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

18th August 2011

Abstract

This paper proposes a new test for simultaneous intraday jumps in a panel of high frequency financial data. We utilize intraday first-high-low-last values of asset prices to construct estimates for the cross-variation of returns in a large panel of high frequency financial data, and then employ these estimates to provide a first-high-low-last price based test statistic to detect 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 Bollerslev et al (2008) return-based co-jump test. 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, Microstructure bias, Nonsynchronous trades, 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

The documentation and study of jumps (discontinuities) in asset prices is important for many financial decisions, and this has led to a burgeoning literature that tests for jumps and characterizes their properties. The jump test developed by Barndorff-Nielsen and Shephard (2006) test (henceforth the BN-S test) plays a leading role in this literature, but since this test is designed for a single asset while financial considerations usually involve many assets, there is a need for tests that can detect simultaneous jumps in many assets (co-jump tests) as well. Progress on this front includes the development of tests for co-jumps in a pair of asset returns (Barndorff-Nielsen and Shephard (2003), Gobbi and Mancini (2007) and Jacod and Todorov (2009)), as well as a co-jump test developed by Bollerslev, Law and Tauchen (2008) that is applicable to a large panel of high-frequency returns. The intuition behind this last test (henceforth called the BLT test) is that idiosyncratic noise in individual returns can hide the presence of a synchronous component, so that a test based on the cross products of returns in a panel can avoid this problem while still being sensitive to systematic movements across all stocks.

The BLT (2008) co-jump test relies on a measure of covariation in the panel that is constructed using the average pair-wise cross product of returns. However, a return-based estimator does not necessarily provide the best estimator of this covariation. Its simplicity facilitates very straightforward construction of the test statistic, but the literature on covariance estimation emphasizes efficiency-bias considerations in high frequency settings, and efficient estimation is particularly desirable when conducting tests within multivariate contexts. There are now many ways of measuring covariation in high frequency settings, and they vary with respect to their computational difficulty, bias and efficiency. Of particular interest here is the work in Bannouh, van Dijk and Martens (2009) that promotes the use of range based estimators of covariance. Their Monte Carlo illustrates that when there are no market frictions, realized co-range estimators can have variances that are up to five times smaller than returns based realized covariance estimators, and that simple bias corrections can be very effective when microstructure noise is present and trading is not synchronous. Thus the use of intraday ranges instead of intraday returns can offer large efficiency gains without substantially increasing the computational burden. The large efficiency gains arise from the fact that the range can be very informative, since it is constructed by looking at the entire price process in the sampled time interval.

This paper proposes a new test for co-jumps in panels of high frequency data that follows the intuition behind the BLT test, but uses intraday first-high-low-last (FHLL) price values to capture cross-variation. FHLL price values were first used by Garman and Klass (1980) in the context of estimating daily volatility. These authors suggested using the minimum variance linear combination of the daily ranges and daily returns to estimate daily volatility, so that the information in returns is augmented by the additional information contained in the range. They showed that the asymptotic variance of this new estimator is 7.4 times smaller than the squared daily return and 1.5 times smaller than the squared daily range for daily volatility estimation. Our use of first-high-low-last (FHLL) price values in the intraday context offers the same efficiency gains, but like range based estimators, it involves the need for bias correction because of the effects of microstructure biases in high frequency settings. We use the correction proposed by Bannouh et al (2009) for this purpose, because it is simple, easy to implement, and it appears to account for the net effect of the many biases that occur in the high frequency context.

Our proposed test statistic is easy to calculate because it neither relies on estimating the entire covariance structure of returns in the panel, nor on any explicit calculations of bivariate products of prices/returns (that might not necessarily be observed at exactly the same time for different assets). In fact, our test statistic can be calculated using existing univariate methodology, because the average pair-wise cross product term in an n-asset context can also be written in terms of the variance of the equally weighted portfolio and each of the n individual asset variances (see Brandt and Diebold (2006), and BLT (2008)). Like the BLT statistic, our test statistic can not only identify jump days, but it can also explicitly pinpoint co-jump times. Given the relative efficiency gains of FHLL estimators, we expect that co-jump tests based on FHLL prices will be more powerful than the BLT (2008) returns based co-jump test.

Our test is based on a multivariate extension of an FHLL estimator for the covariance between two returns, and since an FHLL covariance estimator is novel in itself, we conduct Monte Carlo simulations to assess and compare it (and an associated bias-corrected version) with corresponding range and return based covariance estimators. We simulate frictionless environments, situations in which bid-ask bounce influences price, and situations in which trades are infrequent and not synchronous, and we vary the intraday sampling frequencies as well. As expected the (uncorrected) FHLL estimator is more biased than its range and return based counterparts, but it is more efficient. Once bias corrections are employed, the RMSE of the FHLL estimators are lower than the (bias corrected) range and returns estimators, and often substantially so. Having established the superior properties of FHLL estimation, we then use it to construct our co-jump test for the panel, and demonstrate via simulations that it has higher power than analogously constructed co-jump tests based on realized range and realized variance. Finally, we study an empirical example based on a high-frequency panel of forty stocks on the Chinese mainland stock market. We use this data because our previous research (Liao (2008)) has found that the jumps in individual stocks of this emerging market are more frequent than those in developed financial markets. We find evidence of several co-jumps per month, and note that about half of these can be linked to announcements about changes in monetary policy or stock market regulations.

The rest of the paper is organized as follows. Section 2 introduces our FHLL price based covariance estimator and analyzes its properties. Section 3 uses this new estimator to develop our FHLL 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 6 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 $t_{(j-1)+(0/m)}$, $t_{(j-1)+(1/m),}....t_{(j-1)+(m/m)} = t_j$).[1] 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 $max\{t_{(j-1)+(0/m)}, t_{(j-1)+(1/m),}....t_{(j-1)+(m/m)} = t_j\}$;

[1] M 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 and Shephard (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 $\gamma_{2,m} \rightarrow 4\ln 2$ as $m \rightarrow \infty$.

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 van 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, finding that

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

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 in each intraday period are available for the construction of 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

$$\hat{\sigma_t^2} = 0.511(logH_t - logL_t)^2 - 0.383(logC_t - logC_{t-1})^2$$
$$-0.019((logH_t - logC_t)logH_t + (logL_t - logC_t)logL_t),$$

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

$$\hat{\sigma_t^2} = 0.5(logH_t - logL_t)^2 - (2log(2) - 1)(logC_t - logC_{t-1})^2,$$

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 $0.41\sigma^4$). Simulations by Rogers and Satchell (1991) showed that this estimator (and a modified version of it) performed quite well in a setting that corresponded with typical daily data.

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

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

This estimator is essentially a linear combination of RRV_t and RV_t with weights of $(2ln2)$ and $(1-2ln(2))$ 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. Simulations conducted by Martens and van Dijk (2007) illustrate these efficiency gains.

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

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

[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.

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 the right-hand side of the above equation, realized covariance (see Barndorff-Nielsen and Shephard (2004)) can be calculated using

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

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

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

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 et al (2009) demonstrates that the realized range is robust to market microstructure noise arising from bid-ask bounce, infrequent trading and not synchronous 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

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

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.[3] 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

$$dp^*_{1,(t-1)+h/K} = \sigma_1 dW_{1,(t-1)+h/K}, \qquad h = 1, 2, ..., K$$

$$dp^*_{2,(t-1)+h/K} = \sigma_2(\rho dW_{1,(t-1)+h/K} + \sqrt{1-\rho^2} dW_{2,(t-1)+h/K}), \; h = 1, 2, ..., K$$

where $p^*_{i,(t-1)+h/K}$ is the log price of asset i at the hth point in the time interval $[t-1, t]$, K is the number of time increments over the day, σ_i is the standard deviation of asset i, W_1 and W_2 are two Brownian motion processes, and ρ 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.[4] 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 $\triangle-$minute intervals, which is referred to as the $\triangle - minute$ frequency below. Since we have a four hour trading day, this divides each day into $M = 240/\triangle$ intraday sampling periods. For example, if we sample at a five minute frequency and $\triangle = 5$, 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

[3] 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] True and observed prices are denoted by $p^*_{i,.}$ and $p_{i,.}$respectively. In the next section we set the probability of actually observing the price to be $p_{obs} = 1/\tau$, and use s to denote the bid-ask spread. For the time being, $\tau = 1$ and $s = 0$.

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. 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 $\triangle$ 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

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

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

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

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

where the volatility is a stochastic process, which follows a mean reverting Orstein-Uhlenbeck process with parameters θ_1 and θ_2 as the adjustment speeds, ω_1 and ω_2 as the means of the (log) volatilities, and η_1 and η_2 as the volatilities of the (log) volatilities. W_1 and W_2 are standard Brownian motions with a correlation of ρ_1, and W_3 and W_4 are standard Brownian motions with a correlation of ρ_2. ρ_1 and ρ_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 ρ_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

$(\theta_1, \theta_2, \omega_1, \omega_2, \eta_1, \eta_2, \rho_1, \rho_2, \rho_3) = (2,\, 2,\, \frac{0.04}{14400},\, \frac{0.16}{14400},\, \frac{0.4}{120},\, \frac{0.8}{120},\, -0.6,\, -0.4,\, 0.5)$.[5] 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 α reflects the relative efficiencies of the different estimators, and this holds true regardless of whether σ_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-variation in volatility. Relative to the constant volatility case, time-variation in asset price volatility tends to increase σ^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 not synchronous 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 et al (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 τ seconds. Since price observations for the two assets occur independently, we observe

[5] We set ω_1, ω_2 and ρ equal to the variances σ_1^2 and σ_2^2 of the two assets' prices, and the correlation ρ 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).

[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.

prices at different times

We use the simulations in Table 1 Panel B as a benchmark and change the values of s and τ 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 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 not synchronous 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 et al (2009). These authors define bias-corrected variance estimators as

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

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 not synchronous 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, van 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 not synchronous 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 FHLL prices 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

$$r_{i,t_j} = p_{i,(t-1)+\frac{j}{M}} - p_{i,(t-1)+\frac{j-1}{M}}, \qquad i = 1, 2, ...n, \qquad j = 1, 2, ...M.$$

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

$$r_{EQW,t_j} = \frac{1}{n}\sum_{i=1}^{n} r_{i,t_j}.$$

The daily realized variance for this equiweighted portfolio then satisfies

$$RV_{EQW,t} = \sum_{j=1}^{M}(\frac{1}{n}\sum_{i=1}^{n} r_{i,t_j})^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{M} r_{i,t_j}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\sum_{j=1}^{M} r_{i,t_j} r_{l,t_j}$$

$$\xrightarrow{M \to \infty} \quad \frac{1}{n^2}\sum_{i=1}^{n}\int_{t-1}^{t}\sigma_i^2(s)ds \; + \; \frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{N_{i,t}}\kappa_{i,t,k}^2$$

$$+ \; \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\int_{t-1}^{t}\sigma_i(s)\sigma_l(s))ds \; + \; \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\sum_{k=1}^{N^*_{(i,l),t}}\kappa_{i,t,k}\kappa_{l,t,k},$$

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 $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{M}r_{i,t_j}^2$, and the quadratic covariation of the prices of the n stocks in the portfolio, that is $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\int_{t-1}^{t}\sigma_i(s)\sigma_l(s))ds+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\sum_{k=1}^{N^*_{(i,l),t}}\kappa_{i,t,k}\kappa_{l,t,k}$, which can be estimated using $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\sum_{j=1}^{M}r_{i,t_j}r_{l,t_j}$.

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 equiweighted portfolio's price process as

$$BV_{EQW,t} = \mu_1^{-2}(\frac{M}{M-1})\sum_{j=2}^{M}|\frac{1}{n}\sum_{i=1}^{n}r_{i,t_{j-1}}|\cdot|\frac{1}{n}\sum_{i=1}^{n}r_{i,t_j}|$$

$$\xrightarrow{M\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\int_{t-1}^{t}\sigma_i^2(s)ds+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\int_{t-1}^{t}\sigma_i(s)\sigma_l(s)ds,$$

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

$$RV_{EQW,t} - BV_{EQW,t}\xrightarrow{M\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{N_{i,t}}\kappa_{i,t,k}^2+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\sum_{k=1}^{N^*_{(i,l),t}}\kappa_{i,t,k}\kappa_{l,t,k},$$

$$=\frac{1}{n}\overline{\kappa_t^2}+\frac{n-1}{n}\sum_{k=1}^{N_t^*}\overline{\kappa_{.,t,}\kappa_{..,t,k}}$$

where $\overline{\kappa_t^2}$ is the average jump component for each individual stock on day t, $\overline{\kappa_{.,t,}\kappa_{..,t,}}$ is the average cross-product of co-jumps that occur during that day and N_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

$$RV_{EQW,t} - BV_{EQW,t}\approx\sum_{k=1}^{N_t^*}\overline{\kappa_{.,t,}\kappa_{..,t_k}}\approx\sum_{k=1}^{N_t^*}\overline{co_{t,k}}$$

where $\overline{co_{t,}}$ 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 $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\sum_{j=1}^{M} r_{i,t_j} r_{l,t_j}$ term in the expression for $RV_{EQW,t}$, and it was this observation that motivated Bollerslev et al (2008) to emphasize the cross-product 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 variance 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-product 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. $2\sum_{i=1}^{n-1}\sum_{l=i+1}^{n} r_{i,t_j} r_{l,t_j}$, to measure covariation over the jth intraday interval on day t. Then they defined a mean cross-product (mcp) test statistic to be

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

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 jth intraday interval. The above mcp_{t_j} test statistic can be rearranged to obtain

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

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

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

$$\begin{aligned} mcp_{t_j} &= \frac{n}{n-1}((\frac{1}{n}\sum_{i=1}^{n} r_{i,t_j})^2 - \sum_{i=1}^{n}(\frac{1}{n})^2 r_{i,t_j}^2) \\ &= \frac{n}{n-1}(\widehat{var(r_{EQW,t_j})} - \frac{1}{n^2}\sum_{i=1}^{n}\widehat{var(r_{i,t_j})}), \end{aligned} \tag{7}$$

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

$$z_{mcp,t_j} = \frac{mcp_{t_j} - \overline{mcp_t}}{s_{mcp,t}}, \ j = 1, 2, ...M,$$

where

$$\begin{aligned} \overline{mcp_t} = \frac{1}{M}\sum_{j=1}^{M} mcp_{t_j} &= \frac{1}{M}(\frac{n}{n-1}\sum_{j=1}^{M}(\frac{1}{n}\sum_{i=1}^{n} r_{i,t_j})^2 - \frac{n}{n-1}\sum_{j=1}^{M}\sum_{i=1}^{n}(\frac{1}{n})^2 r_{i,t_j}^2) \\ &= \frac{1}{M}[\frac{n}{n-1}RV_{ew,t} - \frac{1}{n(n-1)}\sum_{i=1}^{n} RV_{i,t}], \end{aligned}$$

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

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 $s_{mcp,t}$. We find that this is the case for our mcp_{t_j} realizations. 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_{t_j} realizations might be masked by the correspondingly large sample mean $\overline{mcp_t}$. 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 information than that contained in returns. 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

$$z_{mcr,t_j} = \frac{mcr_{t_j} - \overline{mcr_t}}{s_{mcr,t}}, \ j = 1, 2, ...M, \tag{8}$$

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

$$\overline{mcr_t} = \sum_{j=1}^{M} mcr_{t_j} = \frac{1}{M}[\frac{n}{n-1}RRV_{ew,t} - \frac{1}{n(n-1)}\sum_{i=1}^{n} RRV_{i,t}] \text{ and}$$

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

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 average cross-product estimator and the daily average FHLL cross-product estimator to replace the cross-product of intraday returns in the Bollerslev et al (2008) $zmcp$ test statistic. Our FHLL co-jump test statistic ($zmfhllc$) is defined by

$$z_{mfhllc,t_j} = \frac{mfhllc_{t_j} - \overline{mfhllc_t}}{s_{mfhllc,t}}, \ j = 1, 2, ...M, \tag{9}$$

where

$$mfhllc_{t_j} = \frac{n}{n-1}(IFHLLC_{ew,t_j} - \sum_{i=1}^{n}(\frac{1}{n})^2 IFHLLC_{i,t_j})),$$

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 $mfhllc_{t_j}$ statistic using

$$\overline{mfhllc_t} = \sum_{j=1}^{M} mfhllc_{t_j} = \frac{1}{M}[\frac{n}{n-1}FHLLCV_{ew,t} - \frac{1}{n(n-1)}\sum_{i=1}^{n} FHLLCV_{i,t}] \quad \text{and}$$

$$s_{mfhllc,t} = \sqrt{\frac{1}{M-1}\sum_{j=1}^{M}(mfhllc_{t_j} - \overline{mfhllc_t})^2},$$

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 (*zmcp*) as defined in (7), the range-based co-jump test statistic (*zmcr*) as defined in (8), and the FHLL co-jump test statistic ($zmfhllc$) 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 *zmcr* test statistic and first-high-low-last price based $zmfhllc$ test statistic, and then compare them with the null distributions of the *zmcp* test statistic of Bollerslev et al (2008). We use the basic Euler scheme to simulate realizations of a stochastic multivariate 20×1 pure diffusion model (without jumps) that has zero drift and a covariance matrix $\Sigma = CC'$, and follows the process given by

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

The variable p_s represents a 20×1 vector process that resembles a vector of logarithmic prices of 20 assets, and $W(s)$ is a 20×1 vector of independent standard Brownian motions at time s. We use empirical calibration and set the covariance matrix Σ 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]

[7]Here we assume that there is no microstructure noise, and comment on how microstructure noise might affect results later.

For each trading day $[0, T]$, we let $(t_h)_{h\epsilon\{0,.....,K\}}$ be an equispaced time discretization of the day, with a time step of $\triangle t = 30$ seconds, where $K = T/\triangle t = 480$, and we generate $Z_{1,} Z_{2,.......} Z_K$ independently $N(0, I)$ distributed random vectors in R^{20}. The 20×1 vector of independent standard Brownian motions at times $0 = t_0 < t_1 < .... < t_K$ is then generated by setting $W(0)$ to be a 20×1 zero vector and calculating

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

The price sample paths for the 20 assets are then simulated by setting $p(0)$ to be a 20×1 zero vector and calculating

$$p_{t_{h+1}} = p_{t_h} + C'W(t_{h+1}), \qquad h = 0, ..., K - 1.$$

We do this for $t = \{1, 2, 3, ..., 90\}$. 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, 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

[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.

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 values 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 λ_i[11] and magnitude $N(0, \sigma_i^2)$,[12] and add them to their corresponding pure diffusion processes. For the common jumps, we simulate one Gaussian compound Poisson process with intensity λ 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 β_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 λ and the common jump size σ_J to empirical data (i.e. $\lambda = 0.05\%$ and $\sigma = 0.005$ in our case), and then change the intensity λ from 0.05% to 1%, and the size of σ_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

[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.

[11] Jump intensity λ_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.

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, λ and σ_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,[13] but the number of subintervals (m) used to calculate 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, λ and σ_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 *zmcp* 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 our multivariate asset price process is constant, and that there is no 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

[13] Our reported results relate to the 5-minute sampling frequency ($M = 48$) , but similar 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.

situation (see Xu and Zhang (2006), Wang, Yao, Fang and Li (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, and Chen at al (2010) study the microstructure of cross listed A and B shares on the Shanghai exchange. 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 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.[15] 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 exclude weekends, public holidays and periods when there are firm specific suspensions from our sample, and we avoid market opening effects 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, when calculating our intraday $zmcp$, $zmcr$, and $zmfhllc$ statistics, 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

[15] Bannouh et al (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.

and daily FHLL variance estimators. 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 $m + 1 = 11$.

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 20×1 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 40×1 diffusion with an appropriately calibrated 40×40 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 several empirically observed test statistics in each panel of data exceed their relevant critical values, providing evidence of co-jumps. Compared with the return-based type test statistics ($zmcp$ statistics), the range-based $zmcr$ statistics and the first-high-low-last price based $zmfhllc$ statistics are statistically significant on more occasions, suggesting that the latter tests are more powerful.

Tables 10 - 12 report the outcomes of the co-jump tests based on the three panels of stock data at the 0.1% significance level. These outcomes include co-jump arrival dates and times. In contrast to our previous research that has found frequent jumps in some of the individual stocks[16], we find relatively few co-jumps in the panels. The return-based test finds six co-jumps on the combined panel of forty stocks, the range based test finds the same co-jumps as well as an additional three co-jumps (9 in total), and the FHLL test finds all of these co-jumps as well as another six co-jumps (15 in total). Many of the detected co-jumps in this market occurred near the morning opening time or the afternoon closing time of trading sessions. Moreover, as noted in the footnote to Table 12, the timing of some of co-jumps coincided with the release of news on stock market regulations or monetary policy.

6 Conclusion

This paper explores the use of first-high-low-last (FHLL) prices in a multivariate high frequency setting. We introduce a first-high-low-last price based covariance estimator and study its properties, and find that after a very simple bias correction, the FHLL covariance estimator has lower root mean squared error than counterparts based on realized range and realized variance. We also use FHLL price data instead of returns data in the Bollerslev et al (2008) co-jump tests, and find an increase in power. When we apply our FHLL based co-jump test to a panel of high frequency data relating to Chinese mainland stock market, we find co-jumps in the stocks from the two stock market exchanges, and we are able to associate many of these co-jumps with announcements about changes in monetary policy or stock market regulations.

Our FHLL estimator of covariance is quite easy to calculate, since it relies only on univariate methodology (i.e. FHLL measures of the variances of two individual stocks and a portfolio containing those stocks). While the computational burden of estimation might not be a primary consideration when working with just two assets, it takes on more importance in a situation in which a test statistic that simply summarizes co-movement in many assets

[16] Details are available upon request.

is required. We have used FHLL measures of variance in our expression for covariance, and found that it compares favorably with analogous covariance measures based on simple range and return based estimates of variance, but if more accuracy or precision were required for a specific purpose, then it would be straightforward in principle to use other more sophisticated univariate measures of variance instead. Some examples include the variance estimators by Zhou (1996) or Zhang et al (2005), as well as methods based on realized kernels (Barndoff-Nielsen et al, 2008) and pre-averaging techniques (Jacod et al (2009)). These, and many other variance estimators could be used not only when estimating the covariance between two assets, but also in the construction of test statistics analogous to the Bollerslev et al (2008) test and our test. Our work in this paper has focussed on the potential benefits of including information about price ranges (in addition to returns) in covariance estimation and in a test for co-jumps in a large panel, but further benefits are likely as additional observed information about the price process and noise structure is employed.

The literature on covariance estimation is growing very rapidly, and it now includes very detailed examinations of the bias effects of microstructure noise and asynchronous trading on covariance estimation, as well as ways of accounting for this. Some recent work on this topic includes Mancino and Sanfelici (2011) and Zhang (2011)). Griffin and Oomen (2011) study several covariance estimators and conclude that the choice between them can depend on the properties of the price process. The bias correction that we employed in our setting of forty stocks was simple and effective, and avoided potential difficulties associated with treating different stocks in the portfolio differently. Nevertheless, we anticipate that further research on bias corrections in high frequency panels will be useful, and advance current understanding of the joint effects of many different types of microstructure noise in truly multivariate contexts.

Thus far we have linked some of the co-jumps that we found to announcements in monetary policy and stock market regulations. We anticipate that it might also be possible to link some of the other co-jumps to political announcements or financial events that occurred overseas. We leave further investigation of possible reasons for co-jumps in China for later work. Meanwhile, the empirical evidence of co-jumps in financial markets suggests that common factor models of jumps have empirical relevance. This lays open the possibility that models of co-jumps might have forecasting potential. We are currently working on this topic (see Liao, Anderson and Vahid (2010)) and have found encouraging results.

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

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

Theorem 1. As $M \longrightarrow \infty$, we have

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

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

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

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

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

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

Therefore,

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

Theorem 2.

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

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

$$P^M(g,h)_t = \sum_{j=1}^{M} g(\triangle_j^M P, \triangle_j^M S) h(\triangle_{j+1}^M P, \triangle_{j+1}^M S), \tag{1}$$

where $\triangle_j^M P = p_{t_j} - p_{t_{j-1}}$, $\triangle_j^M S = h_{t_j} - l_{t_j}$,

$g(\triangle_j^M P, \triangle_j^M S) = \begin{bmatrix} (2log(2)-1)(\triangle_j^M P)^2 & 0 \\ 0 & 0.5(\triangle_j^M S)^2 \end{bmatrix}$ and $h(\triangle_{j+1}^M P, \triangle_{j+1}^M S) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Then equation (1) becomes

$$P^M(g,h)_t = \begin{bmatrix} \sum_{j=1}^{M}(2log(2)-1)(\triangle_j^M P)^2 \\ \sum_{j=1}^{M} 0.5(\triangle_j^M S)^2 \end{bmatrix}.$$

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

$$Y^M(g,h)_t = \frac{1}{M}\sum_{i=1}^{M} g(\sqrt{M}\triangle_j^M Y) h(\sqrt{M}\triangle_{j+1}^M Y), \tag{2}$$

where g and h are two given matrix functions of dimensions $d_1 \times d_2$ and $d_2 \times d_3$, respectively, and $\triangle_j^M Y = p_{t_j} - p_{t_{j-1}}$. 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 $Y^M(g,h)_t$[17]. 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

[17] **See Barndorff-Nielsen et al (2006) for more details.**

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

$j = 1, 2$ 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

$$\rho_\sigma(g^{11}g^{11}) = 3(2log(2)-1)^2\sigma_j^4,\ \rho_\sigma(h^1) = \rho_\sigma(h^1h^1) = 1,$$

$$\rho_\sigma(g^{11}) = \rho_\sigma(g^{11}h^1) = \rho_\sigma(g^{11}h^2) = (2log(2)-1)\sigma_j^2,$$
$$\rho_\sigma(g^{22}g^{22}) = (0.5)^2\lambda_4\sigma_j^4 = \rho_\sigma(h^2) = \rho_\sigma(h^2h^2) = 1,$$

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

Further, since

$$\rho_\sigma(\mathrm{g}^{11}\mathrm{g}^{22}) = \mathrm{E}((\triangle_j^M\mathrm{P})^2(\triangle_j^M\mathrm{S})^2) = \mathrm{E}((\triangle_j^M\mathrm{P})^2(\mathrm{h}_{t_j}\text{-}\mathrm{l}_{t_j})^2)$$

$$= E((\triangle_j^M P)^2 h_{t_{j,\triangle_j}}^2 - 2h_{t_{j,\triangle_j}} l_{t_{j,\triangle_j}} (\triangle_j^M P)^2 + (\triangle_j^M P)^2 l_{t_{j,\triangle_j}}^2)$$
$$= E((\triangle_j^M P)^2 h_{t_{j,\triangle_j}}^2) - 2E(h_{t_{j,\triangle_j}} l_{t_{j,\triangle_j}} (\triangle_j^M P)^2) + E((\triangle_j^M P)^2 l_{t_{j,\triangle_j}}^2),$$

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

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

We then have that

$$A(\sigma,g,h)^{1,1} = (2log(2)-1)^2(3\sigma_j^4 + 2\sigma_j^4 - 3\sigma_j^4) = 0.3\sigma_j^4$$
$$A(\sigma,g,h)^{2,2} = (0.5)^2(\lambda_4\sigma_j^4 + 2\lambda_2^2\sigma_j^4 - 3\lambda_2^2\sigma_j^4) = 0.7825\sigma_j^4$$
$$\text{and } A(\sigma,g,h)^{1,2} = (0.94 - 0.386\times 1.385)\sigma_j^4 = 0.41\sigma_j^4,$$

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

$$\int_{t-1}^{t} (A^{1,1}(\sigma(s),g,h) - 2A^{1,2}(\sigma(s),g,h) + A^{2,2}(\sigma(s),g,h))ds = 0.27\int_{t-1}^{t}\sigma^4(s)ds.$$

References

[1] AïtSahalia, Y., J. Fan and D. Xiu, 2010,"High Frequency Covariance Estimates with Noisy and Asynchronous Financial Data", Journal of the American Statistical Association, 105, 1504- 1517.

[2] Andersen, T.G., T. Bollerslev, F. X. Diebold and P. Labys (2001), "The Distribution of Real Exchange Rate Volatility", Journal of the American Statistical Association, v 96, pp 42 - 55.

[3] Bannouh, K., D. V. Dijk., and M. Martens. (2009), "Range-Based Covariance Estimation Using High-Frequency Data: The Realized Co-Range", Journal of Financial Econometrics 1-32.

[4] Barndorff-Nielsen, O.E. and N. Shephard. (2003), "Measuring the Impact of Jumps in Multivariate Price processes using Bi-Power Covariation", Working Paper, Nuffield College, oxford University.

[5] Barndorff-Nielsen, O.E. and N. Shephard. (2004), "Power and Bipower variation with Stochastic Volatility and Jumps," Journal of Financial Econometrics, 2, 1-37.

[6] Bardnorff-Nielsen, O.E. and N. Shephard. (2006), "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," Journal of Financial Econometrics, 4, 1-30.

[7] Barndorff-Nielsen, O.E., P. R. Hansen, A. Lunde and N. Shephard, (2008). "Designing Realized Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise", Econometrica, 76 (6) 1481 - 1536.

[8] Bollerslev, T., T. H. Law., and G. Tauchen. (2008), "Risk, jumps, and diversification", Journal of Econometrics 144:234-256.

[9] Brandt, M. W., and F. X. Diebold. (2006), "A No-arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations", Journal of Business 79:61-74.

[10] Chen, J., J. Williams and R Buckland (2010), "A Tale of Two Microstructures: Spillovers of Informed Trading and Liquidity for cross listed Chinese A and B Shares. Working Paper, University of Swansea, U.K.

[11] Christensen, K., and M. Podolskij. (2007), "Realized Range-based Estimation of Integrated Variance", Journal of Econometrics 141:323-349.

[12] Garman, M. B. and Klass, M. J. (1980), “On the Estimation of Security Price Volatilities from Historical Data”, Journal of Business 53(1), 67-78.

[13] Gobbi, F., and C. Mancini. (2007), “Identifying the Covariation between the Diffusion Parts and the Co-jumps Given Discrete Observations”, Working Paper, Dipartimento di Matematica per le Decisioni, Universit‘a degli Studi di Firenze.

[14] Griffin, J. E. and R. C. A. Oomen, "Covariance Estimation in the Presence of Non-Synchronous Trading and Microstrucutre Noise", Journal of Econometrics, 160. p58 - 68.

[15] Huang, X., and G. Tauchen. (2005), “The Relative Contribution of Jumps to Total Price Variance”, Journal of Financial Econometrics, 3(4):456-499.

[16] Jacod, J., and V. Todorov. (2009), “Testing For Common Arrivals of Jumps for Discretely Observed Multidimensional Processes”, The Annals of Statistics 37(4): 1792-1838.

[17] Jacod, J., Y. Li, P. A. Mykland, M. Podolskij and M. Vetter, (2009). "Microstructure Noise in the Continuous Case: The Pre-averaging Approach", Stochastic Processes and their Applications, 119, 2249 - 2276.

[18] Liao, Y. (2008), "Does Decomposing Realized Volatility Help in Risk Prediction: Evidence from the Chinese Mainland Stock Market", PhD Dissertation, School of Economics, Australian National University.

[19] Liao, Y., H. Anderson., and F. Vahid (2010), “Do Jumps Matter? Forecasting Multivariate Realized Volatility allowing for Common Jumps”, Department of Econometrics and Business Statistics Working Paper 11, 2010, Monash University.

[20] Ma, C.H and Wang, X.Z. (2009), “The Jump Behavior of China’s Stock Market Prices: A Non-parametric Analysis with the High Frequency Data of the SSE Composite Index”, Working Paper, Xiamen University.

[21] Mancino, M. E. and S. Sanfelici, 2011, "Estimating Covariance via Fourier Method in the Presence of Asynchronous Trading and Microstructure Noise", Journal of Financial Econometrics, p367 - 408.

[22] Martens, M., and D. van Dijk (2007), “Measuring Volatility with the Realized Range”, Journal of Econometrics 138:181-207.

[23] Parkinson, M. (1980), "The Extreme Value Method for Estimating the Variance of the Rate of Return", Journal of Business, 53:61-65.

[24] Podolskij, P., and D. Ziggel. (2008), "A Range-Based Test for the Parametric Form of the Volatility in Diffusion Models", CREATES Research Paper 2008-22.

[25] Rogers, L. L. C. G. and S. E. Satchell (1991), "Estimating Variance from High, Low and Closing Prices", Annals of Applied Probability, 1, 4, pp 504 - 512.

[26] Todorov, V. (2010), "Variance Risk Premium Dynamics: The Role of Jumps", Review of Financial Studies, 23(1):345-383.

[27] Wang, C.F., Yao, N., Fang, Z.M., and Li, Y. (2008), "An Empirical Research on Jump Behavior of Realized Volatility in Chinese Stock Markets", Systems Engineering, 26(2):213-215.

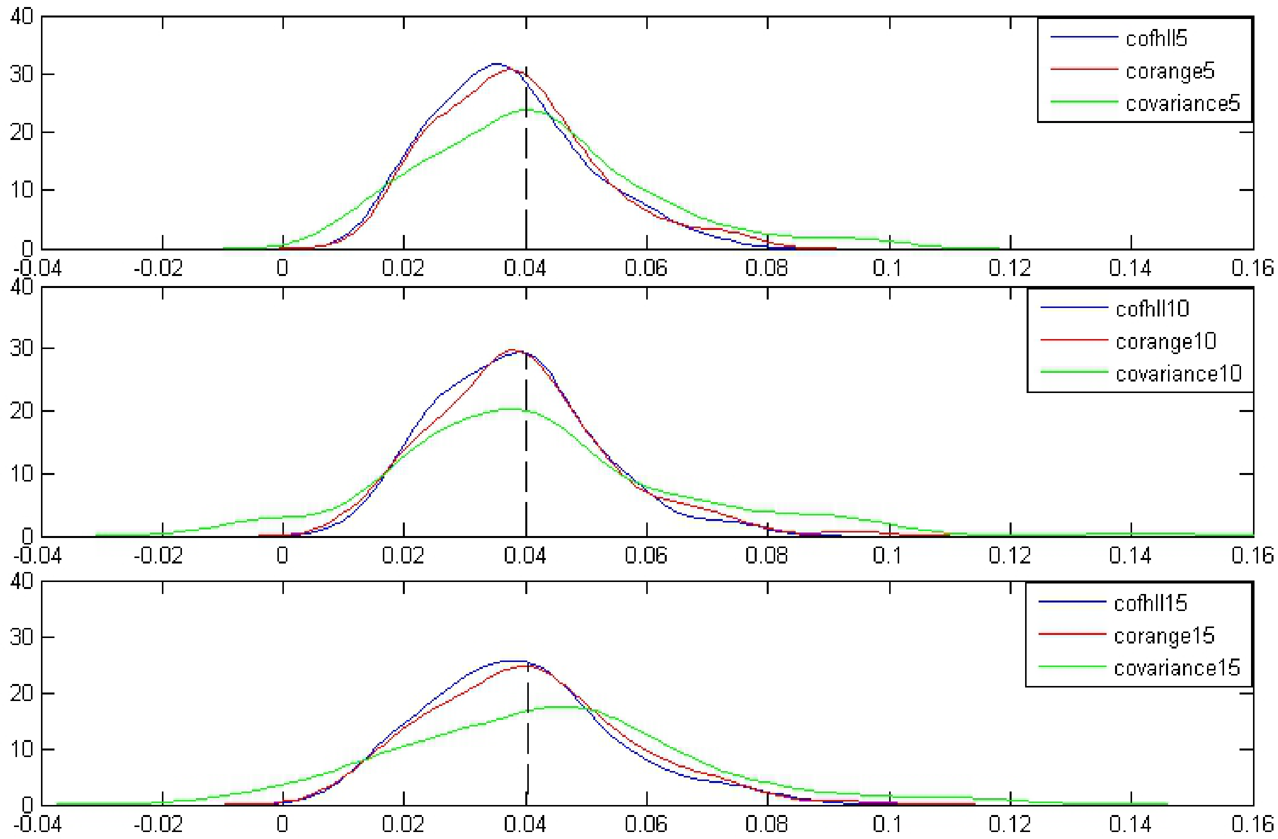
[28] Xu, Z.G., and Zhang, S.Y. (2006), "Research on Modeling Realized Volatility based on High Frequency Data", Journal of Systems Engineering, 21(1):165-167.

[29] Zhang, L, P A Mykland and Y. Aït-Sahalia (2006), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data", Journal of The American Statistical Association, Vol 100, #472, pp 1394 - 1411.

[30] Zhang, L. (2011), "Estimating Covariation: Epps Effect, Microstructure Noise", Journal of Econometrics, 160, 33 - 47.

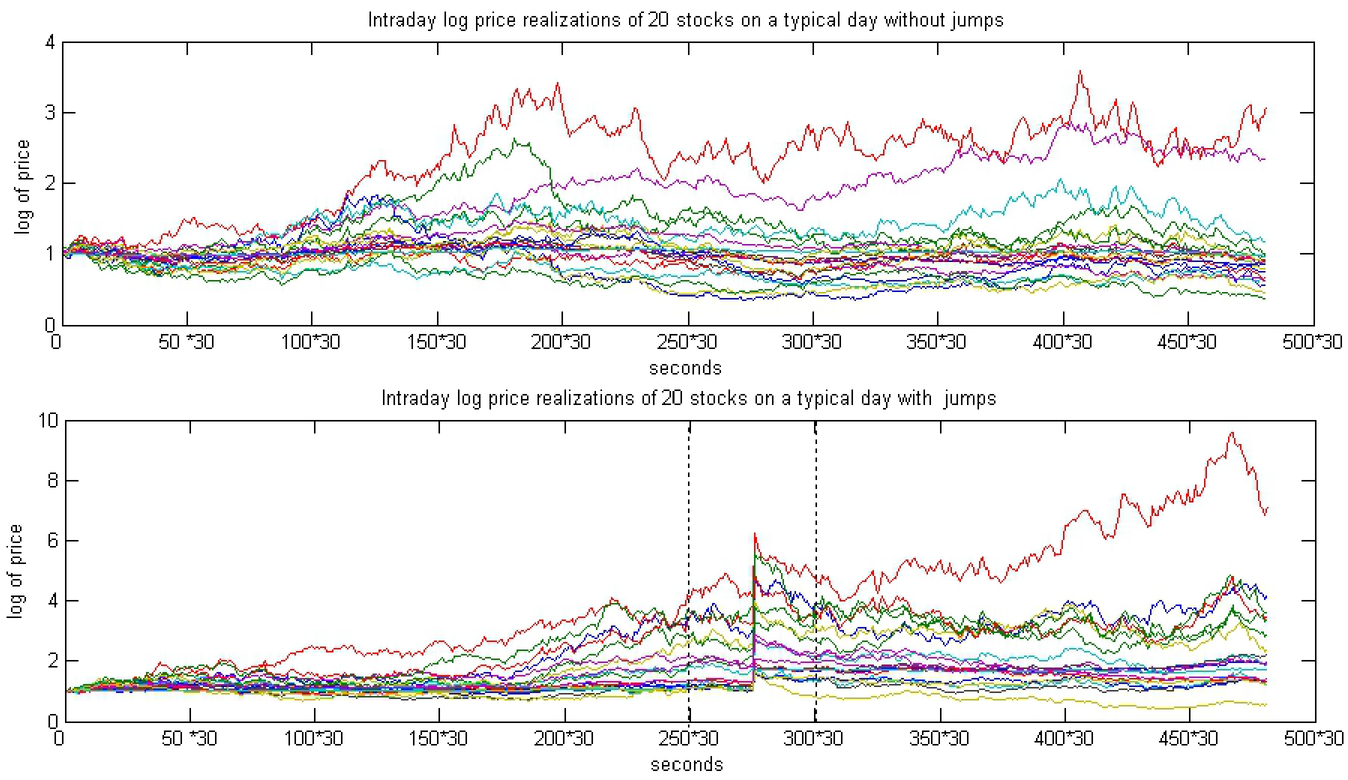
[31] Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign- Exchange Rates", Journal of Business and Economic Statistics, 14(1):45-52.

Figure 1: Density Graphs of the Three Covariance Estimators (Constant Volatility and No Microstructure Noise)



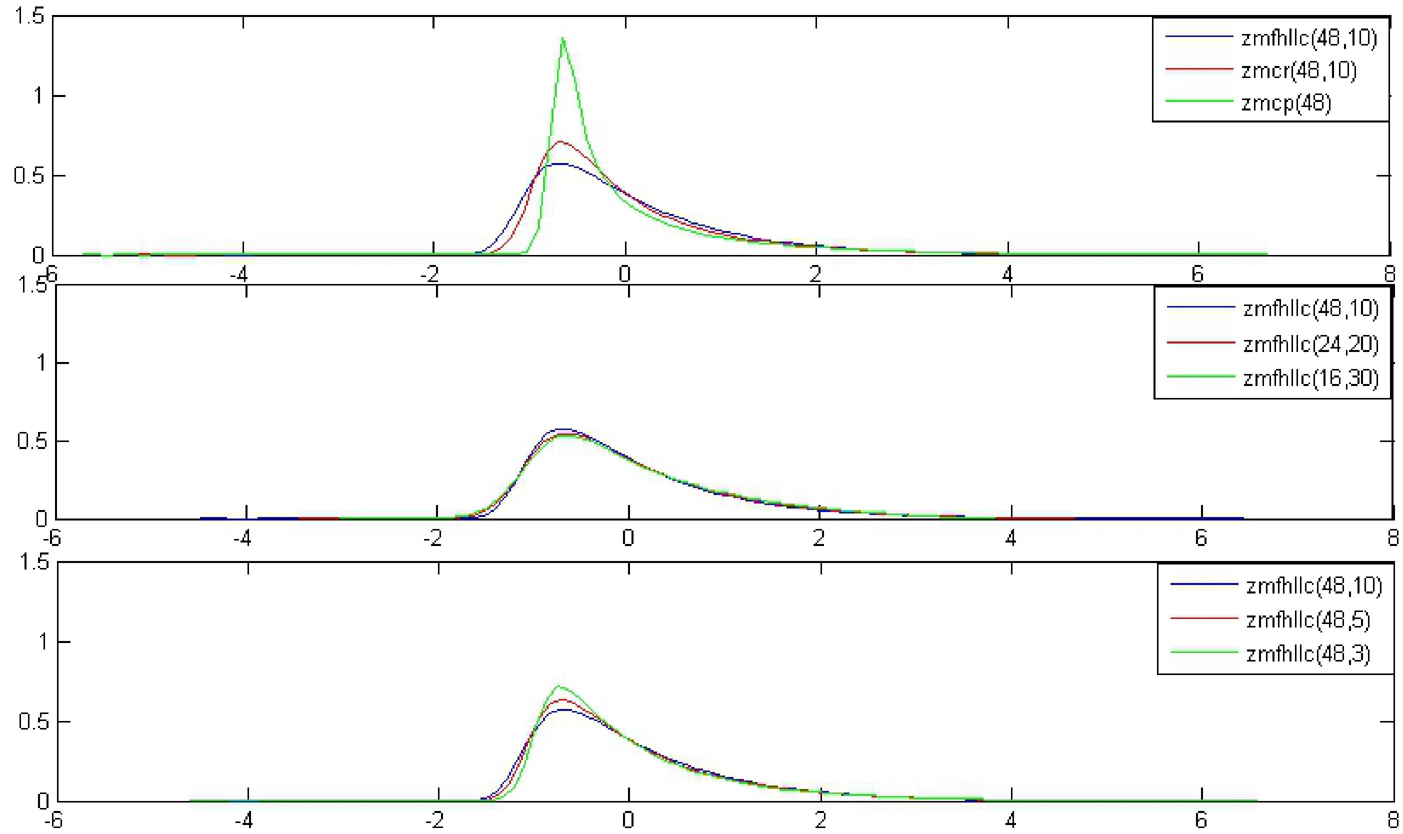
Note: The true covariance is 0.04. The top panel shows the distributions of the three covariance estimators (realized covariance, realized co-range, and first-high-low-last price based estimator) via simulation with 5000 days based on a 5-minute sampling frequency. The middle panel shows the distributions 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



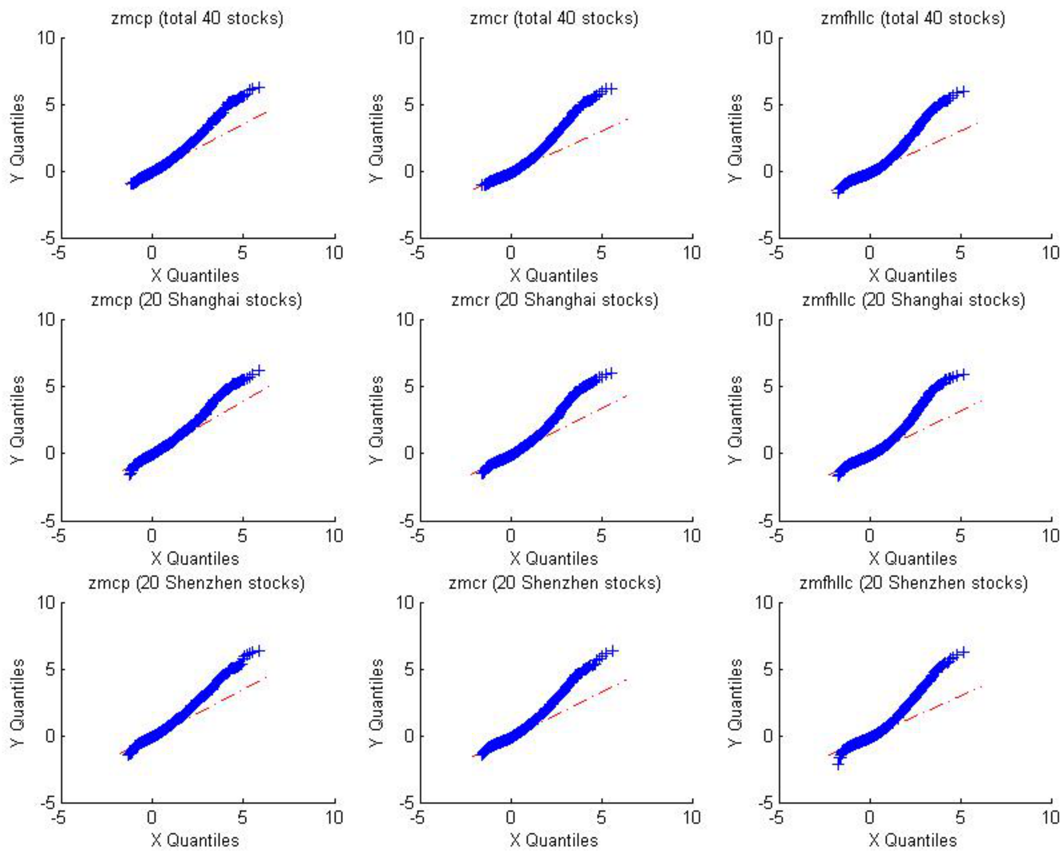
Note: The top panel plot shows simulated intraday price realizations of 20 stocks on a typical day based on a multivariate pure diffusion process with zero drift. The bottom panel plot shows intraday price realizations of 20 stocks which now include idiosyncratic jumps and one co-jump. The interval between 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



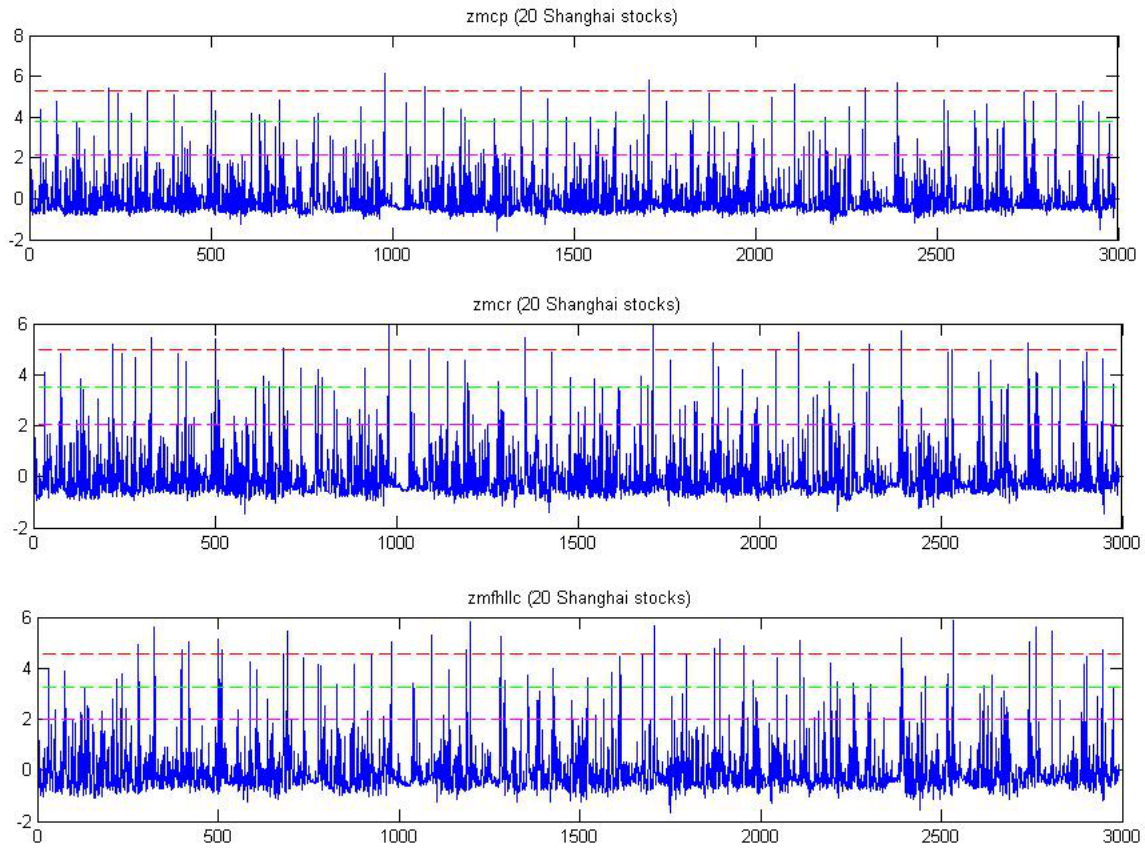
Note: The top panel shows the null distributions of the three test statistics (zmcp, zmcr, and zmfhllc test statistics) when daily variance calculations are based on 5-minute intervals (M=48) and subintervals for range and first-high-low-last price calculations are recorded 10 times (m=10) during each 5-minute interval. The middle panel shows the null distributions of zmfhllc test statistics for different number of intervals (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 4: QQ plots of Empirical vs Bootstrapped Null Distributions of Test Statistics for Chinese Data



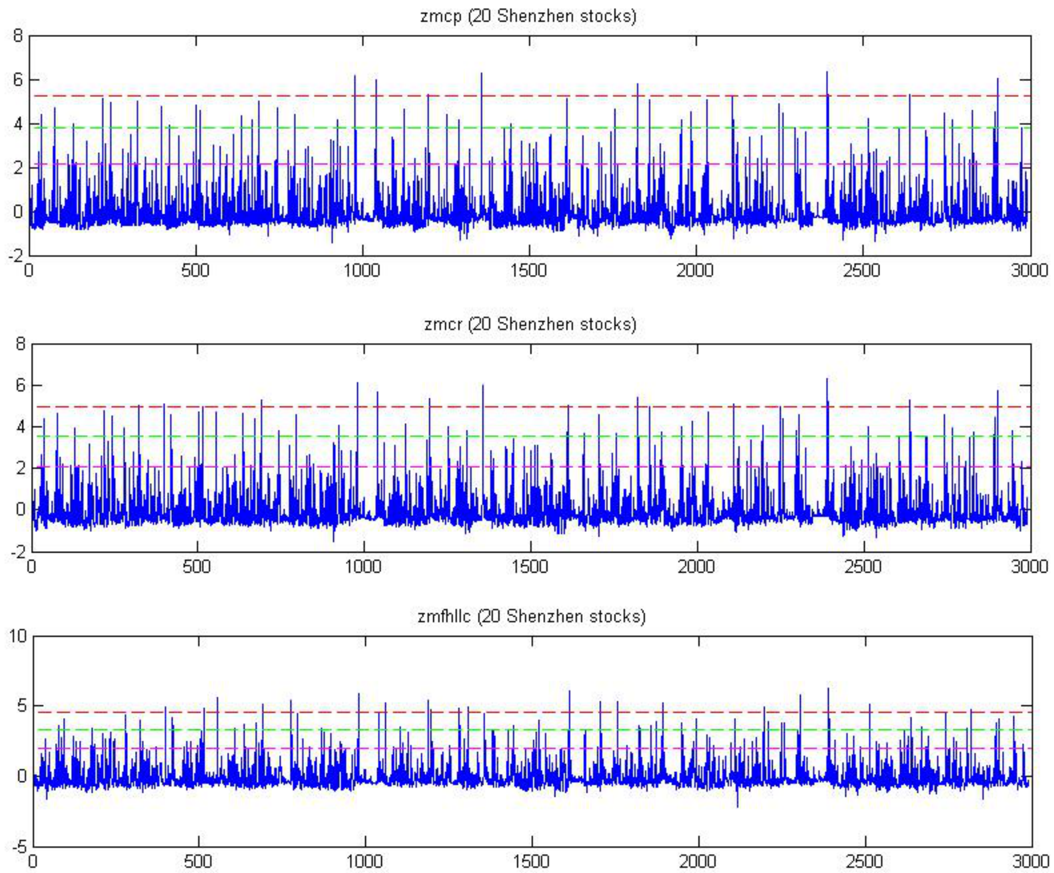
Note: The top panel shows Q-Q plots of empirical vs bootstrapped null distributions of the three test statistics (zmcp, zmcr, and zmfhllc test statistics) for 40 stocks. The middle panel shows Q-Q plots of empirical vs bootstrapped null distributions of the three test statistics (zmcp, zmcr, and zmfhllc test statistics) for 20 stocks on the Shanghai Stock Exchange. The bottom panel shows Q-Q plots of empirical vs bootstrapped null distributions of the three test statistics (zmcp, zmcr, and zmfhllc test statistics) for 20 stocks on the Shenzhen Stock Exchange. The underlying red dotted 45 degree line shows quantiles of the bootstrapped null distributions of the test statistics, and line composed of blue crosses shows quantiles of the empirical distributions of the test statistics.

Figure 5: Plots of Empirical Test Statistics for 20 Stocks on the Shanghai Stock Exchange



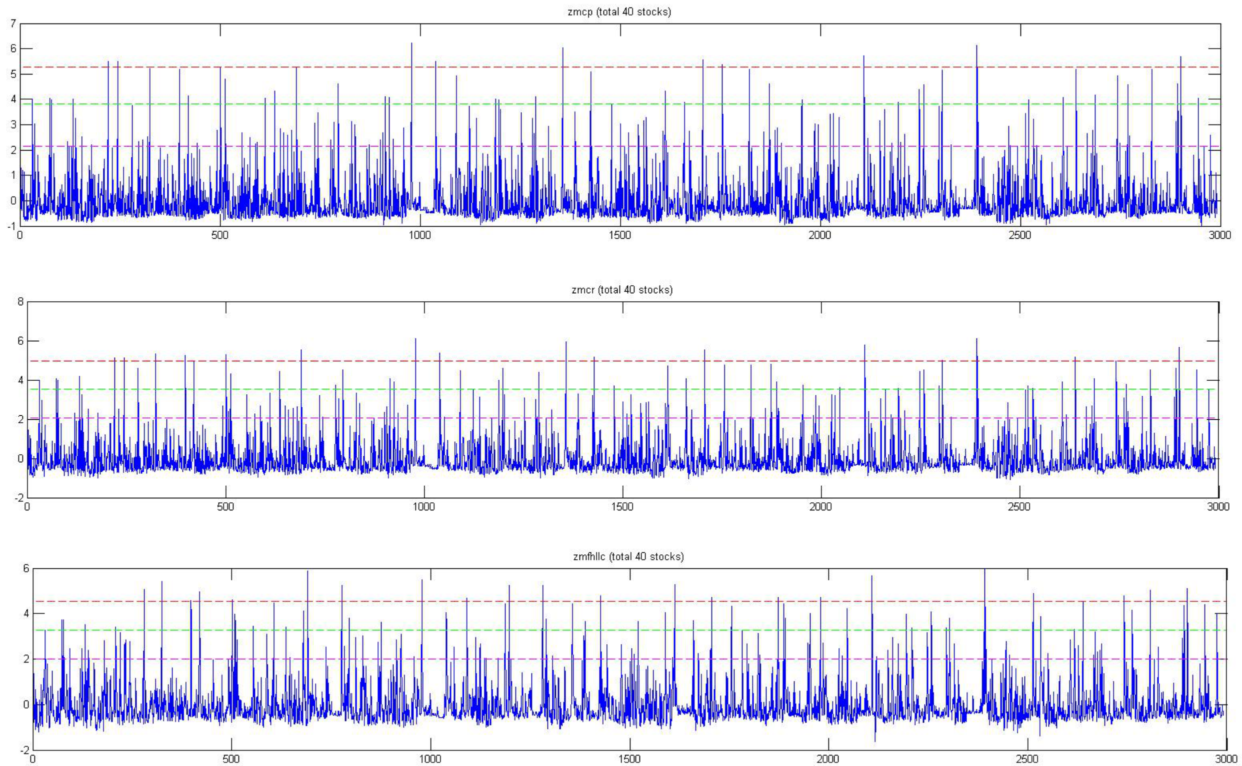
Note: The top plot shows the series of the return-based test statistics (*zmcp*) calculated from the empirical data. The middle plot shows the series of the range-based test statistics (*zmcr*) calculated from the empirical data. The bottom plot shows the series of the first-high-low-last price based test statistics (*zmfhllc*) 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 6: Plots of Empirical Test Statistics for 20 Stocks on the Shenzhen Stock Exchange



Note: The top plot shows the series of the return-based test statistics ($zmcp$) calculated from the empirical data. The middle plot shows the series of the range-based test statistics ($zmcr$) calculated from the empirical data. The bottom plot shows the series of the first-high-low-last price based test statistics ($zmfhllc$) 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 shows the series of the return-based test statistics ($zmcp$) calculated from the empirical data. The middle plot shows the series of the range-based test statistics ($zmcr$) calculated from the empirical data. The bottom plot shows the series of the first-high-low-last price based test statistics ($zmfhllc$) 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

	RCV_t		RCR_t		$FHLLCV_t$	
Sampling Frequency	Bias	RMSE	Bias	RMSE	Bias	RMSE
	Panel A:	Constant Covariance				
1(minute)	-0.0004	0.0229	-0.0007	0.0117	-0.0008	0.0113
5(minutes)	-0.0002	0.0282	-0.0005	0.0130	-0.0006	0.0126
10(minutes)	-0.0001	0.0304	-0.0002	0.0147	-0.0002	0.0132
15(minutes)	0.0001	0.0315	-0.0002	0.0159	-0.0003	0.0153
30(minutes)	-0.0003	0.0367	-0.0003	0.0186	-0.0003	0.0183
60(minutes)	-0.0005	0.0411	-0.0006	0.0204	-0.0006	0.0198
240(daily)	-0.0005	0.0859	-0.0008	0.0411	-0.0009	0.0361
	Panel B:	Stochastic Covariance				
1(minute)	-0.0009	0.0372	-0.0014	0.0134	-0.0016	0.0126
5(minutes)	-0.0006	0.0398	-0.0008	0.0136	-0.0009	0.0127
10(minutes)	-0.0002	0.0420	-0.0006	0.0149	-0.0008	0.0135
15(minutes)	-0.0003	0.0429	-0.0004	0.0154	-0.0004	0.0148
30(minutes)	-0.0003	0.0432	-0.0003	0.0183	-0.0003	0.0177
60(minutes)	-0.0004	0.0487	-0.0005	0.0228	-0.0006	0.0208
240(daily)	-0.0006	0.0881	-0.0007	0.0425	-0.0008	0.0331

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

	RCV_t		$RCV_{C,t}$		RCR_t		$RCR_{C,t}$		$FHLLCV_t$		$FHLLCV_{C,t}$	
Sampling Frequency	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: Stochastic Covariance ($s = 0, \tau = 15$)												
1(minute)	-0.0188	0.0304	-0.0011	0.0298	-0.0203	0.0317	-0.0013	0.0275	-0.0209	0.0334	-0.0020	0.0239
5(minutes)	-0.0159	0.0289	-0.0009	0.0313	-0.0195	0.0291	-0.0011	0.0281	-0.0209	0.0301	-0.0019	0.0254
10(minutes)	-0.0136	0.0273	-0.0009	0.0398	-0.0152	0.0275	-0.0011	0.0294	-0.0158	0.0290	-0.0017	0.0289
15(minutes)	-0.0099	0.0297	-0.0008	0.0408	-0.0104	0.0279	-0.0009	0.0227	-0.0105	0.0278	-0.0014	0.0215
30(minutes)	-0.0087	0.0335	-0.0005	0.0476	-0.0066	0.0280	-0.0009	0.0209	-0.0059	0.0284	-0.0012	0.0197
60(minutes)	-0.0058	0.0413	-0.0003	0.0592	-0.0035	0.0312	-0.0007	0.0242	-0.0026	0.0301	-0.0011	0.0225
240(daily)	-0.0006	0.0862	-0.0006	0.0862	-0.0009	0.0436	-0.0009	0.0436	-0.0010	0.0386	-0.0010	0.0386
Panel B: Stochastic Covariance ($s = 0.075, \tau = 1$)												
1(minute)	-0.0010	0.0367	-0.0001	0.0357	0.0516	0.1133	0.0012	0.0289	0.0719	0.2743	0.0013	0.0322
5(minutes)	-0.0007	0.0402	0.0005	0.0380	0.0570	0.0936	0.0012	0.0299	0.0793	0.2541	0.0013	0.0316
10(minutes)	-0.0005	0.0423	0.0013	0.0488	0.0544	0.0748	0.0011	0.0325	0.0756	0.1678	0.0012	0.0304
15(minutes)	-0.0001	0.0431	0.0015	0.0510	0.0367	0.0654	0.0010	0.0346	0.0509	0.1287	0.0011	0.0312
30(minutes)	0.0007	0.0439	-0.0002	0.0577	0.0255	0.0482	0.0005	0.0389	0.0351	0.0830	0.0008	0.0367
60(minutes)	-0.0001	0.0516	-0.0005	0.0695	0.0173	0.0426	0.0006	0.0414	0.0240	0.0594	0.0007	0.0386
240(daily)	-0.0006	0.0885	-0.0006	0.0885	-0.0006	0.0442	-0.0006	0.0442	-0.0006	0.0429	-0.0006	0.0429
Panel C: Stochastic Covariance ($s = 0.075, \tau = 15$)												
1(minute)	-0.0191	0.0428	-0.0012	0.0420	0.0322	0.1133	0.0009	0.0271	0.0520	0.1025	0.0010	0.0258
5(minutes)	-0.0160	0.0412	-0.0010	0.0415	0.0240	0.0936	0.0008	0.0299	0.0394	0.0925	0.0009	0.0272
10(minutes)	-0.0139	0.0429	-0.0009	0.0438	0.0131	0.0748	0.0006	0.0312	0.0235	0.0714	0.0009	0.0294
15(minutes)	-0.0100	0.0508	-0.0009	0.0516	0.0025	0.0654	0.0006	0.0334	0.0074	0.0601	0.0008	0.0316
30(minutes)	-0.0089	0.0576	-0.0006	0.0598	0.0013	0.0482	0.0005	0.0349	0.0053	0.0461	0.0007	0.0335
60(minutes)	-0.0054	0.0692	-0.0004	0.0713	0.0006	0.0426	0.0003	0.0371	0.0029	0.0496	0.0005	0.0352
240(daily)	-0.0004	0.0894	-0.0004	0.0894	-0.0006	0.0463	-0.0006	0.0463	-0.0007	0.0452	-0.0007	0.0452

Note: The true covariance is 0.04. RCV_t is realized covariance, RCR_t is realized co-range, and $FHLLCV_t$ is the first-high-low-last price covariance estimator. $RCV_{C,t}$ is the bias-corrected realized covariance, $RCR_{C,t}$ is the bias-corrected realized co-range, and $FHLLCV_{C,t}$ is the bias-corrected first-high-low-last price covariance estimator.

Table 3: Critical Values for Co-Jump Statistics in the Monte Carlo Study

	$\alpha = 0.1\%$			$\alpha = 1\%$			$\alpha = 5\%$		
	M=16	M=24	M=48	M=16	M=24	M=48	M=16	M=24	M=48
zmcp	3.608	4.296	5.350	3.215	3.546	3.838	2.235	2.205	2.134
zmcr (m=maximum possible)	3.480	4.096	5.041	2.996	3.274	3.563	2.064	2.066	2.042
zmcr (m=half of maximum possible)	3.491	4.125	5.132	3.012	3.308	3.621	2.078	2.082	2.056
zmcr (m=third of maximum possible)	3.505	4.165	5.198	3.029	3.357	3.684	2.088	2.100	2.067
zmfhllc (m=maximum possible)	3.334	3.852	4.611	2.804	3.025	3.285	1.960	1.980	1.988
zmfhllc (m=half of maximum possible)	3.354	3.905	4.783	2.831	3.075	3.398	1.980	2.002	2.013
zmfhllc (m=third of maximum possible)	3.385	3.983	4.979	2.860	3.162	3.512	1.994	2.034	2.035

Note: α 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

Test Statistics			$\sigma_J = 0.005$	$\sigma_J = 0.01$	$\sigma_J = 0.05$	$\sigma_J = 0.1$
zmcp	$(M = 48)$	$\lambda = 0.05\%$	0.0414	0.0654	0.1078	0.1543
		$\lambda = 0.1\%$	0.0523	0.0967	0.1765	0.2156
		$\lambda = 0.5\%$	0.1245	0.1762	0.2543	0.3059
		$\lambda = 1\%$	0.1698	0.2765	0.3247	0.3976
zmcp	$(M = 24)$	$\lambda = 0.05\%$	0.0394	0.0581	0.0967	0.1325
		$\lambda = 0.1\%$	0.0372	0.0765	0.1524	0.1987
		$\lambda = 0.5\%$	0.1039	0.1434	0.2076	0.2764
		$\lambda = 1\%$	0.1355	0.2076	0.2877	0.3088
zmcp	$(M = 16)$	$\lambda = 0.05\%$	0.0297	0.0455	0.0731	0.1087
		$\lambda = 0.1\%$	0.0308	0.0692	0.1123	0.1568
		$\lambda = 0.5\%$	0.0935	0.1120	0.1567	0.2034
		$\lambda = 1\%$	0.1027	0.1569	0.2033	0.2534
Nominal significance level$= 0.1\%$						

Note: λ denotes co-jump intensity, and σ_J denotes co-jump size.

Table 5: The Power of the Range-based Co-Jump Test Statistic $zmcr$ for Different Sampling Frequencies M

Test Statistics			$\sigma_J = 0.005$	$\sigma_J = 0.01$	$\sigma_J = 0.05$	$\sigma_J = 0.1$
$zmcr$	$(M = 48)$	$\lambda = 0.05\%$	0.1032	0.1765	0.2654	0.3421
		$\lambda = 0.1\%$	0.1421	0.2325	0.3217	0.4236
		$\lambda = 0.5\%$	0.2144	0.3524	0.4762	0.5543
		$\lambda = 1\%$	0.3176	0.4789	0.5632	0.6321
$zmcr$	$(M = 24)$	$\lambda = 0.05\%$	0.0721	0.1415	0.2106	0.2930
		$\lambda = 0.1\%$	0.1034	0.1925	0.2854	0.3507
		$\lambda = 0.5\%$	0.2032	0.2985	0.3987	0.4965
		$\lambda = 1\%$	0.2526	0.3247	0.4123	0.5521
$zmcr$	$(M = 16)$	$\lambda = 0.05\%$	0.0587	0.1065	0.1523	0.1921
		$\lambda = 0.1\%$	0.0705	0.1490	0.2067	0.2542
		$\lambda = 0.5\%$	0.1442	0.2103	0.2869	0.3103
		$\lambda = 1\%$	0.2042	0.3024	0.3976	0.4576
Nominal significance level$=0.1\%$						

Note: λ denotes co-jump intensity, and σ_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

Test Statistics			$\sigma_J = 0.005$	$\sigma_J = 0.01$	$\sigma_J = 0.05$	$\sigma_J = 0.1$
$zmfhllc$	$(M = 48)$	$\lambda = 0.05\%$	0.1247	0.1987	0.2871	0.3589
		$\lambda = 0.1\%$	0.1567	0.2546	0.3564	0.4236
		$\lambda = 0.5\%$	0.2534	0.3865	0.4987	0.5986
		$\lambda = 1\%$	0.3985	0.5321	0.6543	0.7653
$zmfhllc$	$(M = 24)$	$\lambda = 0.05\%$	0.1087	0.1523	0.2456	0.3452
		$\lambda = 0.1\%$	0.1324	0.2357	0.3067	0.3783
		$\lambda = 0.5\%$	0.2462	0.3234	0.4357	0.5346
		$\lambda = 1\%$	0.2985	0.3764	0.4893	0.6341
$zmfhllc$	$(M = 16)$	$\lambda = 0.05\%$	0.0745	0.1236	0.1673	0.2436
		$\lambda = 0.1\%$	0.0843	0.1651	0.2587	0.2791
		$\lambda = 0.5\%$	0.1659	0.2546	0.3466	0.3763
		$\lambda = 1\%$	0.3065	0.3987	0.4671	0.5334
Nominal significance level= 0.1%						

Note: λ denotes co-jump intensity, and σ_J denotes co-jump size.

Table 7: The Power of the Range-based Test Statistic $zmcr$ for Different m ($M = 48$)

Test Statistics			$\sigma_J = 0.005$	$\sigma_J = 0.01$	$\sigma_J = 0.05$	$\sigma_J = 0.1$
$zmcr$	($m = 10$)	$\lambda = 0.05\%$	0.1032	0.1765	0.2654	0.3421
		$\lambda = 0.1\%$	0.1421	0.2325	0.3217	0.4236
		$\lambda = 0.5\%$	0.2144	0.3524	0.4762	0.5543
		$\lambda = 1\%$	0.3176	0.4789	0.5632	0.6321
$zmcr$	($m = 5$)	$\lambda = 0.05\%$	0.0487	0.0921	0.1543	0.2178
		$\lambda = 0.1\%$	0.0842	0.1541	0.2367	0.3219
		$\lambda = 0.5\%$	0.1746	0.2543	0.3167	0.4876
		$\lambda = 1\%$	0.2124	0.2987	0.3542	0.5273
$zmcr$	($m = 3$)	$\lambda = 0.05\%$	0.0326	0.0743	0.1237	0.1876
		$\lambda = 0.1\%$	0.0654	0.1328	0.1965	0.2378
		$\lambda = 0.5\%$	0.1343	0.2145	0.2976	0.3542
		$\lambda = 1\%$	0.1978	0.2856	0.3765	0.4589
Nominal significance level= 0.1%						

Note: λ denotes co-jump intensity, and σ_J denotes co-jump size.

Table 8: The Power of the First-high-low-last Price Based Test Statistic $zmfhllc$ for Different m ($M = 48$)

Test Statistics			$\sigma_J = 0.005$	$\sigma_J = 0.01$	$\sigma_J = 0.05$	$\sigma_J = 0.1$
$zmfhllc$	$(m = 10)$	$\lambda = 0.05\%$	0.1247	0.1987	0.2871	0.3589
		$\lambda = 0.1\%$	0.1567	0.2546	0.3564	0.4236
		$\lambda = 0.5\%$	0.2534	0.3865	0.4987	0.5986
		$\lambda = 1\%$	0.3985	0.5321	0.6543	0.7653
$zmfhllc$	$(m = 5)$	$\lambda = 0.05\%$	0.0503	0.0990	0.1982	0.2527
		$\lambda = 0.1\%$	0.0967	0.1854	0.2765	0.3588
		$\lambda = 0.5\%$	0.1998	0.2876	0.3543	0.4908
		$\lambda = 1\%$	0.2554	0.3234	0.3778	0.5879
$zmfhllc$	$(m = 3)$	$\lambda = 0.05\%$	0.0414	0.0885	0.1534	0.1936
		$\lambda = 0.1\%$	0.0702	0.1534	0.2007	0.2512
		$\lambda = 0.5\%$	0.1538	0.2235	0.3079	0.3774
		$\lambda = 1\%$	0.2017	0.2998	0.3974	0.4836
Nominal significance level= 0.1%						

Note: λ denotes co-jump intensity, and σ_J denotes co-jump size.

Table 9: Critical Values for Co-Jump Statistics in the Empirical Study

	$\alpha = 0.1\%$			$\alpha = 1\%$			$\alpha = 5\%$		
	SH20	SZ20	All40	SH20	SZ20	All40	SH20	SZ20	All40
$zmcp$	5.265	5.268	5.264	3.815	3.813	3.820	2.138	2.132	2.141
$zmcr$	4.957	4.961	4.953	3.535	3.528	3.541	2.043	2.042	2.045
$zmfhllc$	4.538	4.549	4.546	3.257	3.261	3.262	1.987	1.983	1.989

Note: α is the significance level. Critical values in the columns labled SH20 relate to the 20 stocks from the Shanghai Stock Exchange, critical values in the columns labled SZ20 relate to the 20 stocks from the Shenzhen Stock Exchange and critical values in the columns labled All40 relate to the 40 stocks from both exchanges.

Table 10: Co-jumps Dates and Times for the Three Panels of Stocks Based on the Return-based Co-jump Test

20 Stocks on Shanghai Stock Exchange		20 Stocks on Shenzhen Stock Exchange		Total 40 Stocks	
Co-Jumps Date	Co-Jumps Time	Co-Jumps Date	Co-Jumps Time	Co-Jumps Date	Co-Jumps Time
2007/07/09	09:50-09:55	2007/07/31	10:30-10:35	2007/07/09	09:50-09:55
2007/08/02	13:50-13:55	2007/08/01	13:20-13:25	2007/07/31	10:30-10:35
2007/08/10	11:20-11:25	2007/08/24	13:25-13:30	2007/08/02	13:50-13:55
2007/08/22	09:50-09:55	2007/09/11	14:50-14:55	2007/08/22	09:50-09:55
2007/09/03	14:25-14:30	2007/09/12	09:35-09:40	2007/09/03	14:25-14:30
2007/09/10	09:50-09:55			2007/09/11	14:50-14:55
2007/09/11	14:50-14:55				

Table 11: Co-jumps Dates and Times for the Three Panels of Stocks Based on the Range-based Co-jump Test

20 Stocks on Shanghai Stock Exchange		20 Stocks on Shenzhen Stock Exchange		Total 40 Stocks	
Co-Jumps Date	Co-Jumps Time	Co-Jumps Date	Co-Jumps Time	Co-Jumps Date	Co-Jumps Time
2007/07/09	09:50-09:55	2007/07/11	09:40-09:45	2007/07/09	09:50-09:55
2007/07/11	09:40-09:45	2007/07/23	09:35-09:40	2007/07/11	09:40-09:45
2007/07/16	14:30-14:35	2007/07/31	10:30-10:35	2007/07/23	09:35-09:40
2007/07/23	09:35-09:40	2007/08/01	13:20-13:25	2007/07/31	10:30-10:35
2007/07/31	10:30-10:35	2007/08/14	09:35-09:40	2007/08/02	13:50-13:55
2007/08/02	13:50-13:55	2007/08/24	13:25-13:30	2007/08/14	09:35-09:40
2007/08/10	11:20-11:25	2007/09/03	14:25-14:30	2007/08/22	09:50-09:55
2007/08/22	09:50-09:55	2007/09/06	14:20-14:25	2007/09/03	14:25-14:30
2007/09/03	14:25-14:30	2007/09/11	14:50-14:55	2007/09/11	14:50-14:55
2007/09/10	09:50-09:55	2007/09/12	09:35-09:40		
2007/09/11	14:50-14:55				
2007/09/17	09:40-09:45				

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

20 Stocks on Shanghai Stock Exchange		20 Stocks on Shenzhen Stock Exchange		Total 40 Stocks	
Co-Jumps Date	Co-Jumps Time	Co-Jumps Date	Co-Jumps Time	Co-Jumps Date	Co-Jumps Time
2007/07/09	09:50-09:55	2007/07/12	13:35-13:40	2007/07/09[a]	09:50-09:55
2007/07/11	09:40-09:45	2007/07/17	10:05-10:10	2007/07/11	09:40-09:45
2007/07/16	14:30-14:35	2007/07/18	09:40-09:45	2007/07/17[b]	09:50-09:55
2007/07/17	09:50-09:55	2007/07/23	09:35-09:40	2007/07/23[c]	09:35-09:40
2007/07/23	09:35-09:40	2007/07/31	10:30-10:35	2007/07/31[d]	10:30-10:35
2007/07/31	10:30-10:35	2007/08/02	13:50-13:55	2007/08/02	13:50-13:55
2007/08/02	13:50-13:55	2007/08/06	14:10-14:15	2007/08/07	09:35-09:40
2007/08/06	14:15-14:20	2007/08/07	09:35-09:40	2007/08/14	09:35-09:40
2007/08/07	09:35-09:40	2007/08/09	11:00-11:05	2007/08/20	09:45-09:50
2007/08/21	10:55-11:00	2007/08/20	09:45-09:50	2007/08/22[e]	09:50-09:55
2007/08/22	09:50-09:55	2007/08/22	09:50-09:55	2007/08/27	13:45-13:50
2007/08/24	09:35-09:40	2007/08/23	10:10-10:15	2007/08/30	09:35-09:40
2007/08/27	13:45-13:50	2007/09/05	13:50-13:55	2007/09/03	14:25-14:30
2007/08/29	11:15-11:20	2007/09/10	09:50-09:55	2007/09/11[f]	14:50-14:55
2007/09/03	14:25-14:30	2007/09/11	14:50-14:55	2007/09/17[g]	09:40-09:45
2007/09/11	14:50-14:55	2007/09/17	09:40-09:45		
2007/09/17	09:40-09:45	2007/09/25	09:35-09:40		

[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.