	4.54	**2.10**
	0.01	45.24	45.41	**21.07**	45.24	45.41	**21.07**
(3/2,30)	0	0.01	0.01	**(0.02)**	0.08	0.08	**0.08**
	0.001	0.23	0.23	**0.07**	0.25	0.24	**0.10**
	0.01	2.24	2.25	**0.92**	2.25	2.26	**0.93**
(3/2,2)	0	(0.00)	0.00	**(0.17)**	0.02	0.02	**0.17**
	0.001	1.78	1.80	**0.53**	1.78	1.80	**0.53**
	0.01	17.83	17.98	**6.77**	17.84	17.99	**6.78**
(20,30)	0	0.01	0.00	**(0.27)**	0.09	0.07	**0.28**
	0.001	0.16	0.15	**(0.24)**	0.18	0.17	**0.25**
	0.01	1.55	1.55	**0.08**	1.56	1.56	**0.15**

Table 2.2: 2 dimensional covariation matrix - continuous SV (·/100)

		Realized Cov Refresh Time			Realized Cov 5 min			HY	Realized Kernel Refresh Time			**FRK average H**			**FRK opt H**		
NRS		(1,1)	(1,2)	(2,2)	(1,1)	(1,2)	(2,2)	(1,2)	(1,1)	(1,2)	(2,2)	(1,1)	(1,2)	(2,2)	(1,1)	(1,2)	(2,2)
	Sampling: (3/2,30) Balanced																
0	bias	1.0	(1.9)	0.8	**(0.1)**	(7.6)	0.9	1.0	1.0	(1.3)	**0.8**	0.9	**(0.6)**	1.4	0.9	**(0.6)**	1.4
	rmse	**8.0**	7.6	8.1	17.1	17.1	16.8	7.6	8.2	**7.5**	**8.0**	20.9	19.6	20.8	20.9	19.6	20.8
0.001	bias	23.2	7.3	22.9	2.2	(7.7)	3.3	10.3	1.6	0.9	1.6	1.0	0.8	1.7	**0.9**	**(0.1)**	**0.1**
	rmse	24.7	**10.6**	24.6	17.2	17.2	17.6	12.9	**16.7**	15.2	**15.4**	30.2	28.9	32.3	38.2	35.5	41.2
0.01	bias	225	92.3	225	24.5	(6.6)	25.5	96.4	3.5	2.3	4.0	**0.6**	**(1.2)**	**(1.5)**	**0.6**	**(1.2)**	**(1.5)**
	rmse	226	93.6	226	31.6	17.9	33.0	97.6	25.3	23.7	25.5	43.8	40.1	46.7	43.8	40.1	46.7
	Sampling: (3/2,30) Unbalanced																
0	bias	0.1	(2.6)	**0.5**	(0.9)	(8.2)	0.9	0.4	**0.0**	(1.9)	0.6	1.0	(3.1)	1.4	1.1	**(0.1)**	1.3
	rmse	**6.7**	6.9	**7.0**	17.4	18.0	17.8	**6.6**	7.0	7.0	7.3	15.9	14.8	15.9	22.2	21.0	23.1
0.001	bias	23.5	6.6	21.3	1.4	(8.0)	3.8	9.5	2.0	**1.8**	3.1	1.5	(12.5)	5.0	**1.0**	(4.2)	**2.9**
	rmse	24.7	**9.6**	22.6	17.9	18.5	18.7	11.7	15.7	14.6	15.6	**10.5**	15.5	**12.2**	15.2	14.4	15.3
0.01	bias	239	92.5	210	23.5	(6.9)	27.3	95.7	2.6	2.3	**4.6**	1.9	(1.9)	11.6	**1.6**	**0.3**	4.9
	rmse	240	93.9	211	31.7	20.3	35.2	96.8	24.0	23.3	25.9	**18.1**	**16.7**	**21.7**	23.9	22.9	25.8
	Sampling: (20,30) Unbalanced																
0	bias	**0.1**	(27.1)	0.7	0.6	(6.7)	1.3	0.8	0.2	(23.7)	**0.7**	0.9	**(0.4)**	1.8	0.9	**(0.4)**	1.8
	rmse	**8.3**	28.0	**8.2**	17.2	17.5	17.8	**9.3**	8.4	24.8	8.2	20.2	19.1	20.9	20.2	19.1	20.9
0.001	bias	16.8	(23.1)	16.2	3.1	(6.1)	3.7	4.9	2.0	0.1	2.8	1.1	**0.1**	1.9	**0.8**	0.4	**1.5**
	rmse	19.0	24.4	18.6	18.2	18.0	18.8	**11.0**	**16.3**	15.0	**16.2**	23.3	22.1	24.2	26.9	25.9	28.7
0.01	bias	162	10.4	155	26.1	(1.7)	26.0	41.1	3.8	2.2	4.7	1.7	0.6	1.9	**1.0**	**(0.5)**	**0.2**
	rmse	164	**16.5**	156	34.0	20.4	34.3	43.6	**24.3**	23.4	26.7	32.2	30.8	34.5	35.8	34.1	39.4

Table 2.3: 2 dimensional covariation matrix - jump diffusion SV (·/100)

		Realized Cov Refresh Time			Realized Cov 5 min			HY	Realized Kernel Refresh Time			**FRK average H**			**FRK opt H**		
NRS		(1,1)	(1,2)	(2,2)	(1,1)	(1,2)	(2,2)	(1,2)	(1,1)	(1,2)	(2,2)	(1,1)	(1,2)	(2,2)	(1,1)	(1,2)	(2,2)
	Sampling: (3/2,30) Balanced																
0	bias	(0.3)	(2.0)	**(18.7)**	(2.2)	(6.0)	(20.4)	**(0.3)**	(0.2)	(1.5)	(18.6)	**(0.1)**	(0.7)	(18.8)	**(0.1)**	(0.7)	(18.8)
	rmse	**6.7**	**18.1**	20.3	17.2	21.9	25.4	18.6	7.1	18.3	**20.3**	18.0	23.4	24.6	18.0	23.4	24.6
0.001	bias	23.3	8.2	**4.9**	0.5	(5.6)	(17.2)	9.8	1.5	1.3	(16.9)	**0.0**	**0.1**	(19.3)	**0.0**	**0.1**	(19.3)
	rmse	24.7	**19.8**	**9.9**	16.8	22.0	23.0	21.1	**13.4**	22.0	21.8	29.1	29.2	31.9	29.1	29.2	31.9
0.01	bias	233	96.7	211	24.2	(3.6)	**7.0**	98.8	2.6	2.1	(15.8)	1.4	1.3	(18.0)	**10.2**	**(0.7)**	**(10.0)**
	rmse	234	99.6	212	30.8	22.6	19.9	101.8	19.8	24.5	25.3	42.6	38.9	42.8	**17.2**	**21.2**	**17.1**
	Sampling: (3/2,30) Unbalanced																
0	bias	(0.3)	(1.9)	**(18.8)**	(1.6)	(6.6)	(21.2)	**(0.1)**	(0.2)	(1.6)	(18.8)	0.2	(2.5)	(19.1)	**0.0**	(0.5)	(18.9)
	rmse	7.1	**17.7**	**20.4**	16.9	22.0	26.1	18.2	**7.1**	17.9	20.5	13.7	21.1	23.0	20.1	24.4	25.3
0.001	bias	24.0	8.0	**2.6**	0.8	(6.6)	(19.0)	10.0	0.7	**0.5**	(17.5)	1.5	(8.3)	(14.6)	**0.7**	(3.2)	(17.5)
	rmse	25.5	19.7	**9.3**	17.6	22.4	24.7	21.1	14.0	20.7	21.5	9.2	**19.1**	18.0	**12.8**	20.5	21.6
0.01	bias	243	94.1	192	23.1	(6.1)	3.4	97.8	**1.4**	**1.0**	(15.7)	1.8	(1.6)	(8.6)	3.8	(3.3)	**2.6**
	rmse	244	97.1	193	32.1	24.2	20.8	100.7	22.1	24.7	23.8	15.2	21.6	17.3	**12.6**	**19.6**	**14.8**
	Sampling: (20,30) Unbalanced																
0	bias	(0.1)	(16.7)	**(18.0)**	(1.4)	(5.4)	(20.1)	**(0.0)**	0.1	(14.7)	(18.0)	**(0.0)**	(0.9)	(18.9)	(0.2)	(0.1)	(18.9)
	rmse	**6.8**	21.5	**19.5**	16.8	21.7	26.3	**19.2**	6.8	20.4	19.6	17.6	23.3	24.2	24.2	27.0	28.3
0.001	bias	15.3	(13.2)	**(2.7)**	1.4	(4.7)	(17.6)	3.7	1.6	(0.4)	(17.9)	0.1	(0.4)	(18.5)	**0.1**	**(0.0)**	(18.8)
	rmse	17.1	**19.1**	8.9	17.1	21.6	24.7	19.6	**14.5**	21.2	21.7	19.7	24.5	25.1	23.9	26.7	28.0
0.01	bias	159	20.6	136	24.9	**0.1**	**4.9**	40.1	3.0	1.2	(15.9)	0.8	0.9	(17.8)	6.2	**0.7**	**(10.5)**
	rmse	**160**	26.8	137	32.2	**22.3**	20.9	45.5	22.7	25.8	24.1	30.2	30.6	32.6	17.9	**22.6**	**17.7**

Table 2.4: Scalar function of 10 dimensional covariation matrix

		max (eigenvalue)		portfolio	
Noise to Signal Ratio=0		Bias	rMSE	Bias	rMSE
RV_refresh		(2.34)	2.75	1.76	2.74
RV_fixed		**(0.85)**	3.18	**0.14**	4.09
Realized Kernel		(2.21)	2.65	1.66	2.67
Fourier RK		(1.18)	**2.17**	0.26	**2.51**
Noise to Signal Ratio=0.001					
RV_refresh		7.22	7.50	27.20	27.47
RV_fixed		0.40	3.28	4.14	6.31
Realized Kernel		0.38	3.00	1.67	4.16
Fourier RK	minH	**(0.28)**	**1.95**	3.88	4.87
	avgH	(0.47)	2.64	0.73	**3.67**
	maxH	(0.52)	3.13	**(0.20)**	3.99
Noise to Signal Ratio=0.01					
RV_refresh		127.24	127.81	256.02	257.05
RV_fixed		15.42	16.54	40.37	41.99
Realized Kernel		1.29	4.82	3.23	6.86
Fourier RK	minH	1.22	**3.27**	6.88	8.55
	avgH	0.13	4.02	1.76	**5.58**
	maxH	**(0.03)**	4.92	**0.67**	5.96

Figure 2.7: Simulation Result : Balanced, Sampled at $\{3/2, 30\}$

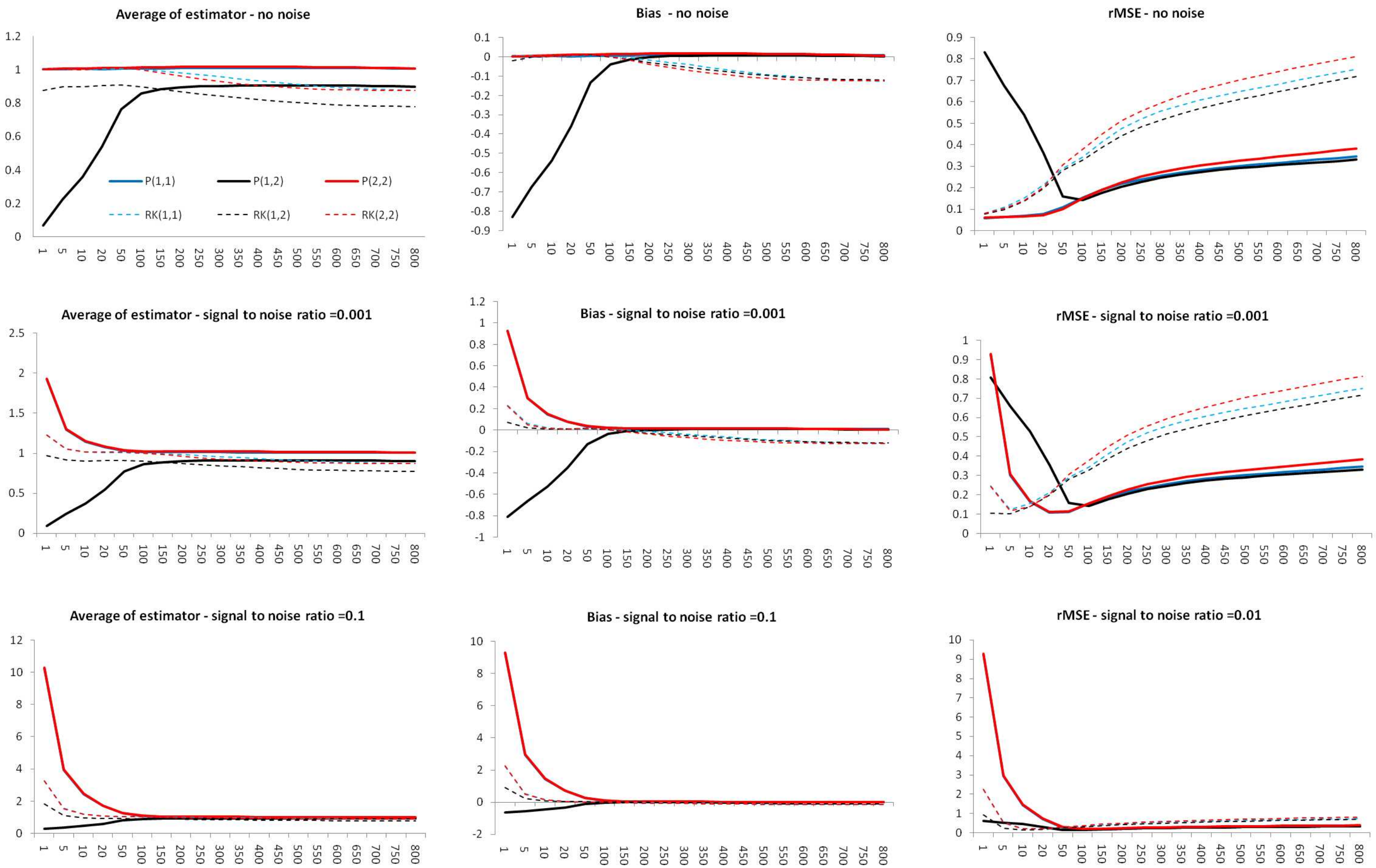


Figure 2.8: Simulation Result : Unbalanced, Sampled at $\{3/2, 30\}$

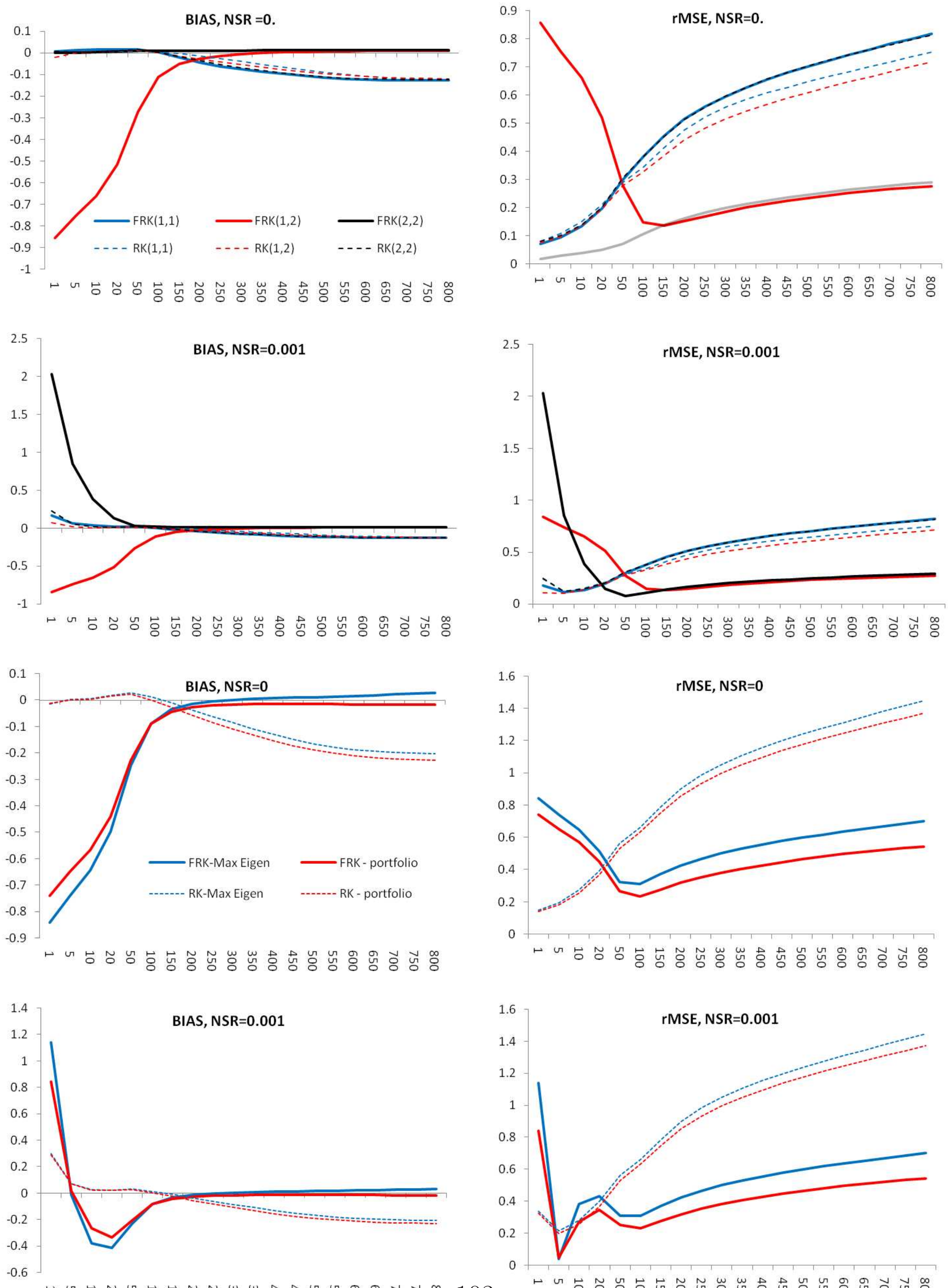


Figure 2.9: Simulation Result : Unbalanced, Sampled at $\{20, 30\}$

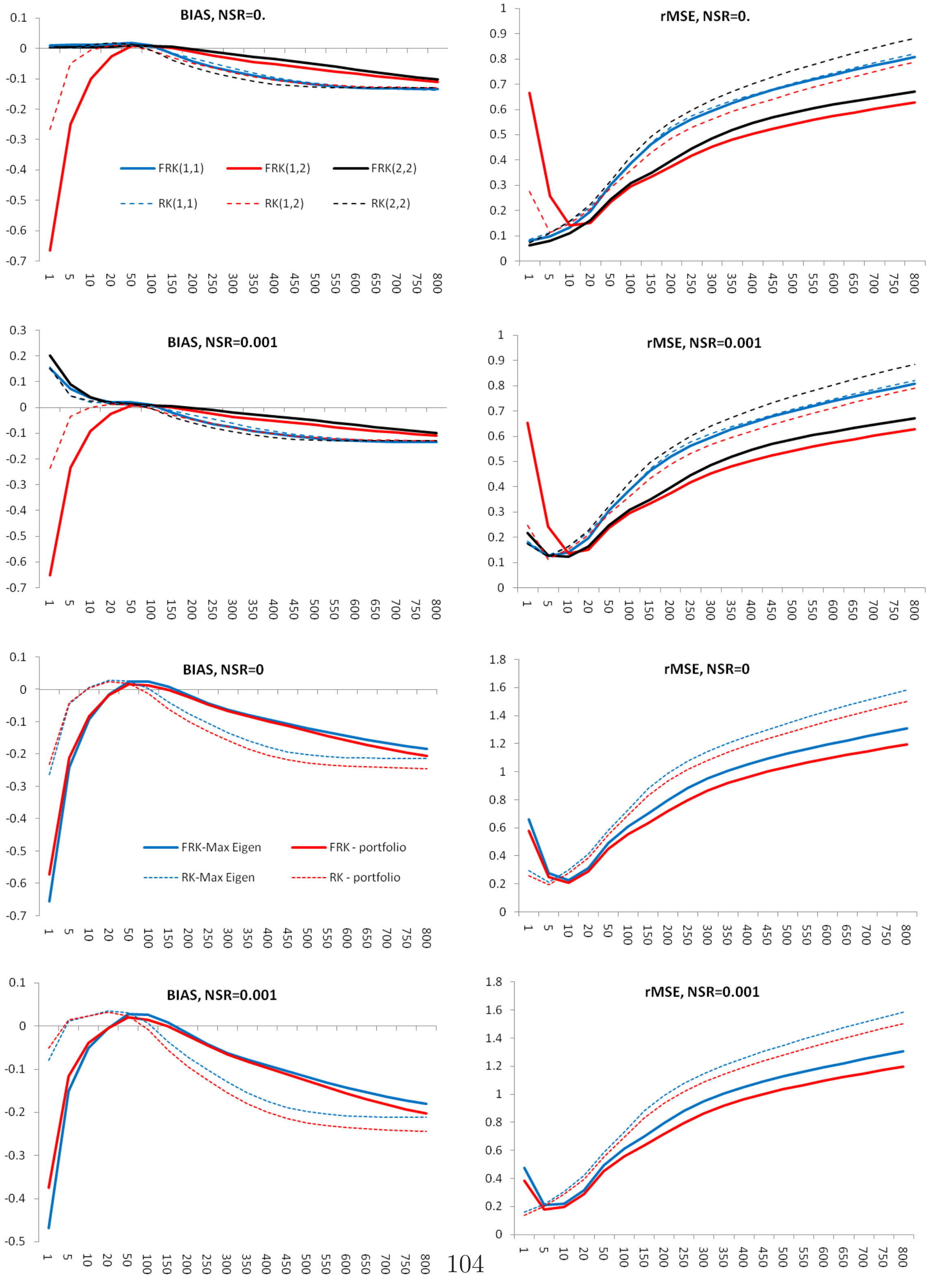


Figure 2.10: Simulation Result : Unbalanced, Sampled at $\{3/2, 2\}$

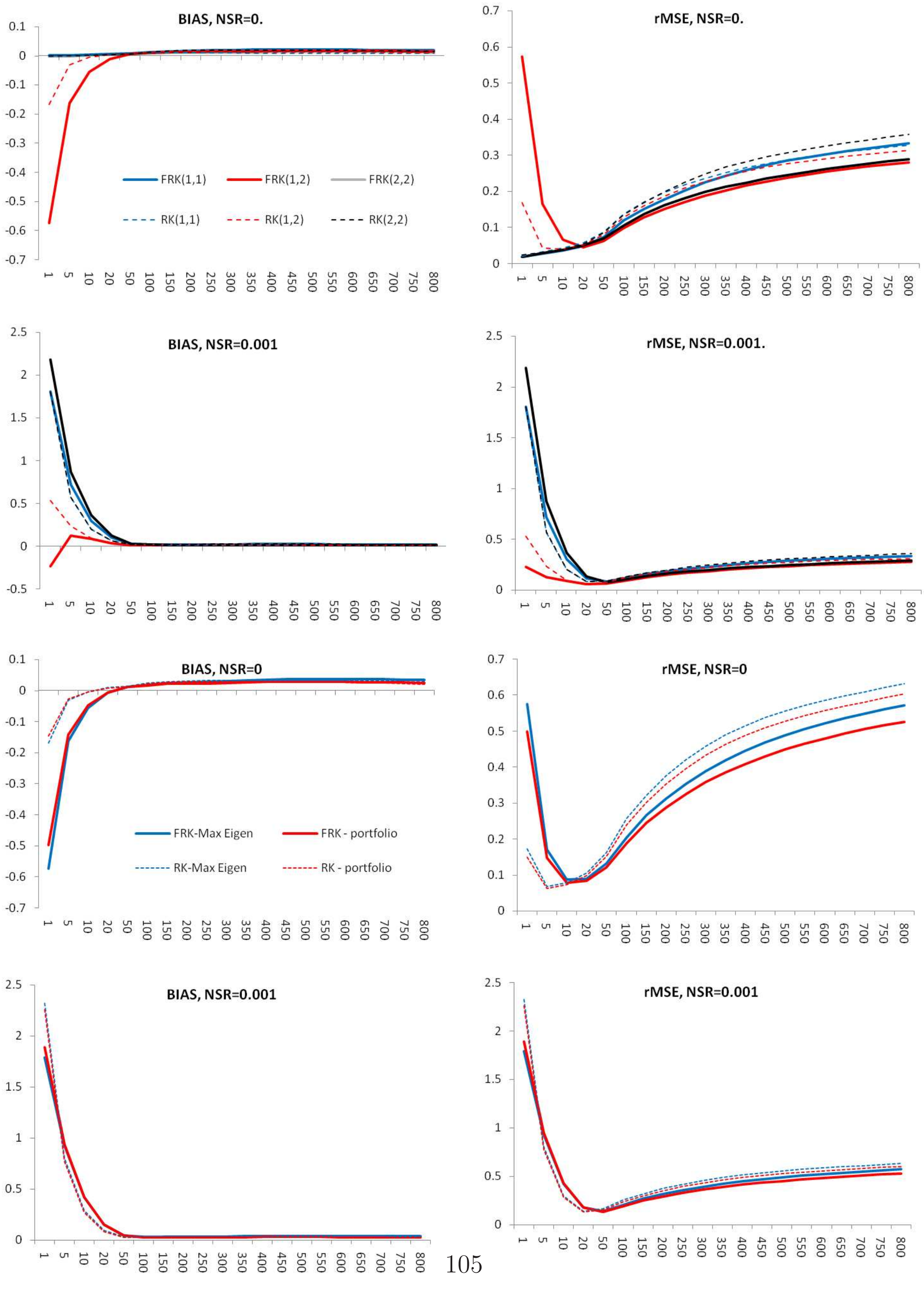


Figure 2.11: Covariation matrix estimates : element $(1, 1)$ to $(2, 5)$

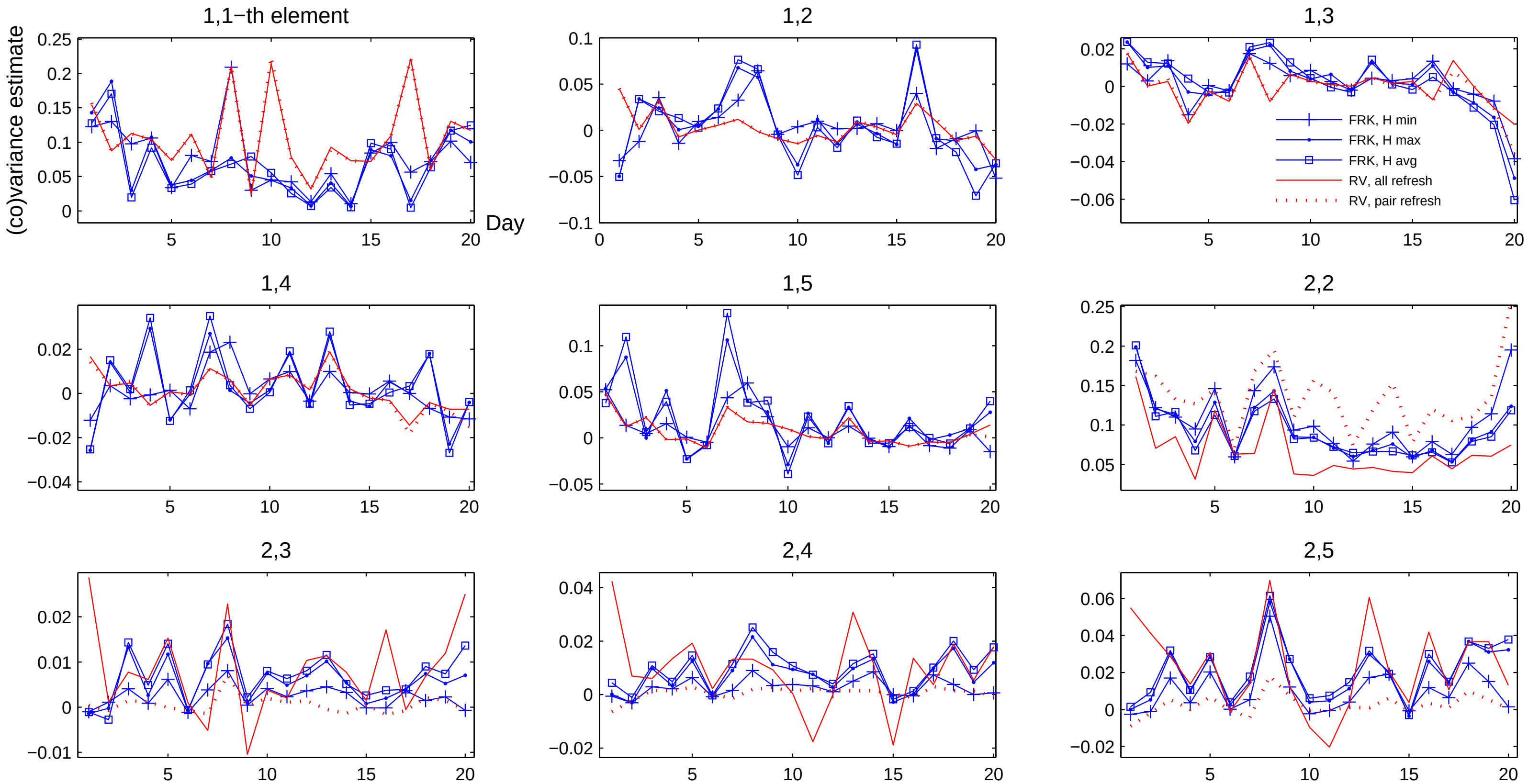
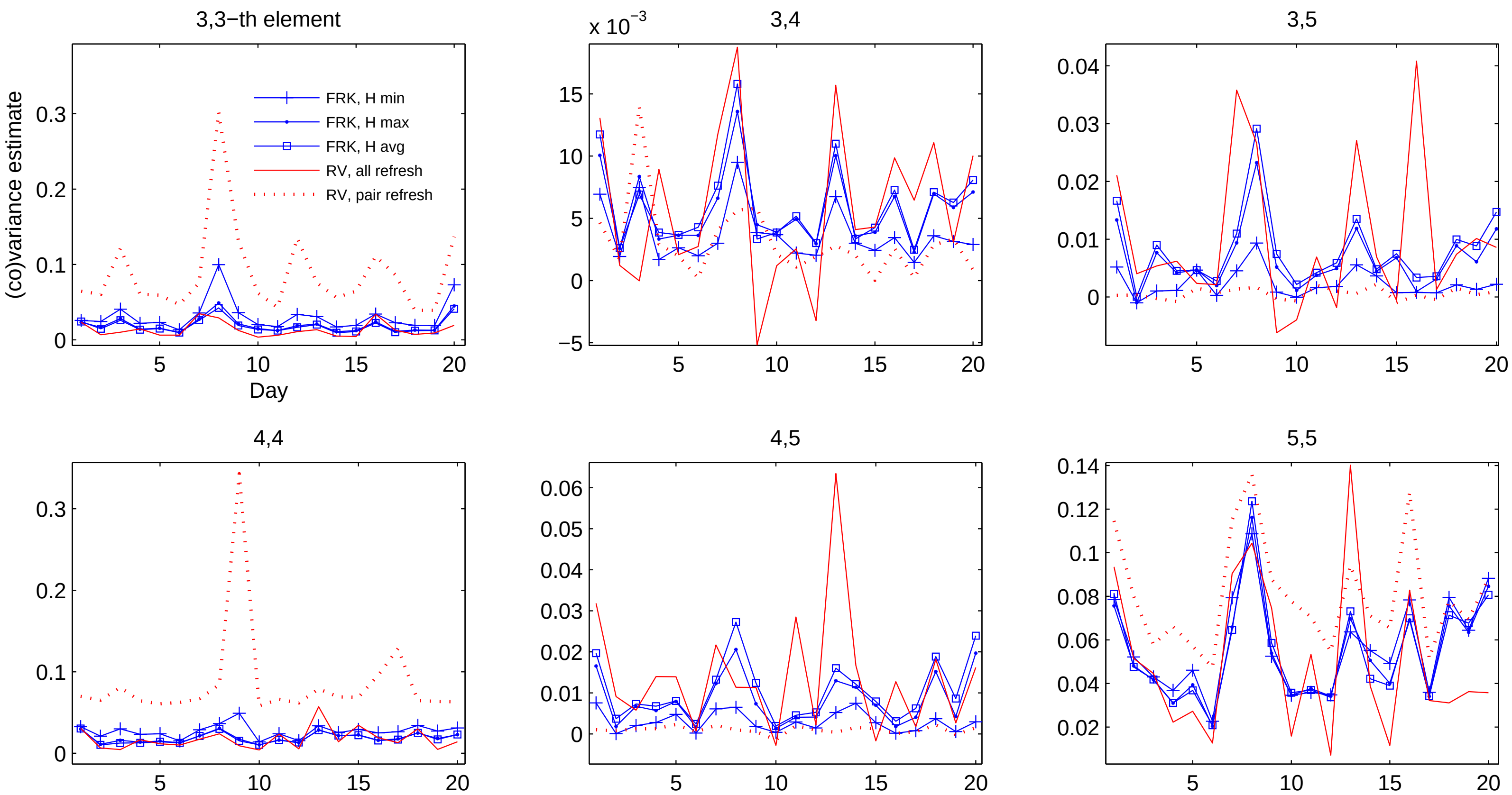


Figure 2.12: Covariation matrix estimates : element $(3,3)$ to $(5,5)$



Chapter 3

Deformation Estimation for High Frequency Data

3.1 Introduction

We propose to model high frequency price series by a time-deformed Lévy process. The deformation function is modeled by a piecewise linear function of a physical time with a slope depending on the marks associated with intra-day transaction data. The performance of a quasi-MLE and an estimator based on a permutation-like statistic is examined in extensive simulations. We also consider estimating the deformation function nonparametrically by pulling together many time series. We show that financial returns spaced by equal elapse of estimated deformed time is homogeneous. The proposed model better recovers the homogeneity than the Realized Variance. We propose an order execution strategy using the fitted deformation time.

3.2 Stylized features of high frequency prices

We first carry out a descriptive analysis of high frequency returns. See Section 2 of Park and Linton (2011) for more detailed description of high frequency data. We analyze the transaction price of NYSE traded J P Morgan stock from the TAQ database. Table 3.1 reports the result. The distribution of returns over a shorter horizon, for

Table 3.1: Sample moments of high frequency returns

Freq	$\hat{\mu}$	$\hat{\sigma}$	Skew	Kurtosis	min	max	size
tick	-1.0E-07	2.4E-08	-0.14	21	-0.005	0.003	151844
1 sec	-3.3E-08	8.0E-09	-0.25	66	-0.005	0.003	467276
30 sec	-1.1E-06	1.9E-07	0.06	10	-0.005	0.005	15565
10 min	-1.2E-07	3.8E-06	0.23	4	-0.006	0.008	760
30 min	2.8E-05	1.3E-05	0.44	6	-0.012	0.015	240

example every transaction returns deviates more from the Gaussian distribution. The fat tailed distribution can be contributed to time varying conditional variance and presence of jumps. The Figure 3.1 shows the daily volatility curve proxied by the squared high frequency returns. It shows a daily repetition of U-shaped pattern. The intra-day activity variables such as number of transactions, trade volume and trade duration also show a similar (inverted) U-shaped pattern. We will term "transaction marks" for such intra-day activity variables. These are candidate variables for modeling a financial clock and their daily cumulative sum is plotted in Figure 3.2. Both Figure 3.1 and Figure 3.2 show that the time series pattern of transaction marks repeats itself on a daily interval. They are highly persistent (Figure 3.3) and cross-correlated (Figure 3.4). Such empirical observation suggests that we may model the financial clock by pulling transaction marks over many days rather than solely using noisy squared returns and model the daily curve separately.

3.3 Methodology

3.3.1 Models

Let $X(t)$ be a log price process defined as

$$X(t) = Y\{h(t)\}, \qquad t \in [0, T], \tag{3.1}$$

where $h(\cdot)$ is an unknown monotonically increasing function, $T > 0$ is a constant, and the latent process $Y(\cdot)$ is defined by a stochastic differential equation driven by a Lévy process

$$dY(t) = \mu(t)dt + dZ(t), \tag{3.2}$$

where $Z(t)$ is a Lévy process, i.e. a process with independent and stationary increments. Note that the characteristic function of $Z(t)$ is of the form

$$E\left[\exp\{i\lambda Z(t)\}\right] = \exp\{ti\varphi(\lambda)\}, \tag{3.3}$$

where $\varphi(\lambda)$ is a characteristic exponent of Lévy process. Hence both the mean and the variance of $Z(t)$, if exist, are linear in t. See, for example, the appendix of Norberg (2004) . We assume $E\{Z(t)^2\} < \infty$, and $Z(t)$ is already centered (i.e. $Z(t)$ is a compensated Lévy process), and its mean is absorbed in the drift $\mu(t)dt$. Examples of the Lévy processes include, for example, Brownian motion, compound Poisson process, and, more generally, the sum of independent Brownian motion and compound Poisson process. In fact the latter is one of most frequently used form for modeling high frequency prices with jumps. We assume $\mu(\cdot)$ in (3.2) is either a smooth deterministic function, or $\mu(t) = aY(t) + b$, where a, b are unknown constant and b stands for the mean constant of the Lévy process. In latter case $Y(t)$ is an Orstein-Uhlenbeck process (driven by a Lévy process), which is a continuous-time version of AR(1) models. A word on notation. For any given function $f : \mathbb{R}^+ \to \mathbb{R}$ if we mean mapping from the generic time index $t \in [0, T]$, we will denote the function by $f(t)$ and if we mean mapping from the deformed time $h(t) \in [0, T]$ then we denote by $f\{h(t)\}$.

We observe the process X at the discrete times $0 \le t_0 < t_1 < \cdots < t_n \le T$ with the observations $X(t_0), X(t_1), \cdots, X(t_n)$. Put $Y_0 = X(t_0)$. For $j = 1, \cdots, n$, let

$$Y_j = Y\{h(t_j)\} = X(t_j), \quad \Delta h_j = h(t_j) - h(t_{j-1}), \quad V_j = Y_j - Y_{j-1}.$$

Suppose that all Δh_j are sufficiently small. It follows from (3.2) and (3.3) that

$$V_j \ \approx \ \mu_j \Delta h_j + Z\{\Delta h_j\} = \mu_j \Delta h_j + \sigma\sqrt{\Delta h_j}\varepsilon_j, \tag{3.4}$$

where $\mu_j = \mu\{h(t_{j-1})\}$, $\sigma > 0$ is a constant and ε_j is a mean 0 and variance 1 random variable. For the Orstein-Uhlenbeck process, $\mu_j = aY_{j-1} + b$. If $Z(t)$ is a Brownian motion, ε_j is $NID(0,1)$. In other cases ε_j is not independent of j through dependence on Δh_j. We first outline some parametric specifications for $h(\cdot)$ in the next section and will propose number of methods to estimate the deformation function $h(\cdot)$.

REMARK To exploit the scaling relationship, we could restrict attention to a self-similar Lévy process which has following characteristics; $Z\{\Delta h_j\} \sim (\Delta h_j)^{\frac{1}{\alpha}}\varepsilon$ where ε is an i.i.d random variable governed by law of $Z(1)$. A self similar Lévy process is either a Brownian motion with $\alpha = 2$ or a Lévy process with a symmetric α-stable density, $\alpha \in (0,2)$. It has a characteristic function, $E(e^{izX(t)}) = \exp -c^{\alpha}|z|^{\alpha}$, $0 \leq c < \infty$. This is derived from

$$E\left[e^{izX(t)}\right] = e^{t\psi(z)} = e^{\psi(t^{1/\alpha}z)},$$

where the first equality is from the definition of Lévy processes and the second equality is using the property of self-similarity. Solving for $\psi(z)$ and exponentiating we have the result. A stable law has infinite j-th moments for $j \geq \alpha$. The sample path of a stable process resembles a compound Poisson for α close to 0 and resembles a Brownian motion for α close to 2.

Before we proceed to the estimation method of our model, we discuss how our model is related to the literature using time changing technique. Geman (2006) provides an extensive survey on the time deformation method in finance. The idea of economic clock was first initiated by a search for an explanation why the financial returns exhibit an excess kurtosis. See Mandelbrot and Taylor (1967) and Clark (1973). By modeling the asset price process as a time-changed Brownian motion, it leads to a representation of asset returns as a mixture of normal distribution where mixing factor is given by the time change. The model accommodates the stochastic volatility where changes in volatility is driven by $h(t)$ in our model, a economic time scale that follows the latent flow of information arriving the market. One noteworthy feature of the time deformation model is that it provides an explanation how time dependencies of volatility can occur.

Theoretical justification of (3.1) is given by Dubins and Schwarz (1965) who showed that any continuous martingale can be expressed as a time-changed Brownian motion, where the time change is given by the quadratic variation. Monroe (1978) extended the class of Brownian motion embeddable process to a semimartingale. There are two distinct approaches in the literature that attempts to recover the financial clock from the financial return series. The first approach, closest to ours, is to estimate the deformed time from the activity data to best recover the normality of returns. Clark (1973) used a cumulative volume and Ane and Geman (2000) used a cumulative number of transactions. The second approach is to theoretically identify $h(t)$. Geman, Madan and Yor (2001) first specified a model for $X(t)$ and showed that we can recover $h(t)$ by matching the characteristic function of $X(t)$ and $B\{h(t)\}$.

We now discuss the time changing technique applied to Lévy processes. The motivation of such approach is two fold: Firstly, it is to construct another Lévy process. A large class of infinite activity jump processes is constructed by time changing a Brownian motion by a subordinator. A subordinator is defined by an almost surely increasing Lévy process that can have positive jumps of finite variation but not a diffusion component. A Lévy process time changed by a subordinator is also a Lévy process. The second motivation, relevant to our approach is to build a more empirically plausible model for the financial returns. Observe that the independent increment property of $X(t) = Z\{h(t)\}$ inherits from that of $h(t)$. This implies that the time change technique can be used to construct a model with stochastic and mean reverting volatility which cannot be achieved by a Lévy process alone. Carr, Gemann, Madan, Yor (2003) considered $h(t) = \int_0^t \tau(u)du$ where $\tau(u)$ can be a mean reverting square root process or a positive OU process.

3.3.2 Deformation functions

In this section, we outline the parametric specifications for deformation function $h(\cdot)$. A simple option in choosing the deformation function is to let $h(\cdot)$ be piecewisely linear in the sense that it is a linear function on each intervals $[t_{j-1}, t_j)$ depending on U_j, where U_j represents the information accumulating on the interval $[t_{j-1},\, t_j)$. For high frequency asset prices, U_j may contain, among others, the number of transactions, trade volume and spread over the interval $[t_{j-1},\, t_j)$. It may also contain a time stamp reflecting the time when a trade takes place during day. Since we deal with the changes in each of intervals, we only need to specify the increments Δh_j; see (3.4). We let $\Delta h_j = \Delta t_j f(U_j;\theta)$ and list below some possible choices for $f(\cdot)$:

$$\Delta h_j(\theta) = (t_j - t_{j-1}) \exp(\theta' U_j), \tag{3.5}$$

$$\Delta h_j(\theta) = (t_j - t_{j-1}) \exp(\theta' U_j) \Big/ \sum_{k=1}^{n} \exp(\theta' U_k), \tag{3.6}$$

$$\Delta h_j(\theta) = (t_j - t_{j-1}) \big/ \{1 + \exp(\theta' U_j)\}. \tag{3.7}$$

Obviously it always holds that $\Delta h_j > 0$ provided $t_j > t_{j-1}$. When $\theta = 0$, $\Delta h_j = (t_j - t_{j-1})$, i.e. no time deformation is involved. (3.6) may be viewed as a normalized version of (3.5), and is the version used in Stock (1988), and Ghysels and Jasiak (1995). Furthermore it satisfies the condition that $\sum_j \Delta h_j = \sum_j (t_j - t_{j-1})$ when t_j are equally spaced. We prefer (3.5) simply for its simpler form, which may be advantageous when we search for the value of θ via solving a nonlinear optimization problem.

3.3.3 Probabilistic properties

In this section we look at the moment properties of the proposed model and check them against empirical stylized features of high frequency returns discussed in Section 3.2. Also we search for the properties of the model that could be used to identify the parameters in the time deformation function. Let $Z(t)$ be a Lévy process on $\mathbb{R}$ with characteristic triplet (A, ν, γ); $\gamma \in \mathbb{R}$ is a drift, $A > 0$ is diffusion coefficient and ν is a Lévy measure. The characteristic triplet of Brownian motion is $(1, 0, 0)$ and a compensated compound Poisson process with intensity λ has a triplet $(0, \lambda f(x), 0)$ where $f(x)$ is a probability density function of jump size. Assuming a finite moment condition $E|Z_t|^m < \infty$ ($\Leftrightarrow \int |z|^m \nu(dz) < \infty$), denote $\kappa_{t,m}$, a $m-$th cumulant of a

Lévy process $Z(t)$. This is given by the Lévy - Khinchine representation theorem,

$$\begin{aligned}
\kappa_{t,1} &= t\left(\gamma + \int_{|z|\geq 1} z\nu(dz)\right), \\
\kappa_{t,2} &= t\left(A + \int_{-\infty}^{\infty} z^2\nu(dz)\right), \\
\kappa_{t,j} &= t\left(\int_{-\infty}^{\infty} z^j\nu(dz)\right), j \geq 3.
\end{aligned} \tag{3.8}$$

Note that cumulants of a Lévy process is proportionate to t. Let θ^* be a true parameter and define

$$\xi_j(\theta) = Z\{\Delta h_j(\theta^*)\}/\sqrt{\Delta h_j(\theta)}. \tag{3.9}$$

Define m-th cumulant of $Z(1)$ by κ_m.

LEMMA 1. *Assume (3.1) and (3.2) and let the deformed time modeled by (3.5). Assume a simplified case when $\mu(t) = 0$, $\kappa_1 = 0$ and observations are equally spaced, $\Delta t_j = 1/n$, $\forall j$. Then the conditional variance, coefficient of skew (C_3) and kurtosis (C_4) of log return series are given by*

$$\begin{aligned}
&\mathrm{Var}\left[\Delta X(t_j)|\Delta h_j\right] = \Delta h_j \kappa_2 \\
&C_3\left[\Delta X(t_j)|\Delta h_j\right] = \{\Delta h_j\}^{-1/2}\kappa_3/\kappa_2^{3/2}, \quad C_4\left[\Delta X(t_j)|\Delta h_j\right] = \{\Delta h_j\}^{-1}\kappa_4/\kappa_2^2.
\end{aligned}$$

It holds that

$$E\left[\xi_j(\theta)^2|\Delta h_j\right] = \kappa_2 \exp\{U_{j-1}(\theta^* - \theta)\}. \tag{3.10}$$

PROOF Under the true parameter,

$$\mathrm{Var}\left[\xi_j(\theta^*)|\Delta h_j\right] = \mathrm{Var}[Z(1)],$$

where $\mathrm{Var}[Z(1)] = A + \int_{-\infty}^{\infty} z^2\nu(dz)$ using the Lévy-Khinchine characterization theorem in (3.8). The cumulant generating function for $\xi_j(\theta)$ is given by

$$K(s) := \ln E\left[\exp\{is\xi_j(\theta)\}|\Delta h_j\right] = \Delta h_j(\theta^*)\varphi(\frac{s}{\sqrt{\Delta h_j(\theta)}}),$$

where φ is a characteristic exponent of Lévy process $Z(t)$. Then the $m-$th cumulant for $\xi_j(\theta)$ is given by

$$\frac{1}{i^m}\frac{d^m K(s)}{ds^m}|_{s=0} = \frac{\Delta h_j(\theta^*)}{\Delta h_j(\theta)^{m/2}}\frac{\varphi^{(m)}(0)}{i^m}. \tag{3.11}$$

By definition $\frac{\varphi^{(m)}(0)}{i^m} = \kappa_m$, the $m-$th cumulant of $Z(1)$. Letting $m = 2$ we have the result. ■

What Lemma 1 says is that the conditional coefficient of skew and kurtosis of log returns are proportionate to the coefficient of skew and kurtosis of the background Lévy process at time 1, $Z(1)$. As $\Delta h \to 0$, the coefficient of skew and kurtosis become large, which is consistent with the observed features of high frequency data shown in Table 3.1. When $Z(t)$ is a Brownian motion, time changing introduces an excess kurtosis. When $Z(t)$ has a skewed and thick tailed distribution, the time deformation introduces a time dependency in skewness and kurtosis. Under the true parameter, the variance of the scaled returns, $\xi_j(\theta^*)$ is constant and equals the variance of $Z(1)$. In the later section, we devise a method to estimate θ exploiting the relationship in (3.10).

Furthermore, we note that the m-th moment of $\xi_j(\theta^*)$, for $m \geq 3$, depends on time through Δt_j and U_j, unless $\kappa_m = 0$ for $m \geq 3$. If we consider U_j as a fixed covariate, then $Z\{h(t_j)\} - Z\{h(t_{j-1})\}$ is an independent sequence for any increasing $h(\cdot)$. Likewise $\xi_j(\theta)$ is an independent sequence regardless of the value of θ. This suggests that we cannot identify θ^* by testing for the independence of $\xi_j(\theta)$ if we regard U_j as deterministic.

Unconditional properties

In this section, we examine the model properties unconditionally of $\{U_j\}$'s.

Corollary 1 *Assume (3.1) and (3.2). Assume a simplified case when $\mu(t) = 0$, $\kappa_1 = 0$ and observations are equally spaced, $\Delta t_j = 1/n$, $\forall j$. Let denote the deformed time $h_t = \int_0^t \tau_s ds$, where $\tau_t = f(U_t;\theta)$ represents an instantaneous time change with $\mu_\tau := E(\tau_t), \omega := var(\tau_t)$ and $\gamma_h := \mathrm{corr}(\tau_t\tau_{t-h})$. Define $\gamma_t^* = \int_0^t \gamma_u du$ and*

$\gamma_t^{**} = \int_0^t \gamma_u^* du$. *Then*

$$
\begin{aligned}
&E\left[\Delta X(t_j)^s\right] = \mu_\tau \kappa_s / n, s = 2, 3, \\
&E\left[\Delta X(t_j)^4\right] = \mu_\tau \kappa_4 / n + 3\kappa_2^2 \{2\omega^2(\gamma^{**}_{\frac{j}{n}} - \gamma^{**}_{\frac{j-1}{n}}) + (\mu_\tau/n)^2\}, \\
&\mathrm{Cov}\left[\Delta X(t_j)^2 \Delta X(t_{j+s})^2\right] = \kappa_2^2 \mathrm{Cov}[\tau_{t_j} \tau_{t_{j+s}}] = \kappa_2^2 \omega^2 \left(\gamma^{**}_{\frac{s+1}{n}} - 2\gamma^{**}_{\frac{s}{n}} + \gamma^{**}_{\frac{s-1}{n}}\right) \\
&= O(n^{-2}).
\end{aligned}
$$

Let $M_j(s)$ denote the moment generating function of U_j. When the discretized time change Δh_j is modeled by (3.5), then it holds that

$$E[\xi_j(\theta)^2] = \kappa_2 M_j(\theta^* - \theta). \tag{3.12}$$

See Barndorff-Nielsen and Shephard (2006) Proposition 2 and 5. ■

The Corollary 1 states that the serial correlation in squared returns are driven by the serial correlation in instantaneous time change. Our model achieves stochastic volatility through time deformation. The model accommodates stylized features of high frequency returns : skewness, excess kurtosis, and serial correlation in squared returns. To interpret the result in (3.12), consider the case $M_j(s) = g(s)$ for some $s \neq 0$ where $g(s)$ is a well defined function. Then the second moment of $\xi_j(\theta)$ is constant for all θ s.t. $\theta = \theta^* + s$. Such example is when $\{U_j\}$ is a stationary AR(1). On the other hand, if we have the case that $M_j(s) = g(j, s), \forall s$ i.e., the moments of activity variables depend on time, then $\xi_j(\theta)$ has a constant second moment only when $\theta = \theta^*$. Such case is when activity variables follow a random walk. This suggests that estimating θ by testing for a constant variance exploiting the relation in (3.12) may work better with non-stationary $\{U_j\}$'s. Also if $\{U_j\}$ is assumed to be stochastic then $Z\{h_j\}$ has an independent increment property if and only if $\{h_j\}$ has an independent increment property. In our model this holds if and only if it holds that $\{U_j\}$'s are independent of $\{U_k\}$ for $\forall j \neq k$ which is empirically unrealistic.

3.4 Estimation methods

In this section we propose set of estimation methods for the parameter in the time deformation function. We progressively relax the assumption on the underlying process. The key quantity of interest is,

$$\xi_j(\theta, \mu) = (Y_j - Y_{j-1})\Big/\sqrt{\Delta h_j(\theta)} - \mu_j \sqrt{\Delta h_j(\theta)}. \tag{3.13}$$

When (θ, μ) take the true value, $\xi_j(\theta, \mu)$ should have mean value 0 and a constant variance σ^2. When we assume that log price is driven by Brownian motion, we can write down the gaussian likelihood function to maximize over the parameter θ. In this case, ξ_j is identically distributed and we may exploit this property to design a cost function to estimate θ. For weaker conditions, we use the property that ξ_j is serially uncorrelated. We lastly propose a distribution free estimation method using the fact that ξ_j has a constant variance.

3.4.1 Maximum likelihood estimation

If $Z(t)$ is a Brownian motion, ξ_j in (3.13) are $N(0, 1)$. This leads to the (negative) log likelihood function

$$l(\theta, \{\mu_j\}, \sigma^2) = n \log(\sigma^2) + \sum_{j=1}^{n} \big[\log\{\Delta h_j(\theta)\} + \frac{1}{\sigma^2 \Delta h_j(\theta)} \{\Delta X_{t_j} - \mu_j \Delta h_j(\theta)\}^2\big]. \tag{3.14}$$

For the Orstein-Uhlenbeck process, $\mu_j = aY_{j-1} + b$. If we take $\Delta h_j(\theta)$ as defined in (3.5), the above likelihood function is reduced to

$$l(\theta, a, b, \sigma^2) = n \log(\sigma^2) + \sum_{j=1}^{n} \big[\theta' U_j + \frac{1}{\sigma^2 \Delta h_j(\theta)} \{\Delta X_{t_j} - (aY_{j-1} + b)\Delta h_j(\theta)\}^2\big]. \tag{3.15}$$

The MLE $(\widehat{\theta}, \widehat{a}, \widehat{b}, \widehat{\sigma}^2)$ is the value which minimizes $l(\theta, a, b, \sigma^2)$. For any fixed θ, the minimizers $a = a(\theta)$, $b = b(\theta)$ and $\sigma^2 = \sigma^2(\theta)$ can be obtained explicitly from (3.15). Hence the MLE $\widehat{\theta}$ can be obtained by minimizing the profile likelihood $l(\theta) \equiv$

$l\{\theta, a(\theta), b(\theta), \sigma^2(\theta)\}$.

For the cases with deterministic $\mu(\cdot)$, we may first replace μ_j in (3.14) by, for example, kernel (or moving-average) smoothing estimators $\widetilde{\mu}_j$ based on V_j. We may simply use a 8-point moving average estimators

$$\hat{\mu}_j = \frac{1}{4}(V_{j-1} + V_j) + \frac{1}{8}(V_{j-2} + V_{j+1}) + \frac{1}{12}(V_{j-3} + V_{j+2}) + \frac{1}{24}(V_{j-4} + V_{j+3}).$$

Then the quasi-MLE for (θ, σ^2) is obtained by minimizing the 'profile' likelihood

$$l(\theta, \sigma^2) \equiv l(\theta, \{\widetilde{\mu}_j\}, \sigma^2).$$

When $Z(t)$ is not a Brownian motion, the above method may be viewed as a version of quasi-MLE. However, since ξ_j then are not identically distributed, the behavior of such an estimator needs to be examined carefully.

3.4.2 Parametric test for i.i.d

We still assume that $Z(t)$ is a Brownian motion and propose an estimation method that exploits the independent increment property of ξ_j. To simplify the statement, we deal with the Orstein-Uhlenbeck process first. Let

$$\xi_j(\theta, a, b) = (Y_j - Y_{j-1})/\sqrt{\Delta h_j(\theta)} - (aY_{j-1} + b)\sqrt{\Delta h_j(\theta)}. \tag{3.16}$$

When (θ, a, b) take the true value, $\xi_j(\theta, a, b)$ should have mean value 0 and a constant variance σ^2. Hence we may set $n^{-1}\sum_{j=1}^n \xi_j(\theta, a, b) = 0$. This implies that

$$b = b(\theta, a) \equiv \frac{\sum_{j=1}^n (Y_j - Y_{j-1})/\sqrt{\Delta h_j(\theta)} - a\sum_{j=1}^n Y_{j-1}\sqrt{\Delta h_j(\theta)}}{\sum_{j=1}^n \sqrt{\Delta h_j(\theta)}}. \tag{3.17}$$

Let $\xi_j \equiv \xi_j(\theta, a) = \xi_j\{\theta, a, b(\theta, a)\}$.

Put $\omega_k = 2k\pi/n$, $n_1 = [n/2]$. For $k = 1, \cdots, n_1$, define

$$I(\omega_k; \theta, a) = \frac{1}{n} \Big| \sum_{j=1}^{n} \xi_j(\theta, a) e^{-ij\omega_k} \Big|^2 = \widehat{\gamma}(0) + 2 \sum_{j=1}^{n-1} \widehat{\gamma}(j) \cos(j\omega_k)$$

$$U_k(\theta, a) = \frac{\sum_{\ell=1}^{k} I(\omega_\ell; \theta, a)}{\sum_{j=1}^{n_1} I(\omega_j; \theta, a)},$$

where $\widehat{\gamma}(j) = n^{-1} \sum_{k=1}^{n-j} (\xi_k - \bar{\xi})(\xi_{k+j} - \bar{\xi})$, and $\bar{\xi} = n^{-1} \sum_{k=1}^{n} \xi_k$. Under the assumption that ξ_j are i.i.d. normal, $U_1, \cdots, U_{n_1-1}$ are distributed as the order statistics of a random sample of size $(n_1 - 1)$ from the uniform distribution on the interval $(0, 1)$; see, for example, Proposition of 10.2.1 of Brockwell and Davis. (1991) Therefore we may search for a monotonic $h(\cdot)$ which minimizes a Cramér-von Mises type of goodness-of-fit statistic:

$$D(\theta, a) \equiv \sum_{k=1}^{n_1-1} \big(\frac{k}{n_1} - U_k(\theta, a) \big)^2. \tag{3.18}$$

Next, We can also exploit a weaker property that ξ_j are serially uncorrelated by minimizing Ljung - Box statistic:

$$D(\theta, a) \equiv n(n+2) \sum_{j=1}^{n-1} \frac{\widehat{\gamma}(j)^2}{n-j}. \tag{3.19}$$

3.4.3 Permutation-like test for constant variance

In this section we propose a estimation method for θ using only the first two moment properties of $\xi_j(\theta, \mu)$: it has mean value 0 and a constant variance σ^2. Since σ^2 is unknown, we partition the index set $\{1, \cdots, n\}$ into two halves I_1 and I_2. The difference of the sample variance between the two half samples may be measured as

$$D(I_1, I_2; \theta, a) = \Big| \frac{1}{|I_1|} \sum_{j \in I_1} \xi_j(\theta, a)^2 \Big/ \frac{1}{|I_2|} \sum_{j \in I_2} \xi_j(\theta, a)^2 - 1 \Big|,$$

Let $\mathcal{I}$ be a collection of the partitions of the set $\{1, \cdots, n\}$ into two subsets of the equal size $n/2$ if n is even, and of the sizes $(n+1)/2$ and $(n-1)/2$ if n is odd. The estimator $(\widehat{\theta}, \widehat{a})$ is defined as the minimizer of the function

$$D(\theta, a) \equiv \frac{1}{|\mathcal{I}|} \sum_{(I_1, I_2) \in \mathcal{I}} D(I_1, I_2; \theta, a). \tag{3.20}$$

For dimension as low as 3 or 4, a grid-search method may be used to find the solution. Consequently the estimator for b is defined as $\widehat{b} = b(\widehat{\theta}, \widehat{a})$, see (3.17). The estimator for σ^2 may be defined as

$$\widehat{\sigma}^2 = \frac{1}{n - 2 - k_0} \sum_{j=1}^{n} \xi_j(\widehat{\theta}, \widehat{a})^2,$$

where k_0 is the number of components of θ.

If n is not large, we let $\mathcal{I}$ consist of all the partitions as specified above. When n is large, we let $\mathcal{I}$ have K_0 partitions, where K_0 is a large integer. We may include in $\mathcal{I}$, for example, the partition with $I_1 = \{1, \cdots, n/2\}$, or all the odd numbers not greater than n. The other partitions may be selected randomly as follows: generate random variables η_i, $1 \leq i \leq (n+1)/2$ from the uniform distribution on $(0, 0.5)$, let $I_1 = \{[n\eta_i] + 1 : 1 \leq i \leq (n+1)/2\}$, where $[x]$ denotes the integer part of x, i.e. $x = [x] + r$ for some $r \in [0, 1)$.

Remark. (i) One added advantage for using this method is that we do not need to estimate σ^2 as far as the estimation for the deformation function is concerned.

(ii) With deterministic $\mu(\cdot)$, the above method still applies with plug-in estimators $\widetilde{\mu}_j$, i.e. we replace (3.16) by

$$\xi_j \equiv \xi_j(\theta) = (Y_j - Y_{j-1}) \Big/ \sqrt{\Delta h_j(\theta)} - \widetilde{\mu}_j \sqrt{\Delta h_j(\theta)}.$$

3.5 Numerical illustration

The numerical study is carried out in a following way. We create equally spaced database of high frequency prices aligned on the 5 second and 30 second fixed time grid. In our model there are two sets of parameters - parameters for the time deformation function and the another set for the homogenous Lévy process. To simulate the intra-day return we need to have a realistic values of these parameters. We carry out 2 stage estimation of the model. First the time deformation function is estimated by quasi-MLE described in Section 3.4.1. Denote the estimated parameter by θ_0 and the estimated deformed time by $\Delta h_0 = \Delta t f(U, \theta_0)$ from (3.5)-(3.7). Estimate the OU drift term by (3.17) and denote it as $\hat{\mu}$. Then the de-meaned return series given by $\Delta X - \hat{\mu}\Delta h_0$ should be distributed as $Z\{\Delta h_0\}$ according to (3.4). Next step is to estimate the parameters for the Lévy density. This problem is equivalent to estimating parameters for a homogenous Lévy process that is unequally spaced. For homogenous Lévy process $Z(t)$, we consider Brownian motion, Merton jump diffusion and NIG process,

$$
\begin{aligned}
&Z(t) = \sigma W(t), \qquad Z(t) = \sigma W(t) + \sum_{i=0}^{N(t)} J_i, J_i \sim NID(\mu_J, \sigma_J) \\
&Z(t) = \mu_Z T(t) + \sigma W\{T(t)\},
\end{aligned}
$$

where $T(t)$ has an Inverse Gaussian distribution. Since these three Lévy processes permit closed form pdfs, we maximize the exact likelihood to estimate the Lévy parameters and simulate K instances of $Z(t)$ given the estimates. The above procedure gives us the parameter values from which we can simulate the return process, ΔX by $\hat{\mu}\Delta h_0 + Z\{\Delta h_0\}$. For each instances of simulation, the parameters for the time deformation function, $\theta_k, k = 1, \cdots, K$ are estimated by quasi-MLE (Section 3.4.1), the nonparametric method (Section 3.4.3) and the test statistic based on IID test (Section 3.4.2) for comparison. The finite sample properties of the estimators are examined by the mean absolute deviation of θ_k from θ_0.

Under different data generating processes considered and at different sampling frequencies, Table 3.2 and Table 3.3 show that the quasi-MLE is most accurate and efficient. The nonparametric estimation method performs better than the the method based on testing for a serial correlation in squared returns or testing for i.i.d. The later methods break down when activity variables are stationary. Even though the nonparametric method is less efficient than quasi-MLE in most of scenarios, when underlying process is a pure jump process without diffuse component, it outperforms quasi-MLE. We conclude that the finite sample property of quasi-MLE is better than any of the proposed methods. The nonparametric permutation test does better than other parametric tests and it is not as much affected by stationarity of U_j's. For pure jump processes, there is a weak evidence that permutation test outperforms the quasi-MLE.

3.6 An extension: dealing with many time series

In most practical situations time deformation is considered not only for one time period. For example, we may be interested in deforming annual sales for several years, or interday returns over many days. If there are reasons to believe that the time deformation remains about the same over those repeated periods, we may use the deformation function in more general form. For example, we may assume that $\theta = \theta(t_j)$, i.e. θ varies with respect to time t. Then the kernel smoothing may be applied to (3.14) or (3.15) to estimate $\theta(t)$ by pulling together the data from different periods (i.e. different years or different days). The local quasi likelihood function is given by

$$l(\theta) = \sum_{d=1}^{D} \sum_{j=1}^{n} l_{j,d}(\theta) K(\frac{t_j - t}{h}), \tag{3.21}$$

$$l(\theta) = \sum_{d=1}^{D} \sum_{j=1}^{n} l_{j,d}(\Delta h_j) K(\frac{t_j - t}{h}).$$

This give us $\widehat{\theta} = \widehat{\theta}(t)$ and $\widehat{\Delta h_j}$ respectively. The Figure 3.7 reports the result for maximizing the local likelihood (3.21) using the log number of transactions and the log volume as explanatory variables. We carried out a local estimation of $\theta(t)$ for every 50 observations. The fitted increment of the deformed time show an expected U-shaped pattern.

3.7 Empirical Application

In this section we apply the proposed model to a real data. We analysed J P Morgan Co. (ticker: JPM) transaction data traded on NYSE during 20 days period in March 2007. The period is selected to represent a benchmark market condition since the period following had a sharp peak in volatility and preceding was characterized by low volatility. To check if the proposed model given by (3.1)-(3.2) and (3.5)-(3.7) describes the high frequency data well, we use the following property of the model. Let non-decreasing function $T(t)$ such that $h(T(t)) = t$. This is the inverse of deformed clock:

$$T(s) := \inf\{t|h(t) \geq s, s \in \mathbb{R}^+\}.$$

Then it holds that

$$X\{T(t)\} = Y(t).$$

Once the deformed clock $h(t)$ is estimated, the returns spaced by $T(t)$ should possess an independent and stationary increment property according to (3.1)-(3.2). The Figure 3.5 and Figure 3.6 show that stock returns spaced by the equal elapse of the deformed clock show less serial correlation. We compared the proposed method to quadratic variation estimate, i.e. $h(t) = [X]_t$ and conclude that the proposed method better recovers homogeneity of stock returns for the samples examined. Table 3.4 reports homogeneity statistics applied on the returns spaced by the deformed time using local quasi-MLE, quasi-MLE, nonparametric permutation test and the Realized Variance. To test for the homogeneity we consider the i.i.d test given in (3.18), the permutation test in (3.20) and portmanteau test in (3.19). In exception of the portmanteau test statistic applied to the densely sampled data, the various tests indicate that the proposed modeling framework yields more homogenous process than using the Realized Variance as a proxy for deformed time. The results also suggest that pulling many time series and estimating the deformed time locally outperforms estimating daily series separately.

3.8 Application to optimal order execution

How to optimally execute a large trade is an important issue for practitioners. Lets describe the bid side of the order book at time t by $\{B_i(t), Q_i(t)\}_{i\geq 1}$ where $B_i(t)$ is a bid price and $Q_i(t)$ is an associated volume in terms of number of shares. At each point of time, it holds that $B_1(\cdot) > B_2(\cdot) > \cdots$ therefore $B_1(\cdot)$ is a best bid. When a trader place a large market sell order of quantity Q at time t, the prices that the order is filled is $\{B_i(t), 1 \leq i \leq K\}$ for each associated $Q_i(t)$ until it holds that $Q = \sum_{j=1}^{K} Q_j(t)$. After execution of the trade, the bid price move down to $B_K(t)$. The market order is filled immediately at the cost of the market impact given by $B_1(t) - B_K(t)$. If the investor is willing to wait in order to minimize the market impact, then the large order can be broken down and executed over a longer time horizon in some optimal fashion. The benchmark for such order execution strategy is the Volume Weighted Average Price (VWAP) which is defined as follows. Given $\{\tau_j\}$ a transaction time stamp, bid-VWAP is defined by

$$\frac{\sum_{j=1}^{N} Q_1(\tau_j) B_1(\tau_j)}{\sum_{\ell=1}^{N} Q_1(\tau_\ell)}.$$

Then VWAP tracking trading strategy is ex-ante strategy that best predicts the volume profile of the instrument to be traded over the relevant horizon. The strategy should deliver the price that is close to ex-post VWAP. We may design an order execution strategy based on the fitted deformation time and compare it with the ex-post VWAP to see which strategy delivers the better execution price. We can implement the strategy in two different ways. First way is to time the trade with equal elapse of fixed clock time where the quantity bought or sold is proportionate to the deformed time progression. The difference from the VWAP strategy is that the order quantity is set according to the time deformation schedule rather than the volume schedule. The second strategy is to time the trade with the equal elapse of the deformed time. We trade equal amount for each trade but more frequently when the financial clock moves fast and vice versa. We may impose an end condition by pre-determining how many times to trade per day and divide the financial clock into equal intervals. The superior order execution strategy should deliver a better price; the order size weighted prices are higher than the bid-VWAP for a sell order and lower than the ask-VWAP for a buy order.

Figure 3.1: Time series plot of transaction marks

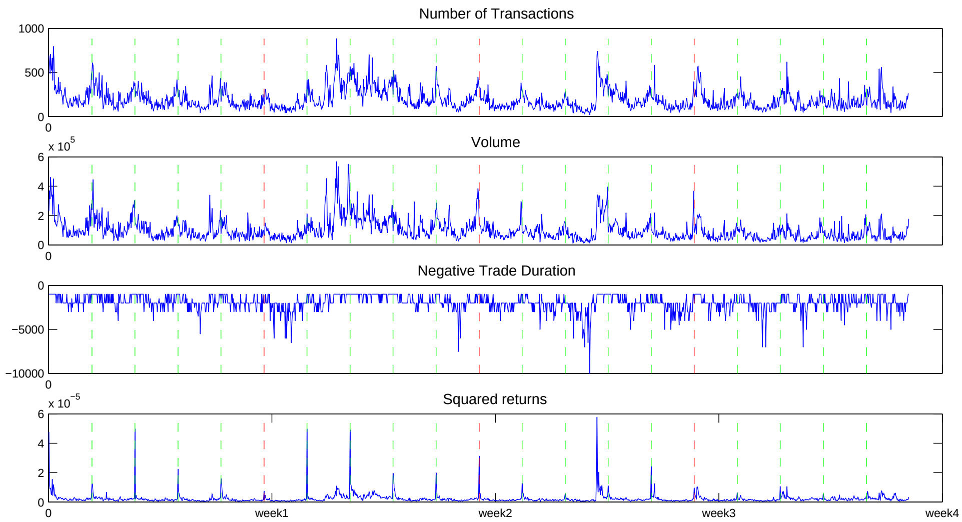


Figure 3.2: Candidate variables for modeling the financial clock : cumulative sum of activity variables for each day

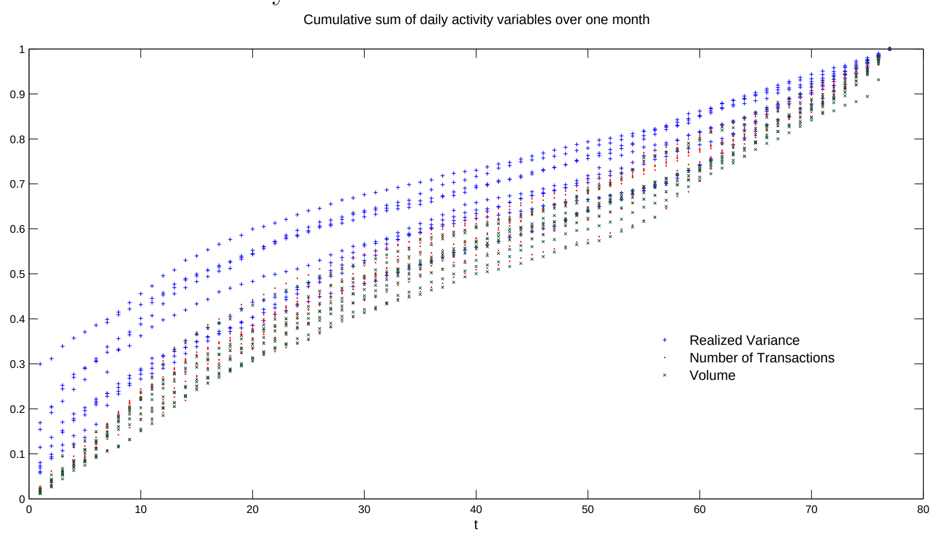


Figure 3.3: Sample autocorrelaiton function of activity variables

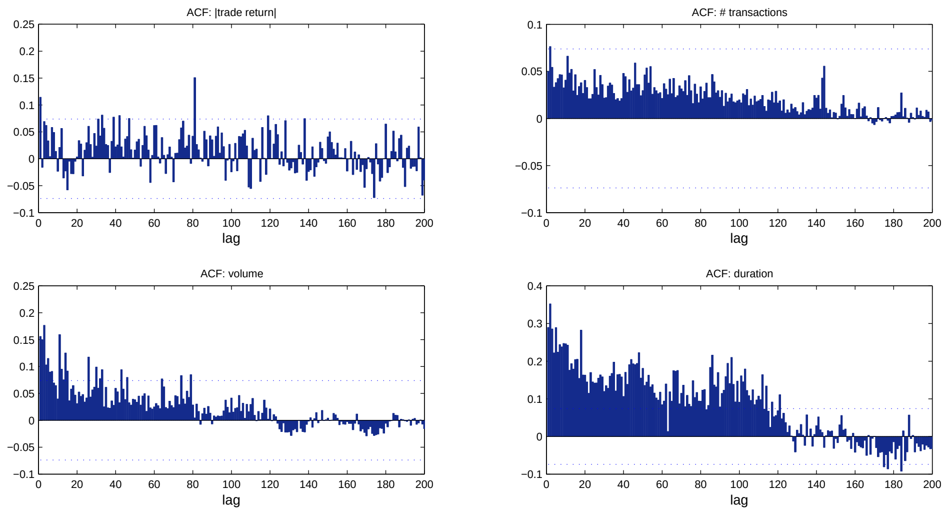


Figure 3.4: Sample cross - correlation function of activity variables

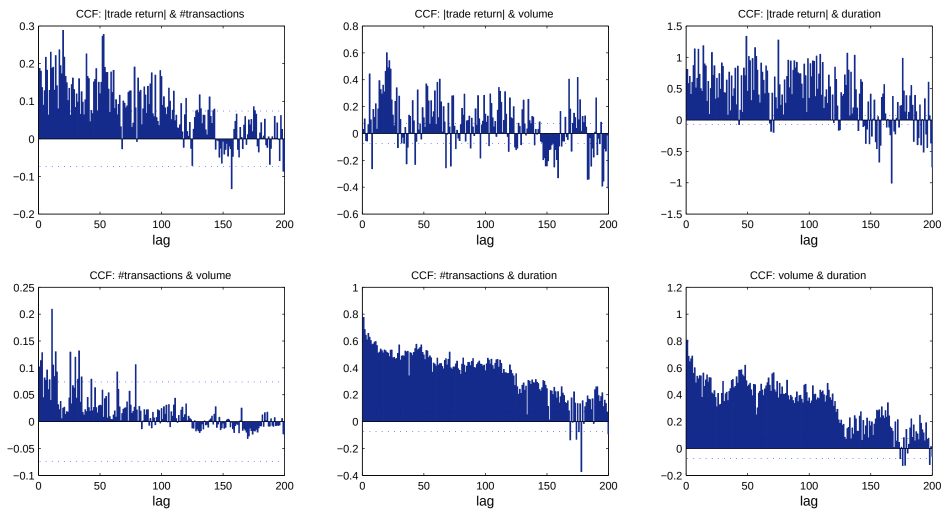


Figure 3.5: Return spaced by equal elapse of financial time, $n = 100$

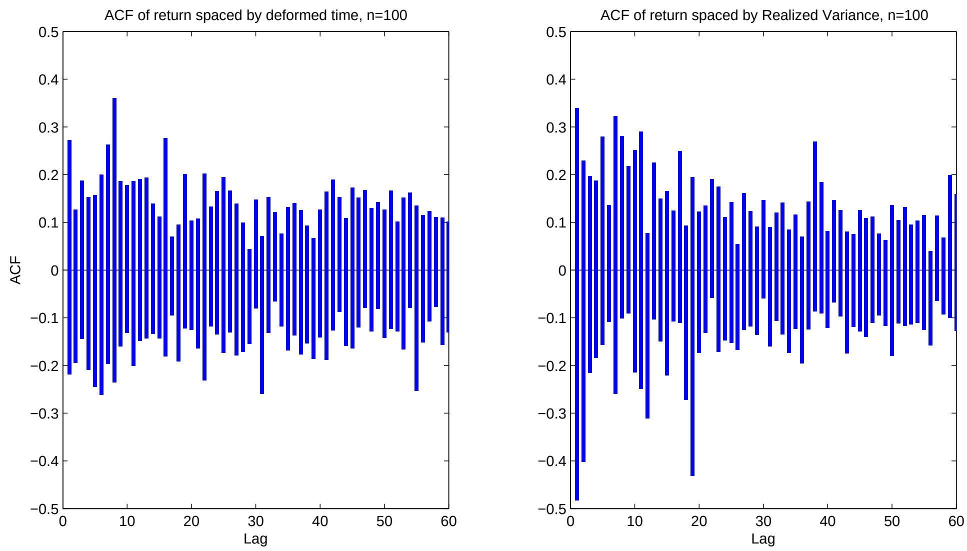


Figure 3.6: Return spaced by equal elapse of financial time, $n = 500$

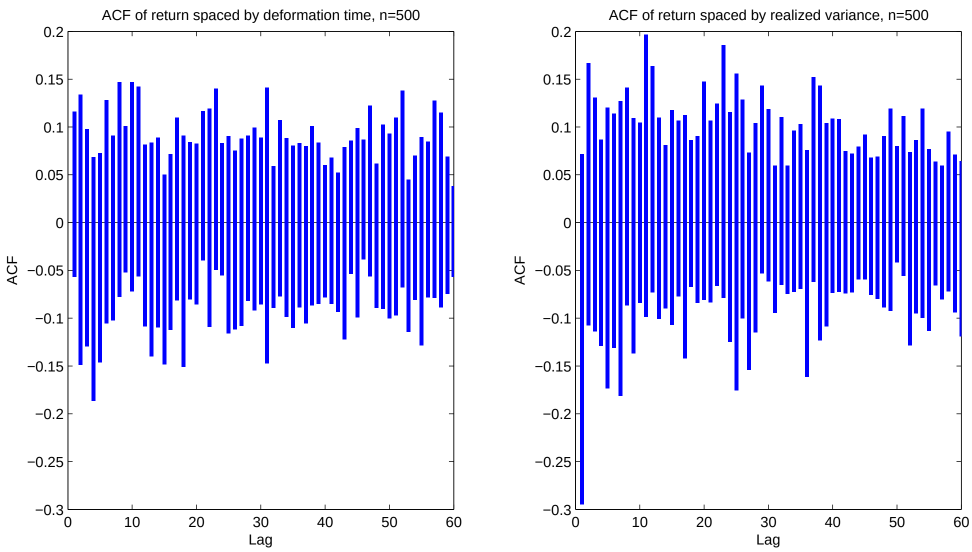


Figure 3.7: Local likelihood estimates of time deformation function

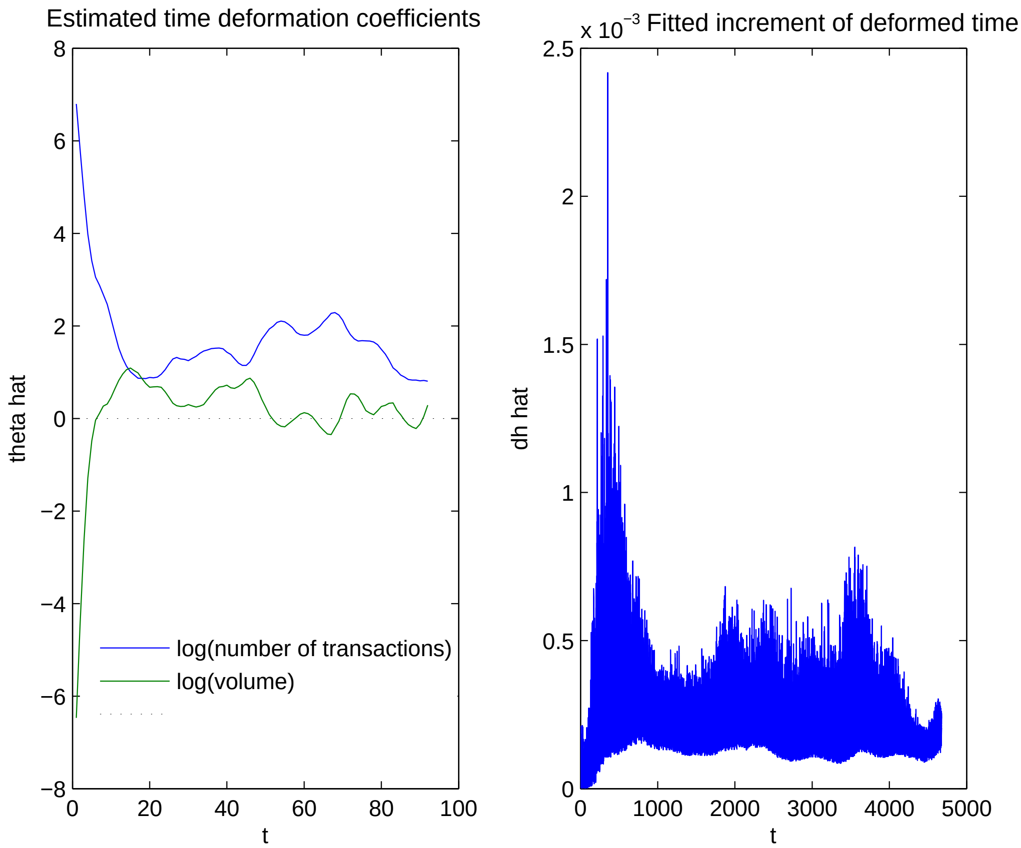


Table 3.2: Accuracy of test statistics: 30 second dataset

Z	U_j	$f(\cdot)$	qMLE					Permutation Test			Portmantau Test			iid Test		
			θ	Var(θ)	MAE	σ	Var(σ)	θ	Var(θ)	MAE	θ	Var(θ)	MAE	θ	Var(θ)	MAE
NID(0,1)	NumTran	expOrg	3	-	-	1.5E-02	0.0E+00	3	-	-	7.5	8.4	3.3	7	6.2	2.7
		expNorm	3	-	-	1.5E-02	0.0E+00	3	-	-	7.5	8.4	3.3	7	6.2	2.7
	Volume	expOrg	3	-	-	1.5E-02	0.0E+00	3	-	-	10	-	7.0	10	-	7.0
		expNorm	3	-	-	1.5E-02	0.0E+00	3	-	-	10	-	7.0	10	-	7.0
	Dur	expOrg	-4	-	-	1.5E-02	0.0E+00	-4	0.05	0.1	-3.5	4.6	1.3	-4	3.7	1.0
		expNorm	-4	-	-	1.5E-02	0.0E+00	-4	0.05	0.1	-3.5	4.6	1.3	-4	3.7	1.0
t dn	NumTran	expOrg	3	0.05	0.10	4.0E-02	1.5E-04	3	0.83	0.7	7	5.7	3.7	7.5	6.8	3.9
2 d.f.		expNorm	3	0.08	0.20	4.0E-02	1.8E-04	3	1.08	0.8	7	5.7	3.7	7.5	6.8	3.9
	Volume	expOrg	3	-	-	3.3E-02	8.7E-04	3	1.20	0.6	10	18.1	4.5	10	16.9	4.4
		expNorm	3	-	-	3.3E-02	8.7E-04	3	0.95	0.5	10	18.1	4.5	10	16.9	4.4
	Dur	expOrg	-4	0.43	0.50	5.8E-02	1.3E-03	-4.5	6.50	2.1	-4.5	2.3	1.4	-4.5	8.6	2.1
		expNorm	-4	0.43	0.50	5.8E-02	1.4E-04	-4.5	5.80	2.0	-4.5	2.3	1.4	-4.5	8.6	2.1
Jump	NumTran	expOrg	3	0.20	0.30	7.5E-03	2.0E-04	3	0.50	0.5	3.5	8.7	1.7	3	17.0	3.1
Diffusion		expNorm	3.5	0.05	0.60	1.8E-02	1.5E-04	4	5.30	1.8	4	6.2	2.1	4.5	10.7	3.2
	Volume	expOrg	3	0.20	0.30	7.5E-03	8.0E-04	3	4.70	1.1	10	9.8	5.6	10	9.8	5.6
		expNorm	3.5	1.25	0.80	1.5E-02	1.9E-06	3.5	4.00	1.4	10	12.8	5.8	10	12.8	5.8
	Dur	expOrg	-4	-	-	1.5E-02	0.0E+00	-4	-	-	-5	5.0	2.1	-3.5	0.6	0.7
		expNorm	-5	0.20	0.70	1.5E-02	1.3E-06	-5.5	8.58	2.2	-4.5	7.4	2.2	-4	4.8	1.6
NIG	NumTran	expOrg	4	0.20	0.80	4.3E-02	1.2E-03	5	0.83	1.8	4	12.1	3.2	4	14.6	3.3
		expNorm	3	0.38	0.40	1.5E-02	8.1E-06	2	0.18	1.1	4	6.3	2.1	3	9.8	2.0
	Volume	expOrg	3.5	0.05	0.40	1.8E-02	3.8E-05	4	3.05	1.3	10	-	7.0	10	-	7.0
		expNorm	3.5	0.43	0.50	1.8E-02	1.4E-05	1.5	1.83	1.3	10	-	7.0	10	-	7.0
	Dur	expOrg	-5.5	-	1.50	7.5E-03	9.4E-37	-5.5	0.13	1.5	-8	0.3	3.6	-7	0.4	3.0
		expNorm	-5	1.68	1.40	1.8E-02	8.1E-06	-2	5.33	2.8	-8	19.6	4.3	-8	1.2	3.4

True θ value is 3 for U_j = number of transactions and volume, -4 for duration

Table 3.3: Accuracy of test statistics: 5 second dataset

Z	U_j	$f(\cdot)$	qMLE					Permutation Test			Portmantau Test			iid Test		
			θ	Var(θ)	MAE	σ	Var(σ)	θ	Var(θ)	MAE	θ	Var(θ)	MAE	θ	Var(θ)	MAE
NID(0,1)	NumTran	expOrg	9	-	-	1.5E-02	0.0E+00	9	-	-	9	24.2	2.2	9	37.5	4.5
		expNorm	9	-	-	1.5E-02	0.0E+00	9	-	-	9	24.2	2.2	9	37.5	4.5
	Volume	expOrg	9	-	-	1.5E-02	0.0E+00	9	-	-	-2	-	11.0	-2	36.3	6.6
		expNorm	9	-	-	1.5E-02	0.0E+00	9	-	-	-2	-	11.0	-2	36.3	6.6
t dn 2 d.f.	NumTran	expOrg	9	0.45	0.30	4.8E-02	6.1E-04	9	5.70	1.8	9	0.4	0.5	9	24.8	2.3
		expNorm	9	0.80	0.40	4.8E-02	6.1E-04	9	6.43	1.9	9	0.4	0.5	9	24.8	2.3
	Volume	expOrg	9	0.05	0.10	4.8E-02	2.2E-04	8.5	1.93	1.1	8.5	56.0	5.7	8.5	23.2	2.4
		expNorm	9	0.05	0.10	4.8E-02	2.2E-04	8.5	1.93	1.1	8.5	56.0	5.7	8.5	23.2	2.4
Jump Diffusion	NumTran	expOrg	9	-	-	7.5E-03	9.4E-37	9	0.05	0.1	-2	36.3	6.6	-2	24.2	8.8
		expNorm	9	0.08	0.20	7.5E-03	3.9E-05	9.5	0.50	1.0	9	0.1	0.1	9	37.5	4.5
	Volume	expOrg	9	-	-	7.5E-03	9.4E-37	9	-	-	-2	36.3	6.6	-2	36.3	6.6
		expNorm	9	-	-	7.5E-03	9.4E-37	9	-	-	9	36.3	4.4	9	36.3	4.4
NIG	NumTran	expOrg	11.5	0.45	2.20	6.8E-02	1.2E-03	9	3.68	1.3	4	48.8	6.9	-1	20.3	7.4
		expNorm	-1.5	51.43	7.40	7.5E-02	9.1E-04	9	0.63	0.6	9	25.3	2.7	9	25.3	2.7
	Volume	expOrg	11.5	36.58	4.20	6.8E-02	5.9E-04	8	11.08	2.5	-0.5	66.4	7.9	2.5	57.0	7.1
		expNorm	10.5	30.08	3.50	1.5E-02	7.3E-04	8.5	1.93	1.3	7.5	11.4	3.4	8	38.5	4.3

True θ value is 9

Table 3.4: Homogeneity test for returns spaced by estimated financial clock

n=100		IID test	Permutation	Portmantau
Local qMLE	mean	0.19	0.49	104984
	std	0.15	0.23	52073
qMLE	mean	0.24	0.47	110526
	std	0.22	0.17	49115
Nonparametric	mean	0.24	0.63	89104
	std	0.24	0.46	58423
RV	mean	0.54	0.83	114763
	std	0.75	0.60	67966

n=500		IID test	Permutation	Portmantau
Local qMLE	mean	0.28	0.23	3234250
	std	0.25	0.05	1262004
qMLE	mean	0.25	0.31	3343456
	std	0.27	0.15	2165752
Nonparametric	mean	0.42	0.31	3962810
	std	0.48	0.17	2156484
RV	mean	0.61	0.30	2659933
	std	0.92	0.16	1933372

Bibliography

[1] Y. Aït-Sahalia, J. Fan, and D. Xiu. High frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 94:9–51, 2010a.

[2] Y. Aït-Sahalia and L. Mancini. Out of sample forecasts of quadratic variation. *Journal of Econometrics*, 147:17–33, 2008.

[3] Y. Aït-Sahalia, P. A. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18:351–416, 2005.

[4] Y. Aït-Sahalia, P. A. Mykland, and L. Zhang. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160:160–175, 2010b.

[5] D. J. Aldous and G. K. Eagleson. On mixing and stability of limit theorems. *The Annals of Probability*, 6:325–331, 1978.

[6] S. Alizadeh, M. W. Brandt, and F. X. Diebold. Range-based estimation of stochastic volatility models. *Journal of Finance*, 57:1047–1091, 2002.

[7] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. Working Paper, University of Toronto, 2001.

[8] T. G. Andersen and T. Bollerslev. Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. *Journal of Finance*, 52:975–1005, 1997.

[9] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Exchange rate returns standardized by realized volatility are (nearly) gaussian. *Multinational Finance Journal*, 4:159–179, 2000.

[10] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55, 2001.

[11] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71:579–625, 2003.

[12] T. G. Andersen, T. Bollerslev, Diebold F. X., and G. Wu. A framework for exploring the macroeconomic determinants of systematic risk. *American Economic Review*, 95:398–404, 2005.

[13] D. W. K. Andrews. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858, 1991.

[14] T. Ané and H. Geman. Order flow, transaction clock, and normality of asset returns. *Journal of Finance*, 55:2259–2284, 2000.

[15] K. Back. Asset pricing for general processes. *Journal of Mathematical Economics*, 20:371–395, 1991.

[16] T. G. Bali and L. Peng. Is there a risk-return tradeoff? *Journal of Applied Econometrics*, 21:1169–1198, 2006.

[17] F. M. Bandi and J. R. Russell. Realized covariation, realized beta and microstructure noise. Working Paper, Graduate School of Business, University of Chicago, 2005.

[18] F. M. Bandi and J. R. Russell. Microstructure noise, realized volatility, and optimal sampling. *Review of Economic Studies*, 75:339–369, 2008.

[19] F. M. Bandi, J. R. Russell, and C. Yang. Realized volatility forecasting and option pricing. *Journal of Econometrics*, 147(1):34–46, 2008.

[20] K. Bannouh, M. Martens, R. Oomen, and D. van Dijk. Realized factor models for vast dimensional covariance estimation. Working paper, University Rotterdam, 2009.

[21] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76:1481–536, 2008a.

[22] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Realized kernels in practice: trades and quotes. *Econometrics Journal*, 4:1–32, 2008b.

[23] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Multivariate realized kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162:149–169, 2011.

[24] O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64:253–280, 2002.

[25] O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High frequency covariance, regression and correlation in financial economics. *Econometrica*, 72:885–925, 2004a.

[26] O. E. Barndorff-Nielsen and N. Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2:1–37, 2004b.

[27] O. E. Barndorff-Nielsen and N. Shephard. Impact of jumps on returns and realized variances: econometric analysis of time-deformed lévy processes. *Journal of Econometrics*, 131:217–252, 2006.

[28] M. S. Bartlett. On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society, Supplement*, 8:27–41, 1946.

[29] G. H. Bauer and K. Vorkink. Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160:93–101, 2011.

[30] T. Bollerslev, M. Gibson, and H. Zhou. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics*, 160:235–245, 2010.

[31] T. Bollerslev, J. Litvinova, and G. Tauchen. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics, No. 3*, 4:353–384, 2006.

[32] T. Bollerslev and Y. B Zhang. Measuring and modeling systematic risk in factor pricing models using high-frequency data. *Journal of Empirical Finance*, 10(5):533–558, 2003.

[33] T. Bollerslev and H. Zhou. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109:33–65, 2002.

[34] M. W. Brandt and F. X. Diebold. A no-arbitrage approach to range-based estimation of return covariances and correlations. *Journal of Business*, 79:61–73, 2006.

[35] M. Britten-Jones and A. Neuberger. Option prices, implied price processes, and stochastic volatility. *Journal of Finance*, 55:839–866, 2000.

[36] C. T. Brownlees and G. M. Gallo. Financial econometric analysis at ultra-high frequency: Data handling concerns. Working Paper, 2006.

[37] C. T. Brownlees and G. M. Gallo. Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics*, pages 29–56, 2010.

[38] P. Carr, H. Geman, D. B. Madan, and M. Yor. Stochastic volatility for lévy process. *Mathematical Finance*, 13:345–382, 2003.

[39] P. Carr, H. Geman, D. B. Madan, and M. Yor. Pricing options on realized variance. *Finance and Stochastics*, 9:453–475, 2005.

[40] P. Carr and L. Wu. Variance risk premia. *Review of Financial Studies*, 22(3):1311–1341, 2009.

[41] A. Chaboud, B. Chiquoine, E. Hjalmarsson, and C. Vega. Rise of the machines: Algorithmic trading in the foreign exchange market. Board of Governors International Finance Discussion paper no 980, 2009.

[42] R. Chiriac and V. Voev. Modeling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26:922–947, 2011.

[43] K. Christensen, S. Kinnebrock, and M. Podolskij. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159:116–133, 2010.

[44] K. Christensen and M. Podolskij. Realized range-based estimation of integrated variance. *Journal of Econometrics*, 141:323–349, 2007.

[45] P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41:135–155, 1973.

[46] F. Corsi. A simple long memory model of realized volatility. *Journal of financial econometrics*, 7:174–196, 2009.

[47] K. Demeterfi, E. Derman, M. Kamal, and J. Zou. More than you ever wanted to know about volatility swaps. Technical report, 1999. Research notes, Goldman Sachs.

[48] L. Dubins and G. Schwartz. On continuous martingale. Proceedings of the National Academy of Sciences of the United States of America, 1965.

[49] E. F. Fama and K. R. French. Common risk factors in the returns to stocks and bonds. *Journal of financial economics*, 33:3–56, 1993.

[50] J. Fan, Y. Li, and K. Yu. Vast volatility matrix estimation using high frequency data for portfolio selection. Working Paper, 2010.

[51] J. Fleming, C. Kirby, and B. Ostdiek. The economic value of volatility timing using realized volatility. *Journal of Financial Economics*, 67:473–509, 2003.

[52] J. Gatheral and R. C. A. Oomen. Zero-intelligence realized variance estimation. *Finance and Stochastics*, 14:249–283, 2010.

[53] H. Geman. *Aspects of Mathematical Finance*, chapter Stochastic Clock and Financial Markets. Spinger, 2006.

[54] H. Geman, D. Madan, and M. Yor. Time changes for lévy processes. *Mathematical Finance*, 11:79–96, 2001.

[55] E. Ghysels and J. Jasiak. Stochastic volatility and time deformation: an application to trading volume and leverage effects. *Available at:http://www.cirano.qc.ca/pdf/publication/95s-31.pdf*, 1995.

[56] E. Ghysels, P. Santa-Clara, and R. Valkanov. Predicting volatility: how to get the most out of returns data sampled at different frequencies. *Journal of Econometrics*, 131:59–95, 2006.

[57] E. Ghysels and A. Sinko. Volatility forecasting and microstructure noise. *Journal of Econometrics*, 160:257–271, 2011.

[58] C. Goodhart and M. O'Hara. High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance*, 4:73–114, 1997.

[59] A. Gosh and O. Linton. Consistent estimation of the risk-return trade off in the presence of measurement error. Working paper, London School of Economics, 2007.

[60] C. Gourieroux, J. Jasiak, and R. Sufana. The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150:167–181, 2009.

[61] J. E. Griffin and R. C. A. Oomen. Sampling returns for realized variance calculations: tick time or transaction time? *Econometric Reviews*, pages 230–253, 2008.

[62] J. E. Griffin and R. C. A. Oomen. Covariance measurement in the presence of non- synchronous trading and market microstructure noise. *Journal of Econometrics*, 160:58–68, 2011.

[63] D. M. Guillaume, M. M. Dacorogna, R. R. Dave, U. A. Muller, R. B. Olsen, and O. V. Pictet. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance Stochastitcs*, 1:95–129, 1997.

[64] B. Hansen. Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica*, 60:967–972, 1992.

[65] P. R. Hansen, Z. A. Huang, and H. H. Shek. Realized garch: a complete model of returns and realized measures of volatility. *Journal of Applied Econometrics*, 2011. forthcoming.

[66] P. R. Hansen and A. Lunde. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24:127–161, 2006.

[67] T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11:359–379, 2005.

[68] T. Hayashi and N. Yoshida. Asymptotic normality of a covariance estimator for non synchronously observed diffusion processes. *Annals of the Institute of Statistical Mathematics*, 60:367–406, 2008.

[69] T. Hendershott, C. Jones, and A. Menkveld. Does algorithmic trading improve liquidity? Working Paper, 2010.

[70] T. Hendershott and R. Riordan. Algorithmic trading and information. NET Institute, working paper #09-08, 2009.

[71] A. Itkin and P. Carr. Pricing swaps and options on quadratic variation under stochastic time change models - discrete observations case. *Review of Derivative Research*, 13:141–176, 2010.

[72] Brockwell P. J. and Davis R. A. *Time series: Theory and Methods.* 1991.

[73] Hannan E. J. Linear regression in continuous times. *Journal of Australian Mathematical Society, Series A*, 19:146–159, 1975.

[74] J. Jacod. On continuous conditional gaussian martingales and stable convergence in law. Technical report, Publ. Lab. Probabilites 339, Univ. Paris-6., 1997.

[75] J. Jacod, Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications*, 119:2249–2276, 2009.

[76] J. Jacod and P. Protter. Asymptotic error distributions for the euler method for stochastic differential equations. *The Annals of Probability*, 26(1):267–307, 1998.

[77] I. Kalnina and O. Linton. Estimating quadratic variation consistently in the presence of correlated measurement error. *Journal of Econometrics*, 147:47–59, 2008.

[78] S. J. Koopman, B Jungbacker, and E. Hol. Forecasting daily variability of the s&p 100 stock index using historical, realized and implied volatility measurements. *Journal of Applied Econometrics*, 20(2):311–323, 2005.

[79] D. Kristensen. Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory*, 26:60–93, 2010.

[80] L. Kyj, B. Ostdiek, and K. Ensor. Covariance estimation in dynamic portfolio optimization: A realized single factor model. Working Paper, 2009.

[81] M. Lanne. A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics*, 4(4):594–616, 2006.

[82] Y. Li and P. A. Mykland. Are volatility estimators robust with respect to modeling assumptions? *Bernoulli*, 13(3):601–622, 2007.

[83] C. Liu and J. M. Maheu. Forecasting realized volatility: a bayesian model-averaging approach. *Journal of Applied Econometrics*, 24:709–733, 2009.

[84] Q. Liu. On portfolio optimization: How do we benefit from high-frequency data? *Journal of Applied Econometrics*, 24:560–582, 2009.

[85] Robinson P. M. Continuous time regression with discrete data. *The Annals of Statistics*, 3(3):688–697, 1975.

[86] P. Malliavin and M. E. Mancino. Fourier series method for nonparametric estimation of multivariate volatility. *The Annals of Statistics*, 37(46):1983–2010, 2009.

[87] B. Mandelbrot and M. Taylor. On the distribution of stock price differences. *Operations Research*, 15:1057–1062, 1967.

[88] Y. Matsuda and Y. Yajima. Fourier analysis of irregularly spaced data on r^d. *Journal of the Royal Statistical Society, Series B*, 71:191–217, 2009.

[89] M. McAleera and M. C. Medeirosb. Realized volatility. *A Review Econometric Reviews*, 27(1-3):10–45, 2008.

[90] I Monroe. Processes that can be embedded in brownian motion. *The Annals of Probability*, 6:42–56, 1978.

[91] P. A. Mykland and L. Zhang. Anova for diffusions and itô processes. *The Annals of Statistics*, 34(4):1931–1963, 2006.

[92] W. K. Newey and K. D. West. A simple positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708, 1987.

[93] R. Norberg. Vasiček beyond the normal. *Mathematical Finance*, 14:585–604, 2004.

[94] S. Park and O. Linton. Realized volatility: theory and application. In L. Bauwens, C. Hafner, and L. Sebastien, editors, *Volatility models and their applications, Wiley*, pages 293–312. London, 2011.

[95] S. Park and O. Linton. Estimating the quadratic covariation matrix for asynchronously observed continuous time signal masked by additive noise. Working Paper, London School of Economics, 2011a.

[96] M. Parkinson. The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53:61–65, 1980.

[97] A. J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160:246–256, 2011.

[98] R. T. Peters and R. G. de Vilder. Testing the continuous semimartingale hypothesis for the s&p 500. *Journal of Business & Economic Statistics*, 24:444–454, 2006.

[99] P. C. B. Phillips and J. Yu. A two-stage realized volatility approach to estimation of diffusion processes with discrete data. *Journal of Econometrics*, 150:139–150, 2009a.

[100] P. C. B. Phillips and J. Yu. Maxmimum likelihood estimation of continuous time models. In T. G. Andersen, R. A. Davis, J. P. Kreib, and T. Mikosch, editors, *Handbook of Financial Time Series*, pages 497–527. Springer, 2009b.

[101] P. Protter. *Stochastic Integration and Differential Equations. A New Approach.* Springer-Verlag, Berlin Heidelberg, Germany, 1990.

[102] N. Shephard and K. Sheppard. Realizing the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 25:197–231, 2010.

[103] J. H. Stock. Estimating continuous-time processes subject to time deformation. *Journal of the American Statistical Association*, 83:77–85, 1988.

[104] D. D. Thomakos and T. Wang. Realized volatility in the futures markets. *Journal of Empirical Finance*, 10:321–353, 2003.

[105] V. Todorov and G. Tauchen. The realized laplace transform of volatility. Working Paper, KGSM, 2010.

[106] V. Todorov, G. Tauchen, and I. Grynkiv. Realized laplace transforms for estimation of jump diffusive volatility models. Working Paper, KGSM, 2010.

[107] V. Voev. Dynamic modeling of large-dimensional covariance matrices. In L. Bauwens, W. Pohlmeier, and D. Veredas, editors, *In High Frequency Financial Econometrics: Recent Developments (1st edn)*, pages 293–312. Berlin, 2007.

[108] V. Voev and A. Lunde. Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics*, 5(1):68–104, 2007.

[109] L. Wu. Variance dynamics: Joint evidence from options and high-frequency returns. *Journal of Econometrics*, 2010. forthcoming.

[110] J. Yu and P. C. B. Phillips. A gaussian approach for estimating continuous time models of short term interest rates. *The Econometrics Journal*, 4:211–225, 2001.

[111] L. Zhang. Stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12:1019–1043, 2006.

[112] L. Zhang. Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160:33–47, 2010.

[113] L. Zhang, P. A. Mykland, and Y. Aït-Sahalia. A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100:1394–1411, 2005.

[114] B. Zhou. High-frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics*, 14:45–52, 1996.