Testing for co-jumps in high-frequency financial data: an approach based on first-high-low-last prices

Yin Liao†
Australian National University
Canberra, ACT, 0200
AUSTRALIA

Heather M. Anderson
Monash University
Clayton, Victoria 3800
AUSTRALIA

Preliminary
25th June 2011

Abstract

This paper proposes a new approach to test for common intraday jumps in a panel of high frequency financial data. We utilize intraday first-high-low-last values of asset prices to construct a novel estimator for the cross variance structure of a large panel of high frequency financial data, and then employ this estimator to provide a first-high-low-last price based test statistic to determine if this covariance is partially attributable to common large discrete movements (co-jumps). We study the finite sample behavior of our first-high-low-last price based test using Monte Carlo simulation, and find that it is more powerful than the existing return-based co-jump test (see Bollerslev et al (2008)) for covariance measured at the same sampling frequency. When applied to a panel of high frequency data from the Chinese mainland stock market, our first-high-low-last price based test identifies more common jumps than the return-based test in this emerging market.

Keywords: Covariance, Co-jumps, High-frequency data, First-High-Low-Last price, Realized covariance, Realized co-range.

JEL classification: C12, C22, C32, G12, G14

*We thank Tom Smith and Farshid Vahid for helpful comments and acknowledge financial assistance from the Australian Research Council, for Discovery Grants # DP0449995 and #DP0665710.
†Corresponding author: Email address: yin.liao@anu.edu.au
1 Introduction

Recent literature on measuring and predicting risk from high-frequency financial data has noted that jumps and the continuous component of an asset’s price process play different roles in the generation of volatility. This finding has important implications for financial endeavours such as option pricing and risk management, and it has led researchers to consider various methods for separately measuring the two components and to test for jumps in the evolution of asset prices. Due to the difficulty in robustly estimating the continuous part of quadratic covariation of a multivariate asset price process in the presence of jumps, most literature relating to jump tests is limited to a univariate or bivariate setting. A recent exception can be found in the work undertaken by Bollerslev et al. (2008), who developed a new co-jump test which is applicable to a large panel of high-frequency returns. These authors recognized the role that covariance can play in common jump detection, and they suggested the use of cross-products in a panel of intraday stock returns to measure this covariance, and to detect co-jumps by testing whether the covariance is partially attributable to common discontinuous movements.

Given that financial economists who work on portfolio theory are generally more interested in the systematic behavior of a large ensemble of asset prices, this newly proposed co-jump test makes a contribution to the literature from the perspectives of portfolio pricing and portfolio risk management. In this article, we make use of the intraday first-high-low-last (FHLL) price values to propose a more efficient estimator for the covariance structure of a large panel of high frequency financial data, which we call the “FHLL covariance estimator” below. Then, we utilize our FHLL covariance estimator instead of the cross-product of returns to construct a more powerful co-jump test.

Testing for jumps in high-frequency data is an on-going active research area. The early work on jumps can be found in work by Barndorff-Nielsen and Shephard (2006), who present a nonparametric toolkit for detecting jumps in the univariate setting. The basic idea is to compare two measures of daily variance for individual asset returns which respectively include and exclude jump contributions, and then to detect the presence of jumps by testing whether the difference between these two measures is statistically significant. This method is not only able to identify the jump days, but it can also explicitly pinpoint the jump arrival times if it is applied to intraday time intervals. Univariate analysis of jumps mostly contributes to the pricing and management of risk associated with individual assets, but extensions that can test for common jumps in multivariate setting are potentially useful, particularly for pricing and managing the systematic risk associated with
holding portfolios.

The early literature detected co-jumps by applying the standard Barndorff-Nielsen and Shephard (2006) univariate test statistic (the BN-S test) to stocks on a one-by-one basis. Researchers such as Gobbi and Mancini (2007) and Jacod and Todorov (2009) then proposed strategies to test for co-jumps in a pair of asset returns. Meanwhile, Bollerslev et al (2008) shifted attention to the covariance structure of a panel of intraday returns, after noting that there was little correspondence between jumps detected by a BN-S test applied to an index and jumps detected by the BN-S test applied to individual returns. These authors realized that idiosyncratic noise in individual returns could hide the presence of a synchronous component, and to overcome this problem they built up a panel based test statistic that was explicitly based on the covariance of all assets in the panel. Using this new test statistic, they identified many modest-sized common jumps that simply escaped detection when using the standard BN-S jump test. The work in Bollerslev et al (2008) provides an attractive starting point for co-jump tests, since it is now clear that cross-covariance is more sensitive to co-jumps than the variance of individual asset returns.

The critical issue relating to the Bollerslev et al (2008) co-jump test is the measurement of covariation in a panel of high-frequency data, and the power of their test relies on how efficiently this comovement is measured. The return-based estimator, which is the cross-product of intraday returns used in Bollerslev et al (2008), may not be the best covariance estimator. Parkinson (1980) first noted that the daily price high-low range reveals more information than the daily price return, because the range is formed by going through the entire price process, whereas the return is calculated from only the first and last price observations. Garman and Klass (1980) suggested further efficiency gains by utilizing the daily open and close prices in addition to the daily high and low prices. They proposed an estimator which is a linear combination of the squared high-low daily range and the squared daily return. With respect to the efficiency of estimation, which is measured by the asymptotic variance of the estimator, this new estimator is 7.4 times more efficient than the squared daily return and 1.5 times more efficient than the squared daily range for daily volatility estimation.

With the increased availability of high-frequency data, there is now a series of papers (see Martens and Dijk (2007) and Christensen and Podolskij (2007)) that has started to make use of intraday price high-low ranges to estimate the quadratic variation in standard continuous time diffusion models of (the logarithm of) an asset’s price. This work introduces a range-based estimator which is the so called “Realized Range”, and it also provides limit theory to show that the precision of the realized range is five times greater
than that of the realized volatility for quadratic variation estimation. Naturally, the extension of the “Garman and Klass estimator” to high-frequency situations for estimating the quadratic variation and quadratic covariation will lead to further efficiency gains.

The application of high-low prices to multivariate covariation estimation is, however, challenging. As discussed in Bannouh, Duk and Martens (2009), the high-low prices of different assets in a given interval are likely to occur at different time points, so that we can’t simply take the cross-products of intraday high-low ranges to estimate covariance as we did for the return based estimator. This challenge was unresolved until the appearance of recent papers by Brandt and Diebold (2006) and Bannouh, Dijk and Martens (2009). Brandt and Diebold (2006) proposed the use of range-based volatility estimates for a portfolio and its components to estimate the covariance of these components. Given that the high-low price observations of the portfolio can be obtained from the market, Brandt and Diebold (2006)'s idea makes it possible for the multivariate issue to be solved using existing univariate methodology. Bannouh, Dijk and Martens (2009) adapted this idea to high-frequency financial data and introduced “Realized Co-range” for quadratic covariation estimation. The Monte Carlo simulation implemented in Bannouh, Dijk and Martens (2009) reveals that in a frictionless market, the realized co-range is five times more efficient than the realized covariance. Furthermore, the realized co-range continues to have attractive properties in the presence of market microstructure noise, because its upward bias from bid-ask bounce and its downward bias from infrequent and nonsynchronous trading partially offset each other, while realized covariance is downward biased due to nonsynchronous trading, without any source of upward bias.

The main contribution of this paper is that we extend the “Garman and Klass estimator” to the multivariate high-frequency context to construct a more efficient quadratic covariation estimator than the existing return-based and range-based estimators, and we call this the “first-high-low-last (FHLL) prices based covariance estimator” below. We employ this new estimator to develop a new co-jump test. Our simulation results for a variety of underlying asset price process specifications strongly suggest that this new FHLL covariance estimator is more efficient than its return-based and range-based competitors. Given the relative efficiency of the first-high-low-last prices based covariance estimator, we expect the first-high-low-last price based co-jump test to be more powerful than the Bollerslev et al (2008) return-based co-jump test.

We conduct Monte Carlo simulations to obtain the null distribution of the first-high-low-last price based co-jump test and its critical values for practical use. Then, we inspect the finite sample properties of our first-high-low-last
price based co-jump test, and compare it to the existing return-based co-jump test in terms of its power properties. Finally, we study an empirical example based on a panel of high-frequency data from the Chinese mainland stock market. We use this data because the institutional settings and trading rules in this market are largely different from that in Western developed stock markets, and our previous research (Liao, 2008) based on three medical stocks has found that the jumps in individual stocks of this emerging market are more frequent than those in developed financial markets. It is therefore interesting to examine the pattern of co-jumps in this particular market.

The rest of the paper is organized as follows. Section 2 introduces our first-high-low-last price based covariance estimator and analyzes its efficiency. Section 3 applies this new estimator to develop our first-high-low-last price based co-jump test. Section 4 conducts a Monte Carlo simulation to study the finite sample performance of our new co-jump test, and compares its power properties with those of the existing return-based co-jump test. Section 5 presents our main empirical findings in a panel of stocks from the Chinese mainland stock market. Section 7 concludes.

2 First-High-Low-Last Price Based Estimators

We let $p_s$ denote the log price of an asset, and assume that it evolves as a standard continuous time diffusion process

$$dp_s = \mu(s)ds + \sigma(s)dW_s,$$

where $\mu(s)$ and $\sigma(s)$ denote the drift and local volatility respectively, and $W_s$ is a standard Brownian motion.

We assume that high-frequency data are available for each day $t$ which runs from time $t - 1$ to $t$, and that we have prices relating to $M$ intraday periods denoted by $\{t_j\}$ for $j = 1, \ldots, M$, where $t_j \in [t - 1, t]$. In addition, we have $m + 1$ equally spaced price observations recorded within each intraday time interval $[t_{j-1}, t_j]$ (i.e. at $t_{(j-1)+(0/m)}$, $t_{(j-1)+(1/m)}$, $\ldots$, $t_{(j-1)+(m/m)} = t_j$). The four extreme price values within each intraday time interval $[t_{j-1}, t_j]$ are:

- $p_{t_{j-1}}$: the first (log) price observed during the time interval $[t_{j-1}, t_j]$;
- $p_{t_j}$: the last (log) price observed during the time interval $[t_{j-1}, t_j]$;
- $h_{t_{j-1}}$: the highest (log) price observed during the time interval $[t_{j-1}, t_j]$;

which is $\max\{t_{(j-1)+(0/m)}, t_{(j-1)+(1/m)}, \ldots, t_{(j-1)+(m/m)} = t_j\}$.

$^1M$ is the number of the intraday periods over a trading day, and each of these $M$ intraday periods is divided into $m$ subintervals.
\( l_{t_j-1} \): the lowest (log) price observed during the time interval \([t_{j-1}, t_j]\), which is \( \min\{t_{(j-1)+(0/m)}, t_{(j-1)+(1/m)}, ..., t_{(j-1)+(m/m)} = t_j\} \).

### 2.1 First-High-Low-Last Price Based Variance Estimator

The most popular approach to estimate the integrated variance \( \int_{t-1}^t \sigma^2(s)ds \) of the above standard continuous time diffusion process is to use “Realized Volatility”, which is constructed using the sum of squared interval returns via

\[
RV_t = \sum_{j=1}^{M} r_{t_j}^2 = \sum_{j=1}^{M} (p_{t_j} - p_{t_{j-1}})^2, \tag{1}
\]

where the return \( r_{t_j} \) of each time subinterval is calculated as the difference between the last price and the first price of that interval. Andersen et al (2001) and Barndorff-Nielsen et al (2004) have proved that realized variance is a consistent estimator for the integrated variation over \([t-1, t]\) in the absence of microstructure noise, and that the asymptotic variance of realized volatility is \( 2 \int_{t-1}^t \sigma^4(s)ds \).

The range of an asset’s price is defined to be the difference between the highest price and the lowest price during a fixed time interval. The use of the high-low price range in volatility estimation dates back to Parkinson (1980). Recently Christensen and Podolskij (2007) and Martens and Dijk (2007) have re-considered the use of price range in a high frequency data context to estimate the integrated variation in a standard continuous time diffusion model of (the logarithm of) an asset’s price as

\[
RRV_t^{M,m} = \frac{1}{\gamma_{2,m}} \sum_{j=1}^{M} s_{p_{t_j}}^2 = \frac{1}{\gamma_{2,m}} \sum_{j=1}^{M} (h_{t_{j-1}} - l_{t_{j-1}})^2, \tag{2}
\]

where \( \gamma_{2,m} = E[s_{w,m}^2] \) and \( s_{w,m} \) is the range of a standard Brownian motion over a unit time interval \([0, 1]\), when we observe \( m \) increments of the underlying continuous time process in each sampling interval \( t_j \). The parameter \( \gamma_{2,m} \) is monotonically increasing in \( m \) with \( \gamma_{2,1} = 1 \), and \( \gamma_{2,m} \to 4 \ln 2 \) as \( m \to \infty \).

Intuitively, the range reveals more information than the return over the same time interval because the highs and lows of asset prices are formed from the entire price evolution path. Parkinson (1980) provided mathematical derivations to show that the daily squared price range is about five times more...
efficient than the daily squared return for estimating daily volatility. Simulations in Martens and Dijk (2007) demonstrated that in a frictionless market without microstructure noise, realized range has a lower mean-squared error than realized volatility. This was corroborated by the asymptotic properties derived by Christensen and Podolskij (2007). They deduced the following central limit theorem for realized range,

$$\sqrt{M}(RRV_t^{M,m} - \int_{t-1}^{t} \sigma^2(s)ds) \longrightarrow MN(0, \frac{\gamma_{4,m} - \gamma_{2,m}^2}{\gamma_{2,m}^2} \int_{t-1}^{t} \sigma^4(s)ds),$$

where $MN(.,.)$ denotes a mixed Gaussian distribution, $\gamma_{r,m} = E[s_{w,m}^r]$ and $\lim_{m \to \infty} \frac{\gamma_{4,m} - \gamma_{2,m}^2}{\gamma_{2,m}^2} \approx 0.4073$. They used this theorem to show that $RRV_t^{M,m}$ is an unbiased and more efficient estimator of integrated variance than realized volatility. If $m = 1$, $RRV_t^{M,m}$ is actually equal to realized volatility $RV_t$ and $\frac{\gamma_{4,m} - \gamma_{2,m}^2}{\gamma_{2,m}^2} = 2$, but when $m \to \infty$ and the entire sample path of the price process is available, $RRV_t^{M,m}$ becomes about five times more efficient than $RV_t$. In practice, inference is typically drawn from discrete data and true ranges are not actually observed. Thus, the efficiency of the $RRV_t^{M,m}$ estimator relative to $RV_t$ depends on how many observations are used to construct the high-low price range measures.

The above two estimators are generated by either the intraday first and last prices or the intraday highest and lowest prices, and it is useful to combine these four types of prices together to further improve estimation efficiency. Garman and Klass (1980) did this in a daily data context, by utilizing daily open, high, low and close prices to derive a minimum variance unbiased estimator for daily volatility given by

$$\hat{\sigma}_t^2 = 0.511(\log H_t - \log L_t)^2 - 0.383(\log C_t - \log C_{t-1})^2$$

$$-0.019((\log H_t - \log C_t)\log H_t + (\log L_t - \log C_t)\log L_t),$$

where $H_t$, $L_t$, and $C_t$ are respectively the highest, lowest and close prices during day $t$. They recommended a simpler version of this estimator for practical use, which is

$$\hat{\sigma}_t^2 = 0.5(\log H_t - \log L_t)^2 - (2\log(2) - 1)(\log C_t - \log C_{t-1})^2,$$

and this latter estimator achieves similar efficiency but eliminates the small cross-product terms. Their calculations showed that the variance of this estimator is $0.27\sigma^4$, which is 7.4 times more efficient for daily volatility estimation than the daily squared return (whose variance is $2\sigma^4$), and 1.5 times more efficient than daily squared range (whose variance is $0.41\sigma^4$).
We can replace intraday returns or intraday ranges in high frequency realized variance or realized range estimators with the intraday versions of “Garman and Klass estimators” to construct a first-high-low-last (FHLL) price based estimator for integrated variance given by

$$FHLLV_t = \sum_{j=1}^{M} (0.5(h_{t_j} - l_{t_j})^2 - (2\log(2) - 1)(p_{t_j} - p_{t_{j-1}})^2).$$  \tag{3}$$

This estimator is essentially a linear combination of \(RRV_t\) and \(RV_t\) with weights of \((2\ln2)\) and \((1 - 2\ln(2))\) respectively.\(^2\) Assuming no microstructure noise and that the entire price path can be observed \((m \to \infty)\), one can derive a central limit theorem for this FHLL variance estimator with respect to \(M\), which is

$$\sqrt{M}(FHLLV_t - \int_{t-1}^{t} \sigma^2(s)ds) \longrightarrow MN(0, 0.27 \int_{t-1}^{t} \sigma^4(s)ds).$$

The derivation details are provided in Appendix 1. This shows that the FHLL price based estimator is a consistent estimator for integrated variance, but it is more efficient than either realized variance or realized range.

### 2.2 First-High-Low-Last Price Based Covariance Estimator

The fact that the FHLL estimator for integrated variance is more efficient than its realized volatility and realized range counterparts suggests that the use of the first, high, low and last values of asset prices might be advantageous in other settings as well. We now apply this idea to covariance estimation.

Assuming that there are two assets \(i\) and \(l\) for simplicity, and a portfolio of them with weights \(w_i\) and \(w_l = 1 - w_i\), Brandt and Diebold (2006) noted that the daily covariance between asset \(i\) and asset \(l\) can be obtained from

$$Cov(r_{i,t}, r_{l,t}) = \frac{1}{2w_iw_l}(Var[r_{p,t}] - w_i^2Var[r_{i,t}] - w_l^2Var[r_{l,t}]),$$

where \(Var[r_{p,t}]\) is the daily variance of the portfolio returns, and \(Var[r_{i,t}]\) and \(Var[r_{l,t}]\) are the daily variances of assets \(i\) and \(l\) respectively. Using the realized variance defined in (1) to estimate the three daily variances on

\(^2\)Note that this is an affine combination (weight coefficients add up to 1), but it is not a convex combination because the weight coefficient of \(RV\) is negative.
the right-hand side of the above equation, realized covariance (see Barndorff-
Neilsen et al (2004) ) can be calculated using

\[ RCV_t = \sum_{j=1}^{M} r_{i,t,j} r_{l,t,j} = \frac{1}{2w_i w_l} (RV_{p,t} - w_i^2 RV_{i,t} - w_l^2 RV_{l,t}) \] (4)

where \( RV_{p,t} \) is the realized variance of the portfolio, \( RV_{i,t} \) and \( RV_{l,t} \) are the realized variances of asset \( i \) and asset \( l \), and \( r_{i,t,j} \) and \( r_{l,t,j} \) are the intraday returns of asset \( i \) and asset \( l \). Using the realized range defined in (2) to estimate the three daily variances on the right-hand side of the above equation, realized co-range (see Bannouh, Dijk and Martens (2009)) can be obtained as

\[ RCR_t = \frac{1}{2w_i w_l} (RRV_{p,t} - w_i^2 RRV_{i,t} - w_l^2 RRV_{l,t}) \] (5)

where \( RRV_{p,t} \) is the realized range of the portfolio, and \( RRV_{i,t} \) and \( RRV_{l,t} \) are the realized ranges of asset \( i \) and asset \( l \). The Monte Carlo work in Bannouh, Dijk and Martens (2009) demonstrates that the realized range is robust to market microstructure noise arising from bid-ask bounce, infrequent trading and nonsynchronous trading, yet it is also highly efficient, delivering up to fivefold efficiency gains relative to realized covariance. Comparison of the theoretical properties of realized covariance and realized co-range is a subject of on-going research.

The first-high-low-last (FHLL) price based covariance estimator can be generated in an analogous fashion to (4) and (5), by using the FHLL variance estimator defined in (3) to estimate the three daily variances in the covariance equation to obtain

\[ FHLLCV_t = \frac{1}{2w_i w_l} (FHLLV_{p,t} - w_i^2 FHLLV_{i,t} - w_l^2 FHLLV_{l,t}) \] (6)

Given the superiority of the FHLL variance estimator over the return-based and range-based competitors, we expect this FHLL covariance estimator to be more efficient than the realized covariance and realized co-range estimators.

### 2.3 Comparison of the Properties of Covariance Estimators

In this section, we use Monte Carlo simulations to investigate the performance of our FHLL covariance estimator given a variety of underlying asset price process specifications. Throughout, we compare the FHLL estimator with the realized covariance estimator and the realized range estimator in these controlled environments.
2.3.1 Constant Volatility without Microstructure Noise

As a baseline case for further analysis, we firstly study the properties of these estimators for a bivariate Brownian motion process with constant volatility. We simulate prices for two correlated assets for 4-hour trading days. For each trading day \( t \), the initial prices for both assets are set equal to one and subsequent log prices for assets are simulated using

\[
dp^*_{1,(t-1)+h/K} = \sigma_1 dW_{1,(t-1)+h/K}, \quad h = 1, 2, \ldots, K
\]

\[
dp^*_{2,(t-1)+h/K} = \sigma_2 (\rho dW_{1,(t-1)+h/K} + \sqrt{1-\rho^2} dW_{2,(t-1)+h/K}), \quad h = 1, 2, \ldots, K
\]

where \( p^*_{i,(t-1)+h/K} \) is the log price of asset \( i \) at the \( h \)th point in the time interval \([t-1, t]\). \( K \) is the number of time increments over the day, \( \sigma_i \) is the standard deviation of asset \( i \), \( W_1 \) and \( W_2 \) are two Brownian motion processes, and \( \rho \) represents the contemporaneous correlation of the two assets’ prices. We set \( \sigma_1 = 0.2 \), \( \sigma_2 = 0.4 \) and \( \rho = 0.5 \), resulting in a constant covariance between the two asset returns which is equal to 0.04 for each day \( t \). We employ the Euler discretization scheme to generate realizations of the above two Brownian motions and record new price observations for the two assets every second, so that \( h \) counts seconds and \( K = 14400 \) (4 \times 60 \times 60). Each of our experiments is based on 5000 simulated days. For the time being our price observations are equidistant and occur synchronously for the two assets. To show the potential merits of using intraday first-high-low-last price for measuring the co-movement of two assets, we compute and compare the bias and root mean squared error (RMSE) of various covariance estimators at different intraday sampling frequencies.

To do this, we divide the trading day \( t \) into \( \Delta \) minute intervals, which is referred to as the \( \Delta \) - minute frequency below. Since we have a four hour trading day, this divides each day into \( M = 240/\Delta \) intraday sampling periods. For example, if we sample at a five minute frequency and \( \Delta = 5 \), then we have \( M = 48 \) intraday sampling periods per day. In our experiment, we vary the sampling frequency, using 1, 5, 10, 15, 30, 60, and 240 minute intervals, and results are reported in Table 1 panel A. The underlying price process \( p^*_{i,(t-1)+h/K} \) in our simulations is assumed to be a pure Brownian motion with constant volatility, but since it is actually discrete (\( h \) can only takes integer values), we see an "infrequent trading" effect, which leads to a downward bias for all estimators when the sampling intervals are relatively short.

\[3\text{We choose to simulate four hour trading days to reflect the trading hours in the Chinese mainland stock market, from which our empirical data are collected.}\]

\[4\text{True and observed prices are denoted by } p^*_i \text{ and } p_i \text{ respectively. In the next section we set the probability of actually observing the price to be } p_{obs} = 1/\tau, \text{ and use } s \text{ to denote the bid-ask spread. For the time being, } \tau = 1 \text{ and } s = 0.\]
We explore the effects of infrequent trading in more detail below. However, the RMSE of the FHLL covariance estimator is always lower than that of realized co-range, and substantially lower than that of realized covariance at the same sampling frequency. Meanwhile, the efficiency improves for all estimators as $\Delta$ decreases and $M$ increases. Figure 1 shows the kernel density graphs of the three covariance estimators at 5-minutes, 10-minutes and 15-minutes sampling frequencies, which further demonstrate that our FHLL estimator is more efficient than the other two, since the kernel density graph of FHLL estimator is narrower than those of the other two estimators.

2.3.2 Stochastic Volatility without Microstructure Noise

The volatility for the two asset prices was assumed to be constant over time in last subsection. In this subsection, we extend the underlying price processes to a system with stochastic volatility, which is closer to reality. The (log) prices now evolve as

$$dp^*_{1,(t-1)+h/K} = \sigma_{1,(t-1)+h/K} dW_{1,(t-1)+h/K}, \quad h = 1, 2, ..., K$$

$$dln\sigma^2_{1,(t-1)+h/K} = \theta_1(\omega_1-ln\sigma^2_{1,(t-1)+h/K})dt + \eta_1 dW_{2,(t-1)+h/K}, \quad h = 1, 2, ..., K$$

$$dp^*_{2,(t-1)+h/K} = \sigma_{2,(t-1)+h/K} dW_{3,(t-1)+h/K}, \quad h = 1, 2, ..., K$$

$$dln\sigma^2_{2,(t-1)+h/K} = \theta_2(\omega_2-ln\sigma^2_{2,(t-1)+h/K})dt + \eta_2 dW_{4,(t-1)+h/K}, \quad h = 1, 2, ..., K$$

where the volatility is a stochastic process, which follows a mean reverting Ornstein-Uhlenbeck process with parameters $\theta_1$ and $\theta_2$ as the adjustment speeds, $\omega_1$ and $\omega_2$ as the means of the (log) volatilities, and $\eta_1$ and $\eta_2$ as the volatilities of the (log) volatilities. $W_1$ and $W_2$ are standard Brownian motions with a correlation of $\rho_1$, and $W_3$ and $W_4$ are standard Brownian motions with a correlation of $\rho_2$. $\rho_1$ and $\rho_2$ represent the “leverage effect”, which reflect the instantaneous correlations between the return process and the corresponding volatility process of each asset. Meanwhile, the $W_1$ and $W_3$ processes are also correlated with a correlation coefficient of $\rho_3$. The price processes of the two assets are simulated via the Euler scheme which we used in the last subsection. The initial prices for the two assets are set to be one, the initial values of the two assets’ volatilities are set to equal to the mean of the volatilities, and the rest of the simulations are based on the configuration $(\theta_1, \theta_2, \omega_1, \omega_2, \eta_1, \eta_2, \rho_1, \rho_2, \rho_3) = (2, 2, 0.04, 0.16, 0.4, 0.8, -0.6, -0.4, 0.5)$.\footnote{We set $\omega_1, \omega_2$ and $\rho$ equal to the variances $\sigma^2_1$ and $\sigma^2_2$ of the two assets’ prices, and the correlation $\rho$ between the two assets’ prices that we used in last subsection. We set the rest of the parameters according to the simulation study in Aït-Sahalia, Fan and Xiu (2010).}
We simulate 5000 days of (log) prices (1 price per second, for \( K = 14400 \)) as before, and then compute and compare the bias and root mean squared error (RMSE)\(^6\) of our various estimators of daily covariance. The data generating process ensures that the daily mean of the covariance between the two assets is 0.04, but it now varies every second.

Table 1 panel B reports the results. Relative to the results in the last subsection, the main difference is that all estimators are now slightly less efficient. From Section 2.1, the asymptotic variance of all three variance estimators is given by \( \alpha \int_{t-1}^{t} \sigma_s^4 ds \), where \( \alpha \) reflects the relative efficiencies of the different estimators, and this holds true regardless of whether \( \sigma_s \) is constant or time varying. Thus, the ranking of the variance (and hence covariance) estimators in terms of efficiency is unaltered once we have time-varying volatility. Relative to the constant volatility case, time-varying in asset price volatility tends to increase \( \sigma^4 \) and hence increase the asymptotic variances of all variance (and covariance) estimators and decrease efficiency, but the FHLL covariance estimator is still more efficient than the other two estimators, at any sampling frequency.

### 2.3.3 Stochastic Volatility with Microstructure Noise

We did not include microstructure noise in the previous experiments, but in this section we compare the three estimators when they are contaminated by the effects of the bid-ask bounce, infrequent trading and nonsynchronous trading. We compare our FHLL covariance estimator with the realized co-range and realized covariance estimators, both with and without corrections for estimation bias resulting from the presence of microstructure noise.

Following Bannouh, Dijk and Martens (2009), we consider the effects of bid-ask bounce by assuming that bid and ask prices occur with equal probability. Hence, the actually observed price \( p_{i,(t-1)+h/K} \) is equal to \( p^*_{i,(t-1)+h/K} + s/2 \) (ask) or \( p^*_{i,(t-1)+h/K} - s/2 \) (bid), where \( s \) is the bid-ask spread and \( p^*_{i,(t-1)+h/K} \) is the true price obtained from subsection 2.3.2. Infrequent trading is simulated by filtering the price sample path \( p^*_{i,(t-1)+h/K} \) simulated from subsection 2.3.2, so that the price of each asset is observed on average only every \( \tau \) seconds. Since price observations for the two assets occur independently, we observe prices nonsynchronously.

We use the simulations in Table 1 Panel B as a benchmark and change the values of \( s \) and \( \tau \) in our simulations to consider three pairs in which \( s = 0.075 \) and \( \tau = 1 \), \( s = 0 \) and \( \tau = 15 \), and \( s = 0.075 \) and \( \tau = 15 \). The first two pairs of settings are used to examine the separate effects of bid-ask

---

\(^6\)Since the covariance is time-varying in this scenario, the bias reported in Table 1 is actually the mean of the covariance estimation bias.
bounce and infrequent trading, while the last setting is used to investigate their joint effects on all the covariance estimators.

It is well known that when continuous underlying price processes are observed only at discrete time points, the intraday range suffers from a downward bias. This is because the observed maximum and minimum prices over a given intraday interval underestimate and overestimate the true maximum and minimum, respectively. Meanwhile, the intraday range also tends to be upward biased due to the presence of bid-ask bounce. For example, when the sampling frequency is relatively high and the intraday time interval is relatively small, the observed high price in a given intraday interval is an ask price and the observed low price is a bid price with probability close to one. The squared intraday range therefore overestimates the true variance of that intraday interval by an amount equal to the squared bid-ask spread $s^2$. Although univariate intraday returns are not effected much by infrequent trading and bid-ask spread, an important concern in a multivariate setting is the presence of nonsynchronous trading. As different assets trade at different times, estimates of their covariance are biased toward zero. This is the so-called “Epps effect”, which becomes worse with an increase of sampling frequency.

We correct this bias by assuming that the observed log price $p_t$ is equal to the underlying log price $p_t^*$ plus an additive noise term, and then employ an additive bias-correction method discussed in Bannouh, Dijk and Martens (2009). These authors define bias-corrected variance estimators as

\[ VE_{C,t}^M = VE_t^M + \frac{1}{Q} \left( \sum_{q=1}^{Q} VE^1_{t-q} - \sum_{q=1}^{Q} VE_t^M \right), \]

where $VE_t^1$ is the daily squared return or daily squared range or daily “Garman and Klass estimator”, and $VE_t^M$ is the realized variance, realized range or our FHLL variance estimator based on $M$ intraday sampling intervals. The number of trading days $Q$ used to compute the correction is a critical choice to make. The RMSE of all the estimators decline as $Q$ increases and we set $Q = 150$, beyond which the RMSEs for the corrected version of all the estimators more or less stabilize.

We set $s = 0$ and $\tau = 15$ to obtain Table 2 panel A, which shows the effects of infrequent (and hence nonsynchronous) trading on all three covariance estimators. As expected, all three non-corrected covariance estimators are downward biased, but realized covariance is downward biased much less than the realized co-range and FHLL covariance estimators. The RMSE first decreases for all the estimators when increasing the sampling frequency and decreasing the length of the sampling interval, but it increases again.
for higher frequencies because the bias associated with microstructure noise outweighs the increase in information from the higher sampling frequency. Without the bias correction, our FHLL estimator has larger RMSE than the other two estimators at most sampling frequencies. This is not surprising because the FHLL estimator is a linear combination of realized covariance and realized co-range, which is contaminated more by the microstructure noise than the other two estimators. However, the correction scheme eliminates the bias to a large extent and reduces the RMSE of our FHLL estimator substantially. More importantly, the bias-corrected FHLL estimator \( FHLLCV_{C,t} \) has the smallest RMSE at all sampling frequencies.

Table 2 panel B demonstrates the influence of bid-ask bounce on the three covariance estimators by setting \( s = 0.075 \) and \( \tau = 1 \). We set \( \tau = 1 \) in this panel, so that results can be compared with those in Table 1 Panel B. Our FHLL covariance estimator and the realized co-range suffer from a strong upward bias in this scenario, which becomes worse with increased sampling frequency, but realized covariance is not affected much by the bid-ask spread. The bias correction reduces the RMSE of the first-high-low-last price based covariance estimator considerably, such that \( FHLLCV_{C,t} \) is more accurate than \( RCV_{C,t} \) and \( RCR_{C,t} \) for all sampling frequencies.

When bid-ask spread and nonsynchronous trading are jointly present, as in the set-up of Table 2 panel C, we find that our FHLL covariance estimator and the realized co-range still have an upward bias, but it is much smaller than that in the case of bid-ask spread only. This finding is consistent with the discussion in Bannouh, Dijk and Martens (2009), who suggest that the upward bias due to the presence of bid-ask bounce has been partially offset by the downward bias due to nonsynchronous trading. As observed in the last two panels, the bias in all estimators has been largely removed by the correction adjustment. Meanwhile, the bias corrected FHLL estimator \( FHLLCV_{C,t} \) has the minimum RMSE at all sampling frequencies.

### 3 The Co-jump Test

In this section, we review the return-based co-jump test proposed by Bollerslev et al (2008), and use our FHLL covariance estimator to develop a first-high-low-last (FHLL) prices based co-jump test statistic.

#### 3.1 Co-jumps in Portfolio Theory

Portfolio theory implies that idiosyncratic jumps can be diversified away in a large portfolio, and only the common jumps remain. In this section we
sketch the derivation provided in Bollerslev et al (2008), that highlights the role of co-jumps in portfolios. This derivation is based on an equiweighted portfolio, but all of the arguments presented below hold with equal force for any well-diversified portfolio.

Bollerslev et al (2008) started with the consideration of a collection of \( n \) stock price processes \( \{p_{i,s}\}_{i=1}^{n} \) evolving in continuous time. Each \( p_{i,s} \) evolves as

\[
dp_{i,s} = \mu_{i}(s)dt + \sigma_{i}(s)dW_{i}(s) + dL_{i}(s),
\]

where \( \mu_{i}(s) \) and \( \sigma_{i}(s) \) refer to the drift and local volatility, \( W_{i}(s) \) is a standard Brownian motion, and \( L_{i}(s) \) is a pure jump process. In practice, the price process is only available at discrete time points. Let \( M+1 \) denote the number of equidistant price observations each day, which is determined by the sampling frequency. Then, the \( jth \) within-day return of the \( ith \) log-price process on day \( t \) is defined by

\[
r_{i,tj} = p_{i,(t-1)+\frac{j}{M}} - p_{i,(t-1)+\frac{j+1}{M}}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, M.
\]

The \( jth \) within-day return on day \( t \) of an equiweighted portfolio of \( n \) stocks can be calculated by

\[
r_{EQW,tj} = \frac{1}{n} \sum_{i=1}^{n} r_{i,tj}.
\]

The daily realized variance for this equiweighted portfolio then satisfies

\[
RV_{EQW,t} = M \sum_{j=1}^{M} \left( \frac{1}{n} \sum_{i=1}^{n} r_{i,tj} \right)^{2} = \frac{1}{n^{2}} \sum_{i=1}^{n} \sum_{j=1}^{M} r_{i,tj}^{2} + \frac{1}{n^{2}} \sum_{i=1}^{n} \sum_{l=1,l\neq i}^{n} \sum_{j=1}^{M} r_{i,tj} r_{l,tj}
\]

\[
\frac{M \rightarrow \infty}{\frac{1}{n^{2}} \sum_{i=1}^{n} \int_{t-1}^{t} \sigma_{i}^{2}(s)ds} + \frac{1}{n^{2}} \sum_{i=1}^{n} \sum_{k=1}^{N_{i,t}} \kappa_{i,t,k}^{2}
\]

\[
+ \frac{1}{n^{2}} \sum_{i=1}^{n} \sum_{l=1,l\neq i}^{n} \int_{t-1}^{t} \sigma_{i}(s)\sigma_{l}(s)ds + \frac{1}{n^{2}} \sum_{i=1}^{n} \sum_{l=1,l\neq i}^{n} \sum_{k=1}^{N_{(i,l),t}} \kappa_{i,t,k} \kappa_{l,t,k},
\]

where \( N_{i,t} \) is the number of jumps occurring in the \( ith \) stock on day \( t \), and \( N_{(i,l),t}^{*} \) is the number of simultaneous co-jumps occurring across any two stocks (the \( ith \) stock and the \( lth \) stock) on day \( t \). It is clear that when \( M \) goes to infinity, the realized variance of this portfolio provides a consistent estimator for the quadratic variation of the equiweighted portfolio’s price process in continuous time. This variation is mostly the quadratic variation of the
equally weighted portfolio of the \( n \) stocks, that is
\[
\frac{1}{n^2} \sum_{i=1}^{n} \int_{t-1}^{t} \sigma_i^2(s) ds + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{k=1}^{N_{i,t}} \kappa_{i,t,k}^2
\]
which can be estimated by
\[
\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{M} \tilde{r}_{i,j,t},
\]
and the quadratic covariation of the prices of the \( n \) stocks in the portfolio, that is
\[
\frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1, l \neq i}^{n} \sigma_i(s) \sigma_l(s) ds + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1, l \neq i}^{n} \sum_{k=1}^{N_{i,l}^*} \kappa_{i,l,k} \kappa_{l,t,k},
\]
which can be estimated using
\[
\frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1, l \neq i}^{n} \sum_{j=1}^{M} r_{i,t} r_{l,t,j}.
\]

As the above equation shows, both the overall quadratic variation of a collection of \( n \) stock price processes \( \{p_{i,s}\}_{i=1}^{n} \) and their quadratic covariation consist of the continuous integrated part plus the sum of the jumps. In order to separately measure the two components, realized bipower variation (see Barndorff-Nielsen and Shephard (2004)) is employed to measure the continuous integrated variance of the equally weighted portfolio’s price process as
\[
BV_{EQW,t} = \mu_1^{-2} \left( \frac{M}{M-1} \right) \sum_{j=2}^{M} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{r}_{i,j-1} \right|^2 \cdot \frac{1}{n} \sum_{i=1}^{n} \tilde{r}_{i,t,j}
\]
\[
M \to \infty \frac{1}{n^2} \sum_{i=1}^{n} \int_{t-1}^{t} \sigma_i^2(s) ds + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1, l \neq i}^{n} \int_{t-1}^{t} \sigma_i(s) \sigma_l(s) ds,
\]
where \( \mu_1 = \sqrt{2/\pi} \). Thus, the contribution of jumps to the total variation can be estimated by taking the difference between \( RV_{EQW,t} \) and \( BV_{EQW,t} \) to obtain
\[
RV_{EQW,t} - BV_{EQW,t} M \to \infty \frac{1}{n^2} \sum_{i=1}^{n} \sum_{k=1}^{N_{i,t}} \kappa_{i,t,k}^2 + \frac{1}{n^2} \sum_{i=4}^{n} \sum_{l=1, l \neq i}^{n} \sum_{k=1}^{N_{i,l}^*} \kappa_{i,l,k} \kappa_{l,t,k},
\]
\[
= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{k=1}^{N_{i,t}} \kappa_{i,t,k}^2 + \frac{n-1}{n} \sum_{k=1}^{N_{i,t}^*} \kappa_{i,t,k} \kappa_{i,t,k}.
\]
where \( \kappa_{i,t} \) is the average jump component for each individual stock on day \( t \), \( \kappa_{i,t} \kappa_{i,t} \) is the average cross-product of co-jumps that occur during that day and \( N_{i,t}^* \) is the number of co-jumps occurring during day \( t \). The first of these terms approaches zero as \( n \) goes to infinity, consistent with the idea that idiosyncratic jumps in a large portfolio are diversified away. The second of these terms remains, since \( \frac{n-1}{n} \approx 1 \) for large \( n \). Thus we have
\[
RV_{EQW,t} - BV_{EQW,t} \approx \sum_{k=1}^{N_{i,t}^*} \kappa_{i,t,k} \kappa_{i,t,k} \approx \sum_{k=1}^{N_{i,t}^*} \kappa_{i,t,k} \kappa_{i,t,k},
\]
where \( \kappa_{i,t} \) denotes the average co-jump on day \( t \), and this shows that in a large portfolio, only co-jumps can cause the price of the portfolio to jump. It is useful to note that these co-jumps originate from the
\[
\frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1, l \neq i}^{n} \sum_{j=1}^{M} \tilde{r}_{i,t} \tilde{r}_{l,t,j}.
term in the expression for $RV_{EQW,t}$, and it was this observation that motivated Bollerslev et al (2008) to emphasize the cross-covariance measures associated with a portfolio when they developed their new co-jump identification procedure. As shown above, the variances of individual stocks (and their individual jump components) play a rather minor role, when considering the covariance structure of the entire portfolio.

### 3.2 Return-based Co-jump Test

The first step in applying the Bollerslev et al (2008) co-jump test to a panel of stocks is to obtain cross-covariance measures for returns that can assess the comovement in these stocks. In terms of Section 3.1’s notation and assuming that there are $n$ individual stocks and an equally weighted portfolio as in Section 3.1, Bollerslev et al (2008) used a return-based estimator that summed the cross-products of intraday individual returns, i.e. $2\sum_{i=1}^{n-1} \sum_{l=i+1}^{n} r_{i,t} r_{l,t}$, to estimate the leading term in the covariance expression over the $j$th intraday interval on day $t$. Then they defined a mean cross-product (mcp) test statistic to be

$$mcp_{tj} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{l=i+1}^{n} r_{i,t} r_{l,t}, \quad j = 1, 2, \ldots, M,$$

where $M$ is the number of intraday returns observed over a trading day, $i$ and $l$ index the $n$ stocks, and $n(n-1)$ is the number of cross-product terms of $n$ stock returns over the $j$th intraday interval. The above $mcp_{tj}$ test statistic can be rearranged to obtain

$$mcp_{tj} = \frac{1}{n(n-1)} \left( \frac{1}{n} \sum_{i=1}^{n} r_{i,t} \right)^2 - \frac{1}{n^2} \sum_{i=1}^{n} \left( \frac{1}{n} r_{i,t}^2 \right)$$

$$= \frac{n}{n-1} \left( \frac{1}{n} \sum_{i=1}^{n} r_{i,t} \right)^2 - \frac{n}{n^2} \sum_{i=1}^{n} \left( \frac{1}{n} r_{i,t}^2 \right).$$

The first term, $(\frac{1}{n} \sum_{i=1}^{n} r_{i,t})^2$ can be used to estimate the variance of the equally weighted portfolio return over the $j$th intraday interval and $r_{i,t}^2$ can be used to estimate the variance of the $i$th asset return over the $j$th intraday interval. Thus, the $mcp_{tj}$ test statistic can also be written as

$$mcp_{tj} = \frac{n}{n-1} \left( (\frac{1}{n} \sum_{i=1}^{n} r_{i,t})^2 - \sum_{i=1}^{n} \left( \frac{1}{n} r_{i,t}^2 \right) \right) = \frac{n}{n-1} (\text{var}(\widetilde{r}_{EQW,t}) - \frac{1}{n^2} \sum_{i=1}^{n} \text{var}(\widetilde{r}_{i,t})).$$
which essentially contains the continuous variation and overall jump component in the portfolio, since the last term relating to individual assets disappears for large $n$. Since this test statistic downplays idiosyncratic risk and is therefore mostly influenced by systematic risk, it should be insensitive to idiosyncratic jumps in individual stocks, and very sensitive to co-jumps which occur simultaneously across all stocks.

The contribution of the continuous co-movement of the $n$ stock returns ensures that the $mcp$-statistic has a non-zero mean, even in the absence of co-jumps. Therefore, Bollerslev et al (2008) studentize the $mcp$ statistic to form a $zmcp$ test statistic given by

$$z_{mcp,t,j} = \frac{mcp_{t,j} - \overline{mcp}_t}{s_{mcp,t}}, \quad j = 1, 2, \ldots, M,$$

where

$$\overline{mcp}_t = \frac{1}{M} \sum_{j=1}^{M} mcp_{t,j} = \frac{1}{M} \left( \frac{n}{n-1} \sum_{j=1}^{M} \left( \frac{1}{n} \sum_{i=1}^{n} r_{i,t,j} \right)^2 - \frac{n}{n-1} \sum_{j=1}^{M} \sum_{i=1}^{n} \left( \frac{1}{n} r_{i,t,j} \right)^2 \right),$$

$$s_{mcp,t} = \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} (mcp_{t,j} - \overline{mcp}_t)^2},$$

and $RV_{ew,t}$ and $RV_{i,t}$ are the daily realized volatilities for the portfolio and individual stocks, obtained by summing all the intraday squared returns within a day. The calculation of $\overline{mcp}_t$ simply takes the sample mean of all the intraday $mcp$ test statistic realizations over the trading day $t$, and $s_{mcp,t}$ is the corresponding sample standard deviation.

The $zmcp$ co-jump test relies on three assumptions. Firstly, the studentization of the $mcp_{t,j}$ test statistic each day relies on an assumption that the location and scale of this statistic remains approximately constant over the day. This assumption may be at odds with the well-known U-shaped pattern associated with intraday stock volatility, but Bollerslev et al (2008) claim that their main conclusions are not influenced by taking this pattern into account. Secondly, the $mcp_{t,j}$ realizations are assumed to be serially uncorrelated, making it appropriate to simply standardize each of the within-day $mcp$ statistics by using the corresponding daily sample standard deviation $s_{mcp,t}$. The empirical data in Bollerslev et al (2008) is consistent with this assumption. Lastly, it is important to note that the sample mean used in
the zmcp test statistic incorporates the co-jump contribution relating to each day, and although the contribution of a few jumps on a day might be negligible, the contribution of several co-jumps is unlikely to be negligible and then relatively large intraday mcp_{ij} realizations might be masked by the correspondingly large sample mean mcp_{it}. Therefore, the test relies on an assumption that co-jumps occur rarely, and in particular that no more than one co-jump occurs during a day. This assumption found empirical support in Bollerslev et al (2008), and also in our empirical study on Chinese data (see Section 5).

It is not easy to derive the asymptotic distribution for this test. Hence, Bollerslev et al (2008) use simulations to obtain the distribution of their test statistic under the null hypothesis of no co-jumps. They find that the critical values of this test statistic are insensitive to the level of correlation between the returns and the number of stocks in the panel, as long as the number of stocks in the portfolio is large. However, they find that their critical values are quite sensitive to the number of intraday returns included in each daily measure. Since this number is determined by the sampling frequency, it is clear that the ability of this test is largely dependent on how much information we can extract from the high-frequency data. This motivates us to construct more powerful co-jump tests by adapting the Bollerslev et al (2008) test to incorporate more efficient estimators of variance and covariance. Accordingly, we propose a range-based co-jump test and a co-jump test based on first-high-low-last price measures of variance and covariance below.

### 3.3 First-High-Low-Last Price Based Co-jump Test

A range-based co-jump test statistics can be obtained when we use the realized co-range and intraday squared range instead of the cross-product of intraday returns and realized volatility in the zmcp test statistic. We refer to this as the zmcr test statistic below. Our range-based co-jump test statistic (zmcr) is defined by

\[
zmcr_{t, j} = \frac{mcr_{t, j} - mcr_{t}}{s_{mcr, t}}, \quad j = 1, 2, \ldots, M, \tag{8}
\]

where

\[
mcr_{t, j} = \frac{n}{n - 1} (ISR_{ew, t, j} - \sum_{i=1}^{n} (\frac{1}{n})^2 ISR_{i, t, j}),
\]

and ISR_{ew, t, j} and ISR_{i, t, j} are intraday squared ranges of the equally weighted portfolio and of each individual stock, which measure the intraday variance.
of this portfolio and each individual stock on day $t$, time $j$. We studentize the $mcr_{t,j}$ statistic using

$$mcr_{t} = \sum_{j=1}^{M} mcr_{t,j} = \frac{1}{M} \left[ \frac{n}{n-1} RRV_{ew,t} - \frac{1}{n(n-1)} \sum_{i=1}^{n} RRV_{i,t} \right]$$

and

$$s_{mcr,t} = \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} (mcr_{t,j} - mcr_{t})^2},$$

where $RRV_{ew,t}$ and $RRV_{i,t}$ measure the daily realized ranges of the equally weighted portfolio and each individual stock, and they are obtained by summing all the intraday squared ranges over the whole trading day.

Similarly, we also develop a first-high-low-last price based co-jump test, and do this by using the intraday FHLL covariance estimator and the daily FHLL covariance estimator to replace the cross-product of intraday returns and realized variance in Bollerslev et al (2008) $zmcp$ test statistic. Our FHLL co-jump test statistic ($zmfhllc$) is defined by

$$z_{mfhllc,t,j} = \frac{mfhll_{ct,j} - mfhll_{ct}}{s_{mfhllc,t}}, \quad j = 1, 2, \ldots, M,$$

where

$$mfhll_{ct,j} = \frac{n}{n-1}(IFHLLC_{ew,t,j} - \sum_{i=1}^{n} \left( \frac{1}{n} \right)^2 IFHLLC_{i,t,j}),$$

and $IFHLLC_{ew,t,j}$ and $IFHLLC_{i,t,j}$ are intraday FHLL variance estimators of the equally weighted portfolio and each individual stock. We studentize the $mfhll_{ct,j}$ statistic using

$$mfhll_{ct} = \sum_{j=1}^{M} mfhll_{ct,j} = \frac{1}{M} \left[ \frac{n}{n-1} FHLLCV_{ew,t} - \frac{1}{n(n-1)} \sum_{i=1}^{n} FHLLCV_{i,t} \right]$$

and

$$s_{mfhllc,t} = \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} (mfhll_{ct,j} - mfhll_{ct})^2},$$

where $FHLLCV_{ew,t}$ and $FHLLCV_{i,t}$ are the daily FHLL variance estimators of the equally weighted portfolio and each individual stock, obtained by summing all the intraday FHLL variance estimators over the whole trading day.

Prior to using these new co-jump test statistics, we conduct a set of Monte Carlo simulations to study their distributions under the null hypothesis of no co-jumps, and compare their finite sample properties with the return-based $zmcp$ test statistic.
4 Monte Carlo Simulation

In this section, we conduct a Monte Carlo simulation to explore the ability of our first-high-low-last price based test to detect co-jumps in a panel of high frequency data. The three test statistics involved in our analysis include the return-based co-jump test statistic ($z_{mcp}$) as defined in (7), the range-based co-jump test statistic ($z_{mcr}$) as defined in (8), and the FHLL co-jump test statistic ($z_{mfhllc}$) as defined in (9).

4.1 The Null Distribution via Simulation

We firstly conduct a series of Monte Carlo simulations to obtain the null distributions of our range-based co-jump $z_{mcr}$ test statistic and first-high-low-last price based $z_{mfhllc}$ test statistic, and then compare them with the null distributions of the $z_{mcp}$ test statistic of Bollerslev et al (2008). We use the basic Euler scheme to simulate realizations of a stochastic multivariate $20 \times 1$ pure diffusion model (without jumps) that has zero drift and a covariance matrix $\Sigma = C \Sigma'$, and follows the process given by

$$dp_s = C'dW(s).$$

The variable $p_s$ represents a $20 \times 1$ vector process that resembles a vector of logarithmic prices of 20 assets, and $W(s)$ is a $20 \times 1$ vector of independent standard Brownian motions at time $s$. We use empirical calibration and set the covariance matrix $\Sigma$ equal to the unconditional covariance matrix of the 30-seconds intraday returns of 20 stocks from the Chinese mainland stock market. This market trades for four hours each day, which corresponds to $T = 14400$ seconds per day, and we set the sample length so as to correspond to 90 days (about three months) and record price observations that correspond to once every 30 seconds.\(^7\)

For each trading day $[0, T]$, we let $(t_h)_{h \in \{0, \ldots, K\}}$ be an equispaced time discretization of the day, with a time step of $\Delta t = 30$ seconds, where $K = T/\Delta t = 480$, and we generate $Z_1, Z_2, \ldots, Z_K$ independently $N(0, I)$ distributed random vectors in $R^{20}$. The $20 \times 1$ vector of independent standard Brownian motions at times $0 = t_0 < t_1 < \ldots < t_K$ is then generated by setting $W(0)$ to be a $20 \times 1$ zero vector and calculating

$$W(t_{h+1}) = W(t_j) + \sqrt{\Delta t}Z_{h+1}, \quad h = 0, \ldots, K - 1.$$

\(^7\)Here we assume that there is no microstructure noise, and comment on how microstructure noise might affect results later.
The price sample paths for the 20 assets are then simulated by setting \( p(0) \) to be a \( 20 \times 1 \) zero vector and calculating

\[
p_{t+h+1} = p_t + C'_t W(t_{h+1}), \quad h = 0, \ldots, K - 1.
\]

We do this for \( t = \{1, 2, 3, \ldots, 90\} \). The top panel of Figure 2 shows 20 simulated price realizations that correspond to a typical day.

We replicate the above procedure 1000 times, and thereby obtain about 43.2 million simulated values under the null hypothesis of no jumps. Then, we calculate the \( zmcp \) test statistics, the \( zmcr \) test statistics and the \( zmfhllc \) statistics from these simulated values using different sampling frequencies, including 5 minutes, 10 minutes and 15 minutes (\( M = 48, 24, 16 \)).\(^8\) The corresponding number (\( m \)) of intraday subintervals used to compute the intraday range and intraday first-high-low-last price is then equal to \( m = 10, 20, \) and 30. We also compute the intraday range and intraday first-high-low-last price by using half of the available price observations and one third of the available price observations within the 5-minute intervals, 10-minute intervals and 15-minute intervals, that is, \( m = 5, 10, 15 \) and \( m = 3, 6, 10 \) to study the sensitivity of these test statistics to \( m \).\(^9\) It is useful to note that when \( m = 1 \), then the range equals the absolute return, so that the Bollerslev et al. (2008) \( zmcp \) test statistic can be regarded as a special case of the range-based \( zmcr \) test statistic. We employ the bias correction discussed in Section 2.3.3 to correct all the covariance estimators for bias arising from infrequent observations, and then calculate the three co-jump test statistics.

Figure 3 presents the simulated probability densities of the \( zmcp \) test statistics, \( zmcr \) test statistics and \( zmfhllc \) test statistics. All of these distributions are obviously non-Gaussian with a strong right skew, regardless of the sampling frequency. The null distributions of the \( zmfhllc \) test statistics have slightly shorter tails and lower peaks than the \( zmcr \) test statistics, and both of them have much shorter tails and lower peaks than the \( zmcp \) test statistics. Meanwhile, the null distributions of the \( zmfhllc \) test statistics have shorter tails and lower peaks as \( M \) decreases, or as \( m \) increases given the same \( M \). This also holds for the \( zmcr \) test statistics.\(^{10}\) Since large val-

\(^8\)\(^K\) is the number of generated intraday prices for each day, while \( M \) is the number of intraday samples taken over the day.

\(^9\)Christensen and Podolskij (2007) note that the entire sample path of the asset price process is unavailable in practice, so that inference is drawn from discrete data and the true price range is often unobserved. Therefore, the range-based estimator has varying degrees of efficiency over the return-based estimator depending on how many observations (it is \( m \) in our paper) that are used to construct the high-low range.

\(^{10}\)We don’t provide the results of sensitivity analysis to \( M \) and \( m \) for \( zmcr \) test statistics in Figure 3. They are available upon request.
ues of the test statistics discredit the null hypothesis of no co-jumps, we are mostly interested in the right tails of these distributions.

Table 3 reports the critical values of all the test statistics at the 0.1%, 1% and 5% significance levels. These results suggest that the critical value is quite sensitive to the sampling frequency ($M$) and the number of subintervals ($m$) involved in forming the intraday high-low prices. In particular, the critical values always rise as the intraday sampling frequency ($M$) increases, and fall as the number of subintervals ($m$) used for each calculation of the high-low price range or first-high-low-last price decreases, with some exceptions at 5% significance level.

### 4.2 Power Comparisons

In this section, we compare the performance of the three test statistics in terms of their power properties. For power comparisons, we add simulated jump components into the above pure diffusion processes. For idiosyncratic jumps, we simulate 20 independent Gaussian Poisson processes with intensity $\lambda_i$, and magnitude $N(0, \sigma_i^2)$, and add them to their corresponding pure diffusion processes. For the common jumps, we simulate one Gaussian compound Poisson process with intensity $\lambda$ and magnitude $N(0, \sigma_J^2)$, and add it to the diffusion process after multiplying each of the twenty components by an estimate of its $\beta_i$ relative to the portfolio. The bottom panel of Figure 2 shows a simulated day in which one co-jump affects all stocks. We calibrate the common jump intensity $\lambda$ and the common jump size $\sigma_J$ to empirical data (i.e. $\lambda = 0.05\%$ and $\sigma = 0.005$ in our case), and then change the intensity $\lambda$ from 0.05% to 1%, and the size of $\sigma_J$ from 0.005 to 0.1 in order to check the sensitivity of the power of these test statistics to these parameters.

Table 4, Table 5 and Table 6 respectively present power calculations relating to the zmcp test statistics, the zmcr test statistics and the zmfhllc test statistics under a nominal significance level of 0.1%, but based on different sampling frequencies ($M$). The number of subintervals ($m$) used to calculate price range and first-high-low-last price correspond to the maximum possible given $M$, and co-jump intensities and co-jump sizes are varied. As expected, the tests have greater ability to find co-jumps as $M, \lambda$ and $\sigma_J^2$ increase. Table 7 and Table 8 report the power of the zmcr test statistics and the zmfhllc test statistics, which are calculated by keeping the sampling frequency $M$ constant, but the number of subintervals ($m$) used to calculate

---

11 Jump intensity $\lambda_i$ is defined here as the percentage of price observations that contain a jump, where $\lambda_i \in [0.005\%, 0.01\%]$ in our simulation.

12 Here $\sigma_i^2 \in [0.0005, 0.001]$ in our simulation.

13 Our reported results relate to the 5-minute sampling frequency ($M = 48$), but similar
the price range or first-high-low-last price vary, as do the co-jump intensities and the co-jump sizes. The nominal significance level is still 0.1%. As expected, the tests have greater ability to find co-jumps as \( m, \lambda \) and \( \sigma_J^2 \) increase. This finding is important, because it shows that the range based statistics lead to increased power relative to the Bollerslev et al (2008) return-based \( zme_p \) test statistic (for which \( m = 1 \)). Furthermore, we find that at all sampling frequencies and all levels of co-jump intensity and co-jump size, the first-high-low-last price based \( zmfhllc \) test statistics lead to further power improvement compared with the range-based \( zmcr \) test statistics.

In our simulation, we assumed that the covariance of a multivariate asset price process is constant without microstructure noise. As section 2.3 shows, the ranking of the three covariance estimators with respect to their estimation efficiency is the same under three different settings (constant volatility without microstructure noise, stochastic volatility without microstructure noise and stochastic volatility with microstructure noise). Therefore, we expect that the ranking of the three types of tests in terms of their power properties should be unaltered after introducing stochastic volatility and microstructure noise.

5 Empirical Application

5.1 Data

Our empirical analysis is based on intraday data relating to 40 very actively traded stocks in the Chinese mainland stock market.\(^{14}\) Twenty of these stocks are traded on the Shanghai Stock Exchange (SSE) and the remaining twenty are traded on the Shenzhen Stock Exchange (SZSE). The existing literature relating to jump detection in this market mostly focuses on the univariate situation (see Zhang and Xu (2006), Wang, Yao, Fang and Li (2006, 2008) and Ma and Wang (2009)), although Liao et al (2010) build factor models of jumps to account for simultaneous jumps in more than one stock. We apply the return-based co-jump test in Bollerslev et al (2008), our range-based co-jump test and our first-high-low-last price based co-jump test to the twenty stocks from the Shanghai Stock Exchange, the twenty stocks from the Shenzhen Stock Exchange and all forty stocks to analyze the co-jump patterns in each stock exchange and co-jumps across the two tendencies are observed at other sampling frequencies.

\(^{14}\)There are two official stock exchanges in the Chinese mainland market, i.e. the Shanghai Stock Exchange (SHSE) and the Shenzhen Stock Exchange (SZSE). These were established in December 1990 and July 1991 respectively. All stocks are A-share stocks.
stock exchanges. The raw transaction prices (together with trading times and volumes) were obtained from the China Stock Market & Accounting Research (CSMAR) database provided by the ShenZhen GuoTaiAn Information and Technology Firm (GTA). Our sample covers the period from July 2nd, 2007 to September 28th, 2007 (three months).

We focus on the active trading period and leave issues associated with overnight jumps for further research. Due to the fact that it is difficult to construct the price sample path of a portfolio from the tick-by-tick data of each individual stock in the case of nonsynchronous trading, we firstly use 30 seconds as the sampling frequency to obtain equally spaced high frequency data for each individual stock, then average the 30-second prices of individual stocks to obtain a price sample path for the equally weighted portfolio. Therefore, the baseline data used in the following analysis is equally spaced high frequency data (observed at thirty second intervals) rather than irregularly spaced tick-by-tick data. We use data from July 2nd, 2007 to September 28th 2007 (excluding weekends, public holidays and firm specific suspensions), and we avoid market opening effects on trading days by only using data from 09:35-11:30 and 13:05-15:00.

Paralleling many previous studies, we attempt to strike a reasonable balance between efficiency and accuracy by using five-minutes as the sampling frequency to construct intraday returns, intraday range, daily realized volatility, daily realized range and daily FHLL estimators of volatility. Moreover, we employ the additive bias-correction method as discussed in section 2.3.3 to correct for microstructure noise bias in daily realized volatility, daily realized range and daily FHLL variance estimators calculated from 5-minutes intraday transactions, and thereby calculate the corresponding intraday $zmcp$, $zmcr$, and $zmfhllc$ statistics. Our sample spans 65 trading days, and each trading day has 462 intraday (30-seconds) price observations. Hence, there are $M = 46$ $zmcp$, $zmcr$, and $zmfhllc$ test statistics for each trading day, and the number of intraday prices that are used to calculate the range for each five minute interval is $m + 1 = 11$.

---

$^{15}$Bannouh, Dijk and Martens (2009) accounted for nonsynchronous trading by updating their portfolio price each time that they observed a new price for one of the constituent assets. They have only three assets in their portfolio. Their procedure becomes relatively complicated when the number of the constituent assets is large, so in our case, we simply sample the raw tick-by-tick data once every 30 seconds to effectively mimic a synchronous trading scenario in which the portfolio price is updated every 30 seconds. This does not lead to a large loss of information because our prices rarely change much in 30 seconds.
5.2 Co-jumps in the Chinese Mainland Stock Market

We simulate the null distributions of the co-jump test statistics for each panel prior to performing the tests. For the tests relating to the twenty stocks from the Shanghai Stock Exchange, we simulate 1000 realizations of a $20 \times 1$ diffusion process with zero drift and a covariance matrix determined by an unconditional estimate of the covariance matrix of the 30-seconds within-day returns for the relevant 20 stocks. The length of each realization is set equal to the sample size (462 per day for 65 days). We use these simulations to obtain observations on each of the $zmcp$, $zmcr$ and $zmfhllc$ test statistics. We repeated this for the twenty stocks from Shenzhen Stock Exchange (which now have a covariance structure calibrated to the Shenzhen stocks) and then again for the set of all forty stocks (which now involved generating test statistics based on realizations of a $40 \times 1$ diffusion with an appropriately calibrated $40 \times 40$ covariance structure). This scheme generated over 30 million simulated values for each of the three test statistics for each panel of data under the null of no jumps. Table 9 reports the critical values at the 0.1%, 1% and 5% significance levels.

Next, we investigate if there are co-jumps in the three panels of high frequency data. Figure 4 presents Q-Q plots of the quantiles of the empirical distributions versus the quantiles of the simulated null distributions. All the empirical distributions of the three test statistics in three panels of data are sharply right shifted relative to the null distribution, which reveals striking evidence for co-jumps in each panel of data. Compared with the return-based co-jump test statistics (the $zmcp$ statistics), the empirical distributions of the range-based co-jump test statistics (the $zmcr$ statistics) and the first-high-low-last price based co-jump test statistics ($zmfhllc$ statistics) deviate more from their corresponding null distributions in the right tail. This indicates that range based test statistics and first-high-low-last price based test statistics might be able to find more co-jumps than return based test statistics.

The above evidence is reinforced by Figures 5-7, which plot the series of the three test statistics calculated from each panel of empirical high frequency data. The horizontal lines represent the 99.9%, 99% and 95% quantiles of the relevant simulated null distributions, and these can be used as critical values. These figures show that there are a number of empirically observed test statistics from each panel of data that exceed their relevant critical values, which suggest that there are co-jumps in each panel of data. Compared with the return-based type test statistics ($zmcp$ statistics), the range-based type test statistics (the $zmcr$ statistics) and the first-high-low-last price based $zmfhllc$ test statistics are statistically significant on more occasions. This
provides further evidence that the range based test and first-high-low-last price based test is more powerful for co-jump detection than the return-based test.

Table 10, Table 11 and Table 12 report the outcomes of the co-jump tests based on the three panels of stock data at 0.1% significance level. These outcomes include co-jump arrival dates and times. Even though our previous research found frequent small jumps in some of the individual stocks of this market, the co-jumps are seemingly not frequent in this market. The return-based test found a few co-jumps. (Those for the Shanghai Stock exchange occurred on 7 out of the 65 days, those for the Shenzhen stocks also occurred on 5 out of 65 days in Shenzhen Stock Exchange, and there were co-jumps on 6 out of 65 days across the two stock exchanges). Based on the range based test, a single co-jump in the Shanghai stocks occurred on 12 out of the 65 days in the sample, a single co-jump in the Shenzhen stocks occurred on 10 out of the 65 days in the sample, and some of these co-jump occurred at exactly the same time. The range based co-jump test for the set of 40 stocks identified nine co-jumps and the timing of these. Based on the first-high-low-last price based test, a single co-jump in the Shanghai stocks occurred on 17 out of the 65 days in the sample, a single co-jump in the Shenzhen stocks occurred on 17 out of the 65 days in the sample, and some of these co-jump occurred at exactly the same time. The first-high-low-last price based co-jump test for the set of 40 stocks identified fifteen co-jumps and the timing of these. Nearly all of these corresponded to a co-jump observed on at least one of the individual exchanges. Overall, co-jumps were experienced, approximately once every three days.

Most of detected co-jumps in this market occurred near the opening or closing time of each trading session (morning session or afternoon session). This provides evidence that there is a strong diurnal pattern in Chinese stock market. Moreover, some of co-jumps are associated with news announcements in this particular market. These news announcements are provided in the footnote to Table 12. Compared with the existing return based test, our range based test and the test based on first-high-low-last price are more sensitive to these co-jumps. The test based on first-high-low-last price is most powerful.

6 Conclusion

We build on a return-based co-jump test proposed in Bollerslev et al (2008) and a newly introduced first-high-low-last price based covariance estimator to construct a more powerful first-high-low-last price based co-jump test.
Our Monte Carlo simulations suggest that the first-high-low-last price based co-jump test outperforms the return-based co-jump test with respect to power. Due to the fact that the first-high-low-last price based covariance estimator is more efficient than the return based one, the first-high-low-last price based test is able to identify more co-jumps than the return based test because it is more sensitive to the medium-sized co-jumps. Applied to a panel of high frequency data relating to Chinese mainland stock market over three months from July 2nd 2007 to September 28th 2007, these test statistics find co-jumps in the stocks from the two stock market exchanges, but unlike the frequent jumps observed in individual stocks, co-jumps in this market occur infrequently. Our first-high-low-last price based test does, however, detects more co-jumps in this market than the return-based test.

There are several directions for ongoing and future research emerging from our study. Firstly, following Bollerslev et al (2008), we conduct Monte Carlo simulation to get the null distribution of the first-high-low-last price based co-jump test. It would be interesting to derive the asymptotic null distribution for both the return-based and first-high-low-last price based co-jump test statistics and compare their asymptotic behavior. Bollerslev et al (2008) showed that their return-based test statistic has the appearance of an affine transformation of a \(\chi^2(1)\) distribution under a simple one-factor representation for the returns. However, this does not hold in empirically more realistic multi-factor situations. It is well known that the joint distribution of first-high-low-last price returns is very different from that of the price return, so one direction for future research would be to firstly work out the asymptotic null distribution of the first-high-low-last price based test statistics under a one-factor model, and then extend this to the multi-factor scenario.

Secondly, the empirical evidence of co-jumps in financial market suggests that common factor models of jumps have empirical relevance, and this lays open the possibility that they might have forecasting potential. We are currently working on a paper related to this topic (see Liao, Anderson and Vahid (2010)).

Thirdly, the return based variance and covariance estimators are not efficient because the standard procedure used to construct these estimators does not use all of the available data information. Zhou (1996) proposed a way to fully make use of the data by combining the subsampling and averaging procedures together. For example, we could subsample every 5 minutes, starting with the first observation, then starting with the second, and so on. Then we average the variance or covariance estimators obtained across those 5-minutes returns. Since our first-high-low-last price based approach also aims to extract information from data as much as possible, it would be in-
teresting to compare our first-high-low-last price based covariance estimator with the covariance estimator constructed using Zhou (1996)'s methodology, with respect to their estimation efficiency and their applications to co-jump tests. Lastly, it would be useful to document and understand the economic determinants of co-jumps.
Appendix 1

Limit Theorem for the FHLL Variance Estimator

Let the log price $p_s$ of an asset follow the continuous process

$$dp_s = \mu(s)ds + \sigma(s)dW_s,$$

and suppose that high frequency data as described in Section 2 is available for each day $t$. Then the FHLL variance estimator is

$$FHLLV_t = \sum_{j=1}^{M} (0.5(h_{t_j} - l_{t_j})^2 - (2\log(2) - 1)(p_{t_j} - p_{t_{j-1}})^2).$$

**Theorem 1.** As $M \to \infty$, we have

$$FHLLV_t \overset{p}{\to} \int_{t-1}^{t} \sigma^2(s)ds$$

Proof. According to Barndorff-Neilsen et al (2002) and Christensen et al (2007), we have that

$$RV_t = \sum_{j=1}^{M} (p_{t_j} - p_{t_{j-1}})^2 \overset{p}{\to} \int_{t-1}^{t} \sigma^2(s)ds,$$

and

$$RRV_t = \frac{1}{\gamma_{2,m}} \sum_{j=1}^{M} (h_{t_j} - l_{t_j})^2 \overset{p}{\to} \int_{t-1}^{t} \sigma^2(s)ds,$$

where $RV_t$ is realized volatility, $RRV_t$ is realized range and $\gamma_{2,m} = E[s_{w,m}^2] = 4ln2$. The FHLL variance estimator can be expressed as

$$FHLLV_t = 0.5 \times \gamma_{2,m} \times RRV_t - (2\log(2) - 1) \times RV_t.$$

Therefore,

$$FHLLV_t \overset{p}{\to} (0.5 \times 4ln(2) - 2\log(2) + 1) \int_{t-1}^{t} \sigma^2(s)ds = \int_{t-1}^{t} \sigma^2(s)ds.$$

**Theorem 2.**

$$\sqrt{M}(FHLLV_t - \int_{t-1}^{t} \sigma^2(s)ds) \overset{d}{\to} MN(0, 0.27 \int_{t-1}^{t} \sigma^4(s)ds)$$
Proof. Since the FHLL variance estimator is a linear combination of realized volatility and realized range, the asymptotic variance of FHLL estimator will be a linear combination of the asymptotic variance of realized volatility, the asymptotic variance of realized range and the asymptotic covariance of realized volatility and realized range. We define a process to study the joint behavior of realized volatility and realized range to be

\[
P^M(g, h)_t = \sum_{j=1}^{M} g(\Delta_j^M P, \Delta_j^M S) h(\Delta_{j+1}^M P, \Delta_{j+1}^M S),
\]

(1)

where \( \Delta_j^M P = p_t - p_{t-1} \), \( \Delta_j^M S = h_t - l_t \).

\[
g(\Delta_j^M P, \Delta_j^M S) = \begin{bmatrix} (2 \log(2) - 1)(\Delta_j^M P)^2 & 0 \\ 0 & 0.5(\Delta_j^M S)^2 \end{bmatrix}
\]

and \( h(\Delta_{j+1}^M P, \Delta_{j+1}^M S) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \).

Then equation (1) becomes

\[
P^M(g, h)_t = \frac{\sum_{j=1}^{M} (2 \log(2) - 1)(\Delta_j^M P)^2}{\sum_{j=1}^{M} 0.5(\Delta_j^M S)^2}.
\]

We can derive the asymptotic variance-covariance matrix for \( P^M(g, h)_t \) using Theorem 2 in Barndorff-Nielsen et al (2006). These authors defined a return-based process given by

\[
Y^M(g, h)_t = \frac{1}{M} \sum_{i=1}^{M} g(\sqrt{M} \Delta_i^M Y) h(\sqrt{M} \Delta_{i+1}^M Y),
\]

(2)

where \( g \) and \( h \) are two given matrix functions of dimensions \( d_1 \times d_2 \) and \( d_2 \times d_3 \), respectively, and \( \Delta_i^M Y = p_t - p_{t-1} \). They pointed out that most of the return-based nonparametric volatility measures used in financial econometrics can be studied within this framework, including realized volatility and realized bipower variation. They then made some assumptions on the functions \( g \) and \( h \) to derive the central limit theorem for \( Y^M(g, h)_t \). Using their results, if \( d_1 = d_2 = 2 \), \( d_3 = 1 \), and \( g \) is diagonal, then the central limit theorem for \( Y^M(g, h)_t \) is

\[
\sqrt{M}(Y^M(g, h)_t - \int_{t-1}^{t} \rho_{\sigma_s}(g) \rho_{\sigma_s}(h) ds) \longrightarrow MN(0, \int_{t-1}^{t} A(\sigma_s, g, h) ds),
\]

where

\[\text{See Barndorff-Nielsen et al (2006) for more details.}\]
we can combine this with the results that

\[
A(\sigma, g, h)^{j, j'} = \rho_\sigma(g^{jj}g^{j'j'})\rho_\sigma(h^{j}h^{j'}) + \\
\rho_\sigma(g^{jj})\rho_\sigma(h^{j'})\rho_\sigma(g^{j'j}) + \rho_\sigma(g^{j'j})\rho_\sigma(h^{j})\rho_\sigma(g^{jj}h^{j'}) \\
-3\rho_\sigma(g^{jj})\rho_\sigma(g^{j'j'})\rho_\sigma(h^{j})\rho_\sigma(h^{j'})
\]

where \(j = 1, 2\) and \(j' = 1, 2\). MN denotes a mixed Gaussian distribution and \(\int_{t-1}^t A(\sigma, g, h)\) denotes an asymptotic variance-covariance matrix.

This theorem can be extended to the range-based nonparametric volatility measures using equation (2) given above. In our case we have

\[
\rho_\sigma(g^{11}) = 3(2\log(2) - 1)^2 \sigma_j^2, \quad \rho_\sigma(h^1) = \rho_\sigma(h^1h^1) = 1,
\]

\[
\rho_\sigma(g^{11}) = \rho_\sigma(g^{11}h^1) = \rho_\sigma(g^{11}h^2) = (2\log(2) - 1)\sigma_j^2,
\]

\[
\rho_\sigma(g^{22}) = (0.5)^2\lambda_4\sigma_j^4 = \rho_\sigma(h^2) = \rho_\sigma(h^2h^2) = 1,
\]

and \(\rho_\sigma(g^{22}) = \rho_\sigma(g^{22}h^1) = \rho_\sigma(g^{22}h^1) = 0.5\lambda_2\sigma_j^2\).

Further, since

\[
\rho_\sigma(g^{11}g^{22}) = E((\Delta_j^M P)^2(\Delta_j^M S)^2) = E((\Delta_j^M P)^2(h_{j} - l_{j})^2)
\]

\[
= E((\Delta_j^M P)^2h_{j,\Delta_j}^2 - 2h_{j,\Delta_j}l_{j,\Delta_j}(\Delta_j^M P)^2 + (\Delta_j^M P)^2l_{j,\Delta_j}^2)
\]

\[
= E((\Delta_j^M P)^2h_{j,\Delta_j}^2) - 2E(h_{j,\Delta_j}l_{j,\Delta_j}(\Delta_j^M P)^2) + E((\Delta_j^M P)^2l_{j,\Delta_j}^2),
\]

we can combine this with the results that \(E((\Delta_j^M P)^2h_{j,\Delta_j}^2) = E((\Delta_j^M P)^2l_{j,\Delta_j}^2) = 2\sigma_j^4\) and \(E(h_{j,\Delta_j}l_{j,\Delta_j}(\Delta_j^M P)^2) = -0.43812\sigma_j^4\) (see Garman et al (1980)) to obtain

\[
\rho_\sigma(g^{11}g^{22}) = (2\log(2) - 1)\times 0.5\times (2\sigma_j^4 + 2\times 0.43812\sigma_j^4 + 2\sigma_j^4) = 0.94\sigma_j^4, \quad \rho_\sigma(h^2) = \rho_\sigma(h^1h^2) = 1.
\]

We then have that

\[
A(\sigma, g, h)^{1,1} = (2\log(2) - 1)^2(3\sigma_j^4 + 2\sigma_j^4 - 3\sigma_j^4) = 0.3\sigma_j^4
\]

\[
A(\sigma, g, h)^{2,2} = (0.5)^2(\lambda_4\sigma_j^4 + 2\lambda_5\sigma_j^4 - 3\lambda_5^2\sigma_j^4) = 0.7825\sigma_j^4
\]

and \(A(\sigma, g, h)^{1,2} = (0.94 - 0.386 \times 1.385)\sigma_j^4 = 0.41\sigma_j^4\).

Therefore, the asymptotic variance-covariance matrix of the FHL estimator is

\[
\int_{t-1}^t (A^{1,1}(\sigma(s), g, h) - 2A^{1,2}(\sigma(s), g, h) + A^{2,2}(\sigma(s), g, h))ds = 0.27 \int_{t-1}^t \sigma^4(s)ds.
\]
Figure 1: Density Graphs of the Three Covariance Estimators (Constant Volatility and No Microstructure Noise)

Note: The true covariance is 0.04. The top panel shows the distributions of the three covariance estimators (realized covariance, realized co-range, and first-high-low-last price based estimator) via simulation with 5000 days based on a 5-minute sampling frequency. The middle panel shows the distributions of the three covariance estimators (realized covariance, realized co-range, and first-high-low-last price based estimator) via simulation with 5000 days based on a 10-minute sampling frequency. The bottom panel shows the distributions of the three covariance estimators (realized covariance, realized co-range, and first-high-low-last price based estimator) via simulation with 5000 days based on a 15-minute sampling frequency.
Figure 2: Simulated Log Price Realizations of 20 Stocks on a Typical Day

Note: The top panel plot shows simulated intraday price realizations of 20 stocks on a typical day based on a multivariate pure diffusion process with zero drift. The bottom panel plot shows intraday price realizations of 20 stocks which now include idiosyncratic jumps and one co-jump. The interval between the two black dotted lines indicates the time interval that contains the co-jump.
Figure 3: The Null Distributions of Co-Jump Test Statistics Based on Monte Carlo Simulation

Note: The top panel shows the null distributions of the three test statistics (zmcp, zmcr, and zmfhllc test statistics) when daily variance calculations are based on 5-minute intervals (M=48) and subintervals for range and first-high-low-last price calculations are recorded 10 times (m=10) during each 5-minute interval. The middle panel shows the null distributions of zmfhllc test statistics for different M, when subintervals for range and first-high-low-last price calculations (m) are recorded the maximum possible number of times. The bottom panel shows the resulting null distributions of zmfhllc test statistics for M=48, when the number of subintervals for range and first-high-low-last price calculations (m) within each of these 48 intervals is varied.
Figure 6: Plots of Empirical Test Statistics for 20 Stocks on the Shenzhen Stock Exchange

Note: The top plot show the series of the return-based test statistics ($zmcp$) calculated from the empirical data. The middle plot show the series of the range-based test statistics ($zmcr$) calculated from the empirical data. The bottom plot show the series of the first-high-low-last price based test statistics ($zmfhlc$) calculated from the empirical data. The red dotted line represents the 0.1% critical value, the green dotted line represents the 1% critical value, and the pink dotted line represents the 5% critical value.
Figure 7: Empirical Test Statistics for all 40 stocks

Note: The top plot show the series of the return-based test statistics ($zmcp$) calculated from the empirical data. The middle plot show the series of the range-based test statistics ($zmcr$) calculated from the empirical data. The bottom plot show the series of the first-high-low-last price based test statistics ($zmflhc$) calculated from the empirical data. The red dotted line represents the 0.1% critical value, the green dotted line represents the 1% critical value, and the pink dotted line represents the 5% critical value.
Table 1: Comparison of Covariance Estimators When There Is No Microstructure Noise

<table>
<thead>
<tr>
<th>Sampling Frequency</th>
<th>( RCV_t )</th>
<th>( RCR_t )</th>
<th>( FHLLCV_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>Panel A: Constant Covariance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(minute)</td>
<td>-0.0004</td>
<td>0.0229</td>
<td>-0.0007</td>
</tr>
<tr>
<td>5(minutes)</td>
<td>-0.0002</td>
<td>0.0282</td>
<td>-0.0005</td>
</tr>
<tr>
<td>10(minutes)</td>
<td>-0.0001</td>
<td>0.0304</td>
<td>-0.0002</td>
</tr>
<tr>
<td>15(minutes)</td>
<td>0.0001</td>
<td>0.0315</td>
<td>-0.0002</td>
</tr>
<tr>
<td>30(minutes)</td>
<td>-0.0003</td>
<td>0.0367</td>
<td>-0.0003</td>
</tr>
<tr>
<td>60(minutes)</td>
<td>-0.0005</td>
<td>0.0411</td>
<td>-0.0006</td>
</tr>
<tr>
<td>240(daily)</td>
<td>-0.0005</td>
<td>0.0859</td>
<td>-0.0008</td>
</tr>
<tr>
<td>Panel B: Stochastic Covariance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(minute)</td>
<td>-0.0009</td>
<td>0.0372</td>
<td>-0.0014</td>
</tr>
<tr>
<td>5(minutes)</td>
<td>-0.0006</td>
<td>0.0398</td>
<td>-0.0008</td>
</tr>
<tr>
<td>10(minutes)</td>
<td>-0.0002</td>
<td>0.0420</td>
<td>-0.0006</td>
</tr>
<tr>
<td>15(minutes)</td>
<td>-0.0003</td>
<td>0.0429</td>
<td>-0.0004</td>
</tr>
<tr>
<td>30(minutes)</td>
<td>-0.0003</td>
<td>0.0432</td>
<td>-0.0003</td>
</tr>
<tr>
<td>60(minutes)</td>
<td>-0.0004</td>
<td>0.0487</td>
<td>-0.0005</td>
</tr>
<tr>
<td>240(daily)</td>
<td>-0.0006</td>
<td>0.0881</td>
<td>-0.0007</td>
</tr>
</tbody>
</table>

Note: The true covariance is 0.04. All points of the simulated data are observed \((\tau = 1)\) and there is no bid-ask bounce \((s = 0)\). \( RCV_t \) is realized covariance, \( RCR_t \) is realized co-range, and \( FHLLCV_t \) is the first-high-low-last price covariance estimator.
Table 2: Comparison of Covariance Estimators When There Is Microstructure Noise

<table>
<thead>
<tr>
<th>Sampling Frequency</th>
<th>$RCV_t$</th>
<th>$RCV_{C,t}$</th>
<th>$RCR_t$</th>
<th>$RCR_{C,t}$</th>
<th>$F H L L C V_t$</th>
<th>$F H L L C V_{C,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>Panel A: Stochastic Covariance ($s = 0, \tau = 15$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(minute)</td>
<td>-0.0188</td>
<td>0.0304</td>
<td>-0.0011</td>
<td>0.0298</td>
<td>-0.0203</td>
<td>0.0317</td>
</tr>
<tr>
<td>5(minutes)</td>
<td>-0.0159</td>
<td>0.0289</td>
<td>-0.0009</td>
<td>0.0313</td>
<td>-0.0195</td>
<td>0.0291</td>
</tr>
<tr>
<td>10(minutes)</td>
<td>-0.0136</td>
<td>0.0273</td>
<td>-0.0009</td>
<td>0.0398</td>
<td>-0.0152</td>
<td>0.0275</td>
</tr>
<tr>
<td>15(minutes)</td>
<td>-0.0099</td>
<td>0.0297</td>
<td>-0.0008</td>
<td>0.0408</td>
<td>-0.0104</td>
<td>0.0279</td>
</tr>
<tr>
<td>30(minutes)</td>
<td>-0.0087</td>
<td>0.0335</td>
<td>-0.0005</td>
<td>0.0476</td>
<td>-0.0066</td>
<td>0.0280</td>
</tr>
<tr>
<td>60(minutes)</td>
<td>-0.0058</td>
<td>0.0413</td>
<td>-0.0003</td>
<td>0.0592</td>
<td>-0.0035</td>
<td>0.0312</td>
</tr>
<tr>
<td>240(daily)</td>
<td>-0.0006</td>
<td>0.0862</td>
<td>-0.0006</td>
<td>0.0862</td>
<td>-0.0009</td>
<td>0.0436</td>
</tr>
<tr>
<td>Panel B: Stochastic Covariance ($s = 0.075, \tau = 1$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(minute)</td>
<td>-0.0010</td>
<td>0.0367</td>
<td>-0.0001</td>
<td>0.0357</td>
<td>0.0516</td>
<td>0.1133</td>
</tr>
<tr>
<td>5(minutes)</td>
<td>-0.0007</td>
<td>0.0402</td>
<td>0.0005</td>
<td>0.0380</td>
<td>0.0570</td>
<td>0.0936</td>
</tr>
<tr>
<td>10(minutes)</td>
<td>-0.0005</td>
<td>0.0423</td>
<td>0.0013</td>
<td>0.0488</td>
<td>0.0544</td>
<td>0.0748</td>
</tr>
<tr>
<td>15(minutes)</td>
<td>-0.0001</td>
<td>0.0431</td>
<td>0.0015</td>
<td>0.0510</td>
<td>0.0367</td>
<td>0.0654</td>
</tr>
<tr>
<td>30(minutes)</td>
<td>0.0007</td>
<td>0.0439</td>
<td>-0.0002</td>
<td>0.0577</td>
<td>0.0255</td>
<td>0.0482</td>
</tr>
<tr>
<td>60(minutes)</td>
<td>-0.0001</td>
<td>0.0516</td>
<td>-0.0005</td>
<td>0.0695</td>
<td>0.0173</td>
<td>0.0426</td>
</tr>
<tr>
<td>240(daily)</td>
<td>-0.0006</td>
<td>0.0885</td>
<td>-0.0006</td>
<td>0.0885</td>
<td>-0.0006</td>
<td>0.0442</td>
</tr>
<tr>
<td>Panel C: Stochastic Covariance ($s = 0.075, \tau = 15$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(minute)</td>
<td>-0.0191</td>
<td>0.0428</td>
<td>-0.0012</td>
<td>0.0420</td>
<td>0.0322</td>
<td>0.1133</td>
</tr>
<tr>
<td>5(minutes)</td>
<td>-0.0160</td>
<td>0.0412</td>
<td>-0.0010</td>
<td>0.0415</td>
<td>0.0240</td>
<td>0.0936</td>
</tr>
<tr>
<td>10(minutes)</td>
<td>-0.0139</td>
<td>0.0429</td>
<td>-0.0009</td>
<td>0.0438</td>
<td>0.0131</td>
<td>0.0748</td>
</tr>
<tr>
<td>15(minutes)</td>
<td>-0.0100</td>
<td>0.0508</td>
<td>-0.0009</td>
<td>0.0516</td>
<td>0.0025</td>
<td>0.0654</td>
</tr>
<tr>
<td>30(minutes)</td>
<td>-0.0089</td>
<td>0.0576</td>
<td>-0.0006</td>
<td>0.0598</td>
<td>0.0013</td>
<td>0.0482</td>
</tr>
<tr>
<td>60(minutes)</td>
<td>-0.0054</td>
<td>0.0692</td>
<td>-0.0004</td>
<td>0.0713</td>
<td>0.0006</td>
<td>0.0426</td>
</tr>
<tr>
<td>240(daily)</td>
<td>-0.0004</td>
<td>0.0894</td>
<td>-0.0004</td>
<td>0.0894</td>
<td>-0.0006</td>
<td>0.0463</td>
</tr>
</tbody>
</table>

Note: The true covariance is 0.04. $RCV_t$ is realized covariance, $RCR_t$ is realized co-range, and $FH L L C V_t$ is the first-high-low-last price covariance estimator.
<table>
<thead>
<tr>
<th></th>
<th>$\alpha = 0.1%$</th>
<th>$\alpha = 1%$</th>
<th>$\alpha = 5%$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M=16$</td>
<td>$M=24$</td>
<td>$M=48$</td>
</tr>
<tr>
<td>$z_{mcsp}$</td>
<td>3.608 4.296 5.350</td>
<td>3.215 3.546 3.838</td>
<td>2.235 2.205 2.134</td>
</tr>
<tr>
<td>$z_{mcrt}$ (m=maximum possible)</td>
<td>3.480 4.096 5.041</td>
<td>2.996 3.274 3.563</td>
<td>2.064 2.066 2.042</td>
</tr>
<tr>
<td>$z_{mcrt}$ (m=half of maximum possible)</td>
<td>3.491 4.125 5.132</td>
<td>3.012 3.308 3.621</td>
<td>2.078 2.082 2.056</td>
</tr>
<tr>
<td>$z_{mcrt}$ (m=third of maximum possible)</td>
<td>3.505 4.165 5.198</td>
<td>3.029 3.357 3.684</td>
<td>2.088 2.100 2.067</td>
</tr>
<tr>
<td>$z_{mftllc}$ (m=maximum possible)</td>
<td>3.334 3.852 4.611</td>
<td>2.804 3.025 3.285</td>
<td>1.960 1.980 1.988</td>
</tr>
<tr>
<td>$z_{mftllc}$ (m=half of maximum possible)</td>
<td>3.354 3.905 4.783</td>
<td>2.831 3.075 3.398</td>
<td>1.980 2.002 2.013</td>
</tr>
<tr>
<td>$z_{mftllc}$ (m=third of maximum possible)</td>
<td>3.385 3.983 4.979</td>
<td>2.860 3.162 3.512</td>
<td>1.994 2.034 2.035</td>
</tr>
</tbody>
</table>

**Note:** $\alpha$ is the significance level, $M$ is the number of intraday test statistics within a day, and $m$ is the number of intraday subintervals used to form the price range and the first-high-low-last price.
Table 4: The Power of the Return-based Co-Jump Test Statistic $zmcp$ for Different Sampling Frequencies $M$

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>$\sigma_J = 0.005$</th>
<th>$\sigma_J = 0.01$</th>
<th>$\sigma_J = 0.05$</th>
<th>$\sigma_J = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$zmcp$ ($M = 48$)</td>
<td>( \lambda = 0.05% )</td>
<td>0.0414</td>
<td>0.0654</td>
<td>0.1078</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.1% )</td>
<td>0.0523</td>
<td>0.0967</td>
<td>0.1765</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.5% )</td>
<td>0.1245</td>
<td>0.1762</td>
<td>0.2543</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 1%  )</td>
<td>0.1698</td>
<td>0.2765</td>
<td>0.3247</td>
</tr>
<tr>
<td>$zmcp$ ($M = 24$)</td>
<td>( \lambda = 0.05% )</td>
<td>0.0394</td>
<td>0.0581</td>
<td>0.0967</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.1% )</td>
<td>0.0372</td>
<td>0.0765</td>
<td>0.1524</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.5% )</td>
<td>0.1039</td>
<td>0.1434</td>
<td>0.2076</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 1%  )</td>
<td>0.1355</td>
<td>0.2076</td>
<td>0.2877</td>
</tr>
<tr>
<td>$zmcp$ ($M = 16$)</td>
<td>( \lambda = 0.05% )</td>
<td>0.0297</td>
<td>0.0455</td>
<td>0.0731</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.1% )</td>
<td>0.0308</td>
<td>0.0692</td>
<td>0.1123</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.5% )</td>
<td>0.0935</td>
<td>0.1120</td>
<td>0.1567</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 1%  )</td>
<td>0.1027</td>
<td>0.1569</td>
<td>0.2033</td>
</tr>
</tbody>
</table>

Nominal significance level= 0.1%

Note: $\lambda$ denotes co-jump intensity, and $\sigma_J$ denotes co-jump size.
Table 5: The Power of the Range-based Co-Jump Test Statistic $zmcr$ for Different Sampling Frequencies $M$

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>$\sigma_J = 0.005$</th>
<th>$\sigma_J = 0.01$</th>
<th>$\sigma_J = 0.05$</th>
<th>$\sigma_J = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$zmcr$ ($M = 48$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.05%$</td>
<td>0.1032</td>
<td>0.1765</td>
<td>0.2654</td>
<td>0.3421</td>
</tr>
<tr>
<td>$\lambda = 0.1%$</td>
<td>0.1421</td>
<td>0.2325</td>
<td>0.3217</td>
<td>0.4236</td>
</tr>
<tr>
<td>$\lambda = 0.5%$</td>
<td>0.2144</td>
<td>0.3524</td>
<td>0.4762</td>
<td>0.5543</td>
</tr>
<tr>
<td>$\lambda = 1%$</td>
<td>0.3176</td>
<td>0.4789</td>
<td>0.5632</td>
<td>0.6321</td>
</tr>
<tr>
<td>$zmcr$ ($M = 24$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.05%$</td>
<td>0.0721</td>
<td>0.1415</td>
<td>0.2106</td>
<td>0.2930</td>
</tr>
<tr>
<td>$\lambda = 0.1%$</td>
<td>0.1034</td>
<td>0.1925</td>
<td>0.2845</td>
<td>0.3507</td>
</tr>
<tr>
<td>$\lambda = 0.5%$</td>
<td>0.2032</td>
<td>0.2985</td>
<td>0.3987</td>
<td>0.4965</td>
</tr>
<tr>
<td>$\lambda = 1%$</td>
<td>0.2526</td>
<td>0.3247</td>
<td>0.4123</td>
<td>0.5521</td>
</tr>
<tr>
<td>$zmcr$ ($M = 16$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0.05%$</td>
<td>0.0587</td>
<td>0.1065</td>
<td>0.1523</td>
<td>0.1921</td>
</tr>
<tr>
<td>$\lambda = 0.1%$</td>
<td>0.0705</td>
<td>0.1490</td>
<td>0.2067</td>
<td>0.2542</td>
</tr>
<tr>
<td>$\lambda = 0.5%$</td>
<td>0.1442</td>
<td>0.2103</td>
<td>0.2869</td>
<td>0.3103</td>
</tr>
<tr>
<td>$\lambda = 1%$</td>
<td>0.2042</td>
<td>0.3024</td>
<td>0.3976</td>
<td>0.4576</td>
</tr>
</tbody>
</table>

Nominal significance level= $0.1\%$

**Note:** $\lambda$ denotes co-jump intensity, and $\sigma_J$ denotes co-jump size.
Table 6: The Power of the First-high-low-last Price Based Test Statistic $zmfhllc$ for Different Sampling Frequencies $M$

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>$\sigma_J = 0.005$</th>
<th>$\sigma_J = 0.01$</th>
<th>$\sigma_J = 0.05$</th>
<th>$\sigma_J = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$zmfhllc$ ($M = 48$)</td>
<td>$\lambda = 0.05%$</td>
<td>0.1247</td>
<td>0.1987</td>
<td>0.2871</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.1%$</td>
<td>0.1567</td>
<td>0.2546</td>
<td>0.3564</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.5%$</td>
<td>0.2534</td>
<td>0.3865</td>
<td>0.4987</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 1%$</td>
<td>0.3985</td>
<td>0.5321</td>
<td>0.6543</td>
</tr>
<tr>
<td>$zmfhllc$ ($M = 24$)</td>
<td>$\lambda = 0.05%$</td>
<td>0.1087</td>
<td>0.1523</td>
<td>0.2456</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.1%$</td>
<td>0.1324</td>
<td>0.2357</td>
<td>0.3067</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.5%$</td>
<td>0.2462</td>
<td>0.3234</td>
<td>0.4357</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 1%$</td>
<td>0.2985</td>
<td>0.3764</td>
<td>0.4893</td>
</tr>
<tr>
<td>$zmfhllc$ ($M = 16$)</td>
<td>$\lambda = 0.05%$</td>
<td>0.0745</td>
<td>0.1236</td>
<td>0.1673</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.1%$</td>
<td>0.0843</td>
<td>0.1651</td>
<td>0.2587</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.5%$</td>
<td>0.1659</td>
<td>0.2546</td>
<td>0.3466</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 1%$</td>
<td>0.3065</td>
<td>0.3987</td>
<td>0.4671</td>
</tr>
</tbody>
</table>

Note: $\lambda$ denotes co-jump intensity, and $\sigma_J$ denotes co-jump size.
Table 7: The Power of the Range-based Test Statistic $z_{mcr}$ for Different $m$ ($M = 48$)

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>$\sigma_J = 0.005$</th>
<th>$\sigma_J = 0.01$</th>
<th>$\sigma_J = 0.05$</th>
<th>$\sigma_J = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_{mcr}$ ($m = 10$)</td>
<td>$\lambda = 0.05%$</td>
<td>0.1032</td>
<td>0.1765</td>
<td>0.2654</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.1%$</td>
<td>0.1421</td>
<td>0.2325</td>
<td>0.3217</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.5%$</td>
<td>0.2144</td>
<td>0.3524</td>
<td>0.4762</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 1%$</td>
<td>0.3176</td>
<td>0.4789</td>
<td>0.5632</td>
</tr>
<tr>
<td>$z_{mcr}$ ($m = 5$)</td>
<td>$\lambda = 0.05%$</td>
<td>0.0487</td>
<td>0.0921</td>
<td>0.1543</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.1%$</td>
<td>0.0842</td>
<td>0.1541</td>
<td>0.2367</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.5%$</td>
<td>0.1746</td>
<td>0.2543</td>
<td>0.3167</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 1%$</td>
<td>0.2124</td>
<td>0.2987</td>
<td>0.3542</td>
</tr>
<tr>
<td>$z_{mcr}$ ($m = 3$)</td>
<td>$\lambda = 0.05%$</td>
<td>0.0326</td>
<td>0.0743</td>
<td>0.1237</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.1%$</td>
<td>0.0654</td>
<td>0.1328</td>
<td>0.1965</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 0.5%$</td>
<td>0.1343</td>
<td>0.2145</td>
<td>0.2976</td>
</tr>
<tr>
<td></td>
<td>$\lambda = 1%$</td>
<td>0.1978</td>
<td>0.2856</td>
<td>0.3765</td>
</tr>
</tbody>
</table>

Nominal significance level = 0.1%

**Note:** $\lambda$ denotes co-jump intensity, and $\sigma_J$ denotes co-jump size.
Table 8: The Power of the First-high-low-last Price Based Test Statistic \( zmfhllc \) for Different \( m \) \((M = 48)\)

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>( \sigma_J = 0.005 )</th>
<th>( \sigma_J = 0.01 )</th>
<th>( \sigma_J = 0.05 )</th>
<th>( \sigma_J = 0.1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( zmfhllc ) ((m = 10))</td>
<td>( \lambda = 0.05% )</td>
<td>0.1247</td>
<td>0.1987</td>
<td>0.2871</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.1% )</td>
<td>0.1567</td>
<td>0.2546</td>
<td>0.3564</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.5% )</td>
<td>0.2534</td>
<td>0.3865</td>
<td>0.4987</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 1% )</td>
<td>0.3985</td>
<td>0.5321</td>
<td>0.6543</td>
</tr>
<tr>
<td>( zmfhllc ) ((m = 5))</td>
<td>( \lambda = 0.05% )</td>
<td>0.0503</td>
<td>0.0990</td>
<td>0.1982</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.1% )</td>
<td>0.0967</td>
<td>0.1854</td>
<td>0.2765</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.5% )</td>
<td>0.1998</td>
<td>0.2876</td>
<td>0.3543</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 1% )</td>
<td>0.2554</td>
<td>0.3234</td>
<td>0.3778</td>
</tr>
<tr>
<td>( zmfhllc ) ((m = 3))</td>
<td>( \lambda = 0.05% )</td>
<td>0.0414</td>
<td>0.0885</td>
<td>0.1534</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.1% )</td>
<td>0.0702</td>
<td>0.1534</td>
<td>0.2007</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 0.5% )</td>
<td>0.1538</td>
<td>0.2235</td>
<td>0.3079</td>
</tr>
<tr>
<td></td>
<td>( \lambda = 1% )</td>
<td>0.2017</td>
<td>0.2998</td>
<td>0.3974</td>
</tr>
</tbody>
</table>

Nominal significance level = 0.1%

Note: \( \lambda \) denotes co-jump intensity, and \( \sigma_J \) denotes co-jump size.
Table 9: Critical Values for Co-Jump Statistics in the Empirical Study

<table>
<thead>
<tr>
<th></th>
<th>SH20</th>
<th>SZ20</th>
<th>All40</th>
<th>SH20</th>
<th>SZ20</th>
<th>All40</th>
<th>SH20</th>
<th>SZ20</th>
<th>All40</th>
</tr>
</thead>
<tbody>
<tr>
<td>zmcp</td>
<td>5.265</td>
<td>5.268</td>
<td>5.264</td>
<td>3.815</td>
<td>3.813</td>
<td>3.820</td>
<td>2.138</td>
<td>2.132</td>
<td>2.141</td>
</tr>
</tbody>
</table>

Note: \( \alpha \) is the significance level. Critical values in the columns labeled SH20 relate to the 20 stocks from the Shanghai Stock Exchange, critical values in the columns labeled SZ20 relate to the 20 stocks from the Shenzhen Stock Exchange and critical values in the columns labeled All40 relate to the 40 stocks from both exchanges.
<table>
<thead>
<tr>
<th>Co-Jumps Date</th>
<th>Co-Jumps Time</th>
<th>Co-Jumps Date</th>
<th>Co-Jumps Time</th>
<th>Co-Jumps Date</th>
<th>Co-Jumps Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007/09/11</td>
<td>14:50-14:55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 11: Co-jumps Dates and Times for the Three Panels of Stocks Based on the Range-based Co-jump Test

<table>
<thead>
<tr>
<th>Co-Jumps Date</th>
<th>Co-Jumps Time</th>
<th>Co-Jumps Date</th>
<th>Co-Jumps Time</th>
<th>Co-Jumps Date</th>
<th>Co-Jumps Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007/07/11</td>
<td>09:40-09:45</td>
<td>2007/07/23</td>
<td>09:35-09:40</td>
<td>2007/07/11</td>
<td>09:40-09:45</td>
</tr>
<tr>
<td>2007/07/16</td>
<td>14:30-14:35</td>
<td>2007/07/31</td>
<td>10:30-10:35</td>
<td>2007/07/23</td>
<td>09:35-09:40</td>
</tr>
<tr>
<td>2007/09/11</td>
<td>14:50-14:55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007/09/17</td>
<td>09:40-09:45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 12: Co-jumps Dates and Times for the Three Panels of Stocks Based on the First-High-Low-Last Price Based Co-jump Test

<table>
<thead>
<tr>
<th>Co-Jumps Date</th>
<th>Co-Jumps Time</th>
<th>Co-Jumps Date</th>
<th>Co-Jumps Time</th>
<th>Co-Jumps Date</th>
<th>Co-Jumps Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007/07/16</td>
<td>14:30-14:35</td>
<td>2007/07/18</td>
<td>09:40-09:45</td>
<td>2007/07/17b</td>
<td>09:50-09:55</td>
</tr>
<tr>
<td>2007/08/06</td>
<td>14:15-14:20</td>
<td>2007/08/07</td>
<td>09:35-09:40</td>
<td>2007/08/14</td>
<td>09:35-09:40</td>
</tr>
<tr>
<td>2007/08/07</td>
<td>09:35-09:40</td>
<td>2007/08/09</td>
<td>11:00-11:05</td>
<td>2007/08/20</td>
<td>09:45-09:50</td>
</tr>
</tbody>
</table>

*a* Regulations of transfer of state-owned shares in the listed companies are issued.

*b* The Insurance Regulatory Commission announced an adjustment to the proportion of insurance funds invested in A-share stocks from 15% to 20%.

*c* The Central Bank announced that the one-year benchmark interest rates are raised by 0.27 percentage.

*d* The Central Bank ordered banks to set aside an additional 0.5 percent of their deposits.

*e* The Central Bank announced that the one-year benchmark interest rates are raised by 0.27 percentage.

*f* Government announced to issue 200 billion yuan (26.7 billion U.S. dollars) of special treasury bonds.

*g* The Central Bank announced that the one-year benchmark interest rates are raised by 0.27 percentage again.
References


Given Discrete Observations”, Working Paper, Dipartimento di Matematica per le Decisioni, Universit'a degli Studi di Firenze.


