Realized Volatility and Correlation for Non-synchronously Traded Financial Assets

Jin-Huei Yeh
Department of Finance
Yuan Ze University

and

Chung-Ming, Kuan
Institute of Economics
Academia Sinica

May 25, 2007
(Preliminary Draft)
Abstract

Though realized volatility has been justified by standard continuous time arguments, so far the construction of realized measures for both volatility and correlation are restricted to actively traded financial assets. We aim at constructing a bias-free realized covariance matrix for assets including both liquid and illiquid equities using high frequency data. The difficulty on how to deal with a set of non-synchronously traded intra-daily returns while at the same time correcting for the other market microstructure noises can be solved through the proposed synchronizing procedure. We use the synchronized set of recovered return values to construct realized volatility and correlations. Our empirical results show the recovering method is effective in correcting for both non-synchronous trading and other market friction effects. We also found non-synchronous trading bias plays a dominant role among the other market microstructure effects for less actively traded equities. Controlling the effect from non-synchronous trading, the effect from the other microstructure is almost negligible in the context of realized measures construction. The new recovering and filtering procedures allow one to get more precise realized measures for volatilities and correlations that are applicable to further financial applications.

JEL classification: C13, C32

Keywords: Realized volatility, high frequency, market microstructure, non-synchronous trading, MCMC
1 Introduction

The modelling of time-varying financial market volatility has long been an intensive research topic of practical relevance in finance. The inherent latency of volatility has led a large number of recent studies to advocate the use of realized volatility from the relevant information of high frequency returns. Using the summation of finely sampled intraday squared returns to approximate daily volatility has been justified by standard continuous time arguments; see for example, Andersen, et al. (2001a, b), Barndorff-Nielson and Shephard (2002a, b). Nonetheless, so far the construction of realized measures for both volatility and correlation are restricted to those actively traded financial assets. This paper aims at disentangling the realized volatility and correlation construction among a group of assets including both liquid and illiquid equities using high frequency data. The difficulty confronted is how to deal with a set of non-synchronously traded intra-daily returns sampled at a given small time interval.

We all know that asynchronous financial return series due to the effect of no-trade complicate or bias many tasks of financial management. Most importantly, the real value of the portfolio is never known at a point in time and consequently results in misleading measures, biases the profit and losses or even distorts the hedging strategy of rational agent. The correlations between returns are generally biased and lead to further inaccuracies in assessing risks or asset allocations. Hence using non-synchronous returns is potentially making systematic errors. This paper tries to resolve this problem by using the methodology of Tsay and Yeh (2004) to recover the virtual return for realized volatility and correlation construction. This is an important issue that has already drawn a lot of attentions from the academics as well as practitioners.

The second issue in construction of realized measures for volatility and correlation is how to circumvent the contamination from market microstructure noises using high frequency data. The typical recipe is to use a moderate sampling frequency to get intra-daily returns instead of sampling returns infinitely often over infinitesimally short time intervals; see Andersen and Bollerslev (1997), Andersen et al. (2005) and Barndorff-Nielson and Shephard (2002). This is to balance the finite sample bias/variance tradeoff induced by microstructure noises; see Bandi and Russell (2003). From the other perspective, since the bias in realized volatility estimator is due to the intraday serial correlation patterns induced by market frictions, some people propose to remove the bias by filtering out the dependence; see for example, Bollen and Inder (2002), French, Schwert and Stambaugh (1987), Zhou (1996), and Hansen and Lunde (2004).
However, when mid- or small cap equities are under consideration in practical financial management, the optimal sampling frequency seeking to balance bias and variance trade-off due to market microstructure sheds no light in studying high dimensional dynamic correlations since it assigns different sampling frequencies for different stocks. Likewise, to encompass the analysis to moderately traded asset given a fixed sampling frequency, the aforementioned filtering methods offer only limited practical relevance in assessing correlations because the trading frequencies among the stocks can be vast.

In this paper, our first goal is to demonstrate the statistical model by Tsay and Yeh (2004) can be designed to accommodate the effect of non-synchronous trading bias in high frequency financial returns series. The basic idea is to recognize that, given a sampling frequency that is high, every asset is subject to a probability of no trade and the real asset values may change due to new information arrivals even trade does not take place for those assets. For the small sampling intervals, information shocks are believed to be locally Gaussian since sampling at high frequency is to approximate the continuous time process of the asset price. Those new values can be recovered (estimated) via MCMC and end up with a synchronized set of series that can be used further to estimate values, value at risks, hedge parameters and correlations for asset allocations in high-frequency. The second goal is to incorporate the effects from the other market microstructure noises in our analysis by allowing the returns to be weakly dependent. Their virtual log return process can then be uncovered after filtering out the dependence attributed to the potential market microstructure effects or noises. Then the uncovered returns are used in construction of realized volatility and correlation. In brief, this approach allows for serial dependence to mimic market microstructure effects while at the same time take into account the non-synchronous trading problem among intraday returns.

Three different versions of realized volatilities and correlations are computed using different return series with different extent of correction for market microstructure noises. We use Kolmogorov-Smirnov test to compare the three different realized measures of volatility and correlation. Our empirical results show the newly proposed method is effective in correcting for both non-synchronous trading and other market friction effects. We also found that non-synchronous trading bias plays a dominant role among the other market microstructure effects for less actively traded equities in the computation of realized volatility and correlation. After controlling the effect from non-synchronous trading, the effect from the other microstructure is evident but negligible in the context of realized measures construction. Applying the new recovering and filtering procedure allows one to get more precise measures for realized daily volatilities as well as correlations.
The rest of this article is organized as follows. Section 2 reviews the theoretical background of realized volatility and the practical problem created by market microstructure noises in computing realized volatility. We then discuss what problem we shall encounter when we try to generalize the construction of realized volatility and correlation to non-synchronously traded assets. In section 3 we introduce the stochastic temporal aggregation model from Tsay and Yeh (2004) that allows for non-synchronous trading. We show that non-synchronous trading can mask the other market microstructure and results in a complicated pattern of dependence in the observed return. To cope with the non-synchronous biases, we thus propose a recovering procedure using Markov chain Monte Carlo (MCMC) method. Coupled with an filtering step in MCMC, the construction of realized volatility is feasible through the microstructure noise-filtered virtual returns. We implement the proposed methodology to construct realized volatility and correlation for several NYSE and DASDAQ equities in section 4. Section 5 concludes this paper.

2 A Theoretical Framework

2.1 Realized Measure for Volatility

Let \( p_t \):= \( \ln(P_t) \) denote the time \( t \) logarithmic price of an asset.

**Assumptions for price process**

1. \( p_t \) is a continuous diffusion process that represented by

\[
dp_t = \mu_t \, dt + \sigma_t \, dW_t, \quad 0 \leq t \leq T,
\]

where \( \mu_t \) is a continuous and local bounded variation process and \( \{W_t : t \geq 0\} \) is a standard Brownian motion. \( \sigma_t \, dW_t \) is therefore a local martingale.

2. The spot volatility process \( \sigma_t \) is càdlàg, independent of \( W_t \) \( \forall t \), and bounded away from zero.

3. The integrated volatility process \( \int_0^t \sigma_s^2 \, ds < \infty, \forall t < \infty \).

Consequently, \( p_t \) is a semi-martingale.

Normalize the daily time interval to unity and label the corresponding discretely sampled
daily returns by $r_t$, the one-period continuously compounded return for $P_t$ is formally given by

$$r_t := p_t - p_{t-1} = \int_{t-1}^{t} \mu_s ds + \int_{t-1}^{t} \sigma_s ds. \quad (1)$$

The quadratic variation for the cumulative return process, $r_t = p_t - p_0$, is then given by,

$$[r, r]_t = \int_0^t \sigma^2_s ds, \quad (2)$$

which is simply the integrated volatility.

Since the drift term is irrelevant at high frequency, the following discussion focuses on functionals of the $\sigma^2_t$ process. Define the daily realized volatility for the generic period $t$ by the summation of the corresponding $1/\delta$ high-frequency intra-daily squared returns,

$$RV_t(\delta) := \sum_{j=1}^{1/\delta} r^2_{t-1+j\delta}, \quad (3)$$

where $\delta$ is the sampling interval and $1/\delta$ is typically assumed to be an integer for simplicity. For example, if $\delta$ is a 5-minute interval then we have $1/\delta = 78$ as the number of observations for intraday return. Then from the theory of quadratic variation, $RV_t$ is employed to estimate the integrated variance, $\int_{t-1}^{t} \sigma^2_s ds$, since the realized volatility measure converges uniformly in probability to the increment of the quadratic variation process as the sampling frequency increases; see a series of recent papers by Andersen et al. (2001, 2002, 2003), Barndorff-Nielsen and Shephard (2002a, b), among others. Specifically, for $\delta \to 0$,

$$RV_t(\delta) \to \int_{t-1}^{t} \sigma^2_s ds. \quad (4)$$

The use of high frequency intraday returns plays a critical role in developing as well as justifying the nonparametric realized measure for volatility estimation. It should be noted that the realized volatility requires only the intraday returns are uncorrelated but not that the intraday return to be homoskedastic. This is evident from the stochastic volatility term $\sigma^2_t$ as shown above. However, the local Gaussianity and uncorrelated nature implied from the virtual diffusion process allow us to summarize that high frequency data are well-approximated by i.i.d. $N(0, \tilde{\sigma}^2)$ when the sampling frequency is high and sampling interval is short and fixed.\(^1\)

Nonetheless, the validity of this realized measure for volatility depends on the observability of the true price process. Unfortunately, it is known that the observed return are contaminated by a diverse array of market microstructure noises. Due to the presence of these frictions or

---

\(^1\)An extreme case is that if the underlying return is generated from a geometric Brownian motion, then the variance for the return sampled at any fixed time interval is indeed a constant.
noises such as bid-ask bounces, discreteness of price changes, inventory control effects, gradual response of prices to a block trade, ..., etc, the $RV_t$ is biased and inconsistent for $\int_{t-1}^t \sigma^2_s ds$; see Andreou and Ghysels (2002), Oomen (2002). This point is clearly addressed in the subsequent subsection.

2.2 Market Microstructure Noises

We use a model borrowed from Bandi and Russell (2003) to demonstrate the effect of market microstructure in the context of realized volatility construction. Suppose that the prices are recorded when transaction happened and the observed prices are mixed with microstructure noises. Denote the contaminated price series as $\tilde{P}_t$ and $\nu_t$ as the market microstructure noise that is independent of $P_t$ and iid $(1, \sigma^2_\nu)$ distributed, the price process is

$$\tilde{P}_t := P_t \cdot \nu_t, \quad t = 1, \cdots, T.$$  (5)

The implied return is obtained from simple log transforms:

$$\tilde{r}_t = \tilde{p}_t - \tilde{p}_{t-1} = p_t - p_{t-1} + \ln(\nu_t) - \ln(\nu_{t-1}), \quad t = 1, \cdots, T,$$

where $\tilde{p}_t := \ln(\tilde{P}_t)$.

Using high frequency returns, denoted by $\tilde{r}_{t-1+j\delta}$, to compute realized volatility $\tilde{RV}_t(\delta)$, we obtain

$$\tilde{RV}_t(\delta) := \frac{1}{\delta} \sum_{j=1}^{1/\delta} \tilde{r}_{t-1+j\delta}^2.$$  (6)

Let $\eta_t := \ln(\nu_t) - \ln(\nu_{t-1})$, then

$$\tilde{r}_t = r_t + \eta_t.$$  (7)

Obviously from equation (7), the market microstructure noise induces serial correlation in $\tilde{r}_t$ since the first order autocovariance is negative and equal to $-\mathbb{E}[\nu^2] = -\sigma^2_\nu$ even if the underlying return process are serially uncorrelated.

To see how the realized volatility is contaminated by market microstructure noise, the realized volatility from the noise-contaminated return is

$$\tilde{RV}_t(\delta) = \sum_{j=1}^{1/\delta} r_{t-1+j\delta}^2 + \sum_{j=1}^{1/\delta} \eta_{t-1+j\delta}^2 + 2 \sum_{j=1}^{1/\delta} r_{t-1+j\delta} \cdot \eta_{t-1+j\delta}.$$  (8)
Two additional terms enter the realized measure for volatility. It is mainly the second term, 
\[ \sum_{j=1}^{1/\delta} \eta_{t-1+j\delta}^2 \], that makes the realized volatility fail to converge to the underlying quadratic 
variation of the log price process since this term diverges to infinity almost surely as \( \delta \to 0 \).
This makes intuitive sense because more and more noises are accumulated as the sampling 
frequency increases in the limit; similar arguments can also be found in Ait-Sahalia, Mykland 
and Zhang (2005).

2.3 A Generalization to Non-synchronously Traded Assets

In view of the fact that the observed prices diverge from their efficient values due to a variety of 
market microstructure effects, some efforts has been devoted to correct for the bias or inconsis-
tency arising from the measurement error problem along different lines. On one hand, Andersen 
and Bollerslev (1997), Andersen et al. (2001) and Barndorff-Nielson and Shephard (2002) sug-
gest to use a moderate sampling frequency to get intra-daily returns instead of sampling returns 
infinity often over infinitesimally short time intervals to compute realized volatility. While 
Andersen and Bollerslev suggest 5-minute sampling interval for actively traded assets, Bandi 
and Russell (2003) show theoretically from the mean-square error (MSE) viewpoint that an 
optimal sampling frequency can be achieved to balance the finite sample bias/variance trade-
off induced by microstructure noises. Zhang, Mykland and Ait-Sahalia (2002) also propose 
an alternative estimator based on subsampling techniques and show its consistency for the 
integrated volatility.

On the other hand, since we know that the bias in realized volatility estimator is due to the 
intraday serial correlation patterns in high frequency returns, some people propose to remove 
the bias (and hence the effects from market frictions) by some filtering techniques. For example, 
Bollen and Inder (2002) use autoregressive models to pre-whiten the intraday returns; French, 
Schwert and Stambaugh (1987), Zhou (1996), and Hansen and Lunde (2004) suggest to correct 
for first order autocorrelation.

This paper attempts to address the issue of realized measures of volatility and correlation 
construction for both liquid and illiquid equities using high frequency intraday returns. When 
mid- or small cap equities are under consideration in practical financial management, one would 
confront the problem of non-synchronous trading. The optimal sampling frequency under MSE 
seeking to balance bias and variance trade-off due to market microstructure sheds no light in 
studying high dimensional dynamic correlations since it assigns different sampling frequencies 
for different stocks.
To encompass the analysis to moderately traded asset given a fixed sampling frequency, the aforementioned pre-whitening methods in the realized volatility literature offer only limited practical relevance in assessing correlations since trading frequencies among the stocks can be vast. We thus propose a new method to synchronize all asynchronously traded assets under a fixed sampling frequency by a recovering procedure. Their virtual log return process can then be uncovered after filtering out the potential market microstructure effects or noises. Then the uncovered returns are used in construction of realized volatility and correlation. In the subsequent section, we introduce a stochastic temporal aggregation model and show how to deal with the non-synchronous trading bias and the other market microstructure noises in high frequency.

3 Synchronizing the Asynchronous High Frequency Returns

In this section, we show that non-synchronous trading can mask the dependence induced by the other market microstructure effects. However, the problem of non-synchronous trading can be dealt effectively with the proposed recovering method in constructing a large dimensional realized covariance matrix, given any fixed frequency. At the same time, we can synchronize the non-synchronously traded equity returns with their underlying returns filtered from market microstructure effects and perform further statistical analysis or inference based on the recovered returns. We use the stochastic temporal aggregation model proposed in Tsay and Yeh (2004) and derive analytical formulae of mean, variance, auto-covariance and cross-covariance for stochastically temporally aggregated time series processes. Since the spirit of stochastic temporal aggregation is in line with the well known phenomenon of infrequent trading or non-synchronous trading, we use the proposed model to investigate its empirical relevance.

Suppose that the observed price keeps track at regular time points of an underlying price process that trades at random times. To see the effect of infrequent trading, let the virtual process \( \tilde{r}_{i,t} \), \( i = p, q \), be the informational shocks regularly transmitted to the market. The underlying information will be realized and reflected in the recorded prices only if a trade happens. For times where no trade takes place, those unrealized information will be accumulated and delivered to the subsequent periods. Once a trade takes place, these information will be realized. Since the market trades randomly at times, it is obvious that returns are actually recorded irregularly spaced in time. It is generally not plausible to disentangle cross correlations between irregularly spaced time series. In order to assess the correlations among series, one may transform the irregularly sampled sequence into a regularly spaced scope by
confronting the phenomenon of infrequent trading or non-synchronous trading. That is to admit there are periods of no trade. The effect of non-synchronous trading under the assumption of i.i.d. daily stock or portfolio returns has been investigated by Atchison, Butler, and Simonds (1987), Lo and MacKinlay (1990), Boudoukh, Richardson, and Whitelaw (1994), and Kadlec and Patterson (1999).

Let \( \{r^o_{i,t}\}, i = p, q \) denote the observed regularly spaced return processes. Table 1 demonstrates a typical example where the two assets are traded at different trading frequencies and thus non-synchronized. Given a fixed sampling frequency, suppose the \( p \) equity is subject to the effect of infrequent trading, informational or not, while the \( q \) equity is an actively traded asset. For no-trading periods, people typically assume price has no change and therefore zero returns are inserted into the actual sampled return for no-trading periods. It has been documented in the literature that it creates some undesired properties in the return sequence. For example, these artificially plugged-in zero returns induce spurious negative auto-correlation in the observed return even if the virtual process is i.i.d. since the sequence is forced to mean revert to zero return for those no trading periods. It will be seen as well that the ad hoc procedure will also bias the covariance between sequences, i.e., \( \text{Cov}(\tilde{r}_{p,t}, \tilde{r}_{q,t}) \) is in general not equal to \( \text{Cov}(r^o_{p,t}, r^o_{q,t}) \).

<table>
<thead>
<tr>
<th>{\tilde{r}_{p,t}}</th>
<th>Market p</th>
<th>Observed {r^o_{p,t}}</th>
<th>{\tilde{r}_{q,t}}</th>
<th>Market q</th>
<th>Observed {r^o_{q,t}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tilde{r}_{p,1} )</td>
<td>Trade ( \tilde{r}_{p,1} )</td>
<td>( \tilde{r}_{q,1} )</td>
<td>Trade ( \tilde{r}_{q,1} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,2} )</td>
<td>No Trade 0</td>
<td>( \tilde{r}_{q,2} )</td>
<td>Trade ( \tilde{r}_{q,2} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,3} )</td>
<td>Trade ( \tilde{r}<em>{p,2} + \tilde{r}</em>{p,3} )</td>
<td>( \tilde{r}_{q,3} )</td>
<td>Trade ( \tilde{r}_{q,3} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,4} )</td>
<td>No Trade 0</td>
<td>( \tilde{r}_{q,4} )</td>
<td>Trade ( \tilde{r}_{q,4} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,5} )</td>
<td>No Trade 0</td>
<td>( \tilde{r}_{q,5} )</td>
<td>Trade ( \tilde{r}_{q,5} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,6} )</td>
<td>Trade ( \tilde{r}<em>{p,4} + \tilde{r}</em>{p,5} + \tilde{r}_{p,6} )</td>
<td>( \tilde{r}_{q,6} )</td>
<td>Trade ( \tilde{r}_{q,6} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,7} )</td>
<td>Trade ( \tilde{r}_{p,7} )</td>
<td>( \tilde{r}_{q,7} )</td>
<td>Trade ( \tilde{r}_{q,7} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,8} )</td>
<td>No Trade 0</td>
<td>( \tilde{r}_{q,8} )</td>
<td>Trade ( \tilde{r}_{q,8} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,9} )</td>
<td>Trade ( \tilde{r}<em>{p,8} + \tilde{r}</em>{p,9} )</td>
<td>( \tilde{r}_{q,9} )</td>
<td>Trade ( \tilde{r}_{q,9} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,10} )</td>
<td>Trade ( \tilde{r}_{p,10} )</td>
<td>( \tilde{r}_{q,10} )</td>
<td>Trade ( \tilde{r}_{q,10} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,11} )</td>
<td>Trade ( \tilde{r}_{p,11} )</td>
<td>( \tilde{r}_{q,11} )</td>
<td>Trade ( \tilde{r}_{q,11} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,12} )</td>
<td>No Trade 0</td>
<td>( \tilde{r}_{q,12} )</td>
<td>Trade ( \tilde{r}_{q,12} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{r}_{p,13} )</td>
<td>Trade ( \tilde{r}<em>{p,12} + \tilde{r}</em>{p,13} )</td>
<td>( \tilde{r}_{q,13} )</td>
<td>Trade ( \tilde{r}_{q,13} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Our focus is on high-frequency log returns for which the underlying return series may have serial dependence since theory suggests some type of market microstructure may induce serial dependence in the return, for example, bid and ask bounce. A formal modelling of non-synchronous trading allowing for other microstructure effects in underlying returns can then be formalized in the proposed stochastic temporal aggregation framework. Here we explore the time series property of a stochastically temporal aggregated time series to understand the effect of non-synchronous trading. We propose a new method to recover the true underlying process that free from the non-synchronous trading bias, and filtering out the other microstructure noises at the same time. The filtered returns are employed for the construction of realized volatility and correlation in the next section.

3.1 Non-synchronous Trading as Stochastic Temporal Aggregation

Let \( \tilde{r}_t \) denote the market microstructure noised log returns of an asset at time \( t \). We assume that the time index \( t \) is discrete in this paper. The actual time interval used depends on the purpose of analysis, but it is typically measured in minutes, e.g., intraday 5-minutes. The observed log return is \( r^o_t \). We assume a time invariant probability of no trade at each time point as \( \pi \), which depends on the asset. However, we assume \( \tilde{r}_t \) to be a stationary time series to exhibit the potential microstructure effects satisfying

\[
E(\tilde{r}_t) = \mu, \quad \text{Cov}(\tilde{r}_t, \tilde{r}_{t-j}) = \gamma_j, \tag{9}
\]

where \( \gamma_j \) exists and depends only on \( j \), and \( \sum_{j=0}^{\infty} \gamma_j < \infty \). The relationship between \( \tilde{r}_t \) and \( r^o_t \) is

\[
r^o_t = \begin{cases} 
0, & \text{with prob. } \pi, \\
\tilde{r}_t, & \text{with prob. } (1 - \pi)^2, \\
\tilde{r}_t + \tilde{r}_{t-1}, & \text{with prob. } (1 - \pi)^2 \pi, \\
\tilde{r}_t + \tilde{r}_{t-1} + \tilde{r}_{t-2}, & \text{with prob. } (1 - \pi)^2 \pi^2, \\
\vdots & \vdots \\
\sum_{i=0}^{k} \tilde{r}_{t-i}, & \text{with prob. } (1 - \pi)^2 \pi^k, \\
\vdots & \vdots 
\end{cases} \tag{10}
\]

That is, \( r^o_t = \sum_{i=0}^{k} \tilde{r}_{t-i} \) when there are trades at \( t \) and \( t - k - 1 \), but no trades from \( t - k \) to \( t - 1 \). The observed return can be represented as a stochastic summation of a random number of terms where \( k \) is the random number of past terms to be aggregated.

It is straightforward to show that \( E(r^o_t) = \mu \). Since

\[
\text{Cov}(r^o_t, r^o_{t-j}) = E(r^o_t r^o_{t-j}) - E(r^o_t)E(r^o_{t-j}) = E(r^o_t r^o_{t-j}) - \mu^2,
\]
it suffices to study \( E(r_{t}^{o}, r_{t-j}^{o}) \) to understand the effect of non-synchronous trading on the return series \( \tilde{r}_t \). Define \( \rho_i = \gamma_i/\gamma_0 \) the lag-\( i \) autocorrelation function of \( \tilde{r}_t \). Consequently, we obtain

\[
\text{Var}(r_{t}^{o}) = \frac{2\pi \mu^{2}}{1 - \pi} + \gamma_0(1 + 2 \sum_{i=1}^{\infty} \pi^i \rho_i).
\] (11)

This result says that the variance of the observed return \( r_{t}^{o} \) is affected by two factors in the presence of non-synchronous trading. The first factor is the expected value of the underlying return \( \tilde{r}_t \) and the probability odd-ratio between no trade and trade. The second factor consists of the discounted serial correlations of \( \tilde{r}_t \) with the probability of no trade serving as the discount factor.

Next, to evaluate \( E(r_{t}^{o}, r_{t-j}^{o}) \) for \( j > 0 \), we have

\[
r_{t}^{o}r_{t-j}^{o} = \begin{cases} 0, & \text{with prob. } 2\pi - \pi^2, \\ (\sum_{i=0}^{k} \tilde{r}_{t-i})(\sum_{i=0}^{\ell} \tilde{r}_{t-j-i}), & \text{with prob. } (1 - \pi)^{4} \pi^{k+\ell}, \ 0 \leq k < j - 1, \ell = 0, 1, 2, \cdots, \\ (\sum_{i=0}^{j-1} \tilde{r}_{t-i})(\sum_{i=0}^{\ell} \tilde{r}_{t-j-i}), & \text{with prob. } (1 - \pi)^{3} \pi^{\ell+j-1}, \ell = 0, 1, 2, \cdots \\ \end{cases}
\] (12)

Taking expectation of the equation, we can obtain the autocovariance function of \( r_{t}^{o} \) below. Recall the results given in theorem 1 in chapter 1, we obtain

\[
\text{Cov}(r_{t}^{o}, r_{t-j}^{o}) = -\pi^j \mu^2 + \gamma_0(1 - \pi)^2 \sum_{i=0}^{j-1} \pi^i \left( \sum_{\ell=0}^{\infty} \pi^\ell \rho_{j+i-\ell} \right),
\]

where \( \gamma_0 \) is the variance of \( \tilde{r}_t \) and \( \rho_i \) is the lag-\( i \) autocorrelation of \( \tilde{r}_t \).

To see how substantial non-synchronous trading may have biasing the covariance of two observed log returns, let \( \{r_{p,t}, r_{q,t}\} \) be the local Gaussian log return of two assets at time \( t \). Denote \( \{\tilde{r}_{p,t}, \tilde{r}_{q,t}\} \) as the returns contaminated by market microstructure noises other than non-synchronous trading that are weakly stationary and \( \{r_{p,t}^{o}, r_{q,t}^{o}\} \) as the final observed return sequences. Assume that the probability of no trade for these two assets at each time point are \( \pi_p \) and \( \pi_q \), respectively. Define the covariance and lagged covariances between \( \{\tilde{r}_{p,t}\} \) and \( \{\tilde{r}_{q,t}\} \) as

\[
\text{Cov}(\tilde{r}_{p,t}, \tilde{r}_{q,t}) = \omega_0, \quad \text{Cov}(\tilde{r}_{p,t+j}, \tilde{r}_{q,t}) = \omega_j, \quad \text{Cov}(\tilde{r}_{p,t}, \tilde{r}_{q,t-j}) = \omega_{-j}, \quad j = 1, 2, \cdots. \ (13)
\]

Again from theorem 2 in chapter 1, we know the covariance of \( r_{p,t}^{o} \) and \( r_{q,t}^{o} \) is

\[
\text{Cov}(r_{p,t}^{o}, r_{q,t}^{o}) = \frac{(1 - \pi_p)(1 - \pi_q)}{(1 - \pi_p \pi_q)} [\omega_0 + \sum_{i=1}^{\infty} (\pi_p^i \omega_{-i} + \pi_q^i \omega_i)].
\]
where $\omega_j$ is defined in Eq. (13). The result shows that non-synchronous trading brings lagged covariances into the covariance of observed returns, but the impact is discounted by the probabilities of no trade as lag increases.

Given the above results, we are ready to demonstrate how the observed correlation between two non-synchronously traded equities can be biased; a direct instance corresponding to the model of Bandi and Russell (2003) we introduced in section 2.2 is as follows. Consider now that both the log return process contaminated with microstructure noises following MA(1) processes:

$$\tilde{r}_{p,t} = \mu_p + \nu_{p,t} - \theta_p \nu_{p,t-1}$$
$$\tilde{r}_{q,t} = \mu_q + \nu_{q,t} - \theta_q \nu_{q,t-1},$$

where $\{(\nu_{p,t}, \nu_{q,t})\}$ is a bivariate white noise series with $\sigma^2_{\nu_i} > 0$ and $\theta_i \neq 0$ with $|\theta_i| < 1$.

For a MA(1) process, $\gamma_0 = (1 + \theta_i^2)\sigma^2_{\nu_i}$, $\gamma_1 = -\theta_i \sigma^2_{\nu_i}$, and $\gamma_j = 0$ for $j > 1$. In this case, Eq. (11) and Theorem 1 give

$$\operatorname{Var}(r_{i,t}) = \frac{2\pi_i \mu_i^2}{1 - \pi_i} + \gamma_0(1 + 2\pi_i \rho_1), \ i = p, q.$$ 

Therefore the variance of the observed return can be inflated or deflated depending on how frequent the equity trades and how strong the effect from the microstructure noise induces. Also, $\omega_0 = (1 + \theta_p \theta_q) \operatorname{Cov}(\nu_{p,t}, \nu_{q,t})$, $\omega_1 = -\theta_q \operatorname{Cov}(\nu_{p,t}, \nu_{q,t})$, $\omega_{-1} = -\theta_p \operatorname{Cov}(\nu_{p,t}, \nu_{q,t})$, and $\omega_i = 0$ otherwise. It is known then from Theorem 2,

$$\operatorname{Cov}(r_{p,t}^o, r_{q,t}^o) = \frac{(1 - \theta_p)(1 - \theta_q)(1 - \pi_p)(1 - \pi_q)}{(1 + \theta_p \theta_q)(1 - \pi_p \pi_q)} \omega_0$$
$$= \frac{(1 - \theta_p)(1 - \theta_q)(1 - \pi_p)(1 - \pi_q)}{(1 - \pi_p \pi_q)} \operatorname{Cov}(\nu_{p,t}, \nu_{q,t}).$$

The sample covariance of the observed returns is a biased estimate of the covariance between the two log returns, but the size of the bias depends on the MA parameters of the models. In general, the sample covariance of the observed returns underestimate the true covariance between the virtual log return process, $\operatorname{Cov}(\nu_{p,t}, \nu_{q,t})$.

### 3.2 Synchronizing by Recovering Virtual Returns

Based on the results of Section 2, it can be seen that the effect of non-synchronous trading on asset returns can be complicated. Therefore, it would be advantageous in applications
if the underlying return series \( r_{i,t} \) were available. While this appears to be an impossible task, the observed return series \( \{ r_{o,i,t} \} \) does indeed contain some useful information about \( \tilde{r}_{i,t} \). In this section, we treat \( \tilde{r}_{i,t} \) for no-trading periods as missing values and use econometric methods to estimate their values under some reasonable assumptions. To formally consider the contamination from market frictions, we assume that \( \tilde{r}_{i,t} \) follows a stationary time series model and we shall estimate the missing \( \tilde{r}_{i,t} \) and hence the underlying \( r_{i,t} \) series based on the observed return \( r_{o,t} \). It is important to note that while we use non-synchronous trading to motivate the need to recover the unobserved log returns, the proposed estimation procedure is independent of the causes of missing values. The econometric method we used is the MCMC method. The implicitly assumed missing at random algorithm in MCMC is in line with the model of non-synchronous trading that subject to the probability of no trade.

We use the Gibbs sampling of MCMC methods to estimate the unobserved return in \( \{ r_{t} \} \) and model parameters. This amounts to treating the missing values as parameters, then evaluating the likelihood function of the data using sequentially conditional posterior distributions of model parameters and missing values.

The observed infrequent trading based log return \( r_{o,t} \) and the noise polluted \( \tilde{r}_{t} \) are related by Eq. (10). Let \( R_{o,n} = (r_{1,1}, r_{2,1}, \ldots, r_{n,n})' \) be the vector of observed returns and \( \tilde{R}_{n} = (\tilde{r}_{1}, \tilde{r}_{2}, \ldots, \tilde{r}_{n})' \) be the vector of log returns subject to the other market microstructure noises, where \( n \) is the sample size. Obviously, \( \tilde{R}_{n} \) contains missing values under non-synchronous trading. Then, the parameter vector of our problem consists of the parameters for the underlying process and all the missing values in \( \tilde{R}_{n} \). For a block of \( k \) consecutive missing values of \( r_{t} \), the prior distribution is multivariate normal with zero mean vector and covariance matrix \( \omega I_{k} \), where \( I_{k} \) is the \( k \times k \) identity matrix and \( \omega \) is a large hyper parameter. These are conjugate priors. In applications, we choose the hyper parameters so that the priors are non-informative.

To implement the Gibbs sampling, we need the conditional posterior distributions of (a) the time series parameters for the noised return process given \( \tilde{R}_{n} \) and \( \sigma^{2}_{\nu} \), (b) the innovational variance \( \sigma^{2}_{\nu} \) given \( \tilde{R}_{n} \) and parameters for contaminated return process, (c) missing values of \( \tilde{R}_{n} \) given \( R_{o,n} \), parameters for noised return process and \( \sigma^{2}_{\nu} \). The first two conditional posterior distributions are standard and can be found in many statistical books, e.g., Tsay (2002). Since no trading can occur in blocks of different sizes, we treat the missing values in \( \tilde{R}_{n} \) sequentially block by block. Consequently, it suffices to consider the conditional posterior distribution of a block of \( k \) consecutive no transactions. In what follows, we provide detailed derivation of the conditional posterior distribution of missing log returns because the distribution plays a key role in this paper.
To increase the efficiency of estimation, we modify the procedure based on the characteristics of high-frequency asset returns. For simplicity, we shall assume that \( \tilde{r}_t \) that contaminated by microstructure noises follows an AR(1) model, i.e.

\[
\tilde{r}_t = \phi_0 + \phi_1 \tilde{r}_{t-1} + \nu_t,
\]

where \( \{\nu_t\} \) is a sequence of independent and identically distributed Gaussian random variables with mean zero and variance \( \sigma^2_{\nu} \). The parameters \( \phi_0, \phi_1 \) and \( \sigma^2_{\nu} \) are unknown. The normality assumption can be relaxed at the expenses of more intensive computation and more complicated formulae, e.g., using heavy-tailed distribution or introducing jumps. Since MA(1) model is typically employed in modelling the pattern we shown in section 2.2 as well as the bid and ask bounce, we use this simple AR(1) model because it provides an adequate approximation for an MA(1) model when the MA coefficient is small.

### 3.3 Posterior Distribution for Recovering

For the case of one nontrading, recall that the observed \( r^o_h \) is actually the sum of two returns \( \tilde{r}_{h-1} \) and \( \tilde{r}_h \) as shown below.

Virtual: \( \tilde{r}_1 \tilde{r}_2 \cdots \tilde{r}_{h-2} \tilde{r}_{h-1} \tilde{r}_h \tilde{r}_{h+1} \cdots \tilde{r}_T \)

Observed: \( r^o_1 r^o_2 \cdots r^o_{h-2} \cdot r^o_h r^o_{h+1} \cdots r^o_T \)

\((\tilde{r}_{h-1} + \tilde{r}_h)\)

With the above restriction and the assumed AR(1) process, we know that \( \tilde{r}_{h-2} = r^o_{h-2}, \tilde{r}_{h+1} = r^o_{h+1} - \tilde{r}_h \) and

for \( t = h - 2 \), \( \tilde{r}_{h-1} = \phi_0 + \phi_1 \tilde{r}^{o}_{h-2} + \nu_{t-1} \)

for \( t = h \), \( \tilde{r}_h = \phi_0 + (1 + \phi_1) r^o_{h-1} + \nu_h \)

for \( t = h + 1 \), \( r^o_{h+1} = \phi_0 + \phi_1 \tilde{r}_h + \nu_{h+1} \).

Slightly re-arranging the equations yields

\[
y = \begin{bmatrix}
-\phi_0 - \phi_1 r^o_{h-2} \\
-\phi_0 + r^o_h \\
-\phi_0 + r^o_{h+1} - \phi_1 r^o_h
\end{bmatrix} = \begin{bmatrix}
-1 \\
(1 + \phi_1) \\
-\phi_1
\end{bmatrix} \begin{bmatrix}
\tilde{r}_{h-1} \\
\nu_{h-1} \\
\nu_h
\end{bmatrix} = X \cdot \tilde{r}_{h,1} + \varepsilon,
\]

which is a linear regression model except that the parameter of interest now is the missing returns. The estimator of \( \tilde{r}_{h-1} \) is normally distributed as

\[
\tilde{r}_{h-1} \sim N(\mu, \Sigma) = N((X'X)^{-1}X'y, \sigma^2_{\nu}(X'X)^{-1}).
\]
The case of two consecutive nontradings is illustrated as

Virtual: \( \tilde{r}_1 \tilde{r}_2 \cdots \tilde{r}_{h-3} \tilde{r}_{h-2} \tilde{r}_{h-1} \tilde{r}_h \tilde{r}_{h+1} \cdots \tilde{r}_T \)

Observed: \( r_1^0 r_2^0 \cdots r_{h-3}^0 \cdots r_h^0 r_{h+1}^0 \cdots r_T^0 \).

Applying similar techniques, one may obtain the sampling distribution for \( (\tilde{r}_{h-1}, \tilde{r}_{h-2})' \) is simply

\[
\begin{bmatrix}
-\phi_0 - \phi_1 r_{h-3}^0 \\
-\phi_0 \\
-\phi_0 + r_h^0 \\
-\phi_0 + r_{h+1}^0 - \phi_1 r_h^0
\end{bmatrix}
= \begin{bmatrix}
-1 & 0 \\
\phi_1 & -1 \\
1 & (1 + \phi_1) \\
-\phi_1 & -\phi_1
\end{bmatrix}
\begin{bmatrix}
\tilde{r}_{h-2} \\
\tilde{r}_{h-1} \\
\nu_{h-1} \\
\nu_h
\end{bmatrix}
+ \begin{bmatrix}
\nu_{h-2} \\
\nu_{h-1} \\
\nu_h \\
\nu_{h+1}
\end{bmatrix} = X \cdot \tilde{r}_{h,2} + \varepsilon. \quad (18)
\]

For \( k \) consecutive nontradings, given the observed log returns \( R_{tn}^0 \) and the model in Eq. (16), including parameters, we assume that there is no trading at time points \( h - k, h - k + 1, \ldots, h - 1 \). That is, \( \tilde{r}_{h,k} = (\tilde{r}_{h-k}, \tilde{r}_{h-k+1}, \ldots, \tilde{r}_{h-1})' \) is not observed. What we have are \( \tilde{r}_{h-k-1} \) and \( r_h^o = \sum_{i=h-k}^{h} \tilde{r}_i \). Notice that we do not include \( \tilde{r}_h \) in the vector of missing values because \( \tilde{r}_h = r_h^o - \sum_{i=h-k}^{h} \tilde{r}_i \).

First, from the AR(1) model in Eq. (16), we have

\[
\tilde{r}_{h-k} = \phi_0 + \phi_1 \tilde{r}_{h-k-1} + \nu_{h-k}. \quad (19)
\]

Since \( \phi_0, \phi_1 \) and \( \tilde{r}_{h-k-1} \) are known, we define \( y_1 = \phi_0 + \phi_1 \tilde{r}_{h-k-1}, \) \( x_1 = (1, 0, \cdots, 0) \) is a \( k \)-dimensional row vector, and \( e_1 = -\nu_{h-k} \). Eq. (19) then gives

\[
y_1 = x_1 \tilde{r}_{h,k} + e_1. \quad (20)
\]

Next, also from the AR(1) model, we have

\[
\tilde{r}_{h-k+1} = \phi_0 + \phi_1 \tilde{r}_{h-k} + \nu_{h-k+1}.
\]

Defining \( y_2 = -\phi_0, x_2 = (\phi_1, -1, 0, \cdots, 0), \) and \( e_2 = \nu_{h-k+1}, \) we can rewrite the above equation as

\[
y_2 = x_2 \tilde{r}_{h,k} + e_2. \quad (21)
\]

Similarly, for \( t = h - k + 2 \), we define \( y_3 = -\phi_0, x_3 = (0, \phi_1, -1, 0, \cdots, 0), \) and \( e_3 = \nu_{h-k+2} \) to obtain the equation

\[
y_3 = x_3 \tilde{r}_{h,k} + e_3. \quad (22)
\]
The same technique continues to apply for \( t = h - k + 3, \cdots, h - 1 \). At \( t = h - 1 \), we define \( y_k = -\phi_0 \), \( x_k = (0, \cdots, 0, \phi_1, -1) \), and \( e_k = \nu_{h-1} \) to obtain

\[
y_k = x_k \tilde{r}_{h,k} + e_k. \tag{23}
\]

Finally, at \( t = h \), we have

\[
\tilde{r}_h = \tilde{r}_{h-k} + \cdots + \tilde{r}_{h-1} + r_h \\
= \phi_0 + \tilde{r}_{h-k} + \cdots + \tilde{r}_{h-2} + (1 + \phi_1)\tilde{r}_{h-1} + \nu_h.
\]

Consequently, defining \( y_{k+1} = \tilde{r}_h - \phi_0 \), \( x_{k+1} = (1, \cdots, 1, 1 + \phi_1) \) and \( e_{k+1} = \nu_h \), we have

\[
y_{k+1} = x_{k+1} \tilde{r}_{h,k} + e_{k+1}. \tag{24}
\]

Combining Equations (20)-(24) together and noting that the distribution of \(-\nu_t\) is the same as that of \(\nu_t\), we have a set of linear regression equations for the unknown parameter \( \tilde{r}_{h,k} \) with \( k + 1 \) observations. That is,

\[
y \begin{bmatrix}
-\phi_0 \\
-\phi_0 \\
-\phi_0 \\
\vdots \\
-\phi_0 \\
-\phi_0 + r_h^0 \\
-\phi_0 + r_{h+1}^0 - \phi_1 r_h^0 \\
\end{bmatrix}
= X \begin{bmatrix}
-1 & 0 & 0 & \cdots & 0 & 0 \\
\phi_1 & -1 & 0 & \cdots & 0 & 0 \\
0 & \phi_1 & -1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & \phi_1 & -1 \\
1 & 1 & \cdots & \phi_1 & -1 \\
\end{bmatrix}
\begin{bmatrix}
\tilde{r}_{h-k} \\
\tilde{r}_{h-k+1} \\
\vdots \\
\tilde{r}_{h-1} \\
\tilde{r}_{h} \\
\phi_0 + r_h^0 \\
\end{bmatrix}
+ \begin{bmatrix}
\nu_{h-k} \\
\nu_{h-k+1} \\
\vdots \\
\nu_{h} \\
\nu_{h+1} \\
\nu_{h+1} \\
\end{bmatrix}
\tag{25}
\]

or rewritten as

\[
y_i = x_i \tilde{r}_{h,k} + e_i, \quad i = 1, \cdots, k + 1, \tag{26}
\]

where \( e_i \) are independent and identically distributed normal random variables with mean zero and variance \( \sigma^2_{\nu} \).

Consequently, we can estimate \( \tilde{r}_{h,k} \) by

\[
\hat{\tilde{r}}_{h,k} = \left( \sum_{i=1}^{k+1} x_i' x_i \right)^{-1} \left( \sum_{i=1}^{k+1} x_i' y_i \right). \tag{27}
\]

Furthermore, the sampling distribution of \( \hat{\tilde{r}}_{h,k} \) is multivariate normal with mean \( \tilde{r}_{h,k} \) and covariance matrix \( \hat{\Sigma} = \sigma^2_{\nu} \left( \sum_{i=1}^{k+1} x_i' x_i \right)^{-1} \).
Combining with prior distribution of $\tilde{r}_{h,k}$, which is multivariate normal with mean zero and covariance matrix $\omega I_k$, we obtain that conditional posterior distribution of $\tilde{r}_{h,k}$ is $k$-dimensional normal with mean $r_{h,k}^*$, and covariance matrix $\tilde{\Sigma}$, where

$$
\tilde{\Sigma}^{-1} = \frac{\sum_{i=1}^{k+1} x_i' x_i}{\sigma_\nu^2} + \frac{1}{\omega} I_k, \quad r_{h,k}^* = \Sigma \left( \frac{\sum_{i=1}^{k+1} x_i' \tilde{r}_{h,k}}{\sigma_\nu^2} \right).
$$

(28)

Clearly, a large hyper parameter $\omega$ will reduce the influence of prior distribution.

### 3.4 Implementation

Using all the conditional posterior distributions in the previous subsection, we estimate the model parameters and the contaminated log returns via Gibbs sampling. Then we are able to trace the serially uncorrelated underlying log returns for realized volatilities and correlations construction. As initial estimates, we use the observed log return series $R_n^o$ to estimate the parameters for the underlying process, say, $\phi$ and $\sigma_\nu^2$. For missing block $\tilde{r}_{h,k}$ in $\tilde{R}_n$, we simply use the local sample average, i.e. $\tilde{r}_t = r_t^o/(k + 1)$ for all elements of $\tilde{r}_{h,k}$. The effect of these initial estimates is negligible. Denote the initial estimates of $\phi$, $\sigma_\nu^2$ and $R_n^o$ by $\phi^{(0)}$, $\sigma_\nu^{2(0)}$ and $R_n^{(0)}$, respectively, where it is understood that the observed log return in $R_n^o$ is given throughout the estimation.

Now, given $\phi^{(i)}$, $\sigma_\nu^{2(i)}$, $\tilde{R}_n^{(i)}$ and priors, the Gibbs sampling procedure updates the estimates as follows.

1. Given $\phi^{(i)}$, $\sigma_\nu^{2(i)}$, and $R_n^{(i)}$, draw a random sample of missing log returns $\tilde{r}_{h,k}$ using the conditional posterior distribution in Eq. (28). This is done block-by-block for all missing log returns. The random draws provide a new estimate $\tilde{R}_n^{(i+1)}$.

2. Given $\sigma_\nu^{2(i)}$ and $\tilde{R}_n^{(i+1)}$, draw a random sample of $\phi$ using its conditional posterior distribution. This gives the new estimate $\phi^{(i+1)}$.

3. Given $\phi^{(i+1)}$ and $\tilde{R}_n^{(i+1)}$, draw a random sample of $\sigma_\nu^2$ from its conditional posterior distribution and denote the new estimate by $\sigma_\nu^{2(i+1)}$. Also recover the microstructure noise free $R_n^{(i+1)} = \{\nu_1, \cdots, \nu_n\}$ using Eq. (27) as the underlying serially uncorrelated returns.

4. Repeat steps 1 to 3 for iterations.
The Gibbs sampling is iterated for many iterations, say $B + N$. We discard random draws of the first $B$ iterations as burn-ins. This step is taken to reduce the effect of initial estimates and to ensure the convergence of the Gibbs sampling. The results of the last $N$ iterations are then used to make inference. For instance, the draws $\{\phi^{(B+i)}|i = 1, \cdots, N\}$ can be used to make inference about the AR(1) model for the log return $\tilde{r}_t$. As a second example, the lag-1 autocorrelation coefficient $\rho_1$ of $R^{(B+i)}_n$ for $i = 1, \cdots, N$ can be used to draw inference about the lag-1 autocorrelation of underlying log return $\nu_t$. We use $R^n_0$, $\tilde{R}^{B+i}_n$, $R^{B+i}_n$ to construct different versions of realized volatility and correlation for comparison. While $R^n_0$ is subject to both non-synchronous trading and other market microstructure effects, $\tilde{R}^{B+i}_n$ is free from non-synchronous trading bias but still suffers from the microstructure noises. The sequence $R^{B+i}_n$ is obtained by filtering $\tilde{R}^{B+i}_n$ with the dependence presented and thus it is free from market frictions if the posterior distribution of $\rho_1$ of $R^{(B+i)}_n$ is found to centered around zero.

In practice, care must be exercised to monitor the convergence of the Gibbs sampling. Our experience indicates that $(B, N) = (1000, 3000)$ is sufficient for the high-frequency asset returns analyzed in this paper.

4 Empirical Analysis

4.1 Data

In this paper, we pick 4 stocks and extract their high frequency transacted tick data from the NYSE TAQ database. The sampling period extends from January 1, 2003, to December 31, 2004. The ticker symbols for the selected stocks are consisted of two types of stocks. The first type are those mid- or small-cap equities that are infrequently traded including ARJ, LDL, PPCO; the other type is the actively traded, highly liquid stock, GE.\(^2\) For each stock, we construct 5-minute intraday log returns as suggested in Andersen et al. (2002) using the following procedure:

1. We focus on transactions with time stamp between 9:30 am and 16:00 pm Eastern Time; transactions occurred outside the normal trading hours are ignored.

2. For each 5-minute time interval starting at 9:30 am, the price is taken to be the first transaction price within that time interval. If there is no transaction within 5-minute

\(^2\)Since our study focus on removing the effect of non-synchronous trading, we do not need to restrict the sample to confined in the group of 214 high-liquidity equities identified through a three-point sample selection scheme as in Andersen et al. (2001).
time interval, the price is treated as missing.

3. We compute observed 5-minute log returns based on the 5-minute transaction prices, i.e. 
\[ r_t^o = 0 \] if no transaction took place in the 5-minute time interval and 
\[ r_t^o = \ln(P_t) - \ln(P_{t-k}) \] if there was a transaction at the 5-minute time interval, where \( P_{t-k} \) denotes the last available 5-minute prior transaction price. The noise contaminated log return \( \tilde{r}_t \) is defined as 
\[ \tilde{r}_t := \ln(P_t) - \ln(P_{t-1}) + \eta_t \] if both transaction prices are available and it is zero otherwise.

4. We ignore the overnight returns from our analysis. Thus, there are 78 intraday 5-minute returns in a normal trading day. The sample size of each intraday 5-minute log return series is 29,328.

5. Finally, all the returns used are in percentages.

This proposed procedure is not restricted to the specific time interval chosen and the other sampling frequency can be used if necessary. For actively traded equities, most 5-minute intraday returns are available. However, for those inactively traded equities, the percentage of no trading ranges from the low at 18% to the high at 45%.

4.2 Realized Volatility

Applying the proposed recovering procedure using MCMC, we get a series of sets of synchronized equities returns. We compute three versions of realized volatility for the selected equities and realized correlation among them based on the three return series \( R_n^o, \tilde{R}_n \) and \( R_n \). Realized volatility based on the observed return \( R_n^o \) is

\[ RV_t := \sum_{j=1}^{78} (r_{t-1+j\delta}^o)^2. \]

It is supposed to be the most noisy estimator since \( R_n^o \) is subject to the biases both from non-synchronous trading and other market microstructure noises. The non-synchronous trading corrected realized volatility (NTCRV), calculated based on \( \tilde{R}_n \) is

\[ \text{NTCRV}_t := \sum_{j=1}^{78} (\tilde{r}_{t-1+j\delta})^2. \]

NTCRV is subject to the other market microstructure biases other than non-synchronous trading. Finally, all noises corrected realized volatility (ACRV) based on the pre-whitened return
The posterior distribution for these parameter of interest are depicted in Figure 1 and summary of results is summarized in Table 2. The first column of Figure 1 are the posterior distributions of $\phi_1$ for the non-synchronous trading adjusted but still noise contaminated returns \( \tilde{R}_{i,t} \) for the four equities. They are overall negatively distributed, even for the actively traded GE. The mean and standard deviation for these $\phi_1$ distributions are given in Table 2. $\bar{\phi}_1$ for all four equities are significantly negative in terms of their standard deviations $\sigma_{\phi_1}$. These results show that even after taking care of the non-synchronous trading bias, the infrequent trading adjusted return is not free from other microstructure effects. That may explain why $\tilde{r}_t$ is serial negatively correlated. This is true for both the actively traded GE and the other illiquid equities whereas the magnitude of negative dependence does not hinge upon the rate of no-trading periods. While we use 5-minute interval as suggested in ABDE (2001) where the sampling frequency is supposed to be optimal in getting rid of the market microstructure effects for actively stocks like GE, our result shows that it is not sufficient to remove these effects by relying on an optimal sampling interval. The second column in Figure 1 shows the first order autocorrelation $\rho_1$ for the noise-filtered $R_n$ return sequences for the four equities. All distributions centering around zero show that the filtered returns are now serially uncorrelated and free from market microstructure noises. Therefore we are ready to obtain an unbiased and consistent estimate of realized volatility using $R_n = \{\nu_1, \cdots, \nu_n\}$

Look at the second part of Table 2 where we summarize the summary statistics for all three versions of realized volatilities, RV, NTCRV, ACRV. It is found that there are only marginal differences among the three RV estimates for GE but there are substantial differences for NTCRV and ACRV against RV for these inactively traded equities. Similar patterns can be observed in the plot of realized volatilities in Figure 2. RV is closely moved with ACRV for GE, but RV in general underestimates the bias-free ACRV for those less actively traded equities, as shown in the plots for ARJ, LDL and PPCO. It is also evident that the larger the missing rate of return, and hence more severe the non-synchronous trading bias, the more downward bias incurred.

To formally justify the proposed procedure in correcting biases, we performed the Kolmogorov-
Table 2: Summary Statistics of Realized Volatility Estimates

<table>
<thead>
<tr>
<th></th>
<th>GE</th>
<th>ARJ</th>
<th>LDL</th>
<th>PPCO</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT Rate</td>
<td>0.0%</td>
<td>22%</td>
<td>45%</td>
<td>18%</td>
</tr>
<tr>
<td>$\hat{\phi}_1$</td>
<td>-0.100</td>
<td>-0.036</td>
<td>-0.131</td>
<td>-0.204</td>
</tr>
<tr>
<td>$\sigma_{\phi_1}$</td>
<td>0.006</td>
<td>0.007</td>
<td>0.008</td>
<td>0.006</td>
</tr>
<tr>
<td>$\bar{\sigma}_\nu^2$</td>
<td>0.033</td>
<td>0.051</td>
<td>0.137</td>
<td>0.266</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>RV</th>
<th>NTRV</th>
<th>ACRV</th>
<th>RV</th>
<th>NTRV</th>
<th>ACRV</th>
<th>RV</th>
<th>NTRV</th>
<th>ACRV</th>
<th>RV</th>
<th>NTRV</th>
<th>ACRV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>37.98</td>
<td>38.00</td>
<td>37.86</td>
<td>21.17</td>
<td>21.67</td>
<td>21.82</td>
<td>46.28</td>
<td>40.23</td>
<td>38.85</td>
<td>163.90</td>
<td>167.77</td>
<td>132.37</td>
</tr>
<tr>
<td>Min</td>
<td>0.66</td>
<td>0.66</td>
<td>0.67</td>
<td>1.02</td>
<td>1.83</td>
<td>1.80</td>
<td>2.24</td>
<td>6.95</td>
<td>6.70</td>
<td>3.87</td>
<td>8.14</td>
<td>7.78</td>
</tr>
<tr>
<td>Median</td>
<td>1.89</td>
<td>1.89</td>
<td>1.88</td>
<td>2.87</td>
<td>3.56</td>
<td>3.53</td>
<td>6.42</td>
<td>9.84</td>
<td>9.64</td>
<td>15.36</td>
<td>18.74</td>
<td>17.96</td>
</tr>
<tr>
<td>Mean</td>
<td>2.60</td>
<td>2.60</td>
<td>2.57</td>
<td>3.36</td>
<td>3.97</td>
<td>3.96</td>
<td>7.44</td>
<td>10.83</td>
<td>10.64</td>
<td>18.79</td>
<td>21.63</td>
<td>20.74</td>
</tr>
<tr>
<td>Stdev</td>
<td>3.21</td>
<td>3.21</td>
<td>2.99</td>
<td>2.22</td>
<td>2.06</td>
<td>2.05</td>
<td>4.62</td>
<td>3.58</td>
<td>3.61</td>
<td>15.44</td>
<td>14.18</td>
<td>12.57</td>
</tr>
</tbody>
</table>

Smirnov test for two samples to compare the bias corrected NTCRV and ACRV with the original unadjusted RV. The Kolmogorov-Smirnov test, aims at testing whether two samples have been drawn from the same population distribution, can be implemented to show the effectiveness in removing market microstructure noises using the proposed new method. If the noise correction procedure is invalid, then the cumulative distributions of both samples of realized volatility may be expected to be close to each other. However, if the distributions of two RV from different stages of correction differs, this suggests that the proposed procedure do make a difference. We report both the K-S statistics and p-values for a paired comparison between two measures of realized volatility for four equities.

From Table 3, we found NTCRV and ACRV come from the same population distribution as that of RV for the actively traded GE. The p-value of testing NTCRV against RV is 1.00 since GE does not suffer from the infrequent trading bias. Though our results in Table 2 shows that the 5-minute sampled intraday returns are contaminated by the other market microstructure noises, our results here say that correcting for the other noises makes no difference. For the other less actively traded equities, we found NTCRV and ACRV differ substantially from RV, that utilizes the observed return directly. Nonetheless, the difference between NTCRV and ACRV tends to not reject the null of both NTCRV and ACRV are from the same population distribution. Hence the additional correction to remove market microstructure effects in $\tilde{r}_t$ seems to be effective (in terms of zero autocorrelation in $\nu_t$) but improves little for realized
Table 3: Comparison of Realized Volatility Estimates via Kolmogorov-Smirnov Test

<table>
<thead>
<tr>
<th></th>
<th>GE</th>
<th>ARJ</th>
<th>LDL</th>
<th>PPCO</th>
</tr>
</thead>
<tbody>
<tr>
<td>RV v.s. NTCRV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-S statistics</td>
<td>0.016</td>
<td>0.317</td>
<td>0.606</td>
<td>0.25</td>
</tr>
<tr>
<td>P-Value</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>RV v.s. ACRV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-S statistics</td>
<td>0.037</td>
<td>0.317</td>
<td>0.577</td>
<td>0.207</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.936</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>NTCRV v.s. ACRV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-S statistics</td>
<td>0.037</td>
<td>0.027</td>
<td>0.072</td>
<td>0.067</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.936</td>
<td>0.998</td>
<td>0.26</td>
<td>0.343</td>
</tr>
</tbody>
</table>

Note: RV abbreviates for realized volatility, NTCRV for non-synchronous trading corrected realized volatility, and ACRV for all microstructure noise corrected realized volatility.

volatility. Cross refer to the results we presented in Table 2, these results imply that non-synchronous trading plays a dominant role over other market microstructure noises for those less liquid equities. Correcting for the other market microstructure other than non-trading is merely marginal for realized volatility even if the returns are believed to subject to the other market microstructure effects.

4.3 Realized Correlations

The realized daily covariance matrix is computed as

$$\text{Cov}_{ij,t} := \sum_{k=1}^{78} r_{i,t-1+k\delta} r_{j,t-1+k\delta}.$$ 

Following the typical definition, the realized daily correlation between equity $i$ and $j$, denoted by $\text{Corr}_{ij,t}$, is obtained by

$$\text{Corr}_{ij,t} := \frac{\text{Cov}_{ij,t}}{\sqrt{\text{RV}_{i,t}} \sqrt{\text{RV}_{j,t}}}.$$ 

Similarly, three versions of realized daily correlations are computed. Descriptive statistics are reported in Table 4. We also plot NTCRC against RC and ACRC against RC among the four equities in Figure 3 and ??.
Table 4: Summary Statistics of Realized Correlation Estimates

<table>
<thead>
<tr>
<th>Realized correlation</th>
<th>Min</th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
<th>Stand.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARJ/GE&lt;sub&gt;RC&lt;/sub&gt;</td>
<td>−0.56</td>
<td>0.12</td>
<td>0.12</td>
<td>0.72</td>
<td>0.16</td>
</tr>
<tr>
<td>ARJ/GE&lt;sub&gt;NTCR&lt;/sub&gt;</td>
<td>−0.41</td>
<td>0.11</td>
<td>0.11</td>
<td>0.68</td>
<td>0.14</td>
</tr>
<tr>
<td>ARJ/GE&lt;sub&gt;ACR&lt;/sub&gt;</td>
<td>−0.40</td>
<td>0.12</td>
<td>0.12</td>
<td>0.68</td>
<td>0.14</td>
</tr>
<tr>
<td>ARJ/LDL&lt;sub&gt;RC&lt;/sub&gt;</td>
<td>−0.35</td>
<td>0.07</td>
<td>0.05</td>
<td>0.82</td>
<td>0.14</td>
</tr>
<tr>
<td>ARJ/LDL&lt;sub&gt;NTCR&lt;/sub&gt;</td>
<td>−0.22</td>
<td>0.04</td>
<td>0.04</td>
<td>0.37</td>
<td>0.08</td>
</tr>
<tr>
<td>ARJ/LDL&lt;sub&gt;ACR&lt;/sub&gt;</td>
<td>−0.22</td>
<td>0.05</td>
<td>0.04</td>
<td>0.41</td>
<td>0.09</td>
</tr>
<tr>
<td>ARJ/PPCO&lt;sub&gt;RC&lt;/sub&gt;</td>
<td>−0.45</td>
<td>0.04</td>
<td>0.03</td>
<td>0.59</td>
<td>0.15</td>
</tr>
<tr>
<td>ARJ/PPCO&lt;sub&gt;NTCR&lt;/sub&gt;</td>
<td>−0.41</td>
<td>0.03</td>
<td>0.03</td>
<td>0.45</td>
<td>0.12</td>
</tr>
<tr>
<td>ARJ/PPCO&lt;sub&gt;ACR&lt;/sub&gt;</td>
<td>−0.37</td>
<td>0.04</td>
<td>0.04</td>
<td>0.48</td>
<td>0.12</td>
</tr>
<tr>
<td>GE/LDL&lt;sub&gt;RC&lt;/sub&gt;</td>
<td>−0.49</td>
<td>0.08</td>
<td>0.07</td>
<td>0.62</td>
<td>0.15</td>
</tr>
<tr>
<td>GE/LDL&lt;sub&gt;NTCR&lt;/sub&gt;</td>
<td>−0.42</td>
<td>0.06</td>
<td>0.06</td>
<td>0.48</td>
<td>0.11</td>
</tr>
<tr>
<td>GE/LDL&lt;sub&gt;ACR&lt;/sub&gt;</td>
<td>−0.40</td>
<td>0.07</td>
<td>0.07</td>
<td>0.49</td>
<td>0.12</td>
</tr>
<tr>
<td>GE/PPCO&lt;sub&gt;RC&lt;/sub&gt;</td>
<td>−0.52</td>
<td>0.06</td>
<td>0.06</td>
<td>0.71</td>
<td>0.18</td>
</tr>
<tr>
<td>GE/PPCO&lt;sub&gt;NTCR&lt;/sub&gt;</td>
<td>−0.50</td>
<td>0.06</td>
<td>0.06</td>
<td>0.65</td>
<td>0.16</td>
</tr>
<tr>
<td>GE/PPCO&lt;sub&gt;ACR&lt;/sub&gt;</td>
<td>−0.51</td>
<td>0.07</td>
<td>0.07</td>
<td>0.65</td>
<td>0.17</td>
</tr>
<tr>
<td>LDL/PPCO&lt;sub&gt;RC&lt;/sub&gt;</td>
<td>−0.33</td>
<td>0.03</td>
<td>0.03</td>
<td>0.61</td>
<td>0.14</td>
</tr>
<tr>
<td>LDL/PPCO&lt;sub&gt;NTCR&lt;/sub&gt;</td>
<td>−0.25</td>
<td>0.02</td>
<td>0.01</td>
<td>0.37</td>
<td>0.09</td>
</tr>
<tr>
<td>LDL/PPCO&lt;sub&gt;ACR&lt;/sub&gt;</td>
<td>−0.31</td>
<td>0.02</td>
<td>0.02</td>
<td>0.39</td>
<td>0.10</td>
</tr>
</tbody>
</table>

In general, the mean of daily realized volatility RV tends to greater than those bias corrected realized correlation estimates like NTCRC or ACRC, except for the case of GE/PPCO. Similar patterns are reflected in the median of these daily correlation estimates. Moreover, the standard deviation of both NTCRC and ACRC estimates are overall dominated by the standard deviation of RC in all cases. In turn, this implies that realized correlation based on observed return sequences tends to fluctuate more widely than NTCRC and ACRC. That is, the obtained bias-corrected realized correlation estimates, NTCRC and ACRC are more precise and stable over time than RC. This observation is clearly depicted in the realized correlation plots of ARJ/LDL, ARJ/PPCO, GE/LDL and LDL/PPCO in Figure 3 and ??, respectively. In these
Table 5: Comparison of Realized Correlation Estimates via Kolmogorov-Smirnov Test

<table>
<thead>
<tr>
<th></th>
<th>ARJ/GE</th>
<th>ARJ/LDL</th>
<th>ARJ/PPCO</th>
<th>GE/LDL</th>
<th>GE/PPCO</th>
<th>LDL/PPCO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RC v.s. NTCRC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-S statistics</td>
<td>0.0559</td>
<td>0.1676</td>
<td>0.0878</td>
<td>0.1463</td>
<td>0.0718</td>
<td>0.1383</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.5579</td>
<td>0.0000</td>
<td>0.0974</td>
<td>0.0005</td>
<td>0.2590</td>
<td>0.0012</td>
</tr>
<tr>
<td><strong>RC v.s. ACRC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-S statistics</td>
<td>0.0452</td>
<td>0.1516</td>
<td>0.0638</td>
<td>0.1223</td>
<td>0.0612</td>
<td>0.1197</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.0799</td>
<td>0.0003</td>
<td>0.3914</td>
<td>0.0060</td>
<td>0.4435</td>
<td>0.0077</td>
</tr>
<tr>
<td><strong>NTCRC v.s. ACRC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-S statistics</td>
<td>0.0399</td>
<td>0.0559</td>
<td>0.0532</td>
<td>0.0559</td>
<td>0.0585</td>
<td>0.0505</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.8981</td>
<td>0.5579</td>
<td>0.6186</td>
<td>0.5579</td>
<td>0.4993</td>
<td>0.6800</td>
</tr>
</tbody>
</table>

Note: RC abbreviates for realized correlation, NTCRC for non-synchronous trading corrected realized correlation, and ACRC for all microstructure noise corrected realized correlation.

panels, NTCRC or ACRC fluctuates within a smaller band as compared to the region RC fluctuates. Calculating the ranges of RC, NTCRC and ACRC from Table 4 implies the similar argument.

To formally perform comparisons among these three realized correlation measures, we apply again the K-S test. The results are summarized in Table 5. For ARJ/LDL, GE/LDL, LDL/PPCO, we found statistically significant differences between the non-synchronous trading corrected realized correlation NTCRC and RV as well as between the all noise corrected realized correlation ACRC and RC. This can be due to the fact that LDL has a high proportionate of no-trading periods up to 45%. For the other cases, there are indeed difference between NTCRC v.s. RC and ACRC v.s. RC since the p-value is far below 1 but they are statistically insignificant though. When we compare further NTCRC and ACRC, it tends not to reject the null of both are from the same population. Therefore the results here are consistent to our observations for realized volatility that after correcting for non-synchronous trading bias, the left market microstructure noises can be minor for realized correlation construction. Non-synchronous trading bias plays a dominant role in correcting for market microstructure
noises when the basket of assets under investigation is not confined to liquid or actively traded equities only.

5 Concluding Remarks and Future Extensions

This chapter serves as a subsequent extension following the chapter concerning the systematic market risk assessment using high frequency beta. That is, after selecting a targeted basket of equities according to high frequency beta, the coming task is then to monitor the performance of the chosen portfolio. Moreover, one may also interest in finding a market timing strategy by employing the dynamic time-varying realized covariance matrix for strategic asset allocation. These tasks are feasible under the proposed synchronizing procedure. The contributions of this paper lie on the following points:

1. We allow the construction of realized volatility to a larger set of traded equities including both actively traded and in-actively traded stocks. This work is intractable under various methods suggested for realized volatility construction in the literature.

2. The idea to conquer the non-synchronous trading bias arisen among a set of asynchronously traded assets by recovering and synchronizing them within a fixed sampling frequency is new.

3. At the same time when we synchronizing the returns by recovering, the proposed procedure allow us to correct for the other market microstructure effects by filtering out the serial dependence observed in the synchronized return series.

4. Since our methodology can separate the effect of non-synchronous trading from the other market microstructure effects, we demonstrated in how bias or spurious the resultant realized volatility and correlation may be if the non-synchronization of trading among equities is not properly taken into account.

Our empirical results show the proposed method is effective in correcting for both non-synchronous trading and other market friction effects. We also found that non-synchronous trading bias plays a dominant role among the other market microstructure effects for less actively traded equities in the computation of realized volatility and correlation. After controlling the effect from non-synchronous trading, the effect from the other microstructure is evident but negligible in the context of realized measures construction. Applying the new recovering
and filtering procedure allows one to get more precise measures for realized daily volatilities as well as correlations that are applicable to further financial applications.

Reference


Bandi, F.M. and J.R. Russell (2003), Microstructure Noise, Realized Volatility and Optimal Sampling, unpublished manuscript, Graduate School of Business, University of Chicago.


Tsay, R.S. and J.H. Yeh (2004), Non-synchronous Trading and High-Frequency Beta, Working

Figure 1: Performance of Filtering Microstructure Noises

Note: This figure reports the posterior density for several parameters we interested for GE, ARJ, LDL and PPCO (from top to bottom). It shows from left to right respectively the posterior density for the first order autoregressive coefficient $\phi_1$, proxy for the other market microstructure noises, for the non-synchronous trading adjusted returns; the 1st order autocorrelation coefficient $\rho_1$ for all noise filtered returns; and $\sigma^2_\nu$ for the underlying driving forces. The first column shows that after controlling for the effect of non-synchronous trading, the return sequences are still subject to the other market microstructure in the presence of dependence in terms of significant negative $\phi_1$ patterns. However, further filtering $\tilde{r}_t$ with the estimated parameters, column 2 demonstrates the posterior distribution of lag one autocorrelation of the noise-filtered $\{\nu_t\}$ sequences. They all centered around zero and show that $\nu_{t+1}$ is about free from market microstructure effects. Hence they are ready for realized volatility and correlation construction.
Figure 2: A Comparison of Realized Volatility
Figure 3: A Comparison of Realized Correlation