CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Testing for Co-Jumps in Financial Markets

Jan Novotny and Giovanni Urga

CEA@Cass Working Paper Series

WP–CEA–06-2017

Testing for Co-Jumps in Financial Markets

Jan NOVOTNY [1]

Giovanni URGA[2]

This version: 24 April 2017

In this paper, we introduce the notion of co-jumps within the co-features framework. We formulate a limiting theory of co-jumps and discuss their discrete sample properties. In the presence of idiosyncratic price jumps, we identify the notion of weak co-jumps. A Monte Carlo exercise assesses the small sample properties of the co-jump test. We illustrate the empirical relevance of the proposed framework via an empirical application using the components of the DJIA30 between 01/01/2010 and 30/06/2012.

Keywords: Co-features, Jumps and Co-jumps, Portfolio Diversification.

J.E.L. Classification Number: C12, C32, G12

[1]Centre for Econometric Analysis, Cass Business School, London (UK) and CERGE-EI, Charles University, Prague (Czech Republic)

[2]Centre for Econometric Analysis, Faculty of Finance, Cass Business School, London (UK) and University of Bergamo (Italy)

Abstract

In this paper, we introduce the notion of co-jumps within the co-features framework. We formulate a limiting theory of co-jumps and discuss their discrete sample properties. In the presence of idiosyncratic price jumps, we identify the notion of weak co-jumps. A Monte Carlo exercise assesses the small sample properties of the co-jump test. We illustrate the empirical relevance of the proposed framework via an empirical application using the components of the DJIA30 between 01/01/2010 and 30/06/2012.

Keywords: Co-features, Jumps and Co-jumps, Portfolio Diversification.

J.E.L. Classification Number: C12, C32, G12

This paper proposes a novel theoretical framework to assess common price jumps in a multivariate framework using the notion of co-features, i.e. the existence of a linear combination of time series in which individual features are eliminated, as originally proposed by Engle and Kozicki (1993) and more recently reconsidered in the special issue of Journal of Business and Economic Statistics (2007).

There is a huge body of literature on the identification of price jumps in the univariate context. Several procedures have been proposed to test for the presence of price jumps defined as discontinuity in the price process. See, for example, Aït-Sahalia and Jacod (2009), Aït-Sahalia and Jacod (2011), Aït-Sahalia and Jacod (2012), Andersen et al. (2011, 2012), Barndorff-Nielsen and Shephard (2004b, 2006), Jiang and Oomen (2008), Lee and Mykland (2008), Lee and Hannig (2010), Huang and Tauchen (2005), and Mancini (2009). Dumitru and Urga (2012) evaluate the performance of alternative non parametric price jump tests.

In contrast, a multivariate framework allows one to identify common jumps between stochastic processes as highlighted in the seminal work by Barndorff-Nielsen and Shephard (2004a). Bollerslev et al. (2008) test for the presence of portfolio-wide systemic price jumps and focus in particular on systemic common jumps without counterparts on the individual time series level. This framework is extended by Liao and Anderson (2011) using the range-based indicators proposed by Bannouh et al. (2009). Jacod and Todorov (2009) propose a procedure to test for the joint occurrence of price jump arrivals at a pair of time series. In an empirical study, Lahaye et al. (2011) estimate the joint probabilities of common price jump arrivals and also suggest a joint statistic for the estimation of common price jumps and map common jumps in response to specific macro-news for a broader range of assets such as USD exchange rates, US Treasury bonds futures and US equity futures. Based on factor regressions techniques in Bollerslev et al. (2013), Bollerslev et al. (2016) relate the identification of co-jumps to estimating factors and loadings for the strict factor model and verify the method on the sensitivity of the stock price jumps of Microsoft to the market jumps. In the case of an unknown factor structure Aït-Sahalia and Xiu (2015) and Pelger

(2016) provide estimators based on principal component analysis. Li et al. (2016) propose a framework to evaluate the dependency between jumps of two processes and to test for the relationship implied by the linear standard factor model. Caporin et al. (2015) introduce a non-parametric test based on the smoothed estimators of integrated variance to provide evidence for statistically significant multivariate jumps in stock prices. Gilder et al. (2014) analyze the contemporaneous co-jumps of US equities and link them to Federal Fund Target Rate announcements. Jiang et al. (2011) conclude that surprises related to macroeconomic news announcements have limited power in explaining jumps for bonds. Aït-Sahalia et al. (2009) use common price jumps for assets in the same sector to evaluate the optimal portfolio in the presence of jumps. Finally, in a recent paper, Bandi and Renò (2016) propose a novel identification strategy for price and volatility co-jumps to relate some significant price changes to volatility jumps.

This paper contributes to the current literature on common price jumps as follows: We propose a novel notion of co-jumps identified within the co-feature framework. In particular, the notion of co-jumps is linked to the diversification of price jumps out of a basket of assets. Thus, co-jumps can be intuitively understood as a possibility to diversify the price jumps completely out of a portfolio. Bollerslev et al. (2008) discuss the case of a portfolio of common jumps which cannot be diversified out, and as such it serves to identify common jumps. We further extend the notion of co-jumps to cases where each asset has idiosyncratic price jumps, implying the absence of co-jumps. We define weak co-jumps as a linear combination of assets with minimum contribution of price jumps to the quadratic covariance. This notion is further supported by the empirical results of Bollerslev et al. (2008) and Lahaye et al. (2011).

We assess the small sample properties using Monte Carlo simulation under several price jump scenarios. We also report an empirical illustration of the co-jump framework using the sample of the DJIA 30 index for January 1, 2010 to June 30, 2012, sampled at a 5-minute frequency. Finally, we perform a sensitivity analysis to understand the role of the spurious jump detection.

The paper is organized as follows: In Section 1, we provide the definition and the main properties of co-jumps, and specify the procedure to test for the presence of co-jumps. In Section 2, we evaluate the small sample properties of the co-jumps test using Monte Carlo simulations. In Section 3, we report an empirical illustration of the co-jumps using constituents of the DJIA 30 index and provide a robustness check with respect to the multiple testing bias. Section 4 concludes.

1. Modeling Co-Jumps

This paper introduces the notions of co-jumps within the co-feature framework. Let us consider an N-dimensional vector of log-prices $\log P = \left(\log P^{(1)}, \ldots, \log P^{(N)}\right)'$, where the vector $\log P = \{\log P_t\}_{0 \leq t \leq 1}$ is defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}\right)$ over the finite time interval $[0, t]$. The vector of log-prices is the semi-martingale $\mathcal{F}_t$-adapted and its continuous-time dynamics can be specified by the following stochastic differential equation

$$d \log P_t = \mu_t dt + \sigma_t dB_t + d\mathcal{J}_t \,, \tag{1}$$

where μ_t is N-dimensional vector of drift processes, σ_t is the $(N \times N)$-dimensional covariance matrix, dB_t is the N-dimensional vector of independent standard Brownian motions, and $d\mathcal{J}_t$ is the N-dimensional vector of pure jump Lévy processes.

The presence of price jumps in (1) implies that a $(N \times N)$-dimensional quadratic variation process Σ_t can be written as

$$\Sigma_t = \Sigma_t^{(c)} + \Sigma_t^{(d)} \,, \tag{2}$$

where $\Sigma_t^{(c)}$ represents the continuous part of the semi-martingale process,

$$\Sigma_t^{(c)} = \int_0^t \sigma_s \sigma_s' ds \,, \text{ with } \left\{\Sigma_t^{(c)}\right\}_{i,j} < \infty \,, \; i, j = 1, \ldots, N \,, \tag{3}$$

and $\Sigma_t^{(d)}$ represents the discontinuous part of the semi-martingale process,

$$\Sigma_t^{(d)} = \sum_{j=1}^{N_t} c_j c_j' \, , \text{ with } \left\{ \Sigma_t^{(d)} \right\}_{i,j} < \infty \, , \, i,j = 1, \ldots, N \, , < \infty \, , \tag{4}$$

where c_j is N-dimensional vector for which there exists at least one $i = 1, \ldots, N$ such that $d \log P_{t_j -}^{(i)} > 0$, and N_t is the number of $t_j \leq t$. The decomposition of the quadratic variance allows us to map the presence of price jumps in terms of quadratic variation.

1.1. Co-jumps

Consider the integrated counterpart of the N-dimensional process described in (1)

$$\log P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s + \sum_{j=1}^{N_t} c_j \, , \tag{5}$$

where each of the components has discontinuities in the interval $[0, t]$. The N-dimensional Brownian semi-martingale process with finite-activity jumps is closed with respect to the stochastic integration under a linear transformation given by a $(p \times N)$-dimensional matrix Ω, where the matrix can in general be time-dependent (see Jacod and Shiryaev, 2003). The p-dimensional process given as a linear transformation of $\log P_t$ can be written as

$$\begin{aligned} \Omega \log P_t &= \int_0^t \Omega \mu_s ds + \int_0^t \Omega \sigma_s dB_s + \sum_{j=1}^{N_t} \Omega c_j \\ &= \int_0^t \mu_s^{(p)} ds + \int_0^t \sigma_s^{(p)} dB_s + \sum_{j=1}^{N_t} c_j^{(p)} \, , \end{aligned} \tag{6}$$

which is a p-dimensional Brownian semi-martingale with finite activity jumps. If $\log P_t$ is a Brownian semi-martingale, the product $\Omega \log P_t$ is Brownian semi-martingale as well.

Our aim is to find a Ω such that $\Omega \log P_t$ is a Brownian semi-martingale with $\Sigma_t^{(d)} \equiv 0$. If Ω exists, $\Omega \log P_t$ does not have any price jumps despite the presence of price jumps in each component. This characteristic is in fact the notion of co-features, as introduced by Engle and Kozicki (1993), and in the special issue of the Journal of Business and Economic Statistics (2007).

Definition of co-jumps:. For the N-dimensional process $\log P_t$ given by (5) with each of the components having a discontinuity in the interval $[0,t]$, *co-jumps* are defined as the existence of the N-dimensional constant vector Ω, different from the zero vector, such that for the process $\Omega \log P_t$ the discontinuous part of the semi-martingale process in the covariance disappears

$$\Sigma_t^{(d)} = \sum_{j=1}^{N_t} \Omega' c_j c_j' \Omega = 0\,.$$

The vector Ω is called the *co-jump vector* and the space of all co-jump vectors spans the *co-jump space.*

1.2. Identification of Jumps

Following Barndorff-Nielsen and Shephard (2006), let us now consider the $\hat{G}_\Omega$-statistic defined as

$$\hat{G}_\Omega = M^{1/2} \frac{\widehat{IV}_M - \widehat{QV}_M}{\widehat{IQ}_M}\,, \tag{7}$$

where $\widehat{IV}_M$ is the estimator of the Integrated Variance $\left(\widehat{IV}_M \xrightarrow{p} \int_0^t \sigma_s^2 ds\right)$, $\widehat{QV}_M$ is the estimator of the Quadratic Variance $\left(\widehat{QV}_M \xrightarrow{p} \int_0^t \sigma_s^2 ds + \sum_{j=1}^{N_t} c_j^2\right)$, $\widehat{IQ}_M$ is the estimator of the Integrated Quarticity $\left(\widehat{IQ}_M \xrightarrow{p} \int_0^t \sigma_s^4 ds\right)$. For a univariate log-price process $\log P_t$ generated by (1), under the null hypothesis of no price jumps, $\hat{G}_\Omega \xrightarrow{\mathcal{D}} N(0, \vartheta)$ with $\xrightarrow{\mathcal{D}}$ denoting a stable convergence in law and ϑ is some known constant depending on the particular choice of estimators used.

Thus, for the N-dimensional process $\log P_t$ in the interval $[0,t]$ there is a co-jump if a vector Ω exists such that the $\hat{G}_\Omega$-statistic for the univariate process $\Omega \log P_t$ does not reject the null hypothesis. The asymptotic properties of the $\hat{G}_\Omega$-statistic under the null hypothesis hold when there is no discontinuous part of the price process $\Omega \log P_t$. We identify co-jumps when the discontinuous part of the quadratic variance disappears, i.e. $\Sigma_t^{(d)} = 0$.

In this paper, we consider a sparse sampling approach to deal with market micro-structure noise since it provides a reasonable trade-off between accuracy and numerical feasibility at chosen sampling frequency. However, our framework can be extended to employ alternative techniques such as the pre-averaging method by Podolskij and Vetter (2009), employed by Aït-Sahalia and Jacod (2009) and Aït-Sahalia et al. (2012), or the combination of different time scales by Zhang et al. (2005), and Zhang (2011).

1.3. An Additional Co-jumps Feature: Weak Co-jumps

The notion of co-jumps introduced above, aims to find a linear combination which eliminates the jumps. When idiosyncratic price jumps are present for each component $\log P_t$ (see, for instance, Jiang et al., 2011; Lahaye et al., 2011; Lee, 2012) co-jumps do not exist as they cannot be fully eliminated. To this purpose, we modify the notion of co-jumps such that we weaken the requirement for the elimination of the jump term in (6).

Definition of Weak Co-jumps. We define *weak co-jumps* as a linear combination which minimizes the presence of price jumps in $\Omega \log P_t$. The minimisation of price jumps is done through the$\hat{G}_\Omega$-statistic. In the presence of price jumps, i.e., when the null hypothesis does not hold, $\hat{G}_\Omega \xrightarrow{p} \infty$, where p means convergence in probability. Thus, we define a weak co-jump as a linear combination(s) Ω, which maximizes the $\hat{G}_\Omega$-statistic. Then, the weak co-jump portfolio (vector) is

$$\Omega^* = argmin_{\{\Omega,|\Omega|=1\}} \hat{G}_\Omega$$

Thus, for an N-dimensional process $\log P_t$ we are statistically evaluate the difference between two vectors $\Omega^{(1)}$ and $\Omega^{(2)}$ via the $\hat{G}_\Omega$-statistic. .

2. Finite Sample Properties of the Co-jumps Test: Monte Carlo Simulation

In this section, we explore the finite sample properties of the co-jump test we propose via a Monte Carlo simulation exercise. We evaluate the performance of the price jump tests for a single time series assessing its size and power. We then evaluate size and power of the co-jump test under seven alternative scenarios.

2.1. Simulation Design

Let us consider an N-dimensional stochastic volatility process given as

$$
\begin{aligned}
d\log P_t^{(i)} &= u_t \sigma_t^{(i)} dW_t^{(i)} + U_{P,t}^{(i)} dJ_{P,t}^{(i)}\,, && (8)\\
d\sigma_t^{(i)2} &= \kappa^{(i)}\left(\theta^{(i)} - \sigma_t^{(i)2}\right)dt + \omega^{(i)}\sigma_t^{(i)} dB_t^{(i)} + U_{\sigma,t}^{(i)} dJ_{\sigma,t}^{(i)}\,, && (9)\\
E\left[dW_t^{(i)} dB_t^{(i)}\right] &= \rho^{(i)} dt\,, && (10)\\
E\left[dW_t^{(i)} dB_t^{(j)}\right] &= 0\,,\ i \neq j\,, && (11)\\
E\left[dW_t^{(i)} dW_t^{(j)}\right] &= \rho_W^{(i,j)} dt\,, && (12)
\end{aligned}
$$

where $i = 1, \ldots, N$, and u_t is the intraday volatility pattern based on Andersen et al. (2012) specified as a sum of two exponential terms leading in the asymmetric U-shape pattern

$$
u_t = c_1 + c_{open} \exp\left(-a_{open}\left(t - t_{open}\right)\right) + c_{close}\left(-a_{close}\left(t_{close} - t\right)\right)\,,
$$

where t denotes the time in the trading day, t_{open} is the opening time, t_{close} is the closing time and parameters are set as $c_1 = 0.8892$, $c_{open} = 0.75$, $c_{close} = 0.25$, $a_{open} = 10$, and $a_{close} = 10$. We consider the same volatility pattern for every asset.

The parameters for every asset are the same: $\kappa^{(i)} = \kappa = 0.0162$, $\theta^{(i)} = \theta = 0.573$, $\omega^{(i)} = \omega = 0.58$, and $\rho^{(i)} = \rho_{vol} = -0.46$. This ensures that the portfolio is composed of N assets of the same type. The correlation coefficients, $\rho_W^{(i,j)}$, are the same for all pairs with values $\rho_W^{(i,j)} = \{0\}$. The parameters are calibrated to one trading day and log-returns data are scaled by 100; therefore they correspond to percentage change.

The magnitude $U_{P,t}^{(i)}$ is set relative to the prevailing level of stochastic volatility as $U_{P,t}^{(i)} = \nu\sigma_{t-}^{(i)}$, where $\sigma_{t-}^{(i)}$ is the prevailing volatility immediately before the jump occurs and ν is the parameter common to all assets, for which we consider $\nu_i = 4$, $\nu_i = 8$, and $\nu_i = 12$, respectively, for all i. The magnitude of the volatility jumps $U_{\sigma,t}^{(i)}$ follows for every i an exponential distribution with mean $\mu_{\sigma}^{(i)} = \mu_{\sigma} = 1.25$. For every asset, we assume that $U_{\sigma,t}^{(i)}$ is drawn independently of each other. Further, for every asset i, the arrival process $dJ_{P,t}^{(i)}$ and $dJ_{\sigma,t}^{(i)}$ is for every asset assumed to be driven by the same stochastic intensity function $d\Lambda_t^{(i)}$; however, the realization of the price jump arrival is drawn independently. We thus use a stochastic volatility model calibrated by Eraker (2004) to equity markets, which is also used by Lee (2012).

For each scenario, we simulate 1,100 trading days. The first 100 trading days represent the initialization period, which are then discarded. We consider five ($N = 5$) time series of log-returns and we allow one price jump per day. The trading day is sampled at 5-minute frequency, starting at 9:00 and lasting up to 16:30.

Scenario 1 (No Jumps) No price jumps.

Scenario 2 (No Co-jumps Liquid Assets) Price jumps occur at random and at distinct times. The size of the jumps is $\nu_i = 4$ corresponding to the liquid assets with small jumps.

Scenario 3 (No Co-jumps Mid-liquid Assets) Price jumps occur at random and at distinct times. The size of the jumps is $\nu_i = 8$ corresponding to the mid-liquid assets with medium sized jumps.

Scenario 4 (No Co-jumps Illiquid Assets) Price jumps occur at random and at distinct times. The size of the jumps is $\nu_i = 12$ corresponding to the illiquid assets with large price jumps, where liquid on the market is very shallow and thus large price moves are likely.

Scenario 5 (Co-jumps Liquid Assets) Price jumps occur at the same random times for all five assets. The size of the jumps is $\nu_i = 4$ corresponding to the liquid assets with small jumps.

Scenario 6 (Co-jumps Mid-liquid Assets) Price jumps occur at the same random times for all five assets. The size of the jumps is $\nu_i = 8$ corresponding to the mid-liquid assets with medium sized jumps.

Scenario 7 (Co-jumps Illiquid Assets) Price jumps occur at the same random times for all five assets. The size of the jumps is $\nu_i = 12$ corresponding to the illiquid assets with large price jumps.

2.2. Simulation Results

Table 1 reports the size and the power of the univariate price jump test using the Bipower-based (BV-based) estimator and three sizes of price jumps. We use the standard 5% level of significance of the test, i.e. $\alpha = 0.05$, and we consider one time series at the timey. Testing for price jumps individually in Scenario 1 (no jumps) produces 12.6% cases of falsely identified price jumps, 13.8%, 80.1%, and 98.8% in Scenarios 2-4 respectively.

[Table 1 should be inserted here]

Table 2 reports the probability to identify co-jumps, either true or spurious, for Scenarios 2-7 . The results show that the probability to detect a true co-jump in Scenarios 5-7 is higher—the test has better power—than to price jump detection in univariate time series. Thus, testing for co-jumps in the case of true co-jumps is more efficient than testing for jumps in the single time series. On the other hand, the probability to detect a spurious co-jumps is higher with respect to the univariate test. The test is oversized as compared to the univariate test, and this is particularly true if we look at the results of Scenario 3, where the probability to detect a co-jump is 89.5% while there is no true co-jump present in the data. This can be explained as the situation where there has to be at least two time series with price jump detected. Size $\nu = 8$ is large enough to have a number of such cases. In fact, in 99.1% cases, we have detected at least two time series with price jumps and thus we have performed a test for co-jumps. When testing for co-jumps, however, the combined time series is likely not to show any price jump as the Integrated Variance and price jumps scale differently. This can be illustrated as follows: Let us consider n time series, each with the same integrated variance, uncorrelated to others, and one price jump taking place at disitnct times. When we form an equally weighted linear combination of the time series, the overall integrated variance scales as $\sqrt{n}$, i.e., if the volatility of the single time series is σ, the volatility of the linear combination is $\sqrt{n}\sigma$. Prie jumps, however, will be suppresed by factor of $1/n$. This comes from the fact that return, which contains a price jump for any of the time series is averaged out with $n-1$ other returns which does not have price jump. We finally obtain a time series, which has increased Integrated Variance and n price jumps of a suppressed magnitude by $1/n$. Thus, the linear combination of time series is evaluated as not having a jump and thus co-jump is detected. Thus, the test detects large number of false co-jumps.

On the other hand, in the case of Scenario 2, the size of the jumps is small and thus the number of cases when at least two time series with price jumps is detected is small and thus only in a small proportion of cases the co-jumps are tested. Conversely, for Scenario 4, the jumps are so large that at least two time series with price jumps are detected so frequently and, at the same time, the size of the individual jumps cannot be averaged out by $1/n$ as argued above so frequently. Thus, test has quite a large size, but very good power.

[Table 2 should be inserted here]

3. Empirical Illustration

In this section, we illustrate the empirical validity of the proposed theoretical framework by evaluating the presence of co-jumps in high-frequency data.

3.1. Data and Index Selection

We use the individual assets of the Dow Jones Industrial Average 30 (DJIA 30) index running from 1 January 2010 to 30 June 2012 provided by the NYSE TAQ database. We use data on trades only and utilize the appropriate cleaning mechanism by Barndorff-Nielsen et al. (2009). As a result, the data are sampled at a 5-minute frequency. Such a sampling frequency filters out the presence of the market micro-structure noise, while preserving the high-frequency features. The trading day starts at 9:30:00 and ends at 16:00:00, which yields 79 log-prices per day. Our sample contains 621 trading days in total. We split the DJIA 30 index into six indices, each with five companies, based on the capitalization at the beginning of the sample. We illustrate the notion of co-jumps using the High-cap index containing the five most capitalized companies, and the Low-Cap Index the the five least capitalized companies in the DJIA 30. Table 3 presents the composition of each of the indices as well as the market capitalization of companies at the beginning of the sample. The results using the indices with the remaining DJIA 30 companies are available upon request.

The descriptive statistics summarized in Table 3 reveal the large kurtosis for each asset and support the deviation from normality at a 5-minute frequency consistently across the set of equities.

[Table 3 should be inserted here.]

3.2. Co-jumps

We now employ the notion of co-jumps with the $\hat{G_{\Omega}}$-statistic calculated for each trading day. We use $\alpha = 0.05$ to test for the null hypothesis that there is no price jump(s) during the given trading day. Following Barndorff-Nielsen and Shephard (2006), we estimate the Integrated Variance, $\left(\widehat{QV}\right)$, the Integrated Variance, $\left(\widehat{IV}\right)$, and the Integrated Quarticity, $\left(\widehat{IQ}\right)$ as:

$$
\begin{aligned}
\widehat{QV}_D &= \sum_{i=1}^{M_D} r_{i,D}^2 \,, && (13)\\
\widehat{IV}_D &= \frac{M_D}{M_D-1}\mu_1^{-2}\sum_{i=2}^{M_D} |r_{i-1,D}|\,|r_{i,D}| \,, && (14)\\
\widehat{IQ}_D &= \frac{M_D}{M_D-3}\frac{1}{M_D}\mu_1^{-4}\sum_{i=4}^{M_D} |r_{i-3,D}|\,|r_{i-2,D}|\,|r_{i-1,D}|\,|r_{i,D}| \,. && (15)
\end{aligned}
$$

where $r_{i,D}$ is the i-th log-return on the day indexed by D, where each day is divided into $M_D = 78$ equally-sized 5-minute buckets, and $\mu_1 = E\left[|z|\right] = \sqrt{2/\pi}$ with $z \sim N(0,1)$. In such a case, the $\hat{G}$-statistic converges as $\hat{G}_D \xrightarrow{\mathcal{D}} N(0, \vartheta)$ with $\vartheta = (\pi^2/4) + \pi - 5 \cong 0.609$. The test for the presence of price jumps during the trading day D at $\alpha = 0.05$ has the form

$$
\begin{aligned}
H_0: &\quad \hat{G}_{\Omega(D)} \geq \sqrt{\vartheta}\Phi^{-1}(\alpha) \quad \text{no jump} \\
H_A: &\quad \hat{G}_{\Omega(D)} < \sqrt{\vartheta}\Phi^{-1}(\alpha) \quad \text{jump(s)}
\end{aligned},
$$

where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution giving $\sqrt{\vartheta}\Phi^{-1}(\alpha) \cong -1.284$.

In Figure 1, Panels (a) and (b) depict the results of the co-jumps exercise for the High-Cap and Low-Cap Indices, the most and the least capitalized set of assets in the DJIA 30 respectively. For every trading day, we find the co-jump vector Ω such that it maximizes the $\hat{G}_{\Omega}$-statistic (red dots). For every trading day and each Index, we test for the presence of co-jumps and confirm the presence of co-jumps as $\hat{G}_{\Omega(D)} \geq -1.284$, which is captured by the black long-dash line. This means that at the given sampling frequency, a linear combination of assets exists in the Index such that the price jumps diversify out.

[Figure 1 should be inserted here.]

Further, each of the two figures depict the range (gray shaded area) of the individual $\hat{G}$-statistics calculated for each asset in the Index. The results show that, for the majority of the trading days, at least one asset exists in the Index such that the null is rejected for both Indices. At the same time, there is no case where the null would be rejected for every asset and, therefore, there is no co-jump for all five assets at the same time.

In addition, the two figures report the $\hat{G_{\Omega}}$-statistic for equally weighted index (blue dots). The results indicate that in the majority of cases, the $\hat{G_{\Omega}}$-statistic for the equally weighted index is in the range implied by the individual assets. However, a significant number of cases show that the equally weighted index may either amplify or suppress the presence of price jumps. Table 4 summarizes the number of cases for each of the instances. Consequently, the figures suggest that the popular "1/N" strategy, or employing the equally weighted index, is not optimal for dealing with price jumps.

[Table 4 should be inserted here.]

To assess how much the individual assets contribute to the co-jumps, Figure 1, panels (c) and (d), present the range of the components of each co-jump vector identified above for the High-Cap and Low-Cap Indices, respectively. We consider co-jump vectors, normalized such that $\sum_{i=1}^{5} \Omega^{(i)2} = 1$. First, the figure depicts the minimum (green) and maximum (red) of the magnitude of the co-jump vectors. In particular, for High-Cap Index, the maximum magnitude oscillated around 0.75, while the minimum oscillated around 0.1 with the least magnitude taking the value of $2.02 \cdot 10^{-5}$ and the largest one $9.87 \cdot 10^{-1}$, taken from all Indices. Therefore each asset is significantly contributing to the co-jump vector and diversification of price jumps is clearly not caused by picking up an asset with few or no price jumps. Low-Cap Index provides the same qualitative conclusion.

The results show the presence of co-jump vectors. From the index perspective, the price jumps can be ex post diversified out at a 5-minute frequency. Further, the equally weighted index is not in general sufficient to eliminate price jumps. In some cases, it amplifies price jumps and thus the deviation from Gaussianity.

4. Conclusions

In this paper, we employed the co-feature framework to introduce the notion of co-jumps defined as a linear combination of assets which is free of price jumps. We extended the notion of co-jumps to assets with idiosyncratic price jumps to define the weak co-jumps as a linear combination which minimizes the price jumps. We then linked the concept to the optimization of in index of assets with respect to price jumps.

Further, we performed a Monte Carlo exercise to evaluate the small sample properties of the test. We found that the power of the test is higher for true co-jumps relative to the univariate counterpart and thus testing for the co-jumps in the case of true co-jumps is more efficient than testing for jumps in an individual time series.

Finally, we evaluated the empirical validity of the proposed framework using assets from the Dow Jones 30 Index from 1 January 2010 to 30 June 2012 sampled at a 5-minute frequency. We considered two indices, the High-Cap Index and the Low-Cap Index, based on the market capitalization and tested for co-jumps. The results showed the presence of co-jumps at 5-minute frequency, meaning that price jumps could be diversified out. However, our analysis showed that such diversification in general could not be achieved by creating equally weighted indices. Thus, the optimization in terms of removing price jumps should be considered as independent criteria.

The findings in this paper suggest some further developments. First, it will be interesting to extend the framework in this paper to the case of a more general price arrival process, e.g., mutually correlated self-exciting price jumps. Second, the sensitivity of the proposed framework, and in particular of the measure of commonality, can be transformed in the proper testing procedure for asynchronicity among the price jumps. Finally, it will also be interesting to develop an extension of the notion of co-arrivals to define the information measures capturing the different features of the multivariate arrival process. This is part of our ongoing research agenda.

Acknowledgments.

This paper is a substantially revised version of a manuscript previously circulated under the title “Co-features in Finance: Co-arrivals and Co-jumps”. We wish to thank participants in the conference "Skewness, Heavy Tails, Market Crashes, and Dynamics" (University of Cambridge, 28-29 April, 2014), 13th OxMetrics User Conference (Aarhus, 5-6 September, 2013), and in the International Conference on "Systemic Risk, Contagion and Jumps" (Cass Business School, 25 January 2013), special thanks go to Yacine Aït-Sahalia and Neil Shephard, the Finance Research Workshop (Cass Business School, 4 March 2013), in particular Richard Payne and Riccardo Pianeti, and the Economics Division Workshop (Nottingham Trent University, 6 March 2013) for useful comments. Special thanks to the Editor, George Tauchen, an Associate Editor, and two anonymous referees for very useful comments and suggestions which greatly helped to improve the paper. The usual disclaimer applies. Jan Novotný acknowledges funding from the European Community's Seventh Framework Program FP7-PEOPLE-2011-IEF under grant agreement number PIEF-GA-2011-302098 (Price Jump Dynamics), the Centre for Econometric Analysis, and GAČR grant 14-27047S.

References

Aït-Sahalia, Y., Cacho-Diaz, J., and Hurd, T. (2009). Portfolio choice with jumps: A closed-form solution. *The Annals of Applied Probability*, 19(2):556–584.

Aït-Sahalia, Y. and Jacod, J. (2009). Testing for jumps in a discretely observed process. *The Annals of Statistics*, 37(1):184–222.

Aït-Sahalia, Y. and Jacod, J. (2011). Testing whether jumps have finite or infinite activity. *The Annals of Statistics*, 39(3):1689–1719.

Aït-Sahalia, Y. and Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature*, 50(4):1007–50.

Aït-Sahalia, Y., Jacod, J., and Li, J. (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics*, 168:207–222.

Aït-Sahalia, Y. and Xiu, D. (2015). Principal component analysis of high frequency data. *National Bureau of Economic Research, N. 21584*.

Andersen, T., Dobrev, D., and Schaumburg, E. (2011). A functional filtering and neighborhood truncation approach to integrated quarticity estimation. Technical report, National Bureau of Economic Research.

Andersen, T. G., Dobrev, D., and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169(1):75–93.

Bandi, F. M. and Renò, R. (2016). Price and volatility co-jumps. *Journal of Financial Economics*, 119(1):107–146.

Bannouh, K., Van Dijk, D., and Martens, M. (2009). Range-based covariance estimation using high-frequency data: The realized co-range. *Journal of Financial Econometrics*, 7(4):341–372.

Barndorff-Nielsen, O., Hansen, P., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3):1–32.

Barndorff-Nielsen, O. and Shephard, N. (2004a). Measuring the impact of jumps in multivariate price processes using bipower covariation. *DP, Nuffield College, Oxford University*, 2:1–37.

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

Barndorff-Nielsen, O. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1–30.

Bollerslev, T., Law, T., and Tauchen, G. (2008). Risk, jumps, and diversification. *Journal of Econometrics*, 144(1):234–256.

Bollerslev, T., Li, S. Z., and Todorov, V. (2016). Roughing up beta: Continuous vs. discontinuous betas, and the cross-section of expected stock returns. *Journal of Financial Economics*, 120:464–490.

Bollerslev, T., Todorov, V., and Li, S. (2013). Jump tails, extreme dependencies, and the distribution of stock returns. *Journal of Econometrics*, 172(2):307–324.

Caporin, M., Kolokolov, A., and Renò, R. (2015). Multi-jumps. *Syrto Working Papers Series*, (3).

Dumitru, A.-M. and Urga, G. (2012). Identifying jumps in financial assets: a comparison between nonparametric jump tests. *Journal of Business & Economic Statistics*, 30(2):242–255.

Engle, R. and Kozicki, S. (1993). Testing for common features. *Journal of Business & Economic Statistics*, 11(4):369–380.

Eraker, B. (2004). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *The Journal of Finance*, 59(3):1367–1404.

Gilder, D., Shackleton, M. B., and Taylor, S. J. (2014). Cojumps in stock prices: Empirical evidence. *Journal of Banking & Finance*, 40:443–459.

Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3(4):456–499.

Jacod, J. and Shiryaev, A. (2003). *Limit Theorems for Stochastic Processes*, volume 288. Springer-Verlag Berlin.

Jacod, J. and Todorov, V. (2009). Testing for common arrivals of jumps for discretely observed multidimensional processes. *The Annals of Statistics*, 37(4):1792–1838.

Jiang, G. J., Lo, I., and Verdelhan, A. (2011). Information shocks, liquidity shocks, jumps, and price discovery: Evidence from the us treasury market. *Journal of Financial and Quantitative Analysis*, 46(02):527–551.

Jiang, G. J. and Oomen, R. C. (2008). Testing for jumps when asset prices are observed with noise: a swap variance approach. *Journal of Econometrics*, 144(2):352 – 370.

Lahaye, J., Laurent, S., and Neely, C. (2011). Jumps, cojumps and macro announcements. *Journal of Applied Econometrics*, 26(6):893–921.

Lee, S. (2012). Jumps and information flow in financial markets. *Review of Financial Studies*, 25(2):439–479.

Lee, S. and Hannig, J. (2010). Detecting jumps from lévy jump diffusion processes. *Journal of Financial Economics*, 96(2):271–290.

Lee, S. and Mykland, P. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6):2535–2563.

Li, J., Todorov, V., and Tauchen, G. (2016). Robust jump regressions. *Journal of the American Statistical Association (forthcoming).*

Liao, Y. and Anderson, H. M. (2011). Testing for co-jumps with high-frequency financial data: an approach based on first-high-low-last prices. *Monash University, Department of Econometrics and Business Statistics, Working Paper*, 09/11.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2):270–296.

Pelger, M. (2016). Large-dimensional factor modeling based on high-frequency observations. *Department of Management Science and Engineering, Stanford University, USA.*

Podolskij, M. and Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3):634–658.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47.

Zhang, L., Mykland, P., and Ait-Sahalia, Y. (2005). A tale of two time scales. *Journal of the American Statistical Association*, 100(472):1394–1411.

Table 1: Univariate price jump test.

Size	
No Jumps	12.6%
Power	
Liquid Assets	13.8%
Mid-liquid Assets	80.1%
Illiquid Assets	98.8%

Note: The table reports the size and the power of the Bipower-based test. Significance level: $\alpha = 0.05$.

Table 2: Probability to detect true/spurious co-jumps

	Design	Probability to detect co-jump
Scenario 1	**No Jumps**	14.1% (spurious detection)
Scenario 2	**No Co-jumps Liquid Assets**	20.0% (spurious co-jumps detection)
Scenario 3	**No Co-jumps Mid-liquid Assets**	89.5% (spurious co-jumps detection)
Scenario 4	**No Co-jumps Illiquid Assets**	28.2% (spurious co-jumps detection)
Scenario 5	**Co-jumps Liquid Assets**	17.3% (true co-jumps detection)
Scenario 6	**Co-jumps Mid-liquid Assets**	97.3% (true co-jumps detection)
Scenario 7	**Co-jumps Illiquid Assets**	100% (true co-jumps detection)

Note: The table reports power and size of the co-jumps test using the Bipower variation. Significance level: $\alpha = 0.05$.

Table 3: Market capitalization and descriptive statistics for DJI30.

ID	Index selection		Descriptive statistics of 5-minute log-returns [%]				
	Market Cap ($bn)	No.	σ	S	K	Min	Max
XOM	360,98		0.129	-0.089	10.970	-1.627	1.586
MSFT	266,46		0.147	-0.069	12.601	-2.177	2.190
WMT	211,16	High-Cap	0.102	0.101	12.735	-1.518	1.222
PG	183,81		0.096	-0.038	15.844	-1.529	1.630
JNJ	175,23		0.096	-0.208	14.461	-1.368	1.238
BA	39,03		0.160	-0.060	9.647	-1.737	1.658
CAT	37,16		0.191	-0.148	9.925	-2.282	1.949
DD	31,72	Low-Cap	0.162	0.011	9.972	-2.035	1.924
TRV	28,74		0.129	0.077	12.942	-1.541	1.496
AA	12,47		0.223	-0.121	9.384	-2.722	2.178

Note: The table contains market capitalization in $bn as the markets closed on 31st December 2009 as retrieved from Bloomberg and standard deviation (σ), skewness (S), kurtosis (K), and minimum (Min) and maximum (Max) log-return of the five most capitalized (XOM=Exxon Mobil Corp, MSFT=Microsoft Corp, WMT=Wal-Mart Stores Inc., PG=Procter & Gamble Co., JNJ=Johnson & Johnson) and the five least capitalized (BA=Boeing Co., CAT=Caterpillar Inc., DD=E.I. DuPont de Nemours & Co., TRV=Travelers Cos. Inc., AA=Alcoa Corp.) members of the DJIA 30.

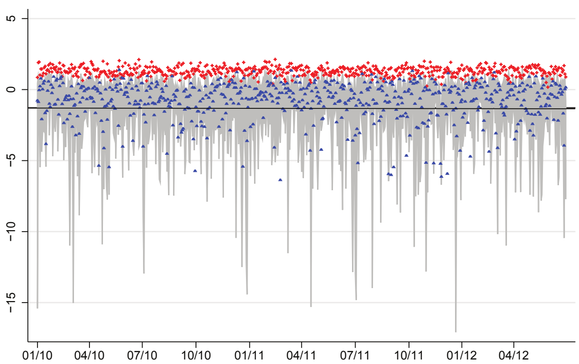
Table 4: Number of co-jumps vs. the individual assets.

	Portfolio	1	2	3	4	5	6
(A)	$\hat{G_\Omega}^{(1/N)} < \min \hat{G_\Omega}^{(i)}$	60	51	47	55	60	70
(B)	$\min \hat{G_\Omega}^{(i)} \leq \hat{G}_\Omega^{(1/N)} \leq \max \hat{G_\Omega}^{(i)}$	494	492	466	484	469	469
(C)	$\hat{G_\Omega}^{(1/N)} > \max \hat{G}_\Omega^{(i)}$	67	78	108	82	92	82

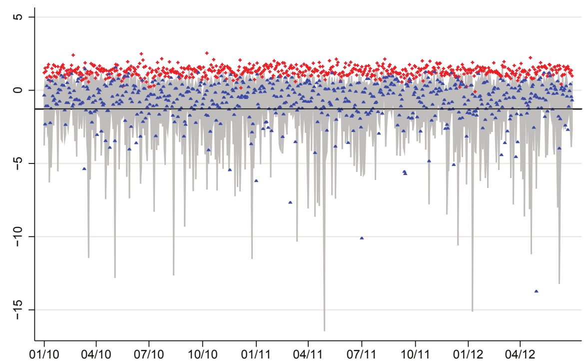
Note: The table evaluates the frequency of: (A) the equally weighted index amplifies price jumps, the $\hat{G_\Omega}^{(1/N)}$-statistic for the equally weighted portfolio is smaller than any individual asset; (B) the price jumps for the equally weighted index are comparable with price jumps at individual assets, the $\hat{G}_\Omega^{(1/N)}$-statistic is in the range implied by the individual assets; (C) the equally weighted index suppresses price jumps, the $\hat{G_\Omega}^{(1/N)}$-statistic is higher than any individual assets.

Figure 1: Co-jumps properties.

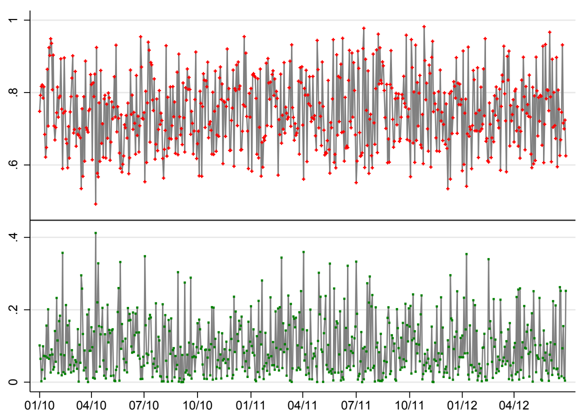
(a) Co-jumps $\hat{G_{\Omega}}$-statistic: High-Cap Index.



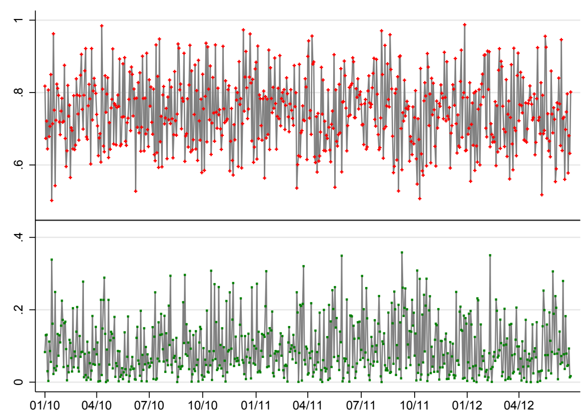
(b) Co-jumps $\hat{G_{\Omega}}$-statistic: Low-Cap Index.



(c) Co-jump vector magnitudes: High-Cap Index.



(d) Co-jump vector magnitudes: Low-Cap Index.



Note: Panels (a) and (b) depict the $\hat{G_{\Omega}}$-statistic for the co-jump vector (red dots), for the equally weighted index (blue dots), and the gray shaded area captures the region in which lies the $\hat{G_{\Omega}}$-statistic for each individual asset in the index. The black long-dash line denotes the $\alpha = 0.05$ critical value to test for the presence of price jumps, $\sqrt{\vartheta}\Phi^{-1}(\alpha) \cong -1.284$. Panels (c) and (d) depict the minimum (green) and maximum (red) of the co-jump vectors. The solid black line corresponds to the value of the equally weighted index. The vectors are normalized as that $\sum_{i=1}^{5} \Omega^{(i)2} = 1$.