Testing Jumps via False Discovery Rate Control

Yu-Min Yen*

Institute of Economics, Academia Sinica, Taipei, Taiwan

Abstract

Many recently developed nonparametric jump tests can be viewed as multiple hypothesis testing problems. For such multiple hypothesis tests, it is well known that controlling type I error often makes a large proportion of erroneous rejections, and such situation becomes even worse when the jump occurrence is a rare event. To obtain more reliable results, we aim to control the false discovery rate (FDR), an efficient compound error measure for erroneous rejections in multiple testing problems. We perform the test via the Barndorff-Nielsen and Shephard (BNS) test statistic, and control the FDR with the Benjamini and Hochberg (BH) procedure. We provide asymptotic results for the FDR control. From simulations, we examine relevant theoretical results and demonstrate the advantages of controlling the FDR. The hybrid approach is then applied to empirical analysis on two benchmark stock indices with high frequency data.

Citation: Yen Y-M (2013) Testing Jumps via False Discovery Rate Control. PLoS ONE 8(4): e58365. doi:10.1371/journal.pone.0058365

Editor: Enrico Scalas, Universita' del Piemonte Orientale, Italy

Received: August 28, 2012; **Accepted:** February 3, 2013; **Published:** April 3, 2013

Copyright: © 2013 Yu-Min Yen. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The author has no support or funding to report.

Competing Interests: The author has declared that no competing interests exist.

* E-mail: YMYEN@econ.sinica.edu.tw

Introduction

Recently many testing procedures have been proposed for detecting asset price jumps [1–8]. These testing procedures use high frequency data to calculate test statistics for a certain period and then use these test statistics to test whether jumps occur in that period. Formally, the null hypothesis for such test at each period $i, i = 1, \dots, m, i = 1, \dots, m$, can be stated as

$$H_i^0 : \text{No jump occurs in period } i. \quad (1)$$

In addition to know whether the inherent price process has a jump component, the “one test statistic for one period” approach for testing (1) also allows us to extract information about when and how frequently jumps occur in the whole sampling period. Such information is even more important for research on event study, derivative pricing and portfolio management.

If the number of periods m is greater than one, the jump test can be naturally viewed as a multiple hypothesis testing problem. Previous research used different test statistics, but often followed a similar decision procedure: Rejecting the null hypothesis if the corresponding p -value is less than the controlled type I error α . Nevertheless, controlling type I error often makes a large proportion of erroneous rejections. Such situation becomes even worse when the jump occurrence is a rare event.

To avoid the problem described above, one may look for a more sensible compound error rate measure. In this paper we focus on false discovery rate (FDR). For testing hypothesis (1), we use a nonparametric jump test procedure proposed by [3,4]. After obtaining the p -value for each single hypothesis test, we use the procedure proposed by [9] to control the FDR when simultaneously carrying out these hypothesis tests.

Several literatures on jump tests also tried to deal with the multiplicity issue. For example, Lee and Mykland [8] set the significance level based on the distribution of the extreme value of

the test statistic under the null. This ensures that the probability of global misclassification on the jumps can achieve zero under some regularity conditions. Bajgrowicz and Scaillet [10] proposed a statistical method which is based on setting an appropriate threshold for the test statistic to eliminate the false detections of jumps. They then applied their method on analyzing relationships between jumps in U.S. stock market and announcements of different kinds of economic news. As for applying the FDR control to jump component detections, it also has been adopted by [5], in which an improved version of the jump test statistic proposed by [1] was used. The main difference between [5] and our paper is that we give theoretical justifications on performance of the jump test statistic in a multiple hypothesis testing context. We also conduct an intensive simulation study to support our theoretical results.

The rest of the paper is organized as follows. In Section Methods, we first briefly describe the Barndorff-Nielsen and Shephard (BNS) nonparametric test and the Benjamini-Hochberg (BH) procedure. We then discuss some asymptotic results for the FDR control. We focus on the case when p -values are calculated based on asymptotic distributions of the test statistics. We show that with some appropriate conditions, the FDR can be asymptotically controlled by the BH procedure when the p -values are obtained via the asymptotic distributions. In addition, magnitude of approximation error of the asymptotic FDR control is bounded by a non-decreasing function of expected number of the true null hypotheses. This property indicates that the more false null hypotheses we have, the better performance the asymptotic FDR control will achieve. In Section Results, we conduct a simulation study to show that performance of the BNS-BH hybrid procedure is positively related to the number of false hypotheses and sampling frequency of the data, and is stable when the number of hypotheses and the required FDR level change. We finally apply the proposed procedure on analyzing jumps in S&P500 index and Dow Jones industrial average index.

Table 1. Number of hypotheses and rejections when a multiple testing is performed.

	Test statistic is not significant	Test statistic is significant	Total number
True null hypotheses	U	V	$\tilde{m}_0$
Non-true null hypotheses	T	S	$\tilde{m}_1$
Total number	$m-R$	R	m

doi:10.1371/journal.pone.0058365.t001

Methods

The BNS nonparametric jump test

Barndorff-Nielsen and Shephard [3,4] proposed a nonparametric test statistic (henceforth the BNS test statistic) which utilizes realized variance and bi-power variation to test jump components of price processes which have continuous sample paths. To begin with, we briefly introduce some important theoretical results of the jump test procedure in financial econometrics. We say a random variable $X(i)$ belongs to the Brownian semimartingale plus jump class if

$$X(i) = \int_0^i \mu(t)dt + \int_0^i \sigma(t)dW(t) + \sum_{j=1}^{N(i)} D(j), \quad (2)$$

where $\mu(t)$ and $\sigma(t)$ are assumed to be càdlàg, $W(t)$ is a standard Brownian Motion, $D(j)$ is the quantity of the j th jump within $(0, i]$, and $N(i)$ is total number of the jumps occurring within $(0, i]$. Here, we assume the number of jumps occurring within the interval $(i-1, i]$, $N(i) - N(i-1)$ is finite for all i . If (2) is without the jump term $\sum_{j=1}^{N(i)} D(j)$, we say $X(i)$ belongs to the Brownian semimartingale without jump class. Some statistical assumptions can be made on $D(j)$ and $N(i)$ for purpose of simplifying the analysis. For example, in empirical finance literatures, magnitude of jump $D(j)$ is often assumed to follow a normal distribution, and the number of jumps within the interval $(i-1, i]$, $N(i) - N(i-1)$ is often assumed to follow a counting process with (finite) intensity parameter $\lambda(i)$ which may be time varying.

The realized variance and the realized bi-power variation in period i are defined as

$$RV_i = \sum_{h=1}^M r_{i,h}^2,$$

$$BV_i = \frac{\pi}{2} \left(\frac{M}{M-1} \right) \sum_{h=1}^{M-1} |r_{i,h}| |r_{i,h+1}|,$$

respectively, where

$$r_{i,h} := \log P\left(i-1 + \frac{h}{M}\right) - \log P\left(i-1 + \frac{h-1}{M}\right)$$

is the intra-period log return in the h th sub-interval of period

$i, h = 1, \dots, M$, and $P(i-1+h/M)$ is the asset price at time point $i-1+h/M$. Assume that for $t \in (i-1, i]$, $\log P(t)$ belongs to the Brownian semimartingale plus jump class. Then it can be shown that under some regularity conditions,

$$RV_i \xrightarrow{P} \int_{i-1}^i \sigma^2(t)dt + \sum_{j=N(i-1)+1}^{N(i)} D^2(j), \quad (3)$$

$$BV_i \xrightarrow{P} \int_{i-1}^i \sigma^2(t)dt \quad (4)$$

as $M \rightarrow \infty$. The term

$$\int_{i-1}^i \sigma^2(t)dt + \sum_{j=N(i-1)+1}^{N(i)} D^2(j)$$

in (3) is called the quadratic variation for the cumulative (log) return process $\log P(i) - \log P(i-1)$, and it is a sum of contributions due to the continuous log price process ($\int_{i-1}^i \sigma^2(t)dt$) and the jump process ($\sum_{j=N(i-1)+1}^{N(i)} D^2(j)$). The result of (3) follows from the theory of quadratic variation (e.g., [11]) and the result of (4) follows from the theory of power variation process which is a generalized version of the theory of the

quadratic variation process [3]. Here $BV_i \xrightarrow{P} \int_{i-1}^i \sigma^2(t)dt$ can hold without any further assumptions on the jump process, the joint distribution of the jump process and $\sigma(t)$. Finally, if $\log P(t)$ belongs to the Brownian semimartingale without jump class, it is easy to see that both RV_i and BV_i will converge in probability to $\int_{i-1}^i \sigma^2(t)dt$ as $M \rightarrow \infty$.

Barndorff-Nielsen and Shephard [3,4] showed that

$$JV_i = RV_i - BV_i$$

can consistently estimate the quantity $\sum_{j=N(i-1)+1}^{N(i)} D^2(j)$. In practice, to guarantee nonnegativity of the estimation, some truncation rules can be applied on JV_i , for example using $\max(JV_i, 0)$ or a shrinkage type estimator like (15) in our empirical analysis. To construct a test statistic to test whether the jump term presents, now suppose for $t \in (i-1, i]$, $\log P(t)$ belongs to the Brownian semimartingale without jump class, and the following conditions hold,

1. The process of $\sigma(t)$ is pathwise bounded away from 0.
2. The joint process of $\mu(t)$ and $\sigma(t)$ is independent of the Brownian motion term $W(t)$ of the log price process,

then conditioning on $\mu(t), \sigma(t)$, the quadratic variation and realized bi-power variation process, Barndorff-Nielsen and Shephard [3,4] showed that joint distribution of RV_i and BV_i will converge asymptotically to a bivariate normal distribution. Then under the null hypothesis when no jumps are present on period i , it can be shown that

Figure 1. Realized FDR, $\hat{S}/m_1$, significance level obtained from the BH procedure and number of rejections. In the graphs, each point is an average value from 1000 simulations.
doi:10.1371/journal.pone.0058365.g001

$$\frac{\sqrt{M}(RV_i - BV_i)}{\sqrt{A \int_{i-1}^i \sigma^4(t) dt}} \xrightarrow{L} \mathcal{N}(0,1), \quad (5)$$

where $A = (\pi/2)^2 + \pi - 5$. The term $\int_{i-1}^i \sigma^4(t) dt$ in the denominator of (5) is called the integrated quarticity, and to consistently estimate it, we can use the realized tri-power quarticity,

$$TP_i = \mu_4^{-3} \left(\frac{M^2}{M-2} \right) \sum_{h=1}^{M-2} (|r_{i,h}| |r_{i,h+1}| |r_{i,h+2}|)^{\frac{4}{3}},$$

where $\mu_a = \mathbb{E}(|Z|^a)$ and $Z \sim \mathcal{N}(0,1)$.

In the following simulation study and empirical applications, instead of using the test statistic shown in (5), we will use three improved test statistics to obtain better performances. The first one is proposed by [3,4], which uses the log transformation and is defined as

$$Z_{\log,i} = \frac{\sqrt{M}(\log(RV_i) - \log(BV_i))}{\sqrt{A \max(1, B)}},$$

where

$$B = \frac{\int_{i-1}^i \sigma^4(t) dt}{\left(\int_{i-1}^i \sigma^2(t) dt \right)^2}$$

The second one is the Box-Cox transformed test statistic with parameter $\rho = -1.5$, which is defined as

$$Z_{-1.5,i} = \frac{\sqrt{M} \left(\int_{i-1}^i \sigma^2(t) dt \right)^3 (BV_i^{-1.5} - RV_i^{-1.5})}{1.5 \sqrt{A \max(1, B)}}.$$

Figure 2. Realized FDR and $\hat{S}/m_i$ of the hybrid method and the conventional procedure. In the graphs, each point is an average value from 1000 simulations.
doi:10.1371/journal.pone.0058365.g002

Here the Box-Cox transformation for a positive number x is defined as

$$g_{\rho}(x) = \begin{cases} \frac{x^{\rho}-1}{\rho} & \text{if } \rho \neq 0 \\ \log(x) & \text{if } \rho = 0. \end{cases}$$

The third one is the ratio type test statistic [12]:

$$Z_{ratio,i} = \frac{\sqrt{M} \frac{RV_i - BV_i}{RV_i}}{\sqrt{A \max(1, B)}}.$$

Under the null hypothesis that there is no jump occurring in period i , the test statistics $Z_{-1.5,i}$, $Z_{log,i}$ and $Z_{ratio,i}$ will have a standard normal distribution as their limiting joint distribution. When jumps occur in period i , the test statistics will approach to infinity as $M \rightarrow \infty$. For more discussions on theoretical properties of the test statistics under the alternative (when jump presents), please see [13].

The FDR and BH procedure

Let H_i^0 and p_i denote the i th null hypothesis and the corresponding p -value, $i = 1, 2, \dots, m$. Among the m hypotheses, suppose there are $\tilde{m}_0$ true hypotheses and $\tilde{m}_1 = m - \tilde{m}_0$ false hypotheses. Note that $\tilde{m}_0$ and $\tilde{m}_1$ are generally unknown to

researchers, so they are assumed to be random variables. On contrary, the total number of hypotheses m is generally known in advance and so is assumed to be nonrandom. Table 1 shows different situations when a multiple testing is performed. The numbers of hypotheses we reject and do not reject are denoted by R and $m - R$. The notations U , T , V and S denote the numbers of hypotheses we correctly accept, falsely accept, falsely reject and correctly reject, respectively. The false discovery rate (FDR) is then defined as the expectation of the false discovery proportion (FDP), i.e.

$$\text{FDR} = \mathbb{E}(\text{FDP}),$$

where

$$\text{FDP} = \begin{cases} 0 & \text{if } R = 0 \\ \frac{V}{R} & \text{if } R \neq 0. \end{cases}$$

In testing jumps, controlling the FDR has several advantages over controlling other compound error rates. First, if the price process really does not have a jump component, i.e., all the null hypotheses are true, then controlling the FDR will be equivalent to controlling $\Pr(V \geq 1)$, the familywise error rate (FWER). Second, if the intensity of the jump process $\lambda \neq 0$, as time goes on (m increases), the proportion of false hypotheses among all hypotheses

Figure 3. Realized FDR and $\hat{S}/m_l$ of the hybrid method with lower frequency data. In the graphs, each point is an average value from 1000 simulations.

doi:10.1371/journal.pone.0058365.g003

will be a nonzero constant with a high probability. Although such proportion may not be large, one may still expect the more (fewer) rejections one has, the more (fewer) erroneous rejections are allowed to occur; or the number of rejections should be proportional to m . In this situation, controlling compound error rates associated with proportion of erroneous rejections, like the FDR, makes sense. In addition, rejection criterion of some compound error rates such as the FWER, are sometimes too stringent to get rejections when the number of hypotheses becomes large. The criterion of the FDR is less conservative in this aspect. Finally, controlling the FDR currently seems to be more acceptable than controlling other compound error rates in many different research fields [14].

Let $p_{(1)} \leq \dots \leq p_{(m)}$ be the ordered p_i 's and $H_{(1)}^0, \dots, H_{(m)}^0$ be the corresponding null hypotheses. Benjamini and Hochberg [9] proposed a stepwise procedure to control the FDR at the required level γ . The BH procedure can be simplified as the following two-step decision rule:

1. Obtain $i^* = \max_{i=1,2,\dots,m} \left\{ i : p_{(i)} \leq \frac{i}{m} \gamma \right\}$.
2. Reject $H_{(i)}^0$ for all $i \leq i^*$.

Some controlling procedures for the FDR need a resampling scheme to construct the rejection region, which relies on intensive computations. The BH procedure, however, requires far less computational sources than those computational intensive methods. As shown above, the only computational burden of the BH

procedure is to rank the p -values. Such advantage becomes even more obvious when the number of hypotheses becomes very large.

It can be shown that there is a relationship between the type I error α and the FDR. That is, if we reject H_i^0 as $p_i \leq \alpha$, $i = 1, \dots, m$, it is possible to know what level of the FDR is controlled for the m hypotheses multiple testing. For example, if the hypotheses are identical and the test statistics are all independent, given the type I error α , the following estimator [25]

$$\widehat{\text{FDR}}_{\kappa}(\alpha) = \frac{\#\{p_i > \kappa\} \alpha}{(1 - \kappa) \max(\#\{p_i \leq \alpha\}, 1)},$$

can be used to estimate the corresponding FDR. Here κ is a turning parameter.

How the BH procedure performs relies on dependence structure of the test statistics. Benjamini and Yekutieli [16] showed that the BH procedure can still control the FDR when the test statistics are not independent, but the positive regression dependency (PRDS) for each test statistic under the true null hypotheses is satisfied. In addition, simulation studies in [14] showed that even if the PRDS condition is violated (e.g., there exist negative common correlations between the test statistics or the covariance matrix has an arbitrary structure), the BH procedure can still provide a satisfactory control of the FDR. Finally, if the test statistics have an arbitrary dependence structure, Benjamini and Yekutieli [16] showed that the BH procedure still guarantees that

Figure 4. Realized FDR and $\hat{S}/m_l$ of the hybrid method when the number of hypotheses varies. Here $m = 50, 100, 200, 500, 800, 1000, 1200, 1500$ and 2000 . In the graphs, each point is an average value from 1000 simulations. doi:10.1371/journal.pone.0058365.g004

$$\text{FDR} \leq \gamma \sum_{k=1}^m \frac{1}{k} \approx \gamma \left(\log(m) + \frac{1}{2} \right).$$

A more detailed discussion on the theoretical properties of the FDR and the BH procedure is provided in next section.

Results

Asymptotically results

Let $\{\mathbf{X}_i = (X_{i,1} \dots X_{i,M}) : i \in \mathbb{N}\}$ be a vector of samples defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let $\{\mathcal{F}_{M,i} : i \in \mathbb{N}\}$ be the smallest sub- σ field of $\mathcal{F}$ such that for $j = 1, \dots, M$, $X_{i,j}$ is $\mathcal{F}_{M,i}$ measurable. Let $\mathcal{F}_i = \lim_{M \rightarrow \infty} \mathcal{F}_{M,i}$. A test statistic for testing marginal hypothesis i with sample size M is a function $\hat{T}_{M,i} : \mathbf{X}_i \rightarrow \mathbb{R}$, and

Figure 5. Realized FDR of the hybrid method under different required γ . We fix $m = 1000$ in the simulation. In the graphs, each point is an average value from 1000 simulations. doi:10.1371/journal.pone.0058365.g005

Table 2. The table shows summary statistics of the price variations, different types of $i_{M,i}$, their corresponding $i_{M,i}$ and mutual correlations of these quantities of SPC500 and DJIA.

	SPC500			DJIA			Corr.
	Mean	Std.	LB.10	Mean	Std.	LB.10	
RV_i	0.4842	0.4817	0.0000	0.4793	0.4505	0.0000	0.9580
BV_i	0.4314	0.4348	0.0000	0.4298	0.4013	0.0000	0.9713
JV_i	0.0368	0.0830	0.0000	0.0326	0.0800	0.0000	0.5311
$\rho = -1.5$							
$\hat{T}_{iM}$	0.7782	1.0842	0.5096	0.7034	1.1005	0.4902	0.6925
$\hat{P}_{iM}$	0.3024	0.2674	0.8873	0.3222	0.2755	0.6606	0.6767
Log Type							
$\hat{T}_{iM}$	0.9390	1.2945	0.3319	0.8568	1.3026	0.4456	0.6814
$\hat{P}_{iM}$	0.2916	0.2697	0.8765	0.3113	0.2777	0.6594	0.6741
Ratio Type							
$\hat{T}_{iM}$	0.8274	1.1427	0.4499	0.7504	1.1564	0.4772	0.6901
$\hat{P}_{iM}$	0.2987	0.2683	0.8837	0.3184	0.2764	0.6598	0.6760

The column LB.10 shows p -values of the Ljung-Box statistic based on autocorrelation coefficients with 10 lagged values. The quantities of price

$\hat{T}_{M,i}$ is $\mathcal{F}_{M,i}$ measurable. Let $\hat{\mathbf{T}}_M = (\hat{T}_{M,1}, \dots, \hat{T}_{M,m})$ denote the vector of the test statistics for testing $m(m \geq 1)$ hypotheses. Given each $\mathcal{F}_i$, suppose that there exists a vector of random variables $\mathbf{T} = (T_1, \dots, T_m)$ such that for $i = 1, \dots, m$, T_i is $\mathcal{F}_i$ measurable. Also assume $\hat{T}_{M,i} \xrightarrow{L} T_i$ for each i as $M \rightarrow \infty$. Let Ψ_i be the limiting distribution function of the test statistic under the null hypothesis i . For one-sided test, p -value of the i th one-sided hypothesis is defined by $p_i(x) = 1 - \Psi_i(x)$ for the i th two-sided hypothesis). Let $p_i = 1 - \Psi_i(T_i)$. A feasible estimated p -value for hypothesis i is then given by

$$\hat{P}_{M,i} = p_i(\hat{T}_{M,i}) = 1 - \Psi_i(\hat{T}_{M,i}).$$

In our case of testing jumps, since our null hypotheses are homogeneous, Ψ_i is the c.d.f. of $\mathcal{N}(0,1)$ for all i .

Let $I_0 = \{i : H_i^0 \text{ is true}\}$ and $I_1 = \{i : H_i^0 \text{ is false}\}$. Let $\Psi_{M,i}$ be the exact distribution of $\hat{T}_{M,i}$ under the null hypothesis i . The p -value under such distribution for hypothesis i is $p_{M,i} = 1 - \Psi_{M,i}(\hat{T}_{M,i})$. For $a \in (0,1)$, as $M \rightarrow \infty$, $\Pr(\hat{P}_{M,i} \leq a) \rightarrow \Pr(p_i \leq a)$, $\Pr(\hat{P}_{M,i} \leq a) \rightarrow \Pr(p_i \leq a)$ if $\hat{T}_{M,i} \xrightarrow{L} T_i$. If $T_1, \dots, T_m$ are continuous random variables, then

$$\Pr(p_i \leq a) = a,$$

for $i \in I_0$. If $\hat{T}_{M,1}, \dots, \hat{T}_{M,m}$ are also continuous random variables, then $\Pr(p_{M,i} \leq a) = a$ for $i \in I_0$.

Before we proceed to the main results, we need to introduce some definitions. Let $\mathcal{B}$ denote the Borel set. Let $(\mathbb{R}^m, \mathcal{B}^m) = \prod_{i=1}^m (\mathbb{R}, \mathcal{B})$ define the m -fold products of the real line $\mathbb{R}$ with the Borel sets $\mathcal{B}$. Let

$$\Sigma_m = \{\{i_1, \dots, i_m\} : i_k \in \{1, \dots, m\} \text{ for } k = 1, \dots, m, i_k \neq i_l \text{ for } k \neq l\}.$$

denote the symmetric set of permutations of integers $1, \dots, m$. Let Q_m be a probability measure on $(\mathbb{R}^m, \mathcal{B}^m)$ where $\mathbf{m} \in \Sigma_m$.

Definition 1. A collection of $\{Q_m\}_{\mathbf{m} \in \Sigma_m}$ is consistent if it satisfies

- Let $\mathbf{m} = \{i_1, \dots, i_m\}$, and $\mathbf{m}' = \{i_{k_1}, \dots, i_{k_m}\} \in \Sigma_m$ but $\mathbf{m}' \neq \mathbf{m}$. Then for each $B_i \in \mathcal{B}$, $i = 1, \dots, m$,

$$Q_m(B_1 \times \dots \times B_m) = Q_{\mathbf{m}'}(B_{k_1} \times \dots \times B_{k_m})$$

- For each $B_i \in \mathcal{B}$, $i = 1, \dots, m$,

$$\begin{aligned} Q_m(B_1 \times \dots \times B_{k-1} \times \mathbb{R} \times B_{k+1} \times \dots \times B_m) \\ = Q_{\mathbf{m}/i_k}(B_1 \times \dots \times B_{k-1} \times B_{k+1} \times \dots \times B_m). \end{aligned}$$

We say a set Θ is decreasing if $\mathbf{X} = (X_1, \dots, X_m) \in \Theta$ implies $\mathbf{Z} = (Z_1, \dots, Z_m) \in \Theta$ when $Z_i \leq X_i$ for any $i = 1, \dots, m$; and a set Λ is increasing if $\mathbf{Y} = (Y_1, \dots, Y_l) \in \Lambda$ implies $\mathbf{Z}' = (Z'_1, \dots, Z'_l) \in \Lambda$ when $Z'_i \geq Y_i$ for any $i = 1, \dots, l$. The concept of increasing and decreasing sets was used in [16] and [17] for introducing the concept of positive regression dependency on each one from a subset (PRDS).

Definition 2. Let $\mathbf{Y} = (Y_1, \dots, Y_l)$, $\mathbf{X} = (X_1, \dots, X_m)$, and I be a collection of index $i \in \{1, \dots, m\}$. For any decreasing set Θ and increasing set Λ , an l -dimensional random vector $\mathbf{Y}$ is said to be positive regression dependency on each one from a subset (PRDS) I of a m -dimensional random vector $\mathbf{X}$ is that $\Pr(\mathbf{Y} \in \Theta | X_i = x)$ is non-increasing in x or $\Pr(\mathbf{Y} \in \Lambda | X_i = x)$ is non-decreasing in x for any $i \in I$.

Let $\mathbf{p} = (p_1, \dots, p_m)$ and $\hat{\mathbf{P}}_M = (\hat{P}_{M,1}, \dots, \hat{P}_{M,m})$. Suppose we want to control FDR at the level γ with the BH procedure with $\mathbf{p}$. Let $\tilde{m}_0$ denote the number of true null hypotheses. In practice, $\tilde{m}_0$ is unknown in advance and so is assumed to be random here. Conditioning on $\tilde{m}_0 = m_0$ true null hypotheses (or equivalently $m_1 = m - m_0$ false hypotheses), the FDR is given by

$$\mathbb{E}\left(\frac{V}{R} | \tilde{m}_0 = m_0\right) = \sum_{s=0}^{m_1} \sum_{v=1}^{m_0} \frac{v}{v+s} \Pr(\mathbf{p} \in \mathbf{D}_{m_0}^{v,s}). \quad (6)$$

Here $\mathbf{D}_{m_0}^{v,s}$ is a well constructed union of m -dimensional cubes such that $\{\mathbf{p} \in \mathbf{D}_{m_0}^{v,s}\}$ is the event that v true and s false null hypotheses are rejected when the BH procedure is implemented with $\mathbf{p}$. Benjamini and Yekutieli [16] and Sarkar [17] showed that if the joint distribution of p_i is PRDS on $I_0 = \{i : H_i^0 \text{ is true}\}$, then $\mathbb{E}(V/R | \tilde{m}_0 = m_0) \leq m_0 \gamma / m$. Since $\tilde{m}_0$ is bounded by m , we can get

$$\mathbb{E}\left(\frac{V}{R}\right) \leq \frac{\gamma \mathbb{E}(\tilde{m}_0)}{m} \leq \gamma. \quad (7)$$

Figure 6. Time series plots for identified jump variation on common jump days with the three different jump test statistics. Left: FDR controlled by using the pool method. Right: FDR controlled by using the separate method. The quantities shown here are all scaled by 10000. doi:10.1371/journal.pone.0058365.g006

If $\hat{\mathbf{p}}_M$ is used, the analogue of (6) is then given by

$$\mathbb{E}_{\hat{\mathbf{p}}_M} \left(\frac{V}{R} | \tilde{m}_0 = m_0 \right) = \sum_{s=0}^{m_1} \sum_{v=1}^{m_0} \frac{v}{v+s} \Pr(\hat{\mathbf{p}}_M \in \mathbf{D}_{m_0}^{v,s}), \quad (8)$$

where $\{\hat{\mathbf{p}}_M \in \mathbf{D}_{m_0}^{v,s}\}$ is the event that v true and s false hypotheses are rejected when the BH procedure is implemented with $\hat{\mathbf{p}}_M$. One

should note that the expectations in (6) and (8) are calculated under different probability distributions. In (6), the expectation is obtained under the joint distribution of $\mathbf{p} = (p_1, \dots, p_m)$, while in (8), the expectation is obtained under the joint distribution of $\hat{\mathbf{p}}_M = (\hat{p}_{M,1}, \dots, \hat{p}_{M,m})$.

Ideally, if we know $\Psi_{M,i}$, and the joint distribution of $p_{M,i}$ is PRDS on I_0 , we can implement the BH procedure directly with $p_{M,1}, \dots, p_{M,m}$. However, such information is often unknown, and

Table 3. The table shows summary statistics of significant daily discontinuous quadratic variation $JV_{i,\gamma}$ (sum of squared intradaily jumps) of SPC500 and DJIA.

SPC500, $m=1247$						
	$r=0.01$			$\gamma=0.05$		
	$\rho=-1.5$	Log Type	Ratio Type	$\rho=-1.5$	Log Type	Ratio Type
$\frac{\hat{p}_M}{m}^\gamma$	2.41e-05	0.0004	8.82e-05	0.0008	0.0058	0.0023
No. of days	3	55	11	19	144	58
Mean	0.1464	0.1816	0.2232	0.2163	0.1376	0.1744
Std.	0.0727	0.1885	0.2863	0.2311	0.1492	0.1861
DJIA, $m=1247$						
	$r=0.01$			$r=0.05$		
	$\rho=-1.5$	Log Type	Ratio Type	$\rho=-1.5$	Log Type	Ratio Type
$\frac{\hat{p}_M}{m}^\gamma$	4.01e-05	0.0004	0.0002	0.0012	0.0044	0.0020
No. of days	5	52	22	30	107	51
Mean	0.1458	0.1649	0.1642	0.1662	0.1324	0.1634
Std.	0.0501	0.1198	0.0896	0.1088	0.1053	0.1205

FDR is controlled at level 0.01 and 0.05. The quantities of price variations shown are all scaled by 10000.
doi:10.1371/journal.pone.0058365.t003

instead only $\hat{\mathbf{p}}_M$ is feasible. In the following, we show that under appropriate conditions, FDR can be asymptotically controlled with $\hat{\mathbf{p}}_M$ under a desired level. Our strategy is to show that under appropriate conditions, $\Pr(\hat{\mathbf{p}}_M \in \mathbf{D}_{m_0}^{v,s}) \rightarrow \Pr(\mathbf{p} \in \mathbf{D}_{m_0}^{v,s})$ as $M \rightarrow \infty$ and then to prove

$$\mathbb{E}_{\hat{\mathbf{p}}_M} \left(\frac{V}{R} \mid \tilde{m}_0 = m_0 \right) \rightarrow \mathbb{E} \left(\frac{V}{R} \mid \tilde{m}_0 = m_0 \right).$$

as $M \rightarrow \infty$. Therefore implementing the BH procedure with $\hat{\mathbf{p}}_M$ is asymptotically equivalent to implementing the procedure with $\mathbf{p}$.

The main results are the following two theorems, and their proofs are given in the supplementary materials.

Theorem 1. Suppose we have m hypotheses to be tested simultaneously. If the following conditions hold,

1. The joint distribution of p_i and the joint distribution of $\hat{\mathbf{p}}_{M,i}$ satisfy the consistency for multivariate distribution.
2. The joint distribution of p_i is PRDS on p_i for $i \in I_0$ and all $m \geq 1$.
3. $\Pr(p_i \leq a) \leq a$ for $i \in I_0$ and $a \in (0,1)$.
4. $\sup_{1 \leq k \leq m} \sup_{i \in I_0} |\Pr(p_i \leq q_k) - \Pr(\hat{\mathbf{p}}_{M,i} \leq q_k)| = O(1/M^\delta)$, where $q_k = k\gamma/m$ and $\delta > 0$.
5. Given m_0 true null hypotheses, let $\mathbf{p}^{(-i)}$ and $\hat{\mathbf{p}}_M^{(-i)}$ denote the random $m-1$ dimensional vectors obtained by eliminating p_i and $\hat{\mathbf{p}}_{M,i}$ from the m -dimensional random vectors $\mathbf{p}$ and $\hat{\mathbf{p}}_M$ respectively. Let $p_{(1)}^{(-i)} \leq \dots \leq p_{(m-1)}^{(-i)}$ and $\hat{p}_{M,(1)}^{(-i)} \leq \dots \leq \hat{p}_{M,(m-1)}^{(-i)}$ denote the ordered components of $\mathbf{p}^{(-i)}$ and $\hat{\mathbf{p}}_M^{(-i)}$ respectively. For every $m \geq 1$,

$$\sup_{1 \leq k \leq m} \sup_{i \in I_0} m \left(\Pr(\hat{\mathbf{p}}_{M,i} \leq q_k, \hat{\mathbf{p}}_{M,(k)}^{(-i)} > q_{k+1}, \dots, \hat{\mathbf{p}}_{M,(m-1)}^{(-i)} > q_m) - \Pr(p_i \leq q_k, p_{(k)}^{(-i)} > q_{k+1}, \dots, p_{(m-1)}^{(-i)} > q_m) \right) = o(1), \quad (9)$$

then the BH procedure implemented with the estimated p -values $\hat{\mathbf{p}}_M$ asymptotically control FDR at the required level γ in the sense that.

$$\lim_{M \rightarrow \infty} \mathbb{E}_{\hat{\mathbf{p}}_M} \left(\frac{V}{R} \right) = \mathbb{E} \left(\frac{V}{R} \right) \leq \gamma.$$

Theorem 2. Suppose we have m hypotheses to be tested simultaneously. If conditions 1 and 4 in Theorem 1 and the following conditions hold,

1. $\Pr(\hat{\mathbf{p}}_{M,i} \leq a) \leq a$ for $i \in I_0$ and $a \in (0,1)$
2. $T_1, \dots, T_m$ are mutually independent and continuous random variables.
3. $\hat{T}_{M,1}, \dots, \hat{T}_{M,m}$ are mutually independent, then the BH procedure implemented with the estimated p -values $\hat{\mathbf{p}}_M$ asymptotically control FDR at the required level γ in the sense that

$$\lim_{M \rightarrow \infty} \mathbb{E}_{\hat{\mathbf{p}}_M} \left(\frac{V}{R} \right) = \mathbb{E} \left(\frac{V}{R} \right) \leq \gamma.$$

Discussions on the asymptotic results

The two theorems say that under some regularity conditions, we can asymptotically control FDR. A key condition making the two theorems different is the requirement on the dependence structure of elements in vector $\mathbf{p}$ and $\hat{\mathbf{p}}_M$. If the dependence structure of p_i satisfies PRDS on I_0 , it ensures that $\mathbb{E}(V/R) \leq \gamma$. Here we only require the PRDS should hold on I_0 , and the dependence structure of p_i on I_1 can be arbitrary. Marginal distributions of p_i and $\hat{\mathbf{p}}_{M,i}$ converging with the rate $O(1/M^\delta)$ simultaneously for all i is also needed for the consistent control. In addition, we also require the convergence of the joint distribution of the ordered

Table 4. The table shows summary statistics of significant daily discontinuous quadratic variation $JV_{i,\gamma}$ (sum of squared intradaily jumps) of SPC500 and DJIA on the common jump days by adopting separate and pool methods.

	Separate					
	$r=0.01$			$r=0.05$		
	$\rho=-1.5$	Log Type	Ratio Type	$\rho=-1.5$	Log Type	Ratio Type
No. of common days	1	22	6	8	55	22
Mean, SPC500	0.2304	0.2289	0.3115	0.2769	0.1690	0.2289
Std., SPC500	N.A.	0.2234	0.3759	0.3244	0.1836	0.2234
Mean, DJIA	0.2101	0.2067	0.2060	0.2021	0.1582	0.2067
Std., DJIA	N.A.	0.1271	0.1459	0.1249	0.1286	0.1271
Corr.	N.A.	0.8503	0.9694	0.9556	0.8889	0.8503
Pool, $m=2494$						
	$r=0.01$			$r=0.05$		
	$\rho=-1.5$	Log Type	Ratio Type	$\rho=-1.5$	Log Type	Ratio Type
	$\hat{p}_m^{\gamma}$			$\hat{p}_m^{\gamma}$		
No. of common days	3.21e-05	0.0004	0.0001	0.0010	0.0051	0.0021
No. of days, SPC500	1	22	7	9	58	22
Mean, SPC500	0.2304	0.2289	0.2877	0.2720	0.1647	0.2289
Std., SPC500	N.A.	0.2234	0.3489	0.3038	0.1797	0.2234
No. of days, DJIA	3	55	12	21	140	55
Mean, DJIA	0.2101	0.2067	0.1989	0.2102	0.1535	0.2067
Std., DJIA	N.A.	0.1271	0.1345	0.1193	0.1270	0.1271
Corr.	N.A.	0.8503	0.9694	0.9249	0.8897	0.8503

The term common jump days used here only means that the two indices both have jumps on these days. The mean and standard deviation of $JV_{i,\gamma}$ are calculated conditional on $\hat{P}_{i,M} \leq \hat{P}_m^{\gamma}/m$. The quantities of price variations shown are all scaled by 10000.
doi:10.1371/journal.pone.0058365.t004

p -values. But as stated in Theorem 2, such condition can be ignored if other conditions hold.

The approximation error $\epsilon = |\mathbb{E}_{\hat{P}_M} (V/R) - \mathbb{E}(V/R)|$ essentially vanishes to zero when $M \rightarrow \infty$. Magnitude of ϵ , as shown in our proof, is bounded by a non-decreasing function of $\mathbb{E}(\hat{m}_0)$. This property indicates that the more false null, the better the convergence.

We then have a look of condition 1 in Theorem 1. This is a sufficient condition to ensure that $\Pr(\mathbf{p} \in \mathbf{D}^{v,s})$ (and $\Pr(\hat{\mathbf{p}}_M \in \mathbf{D}^{v,s})$) exists as $m \rightarrow \infty$. It is due to Kolmogorov's extension theorem [19, p.50]: An extension of any consistent family of probability measures on $(\mathbb{R}^m, \mathcal{B}^m)$ to a probability measures on $(\mathbb{R}^\infty, \mathcal{B}^\infty) = \prod_{i=1}^\infty (\mathbb{R}, \mathcal{B})$ necessarily exists and is unique. Conversely, if we have a probability measure on $(\mathbb{R}^\infty, \mathcal{B}^\infty)$, we can induce a family of finite-dimensional distributions on $(\mathbb{R}^m, \mathcal{B}^m)$, and these induced finite-dimensional distributions all satisfy consistency for multivariate distribution.

Condition 2 in Theorem 1 requires that the joint distribution of the p -values should satisfy PRDS on the subset I_0 . It is a sufficient condition for $\mathbb{E}(V/R) \leq \gamma$ when we implement the BH procedure with $\mathbf{p}$. Since our purpose is to control FDR with $\hat{\mathbf{p}}_M$, if we can guarantee that $\lim_{M \rightarrow \infty} \mathbb{E}_{\hat{\mathbf{p}}_M} (V/R) = \mathbb{E}(V/R)$, only the distribution of $\mathbf{p}$ satisfying the condition is needed. For practically using the BH procedure, Benjamini and Yekutieli [16] listed many situations when the condition holds. For example, if $\mathbf{T} \sim \mathcal{N}(\mathbf{m}, \mathbf{S})$, where $\mathbf{m} = (\mu_1, \dots, \mu_m)$ and $\mathbf{S}$ is a $m \times m$ covariance matrix with element σ_{ij} . Suppose for each $i \in I_0$, and each $j \neq i$, $\sigma_{ij} \geq 0$, then the distribution of $\mathbf{T}$ is PRDS on I_0 , regardless what the covariance

structure of $i \in I_1$ is. Mutual independence of $T_1, \dots, T_m$ can be easily seen as a special case of PRDS on I_0 . As for the nonparametric jump test in this paper, since the limiting distribution of the test statistics is a multivariate normal with $\sigma_{ij} = 0$ for each $i \in I_0$ and each $j \neq i$, it implies PRDS on I_0 .

The condition that $\Pr(p_i \leq a) \leq a$ for $a \in (0, 1)$ is called the distribution of p_i is stochastically dominated by the Uniform(0,1). If $\lim_{M \rightarrow \infty} \Pr(\hat{p}_{M,i} \leq a) \leq a$, it is called that the distribution of $\hat{p}_{M,i}$ is stochastically dominated by the Uniform(0,1) distribution asymptotically. In order to control FDR with the BH method asymptotically, we at least need that $\Pr(p_i \leq a) \leq a$ for $a \in (0, 1)$ and $i \in I_0$. The condition is more liberal than that p_i has the exact Uniform(0,1) distribution for $i \in I_0$, and applies to the case when the test statistics are discrete random variables.

As shown in the proof of Theorem 1,

$$\begin{aligned}
 & \left| \mathbb{E}_{\hat{\mathbf{p}}_M} \left(\frac{V}{R} \mid \tilde{m}_0 = m_0 \right) - \mathbb{E} \left(\frac{V}{R} \mid \tilde{m}_0 = m_0 \right) \right| \\
 &= \left| \sum_{i \in I_0} \sum_{k=1}^m \frac{1}{k} \left(\Pr \left(\hat{p}_{M,i} \leq q_k, \hat{p}_{M,(k-1)}^{(-i)} \leq q_k, \hat{p}_{M,(k)}^{(-i)} > q_{k+1}, \dots, \hat{p}_{M,(m-1)}^{(-i)} > q_m \right) \right. \right. \\
 & \quad \left. \left. - \Pr \left(p_i \leq q_k, p_{(k-1)}^{(-i)} \leq q_k, p_{(k)}^{(-i)} > q_{k+1}, \dots, p_{(m-1)}^{(-i)} > q_m \right) \right) \right| \quad (10)
 \end{aligned}$$

Figure 7. Time series plots for jump intensity estimations when different jump test statistics are used.
doi:10.1371/journal.pone.0058365.g007

$$\leq \left| \sum_{i \in I_0} \sum_{k=1}^{m-1} \frac{1}{k(k+1)} \left(\frac{\Pr(\hat{p}_{M,i} \leq q_k, \hat{p}_{M,(k)}^{(-i)} > q_{k+1}, \dots, \hat{p}_{M,(m-1)}^{(-i)} > q_m)}{\Pr(p_i \leq q_k, p_{(k)}^{(-i)} > q_{k+1}, \dots, p_{(m-1)}^{(-i)} > q_m)} - 1 \right) \right| \quad (11)$$

$$+ \left| \sum_{i \in I_0} \frac{1}{m} (\Pr(\hat{p}_{M,i} \leq q_m) - \Pr(p_i \leq q_m)) \right|. \quad (12)$$

In the first equality, $\Pr(p_i \leq q_k, p_{(k-1)}^{(-i)} \leq q_k, p_{(k)}^{(-i)} > q_{k+1}, \dots, p_{(m-1)}^{(-i)} > q_m)$ is the probability that in addition to rejecting the hypothesis i , we also reject other $k-1$ hypotheses. Sarkar [23] showed that if $m_0 = m$, then

$$\sum_{i \in I_0} \sum_{k=1}^m \frac{1}{k} \Pr(p_i \leq q_k, p_{(k-1)}^{(-i)} \leq q_k, p_{(k)}^{(-i)} > q_{k+1}, \dots, p_{(m-1)}^{(-i)} > q_m) = 1 - \Pr(\hat{p}_{M,(1)} > q_1, \dots, \hat{p}_{M,(m)} > q_m).$$

Therefore if $m_0 = m$, (10) becomes

$$\left| \left(1 - \Pr(\hat{p}_{M,(1)} > q_1, \dots, \hat{p}_{M,(m)} > q_m) \right) - \left(1 - \Pr(p_{(1)} > q_1, \dots, p_{(m)} > q_m) \right) \right|. \quad (13)$$

Equation (13) is the difference between two familywise error rates (FWER, the probability that we at least have one false rejection) which are obtained respectively from using $\mathbf{p}$ and $\hat{\mathbf{p}}_M$ under the BH procedure. The result is not surprising since when all null hypotheses are true, $\text{FDR} = \text{FWER}$.

To make (11) vanish as $M \rightarrow \infty$, (9) in condition 5 of Theorem 1 is one of the sufficient conditions. However, as shown in Theorem 2, such condition is redundant when test statistics are independent and continuous.

We finally have a look of the assumption:

$$\sup_{1 \leq k \leq m} \sup_{i \in I_0} |\Pr(p_i \leq q_k) - \Pr(\hat{p}_{M,i} \leq q_k)| = O\left(\frac{1}{M^\delta}\right). \quad (14)$$

The assumption says that the convergence in law should hold simultaneously at the points q_k for $1 \leq k \leq m$, and for all $i \in I_0$. Such convergence is reasonable for test statistics with limiting normal distribution if we set $\delta = s/2, s = 1, 2, \dots$. Note that if T_i and $\hat{T}_{M,i}$ are continuous,

$$\begin{aligned} & \Pr(p_i \leq q_k) - \Pr(\hat{p}_{M,i} \leq q_k) \\ &= \Pr(\hat{T}_{M,i} \leq \Psi_i^{-1}(1 - q_k)) - \Pr(T_i \leq \Psi_i^{-1}(1 - q_k)). \end{aligned}$$

If $\hat{T}_{M,i}$ and T_i are asymptotically normal, and satisfy $\hat{T}_{M,i} = T_i + O_p(1/M^{\frac{s}{2}})$ for an integer $s \geq 1$, then by theory of Edgeworth expansion of the distributions of $\hat{T}_{M,i}$ and T_i [20, pg.76],

$$\Pr(\hat{T}_{M,i} \leq a) = \Pr(T_i \leq a) + O\left(\frac{1}{M^{\frac{s}{2}}}\right).$$

So $\Pr(\hat{p}_{M,i} \leq 1 - \Psi_i(a))$ can converge to $\Pr(p_i \leq 1 - \Psi_i(a))$ with the rate of $O(1/M^{\frac{s}{2}})$.

In our nonparametric jump test, standard normal is used to approximate $\Pr(\hat{T}_{M,i} \leq a)$ under the null. There are several methods to improve the approximation, for example, the bootstrap approximation and the Box-Cox transformation. Some theoretical results about how the methods perform have been established. Goncalves and Meddahi [21] showed that when no jump presents, distribution of the test statistic for standardised realized volatility can be approximated by $\mathcal{N}(0,1)$ with the rate of convergence $O(1/\sqrt{M})$. They also documented that under some situations, the bootstrap approximation is better than the standard normal approximation, and the error rate can be reduced to $o_p(1/\sqrt{M})$. For the Box-Cox transformation, if there is no jump component, the skewness of the test statistic for realized volatility can be efficiently reduced by optimally choosing the parameter for the Box-Cox transformation [22].

When M and m both go to infinity

In practice, the number of samples M within a hypothesis, may be less than the number of hypotheses m . How such a large m , small M (or in statisticians' view: Large p (number of dimensions), small n (number of samples)) situation affects statistical inferences has been intensively studied recently, especially in simultaneously convergence of the test statistics. For example, when the samples are i.i.d., sufficient conditions for $\hat{p}_{M,i} \xrightarrow{P} p_i$ uniformly for all i already was provided by [23]. Clarke and Hall [24] documented that the difficulties caused by dependence of test statistics can be alleviated when m grows, but the result subjects to that distributions of test statistics should have light tails such as normal or Student's t . Fan et al. [25] proved that if normal or Student's t distribution is used to approximate the exact null distribution, the rejection area is accurate when $\log m = o(M^{1/3})$; but if the bootstrap methods are applied, then $\log m = o(\sqrt{M})$ is sufficient to guarantee the asymptotic-level accuracy.

In practice, high frequency returns might not be i.i.d. distributed. Instead of assuming that samples have certain distributional properties, here we assume that (14) needs to hold. However, by jointly restricting growth rates of M and m , and together with some mild conditions, (14) can also be achieved. It can be seen in the following proposition.

Proposition 1. For all $i \in I_0$ and every a , if there exists some constant $M_0 > 0$ such that for $M \geq M_0$,

$\Pr(\max_{i \in I_0} |M^\delta (T_i - \hat{T}_{M,i})| > a) \leq c_1 \exp(-c_2 a^\rho)$, where c_1 and c_2 are two constants and $\delta > 0$, and $\rho \geq 1$. Also $(\log(m))^{1/\rho}/M^\delta = o(1)$ as $M, m \rightarrow \infty$ holds, then $\sup_{1 \leq k \leq m} \sup_{i \in I_0} |\Pr(\hat{p}_{M,i} \leq q_k) - \Pr(p_i \leq q_k)| = o(1)$.

The proof of proposition 1 can be found in the supplementary materials.

Simulation study

For the simulation study, we consider the following stochastic volatility plus jump model (SVJ):

$$d \log P(t) = \left(\mu - \frac{1}{2} \sigma^2(t) \right) dt + \sigma(t) dW_1(t) + dJ(t),$$

$$d\sigma^2(t) = a(b - \sigma^2(t))dt + \omega\sigma(t)dW_2(t),$$

$$J(t) = \sum_{j=1}^{N(t)} D(t,j), D(t,j) \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1),$$

$$N(t) \stackrel{i.i.d.}{\sim} \text{Poisson}(\lambda dt),$$

where $dW_1(t)$ and $dW_2(t)$ follow the standard Brownian motion and $\sigma^2(t)$ follows the CIR process. $J(t)$ follows a Compound Poisson Process (CPP) with a constant intensity λdt , and $N(t)$ is the number of jumps occurring within the small interval $(t - \Delta t, t]$. We set correlation between $dW_1(t)$ and $dW_2(t)$ equal to zero (no leverage effect). We use the following parameter values for the simulation:

$$\mu = 0.05, a = 0.015, b = 0.2, \text{ and } \omega = 0.05.$$

In the simulation, the unit of a period is one day. We vary the (daily) jump intensity λ at five different levels: 0, 0.02, 0.05, 0.1, 0.15, and 0.2. Note that the intensity parameter λ here is the expected number of jumps occurring per day. Different values of λ tend to have different numbers of jump days over the whole sampling period, therefore result in different numbers of false null hypotheses. This allows us to see how such differences affect outcomes of the simulation.

We mimic the U.S. stock market and generate one minute intraday log prices over 6.5 hours each day. Thus in our simulation, $M = 6.5 \times 60 = 390$, $dt \approx \Delta t = \frac{1}{M}$ and $\lambda dt \approx \frac{\lambda}{M}$. After obtaining a sample path, the jump test statistics $Z_{-1.5,i}$, $Z_{\log,i}$ and $Z_{ratio,i}$ and their corresponding p -values are calculated. We test hypothesis (1) with the test statistics and control the FDR at the level γ with the BH procedure.

Simulation results

We first focus on the case when the FDR control level $\gamma = 0.05$ and the number of null hypotheses $m = 1000$. Figures 1, 2, 3, 4 and 5 show the plots of average values of relevant quantities from 1000 simulation runs. Figure 1 is for performances of the three different test statistics when the FDR is controlled with the BH

procedure. In the top left panel, we show the realized FDR. The solid horizontal line is at the level $\gamma=0.05$. It can be seen that the realized FDR of $Z_{-1.5,i}$ is almost around or under the required level, while $Z_{\log,i}$ has the largest realized FDR for all different values of λ . Overall, as λ increases, no matter which test statistic we use, the desired FDR level can be achieved.

Let $\hat{S}$ denote the realized number of correct rejections. We use $\hat{S}/m_1$ to measure the ability of the test statistics to correctly reject the false hypotheses. As shown in the top right panel of Figure 1, the three test statistics have small differences in $\hat{S}/m_1$. It also can be seen that $\hat{S}/m_1$ increases only slightly as λ increases.

In the bottom left panel of Figure 1, we can see that the significance level $\hat{\pi}^*\gamma/m$ obtained from the BH procedure increases as λ increases. As λ goes up, the number of false hypotheses m_1 tends to increase, and we have less possibility that the test statistic will signal a true null as a false one. Consequently, we do not need a more stringent $\hat{\pi}^*\gamma/m$ to prevent the false rejections, and more rejections can be obtained.

The average number of rejections $\hat{S}$ made by the BH procedure is constantly less than the average value of m_1 , as shown in the bottom right panel of Figure 1. It might be due to that $\gamma=0.05$ is too restricted to obtain more rejections. A remedy is that we can use a more liberal level ($\gamma=0.1$ or 0.15), but tolerate more false rejections. One thing worth to note here is that the average values of m_1/m would in general be less than their corresponding λ , since there may be more than one jump on a day, and this becomes even more obvious when λ becomes large.

We then compare performances of the BH procedure with the conventional procedure of controlling type I error in each hypothesis: H_i^0 is rejected if its realized p -value is no greater than α . Here α we specify are two frequently used levels: 0.01 and 0.05. Relevant results are shown in Figure 2. As can be seen in the first row, when different test statistics are used, the conventional procedure results in a high realized FDR, especially when the jump intensity λ is small (the number of the false null hypotheses tends to be relatively low in the situation). An extremely case is that when there is no jump ($\lambda=0$), rejecting H_i^0 when $\hat{p}_i \leq 0.01$ (or 0.05) results in 100% false rejections. It says that the probability we at least make one false rejection (the familywise error rate, FWER) is one as we follow the conventional procedure. The reason is that when all the null are true and the test statistics for each hypothesis are almost serially independent, if we reject H_i^0 when $\hat{p}_{i,M} \leq \alpha$, on average we would reject αx hypotheses, and all of these rejections are wrong. However, the BH procedure performs far better in this situation. Even in the worst case, on average it only takes about probability 0.276 to make such an error.

Since the specified α 's are on average greater than $\hat{\pi}^*\gamma/m$, it is expected that more rejections can be obtained under the conventional procedure than the BH procedure. This can be seen in the second row of Figure 2. $\hat{S}/m_1$ of the conventional procedure tends to be higher than that of the BH procedure, but as λ goes up, their gap becomes small.

Figure 3 shows performances of the method when lower frequency (5-min, 10-min and 15-min) data is used. $Z_{-1.5,i}$ still has the best ability to satisfy the required FDR levels, but it suffers the greatest loss of $\hat{S}/m_1$ when the data frequency goes lower. $Z_{\log,i}$ does not perform better than the case when 1-min data is used, no matter in satisfying the required FDR level or $\hat{S}/m_1$. For $Z_{ratio,i}$, its performance still is in the middle, but overall its performance is more stable than the other two competitors.

We then have a look at how the method performs when the number of hypotheses changes. We vary m at several different levels, ranging from 50 to 2000 and keep $\gamma=0.05$. The results are shown in Figure 4. It can be seen that when $\lambda \neq 0$ and m is large (no less than 100), the realized FDR and $\hat{S}/m_1$ are stable over different m .

How does the method perform when FDR is controlled at different required levels? Figure 5 shows different required levels γ and the realized FDR. The thick line is a 45-degree line, and the vertical dotted line is for $\gamma=1/2$. Ideally the realized FDR needs to be equal or below the 45-degree line. For $\lambda=0.05$ and 0.15 , the method performs well, especially when γ goes large. However, when $\lambda=0$, there is a significant difference between the three test statistics, and the required FDR level becomes difficult to achieve in this situation.

The above results suggest that performances of the hybrid method are positively related to sampling frequency M and the intensity parameter λ . Although the BH procedure results in quite stringent rejection criteria, it still can keep $\hat{S}/m_1$ at a satisfying level. Fixing rejection region at $\alpha=0.01$ and 0.05 indeed can have better $\hat{S}/m_1$, but it can suffer far higher false rejections when the number of true null is large. In sum, the simulation shows that combining the BNS test with the BH procedure, the FDR can be well controlled and the test statistics also can keep substantial ability to correctly identify jump components. Finally, we also conduct a simulation study with the stochastic volatility plus jump model (SV1FJ) used in [12]. The results can be found in the supplementary materials (Figures S1, S2, S3, S4 and S5 in the supplementary materials) and they are qualitatively similar to those of the SVJ case shown here.

Real data applications

In the following we present some empirical results with real data. The raw data used for the empirical applications are one minute recorded prices of S&P500 (SPC500) index in cash and Dow Jones Industrial Average (DJIA) index. The sample period spans from Jan-02-2003 to Dec-31-2007. In order to reduce estimation errors caused by microeconomic structure noises, we use five minute log returns to estimate RV_i , BV_i and JV_i and the jump test statistics. Figure S6 and S7 in the supplementary materials show volatility signature plots for detecting microstructure noise and time series plots of the price variations. A detail description of the data and discussion on the microstructure issue can be found in the supplementary materials.

Table 2 shows summary statistics of the price variations, different types of $\hat{T}_{i,M}$, their corresponding $\hat{p}_{i,M}$ and mutual correlations of these quantities of the two indices. Results of the Ljung-Box test (denoted by LB.10) indicate that the price variations are highly serially correlated. However, for $\hat{T}_{i,M}$ and $\hat{p}_{i,M}$, the Ljung-Box test instead indicates that they exhibit almost no serial correlation, which suggests that the BH procedure may efficiently control the FDR in this case.

The daily test statistics of the two indices have high mutual correlations. This property is quite different from the daily test statistics between individual stocks and the market index. As shown in [2], the jump test statistics of individual stocks and the market index almost have no mutual correlation, even though their returns are highly correlated. Such low correlation is due to a large amount of idiosyncratic noises in the individual stock returns, which causes a low signal-to-noise ratio in the nonparametric jump test statistics. The high mutual correlation between the jump test statistics of the two benchmark indices suggests that the idiosyncratic noises of returns is not significant and we may have

more reliable results when we perform the jump test at the market level.

Common jump days

To measure daily price variation induced by jumps, we use sum of squared intraday jumps, which can be estimated by the following estimator:

$$JV_{i,\gamma} = JV_i \times 1\left\{\hat{p}_{i,M} \leq \frac{\hat{i}^*}{m}\gamma\right\}, \quad (15)$$

where $JV_i = RV_i - BV_i$. Table 3 shows summary statistics of $JV_{i,\gamma}$ when FDR is controlled at level $\gamma = 0.01$ and 0.05 . The mean and standard deviation of (15) shown here are conditional on $\hat{p}_{i,M} \leq \hat{i}^*/m$. The conditional mean is around 0.14 to 0.22 for SPC500 and 0.13 to 0.16 for DJIA. For SPC500 and DJIA, the significant levels $\hat{i}^*/m$ for the three statistics are all below 0.006 when the FDR control level $\gamma = 0.05$. Depending on different test statistics, the proportion of identified jump days among all days, is around 1.5% to 11.6% for SP500 and around 2.4% to 8.6% for DJIA.

Common components in two highly correlated asset prices are often one of the most widely studied issues in empirical finance. Here we document some relevant empirical findings. Figure 6 shows the time series plots of the identified $JV_{i,\gamma}$ on the common jump days, and Table 4 shows their summary statistics. The term common jump days used here only means that the two indices both have jumps on these days. It does not necessarily mean that the two indices jump exactly at the same time within these days. Since the daily BNS test statistic is obtained by integrated quantities over one day, it cannot tell us how many and what exact time the jumps occur within that day. Nevertheless such test at least let us know what common days they have jumps, and this information is still valuable for further research.

It can be seen that the results from the two methods are very similar. When the FDR control level $\gamma = 0.05$, proportion of the common jump days among all jump days is around 41% for SPC500. This proportion varies from 31% to 51% for DJIA when different test statistics are used. Comparing magnitudes of the variations in Table 4 with those in Table 3, the two indices tend to have larger jumps on the common days. The result seems to imply that a common shock such as announcements of macroeconomic news, may induce a larger jump than other idiosyncratic shocks such as announcements of news of individual stocks.

Jump intensity estimation

Jump intensity of an asset price process is a very crucial parameter for evaluating risks of the asset. As shown in [26] and [27], the jump intensity seems to change over time, which implies that clustering of jump variations is time varying. The time varying jump intensity also demonstrates very different dynamic behavior across different assets. In the previous literatures, the time varying jump intensity is estimated via moving average of the number of identified jump days, but the threshold for identifying these jump days is a fixed type I error. Here, rather than controlling the fixed type I error over the whole sampling period, we try to incorporate the FDR control into the rolling window estimation.

The simple moving average (rolling window) intensity estimator for the k th day is defined as

$$\hat{\lambda}_k^{mov} = \frac{1}{K} \sum_{i=k-K+1}^k 1\{\hat{p}_{i,M} \leq \theta\},$$

where θ is a threshold, and K is length of the rolling window. The estimator can serve as a local approximation for the true intensity of the jump process, if we assume that number of jumps occurring at most once per day. In the following analysis, we set $K = 120$, and θ is chosen based on two different ways: The first one is the FDR criterion using the whole $m = 1247$ hypotheses, and the second one is the FDR criterion using the K hypotheses within that window with the required FDR level $\gamma = 0.15$.

While the first method always has θ fixed, the later method leads to an adaptive FDR criterion which may change over time, since including a new $\hat{p}_{i,M}$ may make a different FDR criterion. Time series plots for the estimations with the three different jump test statistics are illustrated in Figure 7. In the left panel are plots for the SPC500 and the right panel are plots for the DJIA. It can be seen that with $Z_{-1.5,i}$, $\hat{\lambda}_k^{mov}$ tends to be constantly lower than those with the other two test statistics. When θ is chosen adaptively over the whole sampling period, $\hat{\lambda}_k^{mov}$ is more volatile; and it tends to be higher (lower) when more (less) jump days are identified. This phenomenon holds no matter which test statistic is used. On the other hand, with θ fixed, $\hat{\lambda}_k^{mov}$ is less sensitive to inform such large price movements. Finally, one should note that adaptively choosing θ is only meaningful if the control procedure can lead to a different choice of θ as different information appended, which is possible for the BH procedure but can never be achieved via the conventional type I error control.

Conclusion

In this paper, we have tested whether a stochastic process has jump components by the BNS nonparametric statistics, and controlled the FDR of the multiple testing with the BH procedure.

Theoretical and simulation results are presented to support validity of the hybrid method. Under appropriate conditions, the FDR can be asymptotically controlled by the BH procedure if the p -values are obtained via the asymptotical distributions. The simulation results show that the transformed BNS test statistics can perform well in satisfying the required FDR level with the BH procedure. Their ability to correctly reject false hypotheses is also improved as the frequency of jumps increases. By controlling the FDR, we can have a large chance to avoid any wrong rejection when the stochastic process does not have any jump components. Overall, our simulation results suggest that performance of the method is positively related to the jump intensity and sampling frequency, and is stable over different numbers of hypotheses and the required FDR levels.

As for the empirical results, we find the daily nonparametric test statistics and their corresponding p -values almost have no serial correlation, either for the SPC500 or DJIA. But the test statistics between the two indices are highly mutually dependent. The two indices tend to have larger jumps on the common jump days. We also demonstrate different properties of jump intensity estimations from fixed and adaptive threshold methods. The jump intensity estimated from adaptive threshold method is more sensitive to inform large price movements.

Supporting Information

Figure S1 Realized FDR, $\hat{S}/m_I$, significance level obtained from the BH procedure and number of rejections. In the graphs, each point is an average value from 1000 simulations.
(TIF)

Figure S2 Realized FDR and $\hat{S}/m_I$ of the hybrid method and the conventional procedure. In the graphs, each point is an average value from 1000 simulations.
(TIF)

Figure S3 Realized FDR and $\hat{S}/m_I$ of the hybrid method with lower frequency data. In the graphs, each point is an average value from 1000 simulations.
(TIF)

Figure S4 Realized FDR and $\hat{S}/m_I$ of the hybrid method when the number of hypotheses varies. Here $m = 50, 100, 200, 500, 800, 1000, 1200, 1500$ and 2000. In the graphs, each point is an average value from 1000 simulations.
(TIF)

References

- Ait-Sahalia Y, Jacod J (2009) Testing for jumps in a discretely observed process. *The Annals of Statistics* 37: 184–222.
- Bollerslev T, Law TH, Tauchen GE (2008) Risk, jumps, and diversification. *Journal of Econometrics* 144: 234–256.
- Barndorff-Nielsen OE, Shephard N (2004) Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics* 2: 1–37.
- Barndorff-Nielsen OE, Shephard N (2006) Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4: 1–30.
- Fan Y, Fan J (2011) Testing and detecting jumps based on a discretely observed process. *Journal of Econometrics* 164: 331–344.
- Jacod J, Todorov V (2009) Testing for common arrival of jumps for discretely observed multidimensional processes. *The Annals of Statistics* 37: 1792–1838.
- Jiang GJ, Oomen RC (2008) Testing for jumps when asset prices are observed with noise - a swap variance approach. *Journal of Econometrics* 144: 352–370.
- Lee SS, Mykland PA (2008) Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies* 21: 2535–2563.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 57: 289–300.
- Bajgrowicz P, Scaillet O (2011) Jumps in high-frequency data: Spurious detections, dynamics, and news. *Swiss Finance Institute Research Paper No 11–36*.
- Barndorff-Nielsen OE, Shephard N (2002) Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* 17: 457–477.
- Huang X, Tauchen GE (2005) The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3: 456–499.
- Veraart AE (2010) Inference for the jump part of quadratic variation of ito semimartingales. *Econometric Theory* 26: 331–368.
- Romano JP, Shaikh AM, Wolf M (2008) Control of the false discovery rate under dependence using the bootstrap and subsampling. *TEST* 17: 417–442.
- Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64: 479–498.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* 29: 1165–1188.
- Sarkar SK (2002) Some results on false discovery rate in stepwise multiple testing procedures. *The Annals of Statistics* 30: 239–257.
- Karatzas I, Shreve SE (1991) *Brownian Motion and Stochastic Calculus*. New York: Springer, 2nd edition.
- Sarkar SK (1998) Some probability inequalities for ordered m tp2 random variables: A proof of the simes conjecture. *The Annals of Statistics* 26: 494–504.
- Hall P (1992) *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
- Goncalves S, Meddahi N (2009) Bootstrapping realized volatility. *Econometrica* 77: 283–306.
- Goncalves S, Meddahi N (2011) Box-cox transforms for realized volatility. *Journal of Econometrics* 160: 129–144.
- Kosorok MR, Ma S (2007) Marginal asymptotics for the “large p , small n paradigm”: With applications to microarray data. *The Annals of Statistics* 35: 1456–1486.
- Clarke S, Hall P (2009) Robustness of multiple testing procedures against dependence. *The Annals of Statistics* 37: 332–358.
- Fan J, Hall P, Yao Q (2007) To how many simultaneous hypothesis tests can normal, student's t or bootstrap calibration be applied? *Journal of the American Statistical Association* 102: 1282–1288.
- Tauchen G, Zhou H (2011) Realized jumps on financial markets and predicting credit spreads. *Journal of Econometrics* 160: 102–118.
- Andersen TG, Bollerslev T, Diebold FX (2007) Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics* 89: 701–720.

Materials S1 Supplementary Materials.
(PDF)

Author Contributions

Conceived and designed the experiments: YY. Performed the experiments: YY. Analyzed the data: YY. Contributed reagents/materials/analysis tools: YY. Wrote the paper: YY.