Jumps in high-frequency data: spurious detections, dynamics, and news*

Pierre Bajgrowicz† Olivier Scaillet‡ Adrien Treccani§

January 2015

JEL classification: C58, G12, G14.
Keywords: jumps, high-frequency data, spurious detections, jumps dynamics, news releases, cojumps.

*We would like to thank the editor, the associate editor and the referees for constructive criticism and numerous suggestions that have led to substantial improvements over previous versions of the paper. We thank Eric Ghysels, Rajna Gibson, Raul Gonzalez, Per Mykland, Michael Rockinger, Neil Shephard, as well as seminar participants at the Humboldt University of Berlin CASE-QPL Distinguished Lecture Series (2010), Oxford-Man Institute (2010), the Econometric Society World Congress (2010), SoFiE (2010), Banque de France (2010), MAF (2010), Université d'Orleans (2010), Edhec Risk Institute (2010), University College of Dublin (2010), the EC^2 conference (2009), Solvay Business School (2009), Université Paris X–Nanterre (2009), ESEM (2009), and the Swiss Doctoral Workshop in Finance (2009) for helpful comments. The authors received financial support from the Swiss National Science Foundation through the National Center of Competence in Research: Financial Valuation and Risk Management (NCCR FINRISK).

†University of Geneva, and Litasco SA, 40 Bd du Pont d'Arve, 1211 Geneva, Switzerland. Voice: +41 22 705 22 27. Fax: +41 22 379 81 04. E-mail: pierre.bajgrowicz@unige.ch.

‡University of Geneva and Swiss Finance Institute, 40 Bd du Pont d'Arve, 1211 Geneva, Switzerland. Voice: +41 22 379 88 16. Fax: +41 22 379 81 04. E-mail: olivier.scaillet@unige.ch.

§University of Geneva and Swiss Finance Institute, 40 Bd du Pont d'Arve, 1211 Geneva, Switzerland. Voice: +41 22 379 81 66. Fax: +41 22 379 81 04. E-mail: adrien.treccani@unige.ch.

Abstract

Applying tests for jumps to financial data sets can lead to an important number of spurious detections. Bursts of volatility are often incorrectly identified as jumps when the sampling is too sparse. At a higher frequency, methods robust to microstructure noise are required. We argue that whatever the jump detection test and the sampling frequency, a large number of spurious detections remain because of multiple testing issues. We propose a formal treatment based on an explicit thresholding on available test statistics. We prove that our method eliminates asymptotically all remaining spurious detections. In Dow Jones stocks between 2006 and 2008, spurious detections can represent up to 90% of the jumps detected initially. For the stocks considered, jumps are rare events, they do not cluster in time, and no cojump affects all stocks simultaneously, suggesting jump risk is diversifiable. We relate the remaining jumps to macroeconomic news, prescheduled company-specific announcements, and stories from news agencies which include a variety of unscheduled and uncategorized events. The vast majority of news do not cause jumps but may generate a market reaction in the form of bursts of volatility.

1 Introduction

Evidence of stochastic skewness and kurtosis of asset return distributions has led to the development of models with jumps to better incorporate these dynamics. Jumps are rare and larger events than what a continuous diffusion process can explain. Detecting jumps and studying their dynamics is important because of the consequences in applications including derivatives pricing and risk management. However, determining from discrete observations whether we should consider a large return as a jump is no trivial task[1]. Numerous statistical methods to test for the presence of jumps in high-frequency data have been introduced in recent years. When studying the dynamics of jump arrivals, the jump detection tests have to be applied over a period of time, i.e., over a series of days simultaneously. Such a procedure results in performing multiple testing, and leads by construction to making a significant number of spurious detections, regardless of the underlying test. For instance, if the jump tests are carried out at the 5% significance level over a one-year period (i.e., 252 trading days) with no single jump, on average more than 12 days are still going to be erroneously selected as containing a jump. As we show in a Monte Carlo study provided in the supplemental file, the presence of spurious detections can seriously bias the estimated proportion of jump days and results on jump dynamics, e.g., tests of clustering of jump arrivals. In this paper, we propose a new formal methodology to resolve such erroneous detection problems and show its good finite sample performance with Monte Carlo simulations. Our thresholding technique reveals a low frequency of jumps but more frequent bursts of volatility. We study their dynamics and investigate whether news (including stories from news agencies) can explain them.

Our first contribution is to propose a method to eliminate spurious detections due to multiple testing via an explicit thresholding on available test statistics. We are the first to provide a formal treatment of the multiple testing bias with double asymptotics when applying jump detection tests over a sample of many days. We prove that if we consider test statistics above a certain threshold level only, the likelihood of making such spurious detections disappears asymptotically. Monte Carlo results show that our approach behaves well in finite samples. Our theoretical results legitimize the ad hoc response to the multiple testing issue taken in some studies, which is to use very conservative critical values, e.g., at a 0.1% significance level for one-sided tests

[1]The recent literature, e.g., Aït-Sahalia (2004) and Lee and Hannig (2010), distinguishes between big Poisson-type jumps and small possibly infinite-activity jumps. We concentrate on the first kind in order to study the relation with important news announcements.

(Bollerslev, Law and Tauchen (2008), Giot, Laurent and Petitjean (2010)). In our analysis, we focus on very liquid large-capitalization U.S. stocks. We collect high-frequency returns from the Trades and Quotes (TAQ) database for the Dow Jones Industrial Average Index (DJIA) constituents, over the three-year period of January 2006 to December 2008. We also study the presence of jumps in the index, which we proxy with the Diamonds exchange-traded fund (ETF) and a price-weighted portfolio of the 30 Dow Jones constituents. We consider 2 minutes and 10 seconds sampling frequencies, and use the adjusted ratio statistic of Barndorff-Nielsen and Shephard (2006) (BNS) and the estimator of Christensen, Oomen and Podolskij (2014) (COP) as the underlying tests to detect jumps. Our method to eliminate spurious detections can be applied just as easily on other existing jump detection techniques, such as Aït-Sahalia and Jacod (2009), or Andersen, Bollerslev and Dobrev (2007). To summarize our analysis, first, we find significantly less jumps in the 10 seconds case than in the 2 minutes case. This confirms the results of COP, who argue that stock price processes exhibit bursts of volatility that are incorrectly captured as jumps when the sampling grid is too sparse. Second, we find that up to 90% of the jumps found at the 10 seconds frequency are spurious detections due to multiple testing[2]. This illustrates the important bias induced by multiple testing, making our thresholding technique essential for a proper analysis of jumps. Our results bring the high number of jumps detected by existing tests down to an amount more in line with the intuition that jumps are rare events.

The second contribution is the investigation of the dynamic features of durations between jumps in equity prices. As shown in our simulations, we can only uncover true dynamics of jump arrivals once erroneous detections are removed, e.g., using our thresholding technique. The empirical series do not reveal a clustering in time of jump occurrences. Our results are in favor of the hypothesis that jump arrivals follow a simple low intensity Poisson process and, hence, support the jump process used by Merton (1976) to correct the discrepancies between market prices and the Black-Scholes value of options. During the three years of our study, we find no day where the 30 stocks all jump simultaneously and we detect a jump in more than 20% of the stocks only on two occasions. The absence of cojumps affecting all stocks supports the assumption in Merton (1976) that the jump component is nonsystematic or diversifiable. One consequence of the diversifiability is that the jump risk does

[2]We assess the robustness of our results by repeating the procedure at 5 minutes and 20 seconds sampling frequencies with the BNS and COP estimators, respectively. We obtain similar results, but the effect of bursts of volatility is even more pronounced at a 5 minutes frequency.

not require a risk premium in principle. Also, we do not observe a number of cojumps significantly larger than if the stocks jumped independently when we consider industry sectors separately. Not commanding a risk premium seems at odds with the findings of Bollerslev and Todorov (2011). They use 5-minute S&P 500 futures prices and short-term deep out-of-the-money options to infer the path of the equity jump risk premium. They find that the median of the estimated equity jump risk premia is equal to 5.2%. Hence derivatives markets seem to price jump risk significantly. An explanation for such a large value is that to diversify away jump risk is too costly or impossible because of transaction costs or other frictions. Another explanation is that the use of 5-minute interval mistakes a volatility burst for a real jump (as documented in our results below). The use of only short-term deep out-of-the-money options might also incorporate a liquidity or preferred habitat component. We definitely need further research on jump risks to draw a definite conclusion, but we believe that our paper does contribute to such a debate.

The third contribution of the paper is to relate the few jumps that remain after we apply our thresholding technique to news announcements. Early papers have conjectured that jumps are caused by the arrival of important new information, most often specific to the firm, and occasionally more general economic or market news. A hypothesis is that market-level news can cause jumps in many stocks simultaneously which can propagate even to a diversified index. Examining what type of information is dynamically related to jumps helps better explaining market phenomena and improving pricing models. We start by investigating the effect of macroeconomic announcements and observe no significant effect, even when accounting for the surprise component of the news. Our results differ from the findings in other markets, e.g., Dungey, McKenzie and Smith (2009) find that two thirds of cojumps in bond prices coincide with a scheduled US news release, albeit the authors do not control for multiple testing and do not consider the same frequency of sampling. Next, we look at prescheduled announcements specific to each stock. We observe no increase in the occurrence of jumps neither on quarterly earnings nor on dividend announcements. Finally, we consider stories from two news agencies: Reuters and Dow Jones News Service. By examining the content of news stories, we can analyze the impact of a variety of unscheduled and uncategorized events, and are not limited to a predetermined set of event types such as earnings announcements, mergers, or analyst recommendations. We use the Factiva database to retrieve the news stories. To our knowledge, we are the first to perform an extensive analysis (over several years and stocks) of the relation between stories from news agencies and sudden market moves. Given the huge quantity of information archived in the Factiva repository, one major challenge is to get everything relevant while eliminating erroneous and unimportant stories. Our results show that news releases are not likely to cause jumps either. Companies purposefully shift most important announcements after the bell or early in the morning in order to avoid uncontrolled investor reactions and the consequent impact on the stock price. These results are consistent with our finding of a very limited number of actual jumps once we correct for the different sources of bias. Our conclusions differ from the findings of Lee and Mykland (2008) who examine the association of news with jumps on a small sample of stocks over only three months, and find a story for each jump they detect. They also challenge the results of Lee (2012) which are based on the same underlying test. However, both studies use a low sampling frequency of 15 minutes. If we loosely define a burst of volatility as a jump detection at a relatively low frequency that is not captured at high frequency as argued by COP, we also identify a link between news and volatility. For instance, we find that the press releases following scheduled Federal Open Market Committee (FOMC) meetings increase the likelihood of bursts of volatility, although not to a statistically significant extent. Also, announcements concerning share repurchase programs, therefore directly related to the balance sheet of the company, have a significant impact on the volatility of the share price.

2 Eliminating spurious jump detections

2.1 Setting and assumptions

The Black-Scholes option pricing model assumes that stock prices follow a stochastic process that generates a continuous trajectory. This requirement implies that over a short period of time, the stock price cannot suddenly change by much. This assumption is challenged by the too many outliers observed in empirical studies and the behavior of option prices[3]. One solution to capture the skewness and kurtosis of asset returns is to include jumps, i.e., to allow for stock price variations of extraordinary magnitude, no matter how small the interval between successive observations. As explained by Merton (1976), including jumps also allows to solve the discrepancy between market prices of options and their Black-Scholes value. The stock price is then written as a combination of two types of changes. The continuous part models normal vari-

[3]See Aït-Sahalia (2002), Carr and Wu (2003), who test whether a diffusion is sensible to model asset prices.

ations in price. The jump part captures abnormal variations. In Merton (1976), the latter are supposed to be due to the arrival of new important information about the stock. Typically, such information is most often specific to the firm or its industry.

Let X_t for continuous time $t \geq 0$ denote the log-price of the asset. The workhorse model of modern asset pricing theory assumes that the log-price follows an Itô semimartingale. A semimartingale can be decomposed into the sum of a drift, a continuous Brownian-driven part, and a discontinuous, or jump, part:

$$dX_t = b_t dt + \sigma_t dW_t + dJ_t,$$

where W_t denotes a standard Brownian motion and J_t is a pure jump process. We follow the assumption that jumps are relatively rare and large events (Merton (1976), Barndorff-Nielsen and Shephard (2006)), and do not consider infinite-activity jumps (Aït-Sahalia (2004)).

Identifying jumps empirically is difficult because only discretely sampled data are available. In reality, detecting jumps amounts to answering the following question. Given that we observe in discrete data a change in the asset return of a large magnitude, what does that tell us about the likelihood that such a change involves a jump, as opposed to just a large realization of the Brownian part?

2.2 Thresholding technique

Numerous jump detection methods have been developed since high-frequency data have become easily available. In a typical empirical application, the jump tests are applied to detect jump days over a sample period. For each day, a test statistic S is computed to test the null hypothesis of no jump. The performance of S is characterized by its ability to detect actual jumps, yet avoiding rejecting when there is no jump, that is, making a type I error[4]. The problem is that performing the tests for many days simultaneously results in conducting multiple testing, which by nature leads to making a proportion of spurious detections equal to the significance level of the individual tests. For example, if the individual tests are performed at the 5% significance level during

[4]We cannot simultaneously minimize the probability of committing a type I error and maximize the probability of detecting jumps. If the size decreases, the power of the test deteriorates until it reaches zero. The power of a single test is one minus the probability of committing a Type II error. There is nothing wrong per se in choosing a small size in a frequentist world. It only means that we want to be conservative and this decision pertains to the econometrician. The former discussion is well-known and concerns single testing procedures (see, e.g., the review paper of Lehman (1993) on the Fisher, Neyman-Pearson theories of testing hypotheses).

a one-year period with no single jump, by construction on average more than 12 days are going to be erroneously selected as containing a jump. An ad hoc response to this issue has been the use of critical values further in the tails, i.e., 0.1% critical values in Bollerslev, Law and Tauchen (2008) and Giot, Laurent and Petitjean (2010). However, in a multiple testing setting, the notion of size is replaced by the Family Wise Error Rate (FWER, or another type I error measures like the the False Discovery Rate (FDR)), which is the probability of making one or more false rejection[5]. A multiple testing procedure has to achieve a control of the FWER to avoid systematic spurious detections. This requires that the FWER is no bigger than a probability level, at least asymptotically, which is not the case of ad hoc approaches. We illustrate this notion in our simulation experiments and show the importance of controlling the FWER when studying the proportion of jump days and jump dynamics.

The major methodological contribution of the present paper is to propose a formal treatment to the multiple testing issue with a double asymptotics theory. Lee and Mykland (2008) and Lee and Hannig (2010) address the issue of multiple testing in their tests using a thresholding technique but their methods are tailored made for their jump test strategy and are based on single asymptotics. They control the total number of intraday tests on the whole time period. Andersen, Bollerslev and Dobrev (2007) control for the size of the multiple jump tests using a Bonferroni correction. Lee and Mykland (2012) apply a sequential extension of the Bonferroni procedure (Holm (1979)) to control the FWER in multiple tests on IBM stock trades during August 2007 in their empirical section. Our approach separates the number of periods within a day from the number of days, and focuses on the multiple testing control across days. It is therefore more liberal when the sampling frequency is high or the number of periods is large, allowing for more jump detections (see footnote 18 below for a comparison with the conservative approach of Lee and Mykland (2208) based on single asymptotics). To our knowledge, there is no empirical and theoretical literature on the issues arising when applying a jump detection test on a sample containing a large number of days with a double asymptotic analysis, and this for several stocks and their related news.

Our thresholding technique allows to eliminate the spurious detections, based on the following theoretical result developed in detail in the appendix. Denote by N the number of days in the study, and by n the number of observations per day used to compute each individual test statistic. We obtain a series of daily statistics which can be written as $(S_1^n, \ldots, S_N^n)$. For most available tests,

[5]See, e.g., the review paper of Romano, Shaikh and Wolf (2008) on multiple testing procedures and related tools and notions.

under the null hypothesis of no jumps, the statistics converge to independent standard normal random variables. Theorem 1 of the appendix states that, under some technical conditions about the relative rate of convergence of n with respect to N and about the underlying price process, we get, under the null hypothesis of no jumps, $P\left[\sup_t |S_t^n| \leq \sqrt{2\log N}\right] \to 1, \quad \text{as} \quad N, n \to \infty$. This means that, if there are no jumps, the event that the largest and the smallest of the entries of the vector $(S_1^n, \ldots, S_N^n)$ stay within $\left[-\sqrt{2\log N}, \sqrt{2\log N}\right]$ becomes certain for large n and N. The bound $\sqrt{2\log N}$ is the so-called universal threshold for a sample of size N. As explained in Donoho and Johnstone (1994), it is asymptotically a common, i.e., universal, upper bound on the root mean square error of thresholded estimates in multivariate normal decision theory.

Using the theorem, we obtain a method to eliminate spurious detections that we can apply very easily on top of most existing jump detection tests. In the first step, we compute the test statistics individually for each day. In the second step, we discard statistics in the band $\left[-\sqrt{2\log N}, \sqrt{2\log N}\right]$. This way, spurious detections of jumps become negligible with high probability. Our theoretical results legitimize the ad hoc choice of more conservative critical values. The precise statement and the proof of the theorem are in the appendix, for a general test statistic, as well as for the specific example of the BNS statistic. The theorem provides a theoretically appropriate significance level, which depends on the number N of tests, instead of an ad hoc one. Indeed, we can map the universal threshold $\sqrt{2\log N}$ into a significance level α_N by using the characterization of the quantile of a standard Gaussian. We just need to compute $\alpha_N/2 = (1 - \Phi(\sqrt{2\log N}))$, where Φ is the cdf of a standard Gaussian random variable. By solving $0.2\%/2 = (1 - \Phi(\sqrt{2\log N}))$, we deduce that an a priori ad hoc rule based on 0.2% for two-sided tests is misguided when N is different from $\exp((\Phi^{-1}(1 - .002/2))^2/2) \simeq 118$, a little bit less than a half year of data. If $N > 118$, we need to use a larger threshold $\sqrt{2\log N}$ (larger bands), or equivalently a smaller probability level α_N, to get an adequate multiple testing control. One-sided tests give the same size and power if you use half of the significance level, for example 0.1%.

The universal threshold procedure corresponds to a two-sided test in which the level $\alpha_N = 2\Phi(-\sqrt{2\log(N)})$ as explained above. We can approximate this level by $\alpha_N \approx (N\sqrt{\pi \log N})^{-1}$ because $\Phi(-x) \approx \phi(x)/x$, when x is large, where Φ and ϕ denote the cdf and pdf of standard normal distribution. The universal thresholding is thus related to a Bonferroni-type test where the size of the single test is set at $(N\sqrt{\pi \log N})^{-1}$. By construction, it controls the probability of even one erroneous inclusion of a jump, i.e., a false rejection, at level $(\sqrt{\pi \log N})^{-1}$

since the Bonferroni method does so. Indeed, when investigating the properties of a multiple testing procedure, we need to check whether the suggested procedure achieves a control of the FWER. The control of the FWER requires that the FWER is no bigger than a probability level, at least asymptotically, when all null hypotheses are true (weak control) or when some are true and some are false (strong control). The Bonferroni method achieves a strong control at the size used in each single test multiplied by the number of test. If we have 100 one-sided tests, and use a size equal to 0.1%, we get an asymptotic control of the FWER at 10%, namely an upper bound on the FWER of 10%[6]. A multiple testing procedure is considered to be optimal if it maximizes the number of true discoveries, while keeping one of the type I error measures like FWER or FDR at a certain, fixed level. The universal thresholding, the Bonferroni procedure, and the FDR control are known to share such asymptotic optimality properties (see the detailed discussion in the recent paper by Bogdan, Chakrabarti, Frommlet and Ghosh (2011), and the references therein). As in a single test context, we cannot speak about power if we do not control size. A single test with no correct asymptotic size is useless even if its power is large. This is also true in a multiple testing problem.

2.3 FDR thresholding

In addition to the universal threshold, we also report results using the data-adaptive thresholding scheme of Abramovich, Benjamini, Donoho and Johnstone (2006), based on the control of the false discovery rate (FDR). FDR control is a relatively recent innovation in simultaneous testing, which ensures that at most a certain expected fraction of the rejected null hypothesis correspond to spurious detections. Barras, Scaillet and Wermers (2010) use the FDR in the context of mutual fund performance assessment, and Bajgrowicz and Scaillet (2012) to account for data snooping while assessing performance of technical trading rules (see also Harvey, Liu and Zhu (2013) in the analysis of the statistical relevance of newly discovered factors in empirical asset pricing). Throughout the paper, we set the FDR target level at 10%, which results in a less conservative threshold level than with the universal threshold, and eliminates fewer jump detections. When we set the control of the FDR at 10%, we are more liberal and admit that 10% of the rejected null hypotheses, i.e., detected jump days, will be by construction spurious. We obtain qualitatively similar results with an FDR level between 5% and 20%. Setting the FDR target

[6]For a one-sided test with an ad hoc choice of a 0.1% significance level, the Bonferroni procedure only controls asymptotically the FWER at the level $\alpha = N \times 0.1\% = 75\%$ for $N = 756$, which is a very liberal level of control.

level to zero is equivalent to using the universal threshold. The FDR approach results in a threshold inherently adaptive to the data. The FDR threshold is higher when there are few true jumps, i.e., the signal is sparse, and lower when there are many jumps, i.e., the signal is dense.

The choice of which threshold to use—universal or FDR—depends on the application. If, for example, we are interested in the probability of a jump conditional on a news release, the FDR threshold is more appropriate as it reduces the likelihood of missing true jumps. On the other hand, if the goal is to study what kind of news cause jumps, it is better to apply the universal threshold in order to avoid looking vainly for a news when in fact the detection is spurious.

2.4 Jump detection techniques

Our thresholding technique can be applied to most existing jump detection tests. In the present paper, we use the standard tests of BNS and COP, with respective frequencies of 2 minutes and 10 seconds[7].

The essence of the BNS jump detection method is to compare the realized quadratic variation which incorporates volatility originating from jumps (if present) to the realized bipower variation which is robust to jumps. Each day $t = 1, \dots, N$, we observe the log-price process X at the discrete times $i\Delta_n$, $i = 1, \dots, n+1$, where Δ_n is the sampling interval and n is large. We denote by $X_{t,i\Delta_n}$ the i^{th} intraday price observation on day t, and by $\Delta X^n_{t,i} \equiv X_{t,(i+1)\Delta_n} - X_{t,i\Delta_n}$ the i^{th} intraday return on day t, $i = 1, \dots, n$. The realized quadratic variation (RV) and the realized bipower variation (BV) of X are defined as follows and converge in probability to different quantities of the underlying jump-diffusion process.

$$RV^n_t \equiv \sum_{i=1}^{n} (\Delta X^n_{t,i})^2 \quad \xrightarrow[n\to\infty]{p} \quad \int_{t-1}^{t} \sigma^2_s ds + \sum_{i>N_{t-1}}^{N_t} c_i^2,$$

$$BV^n_t \equiv \frac{n}{n-1}\frac{1}{\mu_1^2}\sum_{i=2}^{n} |\Delta X^n_{t,i}||\Delta X^n_{t,i-1}| \quad \xrightarrow[n\to\infty]{p} \quad \int_{t-1}^{t} \sigma^2_s ds,$$

where $\mu_1 = \sqrt{2}/\sqrt{\pi}$, N_t is a simple counting process, and the c_i are nonzero

[7]Determining from high-frequency data whether an asset return process has jumps has been considered by a number of authors, see e.g., Carr and Wu (2003), Barndorff-Nielsen and Shephard (2006), Andersen, Bollerslev, Diebold and Labys (2003), Andersen, Bollerslev and Diebold (2007), Huang and Tauchen (2005), Andersen, Bollerslev and Dobrev (2007), Lee and Mykland (2008), Fan and Wang (2007), Jiang and Oomen (2008), Aït-Sahalia and Jacod (2009), Andersen, Dobrev and Schaumburg (2012), Mancini (2009), Lee and Hannig (2010) and Christensen, Oomen and Podolskij (2014).

random variables so that $J_t = \sum_{i=1}^{N_t} c_i$. If the jumps are of finite activity, the probability of observing jumps in two consecutive returns approaches zero. Consequently, the product of any two consecutive returns is asymptotically driven by the diffusion component only and the contribution of jumps is eliminated in the bipower variation. The assumption underlying the BNS test that jumps are large and rare events makes it particularly well-suited for our analysis of the impact of important news[8].

One of the statistics we use is the adjusted ratio statistic of BNS defined below. It is the preferred test in Huang and Tauchen (2005) who investigate size, jump detection rate, and power properties. Up to a scaling factor, the ratio $\frac{BV_t^n}{RV_t^n} - 1$ converges in law to a standard normal random variable under the null hypothesis of no jumps:

$$\frac{\Delta_n^{-1/2}}{\sqrt{\vartheta \max\left((n\Delta_n)^{-1}, QV_t^n/(BV_t^n)^2\right)}} \left(\frac{BV_t^n}{RV_t^n} - 1\right) \xrightarrow{d} \mathcal{N}(0,1),$$

where $QV_t^n \equiv \frac{n}{n-3}\mu_1^{-4}\Delta_n^{-1}\sum_{i=4}^{n} |\Delta X_{t,i}^n||\Delta X_{t,i-1}^n||\Delta X_{t,i-2}^n||\Delta X_{t,i-3}^n|$, is the realized quadpower variation, and $\vartheta = (\pi^2/4) + \pi - 5$.

COP show that low frequency analysis, i.e., when the sampling interval Δ_n is not sufficiently small, overestimate the number of jumps seriously because of frequent bursts of volatility. They advocate the use of high frequency data with an adequate estimator, which is robust to microstructure noise. The BNS estimator does not have this robustness property and its size is biased at high frequencies, in particular (see, e.g., Huang and Tauchen (2005)). We consider noise-robust versions of RV and BV denoted by $\bar{RV}_t^n$ and $\bar{BV}_t^n$, for which noise is smoothed out with a pre-averaging technique.

We introduce the weighting function $g : [0,1] \to \mathbb{R}$ defined by $g(x) \equiv \min(x, 1-x)$, and set the pre-averaging horizon K such that $K = \lceil \theta\sqrt{n} \rceil$ with a tuning parameter $\theta > 0$. We model the noisy price process as $Y \equiv X + u$, where u is an i.i.d. process with $E[u] = 0$ and $E[u^2] = \omega^2$, $\omega > 0$, and u is independent of X. The noise process u stands for the microstructure noise. The smooth version of the returns is then given by $\bar{Y}_{t,i}^n = \sum_{j=1}^{K-1} g\left(\frac{j}{K}\right) \Delta Y_{t,i+j-1}^n$, where $\Delta Y_{t,i}^n \equiv Y_{t,(i+1)\Delta_n} - Y_{t,i\Delta_n}$ is the i^{th} noisy intraday return on day t, $i = 1, \ldots, n-K+2$. We also define the following constants depending on

[8] The intuition is as follows. Since the probability of contiguous jumps goes to zero as $\Delta_n \to 0$, jump terms are always multiplied by another return not including a jump asymptotically. These products do not play a role in the asymptotics since there are only a countable number of jumps and each return without a jump goes to zero in probability. Consequently, the realized bipower variation used in the BNS test converges to the integrated variance. For detailed technical arguments, see the proof of Theorem 5 in Barndorff-Neilsen and Shephard (2004).

the function g: $\psi_1 \equiv \int_0^1 (g'(s))^2 ds$, $\psi_2 \equiv \int_0^1 g^2(s)ds$ and compute these quantities using the Riemann approximations $\psi_1^K \equiv K\sum_{j=1}^{K}\left[g\left(\frac{j}{K}\right) - g\left(\frac{j-1}{K}\right)\right]^2$, $\psi_2^K \equiv \frac{1}{K}\sum_{j=1}^{K-1} g^2\left(\frac{j}{K}\right)$ to improve finite sample accuracy. Next, we define the robust versions of the RV and BV as

$$\begin{aligned}
\bar{RV}_t^n &\equiv \frac{n}{n-K+2}\frac{1}{K\psi_2^K}\sum_{i=1}^{n-K+2}|\bar{Y}_{t,i}^n|^2 - \frac{\psi_1^K}{\theta^2\psi_2^K}\hat{\omega}^2, \\
\bar{BV}_t^n &\equiv \frac{n}{n-2K+2}\frac{1}{K\psi_2^K}\frac{1}{\mu_1^2}\sum_{i=1}^{n-2K+2}|\bar{Y}_{t,i}^n||\bar{Y}_{t,i+K}^n| - \frac{\psi_1^K}{\theta^2\psi_2^K}\hat{\omega}^2,
\end{aligned}$$

where $\hat{\omega}^2 = -\frac{1}{n-1}\sum_{i=2}^{n}\Delta Y_{t,i-1}^n \Delta Y_{t,i}^n$ estimates ω^2, as in Oomen (2006). These noise-robust estimators have the same respective probability limits as the noise-free analogs RV_t^n and BV_t^n. For this reason, we can use the distance between $\bar{RV}_t^n$ and $\bar{BV}_t^n$ to test for the presence of jumps. Precisely, COP show that, under some regularity conditions, the following convergence in distribution holds:

$$\frac{n^{1/4}}{\sqrt{\bar{\Sigma}_{11} + \bar{\Sigma}_{22} - 2\bar{\Sigma}_{12}}/\bar{BV}_t^n}\ln\left(\frac{\bar{RV}_t^n}{\bar{BV}_t^n}\right) \xrightarrow{d} \mathcal{N}(0,1),$$

where $\left(\bar{\Sigma}_{ij}\right)_{1\le i,j\le 2}$ denote individual entries of the asymptotic covariance matrix of the bivariate vector $n^{1/4}(\bar{RV}_t^n - \int_{t-1}^t \sigma_s^2 ds, \bar{BV}_t^n - \int_{t-1}^t \sigma_s^2 ds)$. In practice, this matrix is not known and we estimate it using subsampling. Additionally, we implement the threshold filter they detail in order to estimate $\bar{BV}_t^n$. Our Monte Carlo simulation study confirms that it is primordial to pre-trim the data to reduce the small sample bias[9].

3 Empirical results on true proportion of jump days

We conduct our analysis over the three-year period from January 2006 to December 2008, on the 30 stocks composing the Dow Jones Industrial Average (DJIA) index between November 21, 2005 and February 19, 2008. Most stocks are listed on the NYSE, except for Microsoft and Intel which are listed on the NASDAQ. The cleaning of high-frequency data has been highlighted in

[9] We refer to COP for a detailed explanation of pre-averaging technique, the asymptotic covariance matrix estimation as well as the threshold filter used to estimate the BV. In the rest of the paper, we precisely follow their methodology. We set the pre-averaging parameter to $\theta = 0.5$ and take 20 subsamples of length 25 for the estimation of the covariance matrix. We use the same parametrization as COP for data trimming. Podolskij and Vetter (2009a) propose an alternative element-by-element estimator of the asymptotic covariance matrix, but their approach does not guarantee positive semi-positiveness, which is particularly problematic with small samples.

e.g. Dacorogna, Gencay, Muller, Olsen and Pictet (2001), and Hansen and Lunde (2006). We closely follow Barndorff-Nielsen, Hansen, Lunde and Shephard (2009) and also discard 'bounce back' outliers as defined in Aït-Sahalia, Mykland and Zhang (2011). The supplemental file gives a detailed description of the data and the cleaning procedure.

[Figure 1]

Table 1 shows the average number of detections per year for each stock at 2 minutes and 10 seconds frequencies. For the latter, we give the results respectively before thresholding, after applying the universal threshold, and after applying the FDR threshold. The table reveals two different sources of bias when we try to detect jumps. First, we find many less jumps at the higher frequency than at the lower frequency. This result is consistent with the findings of COP, who argue that the frequent bursts of volatility in asset prices are incorrectly interpreted as jumps when the sampling grid is too sparse. Indeed, financial series tend to have a highly dynamic conditional volatility, which leads to fast but continuous price changes. A high frequency sampling captures intermediary price steps and avoids these spurious detections. Second, Table 1 reveals an even more severe source of spurious detections due to multiple testing. In average, the FDR threshold removes more than 75% of the residual jump dates of the 10 seconds case. The universal threshold gets rid of even more jump dates. The average number of actual jumps per year amounts to less than five[10]. We assess the robustness of this result by applying the alternative test of Lee and Mykland (2012) on our 10-second data set. The authors propose to identify jumps by comparing realized returns to corresponding estimated local volatilities on sequential blocks of intraday returns. They use a pre-averaging technique to correct for microstructure imperfections at high frequency, and use the method of Holm (1979) to control the FWER. We slightly diverge from their approach and experiment with the MedRV estimator of integrated volatility of Andersen, Dobrev and Schaumburg (2012) and Andersen, Dobrev and Schaumburg (2014). The estimator shows good finite-sample robustness to jumps as confirmed by Dumitru and Urga (2012), making it ideal for local volatility measures using short return blocks. We obtain an average number of jumps of 4.8 per year with this approach[11], validating the very small number

[10] Jump dates for PWI are the following: 06-Feb-2006 14:53, 20-Mar-2006 09:36, 24-Mar-2006 14:17, 28-Apr-2006 09:43, 19-Sep-2006 09:45, 13-Feb-2007 09:59, 30-Mar-2007 09:42, 11-Jun-2007 11:40, 19-Jun-2007 14:31, 05-Oct-2007 14:00, 03-Dec-2007 10:56, 15-Feb-2008 09:45, 29-May-2008 10:37 and 15-Jul-2008 10:00. Full list of jump detection dates with COP estimator at 5, 10 and 20 seconds frequency is available on request.

[11] The MedRV test gives respective minimum, 25%-quantile, median, 75%-quantile and maximum statistics of 0.3, 3.0, 4.3, 5.6 and 15.3 jump detections per year. We find no significant

of jumps identified using the test of COP with our thresholding techniques. Finally, we implement the local volatility MedRV with the more liberal FDR threshold in place of the conservative technique of Holm (1979) used by Lee and Mykland (2012). This approach allows for more spurious detections and favors power rather than a strict control of the FWER. We still find less than six jumps per year in average. This shows the scarcity of jumps when a properly controlled multiple testing procedure is used.

The Monte Carlo study in the supplemental material also confirms the good properties of the underlying jump detection method and of our thresholding technique. Although the power deteriorates with diminishing jump size and sampling frequency, the simulation results are convincing. In practice, the results can be heavily influenced by different phenomena acting simultaneously. Not knowing which effect is stronger (e.g., very small jumps, microstructure noise) renders the analysis of the results even more difficult. One illustration of the difficulty to run the tests on real data is the low intersection between jumps detected by different tests. For example, Gilder (2009) shows that the methods of Andersen, Bollerslev and Dobrev (2007) and BNS agree on only 50% of detected jump days, and COP get very limited jump days in common with BNS. Part of this discrepancy is due to erroneous detections.

[Table 1]

Consequently, we need to use high frequency sampling together with a control for spurious detections for a correct inference in jump analysis. In the following, we consider a 10 seconds frequency and a FDR threshold. We always present results at the 2 minutes frequency and without thresholding for comparison purpose. As an illustration, Figure 1 shows the thresholding process for Boeing (BA) during the first six months of 2007. For each day in the sample, the points show the value of the COP statistic at the 10 seconds frequency. Dashed lines show the critical value of the individual tests, the FDR threshold, and the universal threshold. The dates selected after applying the FDR threshold are shown by asterisks, and the corresponding spurious detections are depicted by circles.

4 Dynamics of jump occurrences

In this section, we study the dynamics of jump arrivals and show how important it is to remove the spurious detections in order to obtain correct results.

difference using either MedRV or MinRV.

Since the work of Merton (1976) on the application of jump processes in option pricing, the inclusion of jumps in financial modeling has gained a lot of attention amongst academics and practitioners. The empirical literature shows that deep in-the-money, deep out-of-the-money, and shorter-maturity options tend to sell for more than their Black-Scholes price, and longer-maturity and marginally in-the-money options sell for less. Merton (1976) suggests to correct the discrepancies between market prices and the Black-Scholes value of options by including a jump component. For deep out-of-the-money call options, there is relatively little probability that the stock price exceeds the strike price prior to expiration if we exclude the possibility of jumps. However, the possibility of a jump in price significantly increases this probability, and hence, makes the option more valuable. Similarly, for deep in-the-money call options, there is little chance that the stock will decline below the exercise price prior to expiration if the underlying process is continuous. However, this event becomes non-negligible if we allow for the possibility of jumps. The phenomenon is exacerbated with short-maturity options

The widely used assumption is that jump arrival times follow a simple Poisson process, or equivalently that durations between successive jumps are independent and exponentially distributed. In the present section, we study the dynamics of jump arrivals to assess whether this assumption is realistic, or whether there is a dependency between successive jump arrivals. The jump tests of BNS and COP indicate whether one or more jumps occurred on a given day but do not give the exact number of jumps. As a result, we cannot observe the durations between successive jumps and are unable to test whether they follow an exponential distribution. For the same reason, because we do not know the probability of more than one jump in a day, we cannot use the standard methods to test whether jump occurrences are driven by a simple Poisson process. To circumvent this difficulty, we use the runs test developed by Mood (1940)[12]. As we show in the supplemental file by performing a Monte Carlo study, the runs test is a powerful method to detect clustering of jumps in time. Removing the spurious detections, e.g., with our thresholding technique, is essential in order to get a correct picture of the jumps dynamics. Table 2 reports

[12]The runs test compares the number of sequences of consecutive days with jump and without jump, or runs, against its sampling distribution under the hypothesis of random arrival. For example, a particular sequence of 10 jump tests may be represented by 0011101001, containing three runs of 1s, and three runs of 0s. In contrast, the sequence 1111100000 contains the same number of 0s and 1s, but only two runs. Too few runs indicate the presence of clustering. Too many runs indicate an oscillation. The runs test has been used in Fama (1965) to test the random walk hypothesis of stock returns. See Section 2.2.2 of Campbell, Lo and MacKinlay (1996) for details and the exact test statistic. We use the runstest function from the MATLAB Statistics Toolbox.

the results of the runs test for the 30 Dow Jones stocks over the period from January 2006 to December 2008. With no account for spurious detections, approximately 10% of the stocks reveal clustering in jumps at both 2 minutes and 10 seconds frequencies but they do not agree on the stocks because of spurious jump detections. In the 2 minutes case, American International Group, Microsoft and AT&T are rejected by the runs test, whereas Boeing and Procter & Gamble are rejected at the 10 seconds frequency. This is not surprising as COP show that the jump dates are very different depending on the estimator and the frequency considered. Additionally, many detections are spurious because of multiple testing. Applying the FDR threshold or the universal threshold in the high frequency case corrects for these errors and removes remaining rejections of the runs test. Additionally, looking at our two index proxies, the runs test indicates that neither DIA nor PWI do cluster in time. Since jump occurrences are very rare after controlling for multiple testing, we observe a strong non-rejection of the null hypothesis of no clustering. Overall, our results do not invalidate the assumption that jump are driven by a simple Poisson process.

[Table 2]

Even if we do not observe exactly the durations between successive jumps, in particular if there are many jumps within the same day, we can still estimate the parameters of the simple Poisson process that would have most likely generated the observations. If we suppose that the durations between jumps follow an exponential distribution with parameter λ, then the probability of one or more jumps occurring on a given day is $1 - e^{-\lambda}$. Hence, even if we do not observe the exact number of jumps within days, we can estimate λ as $\widehat{\lambda} = -\ln(1 - \widehat{p})$, where $\widehat{p}$ denotes the estimated probability of occurrence of a jump, obtained as the ratio of the number of days with jumps over the total number of days. We find intensities between 0.0040 and 0.0244, or equivalently average durations of 250.0 and 41.0 days.

5 Cojumps

Among other explanations, jumps in individual stocks can be due to stock-specific news or common market-level news. Market-level news can cause jumps in many stocks simultaneously, which can in theory propagate even to a diversified index. In this section, we study simultaneous jumps (cojumps) in the Dow Jones stocks and their relation to jumps in the index. We examine in detail the relation between jumps and news announcements in the next section.

Other empirical studies of cojumps include Bollerslev, Law and Tauchen (2008) who examine the relationship between jumps in a sample of forty large-cap U.S. stocks and the corresponding aggregate market index, Lahaye, Laurent and Neely (2011*a*) who investigate cojumps between stock index futures, bond futures, and exchange rates, and Dungey, McKenzie and Smith (2009) who consider simultaneous jumps across the term structure.

We define cojumps with the univariate tests as simultaneous significant jumps, i.e., occurring on the same day, rather than using the multivariate tests proposed e.g. by Bollerslev, Law and Tauchen (2008), or Jacod and Todorov (2009). We detect a jump in more than 20% of the stocks on only two occasions when applying the FDR threshold or the universal threshold (February 15, 2008 and May 29, 2008). Those two events happen during the subprime crisis, but interestingly, there appears to be no date where more than 20% of the stocks jump together during the highly volatile second semester of 2008. Considering a 2 minutes frequency with no thresholding gives a totally different picture of the joint structure of jump events. Because of the positive bias inherent to this method, the results incorrectly suggest that cojumps involving a large portion of the market constituents are frequent. For instance, jumps occurring in more than 40% of the data occur in 17% of the jump detection dates, whereas the COP estimator corrected for multiple-testing detects no single cojump affecting more than 40% of Dow Jones constituents during the three years of our study.

The same pattern is revealed when we group stocks by industry sectors. Many cojumps are identified at a 2 minutes frequency without multiple testing correction but disappear at higher frequency and with thresholding. Table 3 shows the repartition of our thirty stocks among the different Global Industry Classification Standard[13] (GICS) sectors, and Table 4 displays the number of cojumps within each sector for respectively, no account for spurious detections at 2 minutes and 10 seconds frequencies, and use of the universal threshold and FDR threshold at 10 seconds frequency. An asterisk indicates that there are significantly more cojumps than if the stocks jumped independently[14]. For all but one sectors the number of cojumps is significant at a 2 minutes frequency. However, correcting for both biases using the thresholded COP estimator at a 10 seconds frequency again confirms that cojumps are extremely rare. The data reveal no day where there are significantly more cojumps than if the stocks

[13]The Global Industry Classification Standard (GICS) is an industry taxonomy developed by Morgan Stanley Capital International (MSCI) and Standard & Poor's.

[14]Under the null hypothesis that stocks jump independently, the probability that the stocks jump simultaneously on a given day is the product of the jump probabilities of the individual stocks. The distribution of the corresponding test statistic is obtained from a simple application of the Central Limit Theorem and the Delta method.

jump independently.

[Tables 3 and 4]

Then, we investigate the relation between jumps in our index proxies and jumps in the individual Dow Jones stocks. Table 5 shows the likelihood of a jump in DIA or PWI conditional on the proportion of stocks cojumping. We find that cojumps affecting less than 10% of the stocks are very unlikely to have a market-wide impact. DIA and PWI jump in 1.1% and 1.5% of these dates when we apply the FDR threshold, respectively; the universal threshold is even more severe as it detects less than 0.5% of market jumps in this case. This is a significant result as it means that not only jumps are extremely rare events, but they also generally do not induce market-wide jumps. This is a strong argument in favor of diversifiability of the jump risk.

Table 6 displays information on the distribution of the proportion of stocks jumping simultaneously, depending on whether or not there is a jump in the index. When applying the FDR threshold, the average percentage of stocks jumping raises from 1.0% to 3.3% when there is a jump in DIA, and can reach 23.3%. With no jump in the index, the percentage of stocks jumping on the same day never exceeds 16.7%. Our results agree with the findings of Bollerslev, Law and Tauchen (2008), who report that the index jumps less often than the individual stocks, and conclude that the idiosyncratic jumps are diversified away in the aggregate portfolio. They also examine the puzzling fact that jumps in the index are uncorrelated with jumps in its constituents.

[Tables 5 and 6]

6 Relation to news releases

Having removed the spurious detections with our thresholding technique, we now investigate to which extent the few residual jumps are caused by the arrival of news. We cover three different categories of news. We first consider macroeconomic news, which can in theory explain the rare simultaneous jumps in multiple stocks. Next, we look at prescheduled announcements specific to each stock. Finally, we analyze the impact of stories from news agencies, i.e., Reuters and Dow Jones News Service. We show thanks to our methodology that the popular belief in a relationship between jumps and these three categories is not supported by high-frequency data on the period considered.

6.1 Relation to macroeconomic news release

In this section, we investigate the impact of macroeconomic announcements, which are the most likely sort of news to cause simultaneous jumps among many stocks. Our results confirm the findings of the previous section that co-jumps are a very rare event. There is a long literature on the market reaction to macroeconomic news. Cutler, Poterba and Summers (1989) estimate the fraction of the variance in aggregate stock returns that can be attributed to various kinds of news, including major political and world events. Ederington and Lee (1993) and Ederington and Lee (1996) are the first to investigate the intraday reaction of bond prices to macro announcements. More recently, Andersen, Bollerslev, Diebold and Vega (2007) show using high-frequency data that reaction times to news are very short, and Aït-Sahalia, Andritzky, Jobst, Nowak and Tamirisa (2012) examine the market response to policy initiatives during the recent financial crisis. To our knowledge, however, the only papers studying the link between jumps in assets and macroeconomic news are Dungey, McKenzie and Smith (2009), Lahaye, Laurent and Neely (2011*b*), Huang (2007) and Lee (2012). Numerous other studies which mention the relation of jumps to macroeconomic announcements merely investigate the timing of jumps to see whether an unusual pattern corresponds to a regularly scheduled news announcement.

For all announcements except the target Fed funds rate, we use the International Money Market Services (MMS) data on expected (surveyed) and realized (announced) macroeconomic fundamentals. MMS conducts a Friday telephone survey of about 40 money managers, collects forecasts of all indicators to be released during the next week, and reports the median forecasts from the survey. One of the first article to use the MMS survey data is Andersen, Bollerslev, Diebold and Vega (2003). The authors study the effect of macro announcements on U.S. dollar spot exchange rates but do not look at jumps. The target Fed funds rate forecasts are obtained from Action Economics, which also gathers estimates on economic data once a week from economists, strategists, and a few traders. We obtain the data from Haver Analytics. As of December 16, 2008, the funds target rate is a range, i.e., zero to 0.25%, rather than a specific rate. The Federal Open Market Committee (FOMC) can also surprise the market by changing the Fed funds target between scheduled meetings. In our sample, the decisions following such unscheduled meetings are always released early on the next morning and therefore do not cause jumps during market hours.

We consider only the announcements released during the trading hours. We provide the list of the release times in the supplemental file. Table 7 presents the results with the BNS estimator at 2 minutes frequency, and the COP estimator at 10 seconds frequency with respectively no thresholding, use of the universal threshold, and use of the FDR threshold. For each macroeconomic news, the table displays the number of announcement days in our sample, the average probability of a jump in individual stocks on an announcement day, the probability of a jump in the Diamonds ETF, and the probability of a jump in our PWI portfolio. The last row presents corresponding results based on all days in our sample independently on the presence or absence of news.

[Table 7]

The impact of the sampling frequency is notable. At a 2 minutes frequency, most macroeconomic news announcement types generate a significant probability of jump in stocks and also at the index level. Hence, we confirm the finding of Lee (2012) based on a 15 minutes frequency. On the contrary, the significance disappears at a 10 seconds frequency even before correcting for multiple testing. The FDR threshold and universal threshold only make this result stronger by removing most of the residual jump detections on announcement days. This result suggests that macroeconomic news generate financial reactions of the form of sudden rises of volatility, which are incorrectly interpreted as jumps when the sampling is not fine enough. In order to go further with this intuition, we define a burst of volatility as a jump that is identified at a relatively low frequency but is not at a high frequency. This concept is already put forward by COP. We consider the thresholded version of the BNS estimator at a 2 minutes frequency and compare the jump dates with the thresholded version of COP at a 10 seconds frequency. We find that the only announcement which actually increases the likelihood of a burst of volatility is the target Fed funds rate.

It may not be the act of releasing information to the market itself that is important. Rather, it may be the extent to which the actual announcement differs from the market expectation, i.e., the surprise content of each announcement, that determines whether assets jump in reaction to the information release. We capture the surprise content of the announcements using the survey data from MMS and Action Economics. To account for the discrepancies across the various news items, we compute the standardized surprise, defined as the difference between expectations and realizations, divided by the standard deviation. We do not observe any effect caused by the surprise component of macro news announcements, even if we consider separately surprises above and below expectations. The detailed analysis of the above results is available upon request.

6.2 Relation to scheduled company-specific announcements

In this section, we look at two types of scheduled company-specific announcements. First, we investigate whether dividends can cause the stock price to jump. We obtain data from COMPUSTAT and CRSP (for the declaration date). We do not observe significantly more jumps on the ex-dividend date. This result is not surprising, given that companies usually commit to a dividend policy for the long run, that the amounts are known in advance, and that dividends are settled after the bell. The likelihood of a jump increases slightly on the dividend declaration date, but this is not statistically significant. Pooling all the stocks together, the observed probability of a jump on a dividend declaration day is 2.0% against 1.2% when applying the FDR threshold.

Second, using data from I/B/E/S, we perform a similar analysis for quarterly earnings announcements. Patton and Verardo (2012) show that the beta of individual stocks increases by an economically significant amount on quarterly earnings announcement days. We do not, however, detect any effect on the likelihood of a jump in the price. This is explained by the fact that earnings are most often published outside of the trading hours. Using the Factiva database, Bagnoli, Clement and Watts (2005) find that between 2000 and 2003, only 27% of earnings announcements occur during trading hours on the major New York stock exchanges. That figure was higher in the past, i.e., 67% in the 1970s. Managers choose the release time of news strategically to minimize the impact of the news on share prices. Therefore, managers attempt to release bad news when investors have limited opportunities to act on it. Another explanation is that managers delay the release of bad news so investors anticipate it, thus mitigating the drop in stock price at the announcement itself. Finally, we focus on the 15 announcements made during trading hours, which represent only 4.2% of all the announcements for the 30 stocks during our three-year sample, and find that it does not increase the likelihood of a jump neither. The detailed analysis of the impact of dividends and quarterly earnings announcements is available upon request.

6.3 Relation to stories from news agencies

We investigate whether jumps can be explained by news stories from two major newswires, i.e., Dow Jones News Service (DJNS) and Reuters News. By examining the content of news stories, we can analyze the impact of a variety of unscheduled and uncategorized events and are not limited to a predetermined set of event types such as earnings announcements, mergers, or analyst recommendations. To our knowledge, the present study is the first to perform an extensive study of news stories and investigate whether they can cause jumps in stock prices. Previous attempts to study the impact of news are much less detailed, e.g., Cutler, Poterba and Summers (1989) or Lee and Mykland (2008) who consider a small sample of three stocks over three months. Tetlock (2007) and Tetlock, Saar-Tsechansky and Macskassy (2008) also analyze financial news stories, though their goal is to quantifying the language in an attempt to extract investor sentiment. Lee (2012) does not consider news agencies.

We access the DJNS and Reuters News newswires through Factiva. Factiva is a news database that aggregates content from thousands of leading news and business sources. Retrieving information effectively from such a huge repository is a difficult task. The perfect mix of getting everything and avoiding irrelevant or erroneous stories is difficult to achieve. The technology to automatically quantifying language content is not ripe for the scope of our study[15]. Therefore, we rely on the taxonomy applied by Factiva which provides a hierarchy of company names, industries, regions, and subjects. Such an indexing allows to narrow search results on a specific topic, or retrieve stories which are actually about a particular company, and not all the stories where the company name merely occurs.

The Factiva web interface does not allow to perform queries on the "publication time" field, and it is not possible to automate or customize queries. To circumvent this problem, we export all the news stories in XML format. We then parse the XML files and reconstruct our own database. Although the "publication time" field is not searchable using the web interface, it is encoded properly when exporting documents in XML format. As we can download the full articles with indexing, we do not loose any information. This process also allows us to perform text analysis inside the articles, and run custom searches efficiently. Keeping news published in the US only, we are left with 30,071 DJNS stories and 31,228 Reuters stories about our thirty companies during our three-year sample[16]. The stocks we consider are large multinational companies and are the subject of one or more important stories almost every day. We further eliminate irrelevant stories by selecting news published during market hours only and by requiring that the company name appears in the headline[17].

[15]Tetlock (2007) and Tetlock, Saar-Tsechansky and Macskassy (2008) are only able to construct a simple indicator of media pessimism, or look at the fraction of negative words. In the industry, the Thomson Reuters News Analytics service claims to be able to interpret news by providing sentiment analysis. However, each news is merely attributed a -1, 0, or 1 sentiment indication and no study is available on how relevant this indication really is.

[16]These numbers are obtained by using the Factiva option to remove duplicates and exclude republished news, recurring pricing and market data, and non-business stories such as obituaries, sports or calendars.

[17]The Factiva indexing system does not solve the aboutness vs occurrence issue perfectly.

This allows us to reduce the number of stories to 8,498 for DJNS and 6,520 for Reuters News, which corresponds to around one story every three days for each stock.

Having eliminated the irrelevant stories, we analyze the probability of jumps occurring on specific news types using the Factiva indexing hierarchy. We also investigate the impact of news flagged as "Dow Jones/Reuters Top Wire News" in order to capture any uncategorized and unusual story. An important proportion of the "Top Wire News" are stories about earnings. The majority of them is discarded, however, when we eliminate news released outside market hours. Table 8 presents results for a selection of news types susceptible to cause jumps. Results for further kinds of news are available upon request. As one additional precaution, we require that a particular news appears simultaneously on both the DJNS and Reuters News wires. For each news type, the first two columns indicate the total number of stories and the number of stories published during market hours. 78 percent of the announcements are made outside market hours. Once again, only by importing the news stories into our own database are we able to filter out news outside market hours automatically. The remaining columns show the conditional pooled probability of a jump on days a news is released, for each type of news. After applying the FDR threshold, the unconditional probability of a jump computed over all days and stocks is 1.4 percent. The news types for which we observe an increased probability of jumps are "Government Contracts" (4.5%), "Dividends" (2.2%) and "Divestitures / Asset Sales" (7.1%) (a subcategory of "Ownership Changes"). However, none of these increases are statistically significant, which shows that news stories are not very likely to cause jumps. Companies purposefully shift most important announcements after the bell or early in the morning in order to avoid uncontrolled investor reactions and the consequent impact on the stock price.

[Table 8]

In the preceding sections, we have investigated the likelihood that a news release causes a jump, i.e., P(jump|news). In Table 9, we report what proportion of jumps is associated with a particular type of news, i.e., P(news|jump). When searching for news on jump days, we consider the news investigated in the

For instance, an article containing a "Top Wire News" story about Microsoft and secondarily mentioning Intel will also be retrieved in a search for "Top Wire News" and Intel, although the information might be not very important for Intel. When imposing that the headline mentions the company name, we must account for the fact that one company can have different denominations. For example, Bank of America appears as BofA, Bank of Amer, or B of A.

preceding sections, i.e., macroeconomic announcements, eanrnings announcements, dividend declaration dates, and Reuters or DJNS news. The number of announcements corresponds to the total over the 30 stocks. Considering the number of stocks affected, macroeconomic news are much more frequent than other types of announcements. As a consequence, we find a macroeconomic announcement on approximately 30% of jump days. However, no single type of macroeconomic news we consider significantly increases the likelihood of a jump.

Our findings differ from the conclusions of Lee and Mykland (2008) and the results of Lee (2012) mainly for two reasons. First, we sample at a frequency of 10 seconds whereas the authors use a low 15 minutes frequency. As evidenced in COP, our empirical application, and our Monte-Carlo simulations, it is crucial to sample at a high frequency to discriminate between jumps and continuous paths subject to bursts of volatility[18]. Second, the authors keep the opening transactions, which leads them to systematically detect jumps in the first returns of a day. The opening transactions of each day are erratic and do not correspond to normal returns as they result from information accumulated over the night. As the companies under consideration are subject to news articles every day, it is not surprising that the authors are able to find a story for each day they detect a jump. If all such events would systematically induce jumps, we should observe jumps scattered across the day, and not just when the market opens.

The belief in a relationship between jumps and the three categories of news is not supported by the data on the period considered. The few remaining jumps identified in Section 3 must be explained by other unpublished market events. One possible research direction is to relate residual jumps to microstructure phenomenons. For instance, the impact of high frequency trading on the stability of financial markets is largely unknown, and algorithmic traders are regularly accused of playing a major role in flash crashes (see Kirilenko, Kyle, Samadi and Tuzun (2014) for an extensive review and analysis). Also, an ex-

[18]Our multiple testing control has double asymptotics, that is, it separates n, the number of periods in a day, from N, the number of days. Lee and Mykland (2008) control the total number of intraday tests on the whole time period, whereas we only need to control multiple testing across days. Our approach is less conservative when the sampling frequency is high or the number of intervals is large. We investigate a period of three years at a 10-second frequency, that is, more than 1.5 million test points per asset with the Lee and Mykland (2008) approach. Lee and Mykland (2008) consider a period of three months at a 15-minute frequency, or approximately two thousand test points. Because of the intrinsic link between the number of tests and the sampling frequency of their approach, the multiple testing control of Lee and Mykland (2008) is excessively conservative when applied to our data set. We would get a bound of $\sqrt{2 \log N} = 5.33$ with $N = 1.5$ million, which translates into a very conservative probability level $\alpha_N = 2(1 - \Phi(5.33)) = 9.65 \times 10^{-8}$ for their approach.

cessive demand for trading in a relatively illiquid market with imbalanced order books may cause jumps even in the absence of news. Yet, our preliminary results trying to relate jumps to illiquidity based on volume information only do not seem to show an obvious relationship. Hence, we do not report them, and we leave these aspects for further research.

[Table 9]

7 Conclusion

This paper introduces a method to eliminate spurious detections of jumps in high-frequency data via an explicit thresholding on available test statistics. Our theoretical result is the first to provide a formal treatment of the multiple testing issue when identifying jumps over a long period of time. A Monte Carlo study shows that our technique behaves well in finite sample, and illustrates the importance of removing spurious detections when investigating the dynamics of jump arrivals. Applying our method on high-frequency data for the 30 Dow Jones stocks over the three-year period between 2006 and 2008, we find that up to 90% of days selected initially as containing a jump are spurious detections. Overall, our tests do no detect time clustering phenomena of jumps arrivals, and, hence, do not reject the hypothesis that jump arrivals are driven by a simple Poisson process. We do not detect cojumps affecting all stocks simultaneously, which supports the assumption in Merton (1976) that jump risk is diversifiable. The main empirical contribution of the paper is to study the relation between jumps and information arrival. We find that scheduled macroeconomic announcements and company-specific announcements do not increase the likelihood of a jump to a statistically significant extent. Using the Factiva database, we also study the impact of Reuters and Dow Jones News Service news and find that it does not cause jumps neither. Those results show that the conjecture about jumps coming from announcements is not supported by the data. This does not mean announcements have no market impact. Our results indicate that they may induce bursts of volatility. We consider only U.S. large capitalization stocks in our empirical study. It would be interesting to investigate the jump behavior of stocks with different characteristics, especially when studying liquidity issues as an explanation for jumps.

Appendix: Proof of asymptotic control

We prove Theorem 1 with BNS for simplicity here but it can be applied to COP[19]. We can also apply it to other strategies yielding jump detection tests, such as Aït-Sahalia and Jacod (2009) in a similar way. The proof is available on request. Under the null hypothesis of no jumps, Section 3 of BNS shows that the asymptotic distribution of jump test statistics converges to independent standard normal random variables from a classical use of infill asymptotics. These standard results follow from showing asymptotic negligibility of the drift contributions and application of a CLT for triangular arrays of martingale differences to derive the joint asymptotic behaviour of the realized quadratic variation and the realized bipower variation.

For each integer $n \geq 1$, let the real-valued random variables $Y^n_{t,i}$, $1 \leq t \leq N$, $1 \leq i \leq n$, form N square integrable martingale difference sequences w.r.t. the σ-fields $\mathcal{F}^n_{t,0} \subset \mathcal{F}^n_{t,1} \subset \ldots \subset \mathcal{F}^n_{t,n}$, that is, suppose that $Y^n_{t,i}$ is measurable w.r.t. $\mathcal{F}^n_{t,i}$ with $\mathrm{E}[(Y^n_{t,1})^2] < \infty$ and $\mathrm{E}[Y^n_{t,i}|\mathcal{F}^n_{t,i}] = 0$ a.s. for all n, i and t. We apply a CLT to quantities written as $S^n_t = \sum_{i=1}^{n} Y^n_{t,i}$. In the following theorem, we show that the event that the largest and the smallest of the entries of the vector $(S^n_1, \ldots, S^n_N)$ stay within $[-\sqrt{2\log N}, \sqrt{2\log N}]$ becomes certain for large n and N. We use two conditions on higher moments, which imply the conditions to apply the CLT for triangular arrays of martingale differences when n goes to infinity, and require that N is not too large w.r.t. the asymptotics in n.

Theorem 1. *Let $S^n_t = \sum_{i=1}^{n} Y^n_{t,i}$, $1 \leq t \leq N$. If, for $0 < \gamma < \infty$,*

$$L^n_{t,2\gamma} = E\left[\sum_{i=1}^{n} |Y^n_{t,i}|^{2+2\gamma}\right] \to 0, \quad as \quad n \to \infty, \tag{1}$$

$$M^n_{t,2\gamma} = E\left[\left|\sum_{i=1}^{n} E\left[(Y^n_{t,i})^2|\mathcal{F}^n_{t,i}\right] - 1\right|^{1+\gamma}\right] \to 0, \quad as \quad n \to \infty, \tag{2}$$

[19] BNS and COP use the difference of the realized power and bipower variations in their original tests. We use the adjusted ratio statistic of BNS and the log-ratio statistic of COP in our empirical applications and Monte-Carlo simulations. These modified test statistics are advocated by Huang and Tauchen (2005) and COP as they provide better finite sample properties than the original difference statistics. They are respectively the same in the limit as we can always apply the delta method to the modified version to get an asymptotically equivalent linear test statistic (see, for example, BNS for an early application to realized power and bipower variations). Hence, Theorem 1 holds for the linear, ratio, and log-ratio test statistics of BNS and COP.

and

$$(1+\sqrt{2\log N})^{3+6\gamma} N \leq \alpha(L^n_{t,2\gamma}+M^n_{t,2\gamma})^{-1}, \tag{3}$$

with $\alpha > 0$. *Then,*

$$P\left[\sup_t |S^n_t| \leq \sqrt{2\log N}\right] \to 1, \quad \textit{as} \quad N, n \to \infty. \tag{4}$$

Proof. Conditions (1) and (2) imply the conditions of the CLT for triangular arrays of martingale differences, and we get the weak convergence of the distribution $P\left[S^n_t \leq x\right]$ to the standard normal distribution $\Phi(x)$ as $n \to \infty$. Now $P\left[\sup_t |S^n_t| \leq \sqrt{2\log N}\right] = P\left[|S^n_1| \leq \sqrt{2\log N}, \ldots, |S^n_N| \leq \sqrt{2\log N}\right] = \prod_{t=1}^{N} P\left[|S^n_t| \leq \sqrt{2\log N}\right]$ by independence. We have independence since a martingale difference sequence has no serial correlation by construction (see e.g. Hayashi (2000) p. 104 for a proof), the entries S^n_t are linear combinations of $Y^n_{t,i}$ forming martingale difference sequences, all σ-fields indexed by i and t are increasing, and the equivalence between zero correlation and independence for Gaussian vectors. From Grama (1997) Theorem 2.1, Condition (3) ensures that we can use exact bounds for the departure from normality of $P\left[S^n_t \geq \sqrt{2\log N}\right]$ and $P\left[S^n_t \leq -\sqrt{2\log N}\right]$ (see also Hauesler (1988) Theorem 2 for exact uniform bounds, and Lipster and Shiryayev (1989) Section 5.7 Theorems 1 and 2 for uniform bounds, i.e., Berry-Esseen type bounds, instead of the exact nonuniform bounds for moderate deviations that we use here). We get $\prod_{t=1}^N P\left[|S^n_t| \leq \sqrt{2\log N}\right] = \prod_{t=1}^N \left[1 - 2\Phi(-\sqrt{2\log N})\left\{1 + R_t(\alpha,\gamma,N)\right\}\right]$, where $R_t(\alpha,\gamma,N) = \theta C(\alpha,\gamma)\left\{(1+\sqrt{2\log N})^{3+6\gamma} N (L^n_{t,2\gamma}+M^n_{t,2\gamma})\right\}^{1/(3+2\gamma)}$ is the remainder term, with $|\theta| < 1$ and $C(\alpha,\gamma)$ being a constant only depending on α and γ. Using $\Phi(-\sqrt{2\log N}) \leq \phi(\sqrt{2\log N})/\sqrt{2\log N}$ with ϕ denoting the density of the standard normal distribution, we deduce the stated result from $\prod_{t=1}^N \left[1 - 2\Phi(-\sqrt{2\log N})\right] \to 1, \quad$ as $\quad N \to \infty$, and the asymptotic negligibility of the contribution of the remainder term as $N, n \to \infty$ since $R_t(\alpha,\gamma,N)$ is bounded by $\theta C(\alpha,\gamma)\alpha^{3+6\gamma}$ because of (3). □

Condition (3) is rather weak as clearly illustrated in the case of independent random variables by Grama (1997). Let $Y^n_{t,i} = \eta_{t,i}/\sqrt{n}$, where $\eta_{t,i}$ form N given independent sequences of i.i.d. random variables which satisfy $E[\eta_{t,1}] = 0$, $\mathrm{E}[(\eta_{t,1})^2] = 1$, $m_{2\gamma} = E[|\eta_{t,1}|^{2+2\gamma}] < \infty$ with $0 < \gamma < \infty$. In this case $M^n_{t,2\gamma} = 0$ and $L^n_{t,2\gamma} = n^{-\gamma} m_{2\gamma}$. Thus for standard Gaussian $\eta_{t,1}$, condition (3) is easily met for various (n, m, α, γ) since $m_{2\gamma} = \frac{(2+2\gamma)!}{2^{1+\gamma}(1+\gamma)!}$.

BNS test: We can write the linear test statistic of Barndorff-Nielsen and Shephard (2006) (based on the difference $\mu_1^{-2} BV_t^n - RV_t^n$) in the above form using $Y_{t,i}^n := (\vartheta \mu_1^{-4} \Delta_n QV_t^n)^{-1/2} \left(\mu_1^{-2} |\Delta X_{t,i}^n||\Delta X_{t,i-1}^n| - |\Delta X_{t,i}^n|^2\right)$. Since

$$|Y_{t,i}^n|^{2+2\gamma} \leq (\vartheta \mu_1^{-4} \Delta_n QV_t^n)^{-(1+\gamma)} \sum_{l=0}^{\infty} \begin{pmatrix} 2+2\gamma \\ l \end{pmatrix} \left(|\Delta X_{t,i}^n||\Delta X_{t,i-1}^n|\right)^{2(2+2\gamma-l)} |\Delta X_{t,i}^n|^{2l},$$

where $\begin{pmatrix} 2+2\gamma \\ l \end{pmatrix} = \frac{1}{l!} \prod_{k=0}^{l-1} (2+2\gamma-k)$, Condition (1) holds from the convergence of $\Delta_n^{-1/2} \Delta_n^{1-2(1+\gamma)} \sum_{i=2}^{n} \left(|\Delta X_{t,i}^n||\Delta X_{t,i-1}^n|\right)^{2(2+2\gamma-l)} |\Delta X_{t,i}^n|^{2l}$ in law to Gaussian variables for X continuous, the equality $\Delta_n^{-(1+\gamma)} = \Delta_n^{-1/2} \Delta_n^{1-2(1+\gamma)} \Delta_n^{2+\gamma-3/2}$, and $\Delta_n \to 0$. Condition (2) holds since $\sum_{i=2}^{n} \mathrm{E}\Big[(\mu_1^{-2} |\Delta X_{t,i}^n||\Delta X_{t,i-1}^n| - |\Delta X_{t,i}^n|^2)^2 | \mathcal{F}_{t,i}^n\Big]$ converges to $\int_{t-1}^{t} \sigma_s^4 ds$ (see e.g. Barndorff-Nielsen, Graversen, Jacod and Shephard (2006)). The reasoning is similar for the adjusted ratio statistic of BNS and the log-ratio test statistic of COP.

References

Abramovich, F., Benjamini, Y., Donoho, D. L. and Johnstone, I. M. (2006), 'Adapting to unknown sparsity by controlling the false discovery rate', *The Annals of Statistics* **34**(2), 584–653.

Aït-Sahalia, Y. (2002), 'Telling from discrete data whether the underlying continuousï¿½time model is a diffusion.', *Journal of Finance* **57**(5), 2075 – 2112.

Aït-Sahalia, Y. (2004), 'Disentangling diffusion from jumps', *Journal of Financial Economics* **74**, 487–528.

Aït-Sahalia, Y., Andritzky, J., Jobst, A., Nowak, S. and Tamirisa, N. (2012), 'How to stop a herd of running bears? market response to policy initiatives during the global financial crisis', *Journal of International Economics* **87**(4), 162–177.

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., Mykland, P. A. and Zhang, L. (2011), 'Ultra high frequency volatility estimation with dependent microstructure noise', *Journal of Econometrics* **160**(1), 160–175.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007), 'Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility', *Review of Economics and Statistics* **89**(4), 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C. (2003), 'Micro effects of macro announcements: Real-time price discovery in foreign exchange', *The American Economic Review* **93**(1), 38–62.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C. (2007), 'Real-time price discovery in global stock, bond and foreign exchange markets', *Journal of International Economics* **73**(2), 251 – 277.

Andersen, T. G., Bollerslev, T. and Dobrev, D. (2007), 'No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications', *Journal of Econometrics* **138**, 125–180.

Andersen, T. G., Dobrev, D. and Schaumburg, E. (2012), 'Jump-robust volatility estimation using nearest neighbor truncation', *Journal of Econometrics* **169**(1), 75–93.

Andersen, T. G., Dobrev, D. and Schaumburg, E. (2014), 'A robust neighborhood truncation approach to estimation of integrated quarticity', *Econometric Theory* **30**, 3–59.

Bagnoli, M., Clement, M. and Watts, S. G. (2005), Around-the-clock media coverage and the timing of earnings announcements, Working paper.

Bajgrowicz, P. and Scaillet, O. (2012), 'Technical trading revisited: False discoveries, persistence tests, and transaction costs', *Journal of Financial Economics* **106**(3), 473–491.

Barndorff-Neilsen, O. E. and Shephard, N. (2004), 'Power and bipower variation with stochastic volatility and jumps', *(with discussion) Journal of Financial Econometrics* **2**, 1–48. Reprinted in "Financial Risk Measurement and Management" (editor Frank X. Diebold), in Edward Elgar Publishers, 2012.

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J. and Shephard, N. (2006), 'Limit theorems for bipower variation in financial econometrics', *Econometric Theory* **22**(04), 677–719.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2009), 'Realized kernels in practice: Trades and quotes', *Econometrics Journal* **2**, C1–C32.

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

Barras, L., Scaillet, O. and Wermers, R. (2010), 'False discoveries in mutual fund performance: Measuring luck in estimated alphas', *Journal of Finance* **65**, 179–216.

Bogdan, M., Chakrabarti, A., Frommlet, F. and Ghosh, J. (2011), 'Asymptotic bayes-optimality under sparsity of some multiple testing procedures', *Annals of Statistics* **39**(3), 1551–1579.

Bollerslev, T., Law, T. H. and Tauchen, G. (2008), 'Risk, jumps, and diversification', *Journal of Econometrics* **144**, 234–256.

Bollerslev, T. and Todorov, V. (2011), 'Tails, fears, and risk premia', *Journal of Finance* **66**(6), 2165–2211.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1996), *The Econometrics of Financial Markets*, Princeton University Press.

Carr, P. and Wu, L. (2003), 'What type of process underlies options? a simple robust test', *Journal of Finance* **58**, 2581–2610.

Christensen, K., Oomen, R. and Podolskij, M. (2014), 'Fact or friction: jumps at ultra high frequency', *Journal of Financial Economics* (forthcoming).

Cutler, D., Poterba, J. and Summers, L. H. (1989), 'What moves stock prices?', *Journal of Portfolio Management* **15**, 4–12.

Dacorogna, M. M., Gencay, R., Muller, U. A., Olsen, R. B. and Pictet, O. V. (2001), *An Introduction to High-Frequency Finance*, Academic Press, San Diego.

Donoho, D. L. and Johnstone, J. (1994), 'Ideal spatial adaptation by wavelet shrinkage', *Biometrika* **81**(3), 425–455.

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

Dungey, M., McKenzie, M. and Smith, L. V. (2009), 'Empirical evidence on jumps in the term structure of the us treasury market', *Journal of Empirical Finance* **16**(3), 430 – 445.

Ederington, L. H. and Lee, J. H. (1993), 'How markets process information: News releases and volatility', *Journal of Finance* **48**(4), 1161–1191.

Ederington, L. H. and Lee, J. H. (1996), 'The creation and resolution of market uncertainty: The impact of information releases on implied volatility', *Journal of Financial and Quantitative Analysis* **31**(4), 513–539.

Fama, E. (1965), 'The behavior of stock market prices', *Journal of Business* **38**, 34–105.

Fan, J. and Wang, Y. (2007), 'Multi-scale jump and volatility analysis for high-frequency financial data.', *Journal of the American Statistics Association* **102**, 1349–1362.

Gilder, D. (2009), An empirical investigation of intraday jumps and cojumps in us equities, Working paper.

Giot, P., Laurent, S. and Petitjean, M. (2010), 'Trading activity, realized volatility and jumps', *Journal of Empirical Finance* **17**(1), 168 – 175.

Grama, I. (1997), 'On moderate deviations for martingales', *Annals of Probability* **25**, 152–183.

Hansen, P. R. and Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**(2), 127–161.

Harvey, C., Liu, Y. and Zhu, H. (2013), ...and the cross-section of expected returns, Working paper.

Hauesler, I. (1988), 'On moderate deviations for martingales', *Annals of Probability* **16**, 275–299.

Hayashi, F. (2000), *Econometrics*, Princeton University Press.

Holm, S. (1979), 'A simple sequentially rejective multiple test procedure', *Scandinavian Journal of Statistics* **6**(1), 65–70.

Huang, X. (2007), Macroeconomic news announcements, financial market volatility and jumps, Unpublished manuscript.

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

Jacod, J. and Todorov, V. (2009), 'Testing for common arrival of jumps in discretely-observed multidimensional processes', *Annals of Statistics* **37**, 1792–1938.

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

Kirilenko, A. A., Kyle, A. S., Samadi, M. and Tuzun, T. (2014), 'The flash crash: The impact of high frequency trading on an electronic market'.

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

Lahaye, J., Laurent, S. and Neely, C. J. (2011*b*), '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 lï¿$\frac{1}{2}$vy jump diffusion processes', *Journal of Financial Economics* **96**, 271–290.

Lee, S. and Mykland, P. (2012), 'Jumps in equilibrium prices and market microstructure noise', *Journal of Econometrics* **168**(1), 396–406.

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

Lehman, E. (1993), 'The fisher, neyman-pearson theories of testing hypotheses: one theory or two', *Journal of the American Statistical Association* **88**(424), 1242–1249.

Lipster, R. and Shiryayev, A. (1989), *Theory of Martingales*, Kluwer Academic Publishers, Dordrecht.

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

Merton, R. C. (1976), 'Option pricing when underlying stock returns are discontinuous', *Journal of Financial Economics* **3**, 125–144.

Mood, A. (1940), 'The distribution theory of runs', *Annals of Mathematical Statistics* **11**, 367–392.

Patton, A. J. and Verardo, M. (2012), 'Does beta move with news? firm-specific information flows and learning about profitability', *Review of Financial Studies* **25**(9), 2789–2839.

Romano, J. P., Shaikh, A. M. and Wolf, M. (2008), 'Formalized data snooping based on generalized error rates', *Econometric Theory* **24**, 404–447.

Tetlock, P. C. (2007), 'Giving content to investor sentiment: The role of media in the stock market', *Journal of Finance* **62**(3), 1139–1168.

Tetlock, P. C., Saar-Tsechansky, M. and Macskassy, S. (2008), 'More than words: Quantifying language to measure firms' fundamentals', *Journal of Finance* **63**(3), 1437–1467.

Ticker	Company name	2-minute No thresholding	10-second No thresholding	Universal threshold	FDR threshold
Dow Jones stocks:					
AA	Alcoa	73.3	16.3	2.0	1.7
AIG	American International Group	75.3	21.3	3.7	4.3
AXP	American Express	72.0	22.7	2.7	3.0
BA	Boeing	57.7	21.7	5.0	6.0
C	Citigroup	57.7	11.7	3.0	3.3
CAT	Caterpillar	59.0	15.7	2.7	3.7
DD	DuPont	81.0	19.7	4.3	5.7
DIS	Walt Disney	86.7	20.0	3.7	4.0
GE	General Electric	76.7	18.0	2.7	3.3
GM	General Motors	77.0	31.0	2.7	4.3
HD	The Home Depot	70.7	16.7	2.0	2.7
HON	Honeywell	61.7	22.0	0.3	1.3
HPQ	Hewlett-Packard	66.7	18.3	2.0	2.0
IBM	IBM	55.0	22.3	2.7	2.7
INTC	Intel	89.0	17.7	3.3	3.3
JNJ	Johnson & Johnson	68.0	20.7	3.3	3.0
JPM	JPMorgan Chase	60.0	15.0	1.0	1.0
KO	Coca-Cola	73.7	19.3	1.3	1.3
MCD	McDonald's	75.7	22.3	1.7	2.0
MMM	3M	62.7	16.7	2.0	2.7
MO	Altria Group	73.3	20.7	2.3	3.0
MRK	Merck	69.3	21.3	2.7	2.3
MSFT	Microsoft	94.0	20.0	2.7	2.0
PFE	Pfizer	93.0	17.3	2.7	2.7
PG	Procter & Gamble	61.0	19.0	2.0	1.7
T	AT&T	82.0	20.7	3.3	4.0
UTX	United Technologies Corporation	59.3	15.7	2.0	2.3
VZ	Verizon Communications	75.0	19.0	3.0	3.0
WMT	Wal-Mart	50.3	15.3	2.0	2.3
XOM	ExxonMobil	43.7	15.7	3.0	4.7
Index:					
DIA	Diamonds Trust	91.7	18.3	1.3	3.0
PWI	Price-weighted index	72.7	15.3	1.6	4.7
Summary for stocks:					
Mean		70.0	19.1	2.7	4.7
Median		71.3	19.1	2.7	3.0
Minimum		43.7	11.7	0.3	1.0
Maximum		94.0	31.0	5.0	6.0

Table 1: **Average number of jumps per year.** This table reports the average yearly number of jumps over the three-year period between 2006 and 2008, for the 30 Dow Jones stocks, the Diamonds ETF (DIA), and a price-weighted index portfolio of the 30 stocks (PWI). We use the BNS jump detection test at a 2-minute sampling frequency and the COP test at a 10-second frequency. Results for the latter are reported for respectively, no account for spurious detections, use of the universal threshold, and use of the FDR threshold.

Ticker	Company name	2-minute No thresholding	10-second No thresholding	 Universal threshold	 FDR threshold
Dow Jones stocks:					
AA	Alcoa	0.93	1.00	1.00	1.00
AIG	American International Group	0.00*	0.63	1.00	1.00
AXP	American Express	0.14	1.00	1.00	1.00
BA	Boeing	0.89	0.03*	1.00	1.00
C	Citigroup	0.08	0.34	1.00	1.00
CAT	Caterpillar	0.49	0.35	1.00	1.00
DD	DuPont	0.65	0.61	1.00	1.00
DIS	Walt Disney	0.40	1.00	1.00	1.00
GE	General Electric	0.40	0.44	1.00	1.00
GM	General Motors	0.09	0.71	1.00	1.00
HD	The Home Depot	0.15	0.45	1.00	1.00
HON	Honeywell	0.14	0.70	1.00	1.00
HPQ	Hewlett-Packard	0.55	0.73	1.00	1.00
IBM	IBM	0.51	1.00	1.00	1.00
INTC	Intel	0.23	1.00	1.00	1.00
JNJ	Johnson & Johnson	0.36	0.24	1.00	1.00
JPM	JPMorgan Chase	0.29	0.43	1.00	1.00
KO	Coca-Cola	0.45	0.29	1.00	1.00
MCD	McDonald's	0.52	0.05	1.00	1.00
MMM	3M	0.70	0.84	1.00	1.00
MO	Altria Group	0.16	0.77	1.00	1.00
MRK	Merck	0.92	1.00	1.00	1.00
MSFT	Microsoft	0.00*	0.50	1.00	1.00
PFE	Pfizer	0.12	0.20	1.00	1.00
PG	Procter & Gamble	0.49	0.01*	1.00	1.00
T	AT&T	0.02*	1.00	1.00	1.00
UTX	United Technologies Corporation	0.17	0.35	1.00	1.00
VZ	Verizon Communications	0.49	0.50	1.00	1.00
WMT	Wal-Mart	0.47	0.60	1.00	1.00
XOM	ExxonMobil	0.35	0.66	1.00	1.00
Index:					
DIA	Diamonds Trust	0.10	0.14	1.00	1.00
PWI	Price-weighted index	0.21	1.00	1.00	1.00
Summary for stocks:					
Percentage of stocks with clustering		10.0	6.7	0.0	0.0

Table 2: **Runs tests.** This table reports p-values from the runs test of the Null hypothesis that jump arrivals do not cluster in time. Results are presented for the 30 Dow Jones stocks, the Diamonds ETF (DIA), and a price-weighted index portfolio of the 30 stocks (PWI), over the three-year period between 2006 and 2008. We use the BNS jump detection test at a 2-minute sampling frequency and the COP test at a 10-second frequency. Results for the latter are reported for respectively, no account for spurious detections, use of the universal threshold, and use of the FDR threshold.

Industry sector	Dow Jones constituents
Energy	ExxonMobil
Materials	Alcoa, DuPont
Industrials	Boeing, Caterpillar, General Electric, Honeywell, 3M, United Technologies Corporation
Consumer Discretionary	Walt Disney, General Motors, The Home Depot, McDonald's
Consumer Staples	Coca-Cola, Altria Group, Procter & Gamble, Wal-Mart
Health Care	Johnson & Johnson, Merck, Pfizer
Financials	American International Group, American Express, Citigroup, JPMorgan Chase
Information Technology	Hewlett-Packard, IBM, Intel, Microsoft
Telecommunication Services	AT&T, Verizon Communications

Table 3: **Global Industry Classification Standard industry sectors.** This table shows the repartition of the 30 Dow Jones stocks among the different sectors.

		Number of cojumps			
		2-minute	10-second		
Sector	Nb stocks in sector	No thresholding	No thresholding	Universal threshold	FDR threshold
Materials	2	76*	10*	1	1
Industrials	6	3*	0	0	0
Consumer Discretionary	4	9*	0	0	0
Consumer Staples	4	3	0	0	0
Health Care	3	31*	1	0	0
Financials	4	10*	1	0	1
Information Technology	4	18*	0	0	0
Telecom. Services	2	97*	12*	0	0

Table 4: **Cojumps within industry sectors.** This table shows the number of cojumps within each industry sector. We use the BNS jump detection test at a 2-minute sampling frequency and the COP test at a 10-second frequency. Results for the latter are reported for respectively, no account for spurious detections, use of the universal threshold, and use of the FDR threshold. An asterisk indicates that there are significantly more cojumps than if the stocks were independent.

Proportion of stocks jumping simultaneously	Number of occurrences	P(jump in DIA) (%)	P(jump in PWI) (%)
2-minute, no thresholding:			
0–10%	20	5.0	0.0
10–20%	128	15.6	10.9
20–40%	469	36.0	27.3
40–60%	121	63.6	55.4
60–80%	8	87.5	100.0
80–100%	1	100.0	100.0
10-second, no thresholding:			
0–10%	438	3.2	1.6
10–20%	273	11.7	10.6
20–40%	35	22.9	25.7
40–60%	0	-	-
60–80%	1	100.0	100.0
80–100%	0	-	-
10-second, universal threshold:			
0–10%	736	0.4	0.3
10–20%	9	0.0	11.1
20–40%	2	50.0	100.0
40–60%	0	-	-
60–80%	0	-	-
80–100%	0	-	-
10-second, FDR threshold:			
0–10%	733	1.1	1.5
10–20%	12	0.0	8.3
20–40%	2	50.0	100.0
40–60%	0	-	-
60–80%	0	-	-
80–100%	0	-	-

Table 5: **Probability of a jump in the index conditional on the proportion of stocks jumping simultaneously.** This table shows the probability of a jump in the Diamonds ETF (DIA), and in the price-weighted index portfolio of the 30 Dow Jones stocks (PWI), conditional on the proportion of stocks jumping simultaneously. We use the BNS jump detection test at a 2-minute sampling frequency and the COP test at a 10-second frequency. Results for the latter are reported for respectively, no account for spurious detections, use of the universal threshold, and use of the FDR threshold.

Proportion of stocks jumping simultaneously	Jump in DIA: No	Jump in DIA: Yes	Jump in PWI: No	Jump in PWI: Yes
2-minute, no thresholding:				
Mean (%)	25.1	33.4	25.3	35.0
Median (%)	23.3	30.0	26.7	33.3
Maximum (%)	60.0	93.3	53.3	93.3
10-second, no thresholding:				
Mean (%)	7.2	13.6	7.2	15.3
Median (%)	6.7	13.3	6.7	13.3
Maximum (%)	36.7	63.3	33.3	63.3
10-second, universal threshold:				
Mean (%)	0.8	3.0	0.8	4.2
Median (%)	0.0	3.3	0.0	3.3
Maximum (%)	16.7	20.0	13.3	20.0
10-second, FDR threshold:				
Mean (%)	1.0	3.3	1.0	4.3
Median (%)	0.0	3.3	0.0	3.3
Maximum (%)	16.7	23.3	13.3	23.3

Table 6: **Proportion of stocks jumping simultaneously conditional on a jump in the index.** This table displays the distribution of the proportion of stocks jumping simultaneously, depending on whether or not we detect a jump in the index. Results are reported for the Diamonds ETF (DIA), and the price-weighted index portfolio of the 30 Dow Jones stocks (PWI). We use the BNS jump detection test at a 2-minute sampling frequency and the COP test at a 10-second frequency. Results for the latter are reported for respectively, no account for spurious detections, use of the universal threshold, and use of the FDR threshold.

Announcement	Nb of ann.	P(jump in stocks)	P(jump in DIA)	P(jump in PWI)
2-minute, no thresholding:				
Consumer credit	35	28.6 (7.6)	42.9 (8.4)	34.3 (8.0)
Construction spending	35	27.9 (7.6)	20.0 (6.8)	11.4 (5.4)
Factory orders	35	29.7 (7.7)	40.0 (8.3)	31.4 (7.8)
Business inventories	36	28.7 (7.5)	41.7 (8.2)	25.0 (7.2)
Government budget deficit	36	25.3 (7.2)	36.1 (8.0)	33.3 (7.9)
Consumer confidence index	36	24.6 (7.2)	33.3 (7.9)	27.8 (7.5)
ISM manufacturing composite index	35	28.4 (7.6)	34.3 (8.0)	25.7 (7.4)
Target federal funds rate	23	29.3 (9.5)	34.8 (9.9)	34.8 (9.9)
All days	747	28.1 (1.6)	36.9 (1.8)	29.2 (1.7)
10-second, no thresholding:				
Consumer credit	35	6.0 (4.0)	11.4 (5.4)	2.9 (2.8)
Construction spending	35	6.9 (4.3)	5.7 (3.9)	8.6 (4.7)
Factory orders	35	7.6 (4.5)	2.9 (2.8)	5.7 (3.9)
Business inventories	36	9.1 (4.8)	5.6 (3.8)	11.1 (5.2)
Government budget deficit	36	7.0 (4.3)	2.8 (2.7)	-
ISM manufacturing composite index	35	8.1 (4.6)	8.6 (4.7)	8.6 (4.7)
Target federal funds rate	23	3.9 (4.0)	-	-
All days	747	7.7 (1.0)	7.4 (1.0)	6.2 (0.9)
10-second, universal threshold:				
Consumer credit	35	0.2 (0.7)	-	2.9 (2.8)
Construction spending	35	1.0 (1.7)	-	-
Factory orders	35	1.1 (1.8)	-	-
Business inventories	36	1.4 (2.0)	-	2.8 (2.7)
Government budget deficit	36	1.0 (1.7)	-	-
Consumer confidence index	36	0.9 (1.6)	-	-
ISM manufacturing composite index	35	1.0 (1.7)	2.9 (2.8)	-
Target federal funds rate	23	0.4 (1.4)	-	-
All days	747	1.0 (0.4)	0.5 (0.3)	0.7 (0.3)
10-second, FDR threshold:				
Consumer credit	35	0.3 (0.9)	-	2.9 (2.8)
Construction spending	35	1.3 (1.9)	-	2.9 (2.8)
Factory orders	35	1.3 (1.9)	2.9 (2.8)	-
Business inventories	36	1.3 (1.9)	-	2.8 (2.7)
Government budget deficit	36	1.5 (2.0)	-	-
Consumer confidence index	36	0.9 (1.6)	-	-
ISM manufacturing composite index	35	1.2 (1.9)	5.7 (3.9)	2.9 (2.8)
Target federal funds rate	23	0.4 (1.4)	-	-
All days	747	1.2 (0.4)	1.2 (0.4)	1.9 (0.5)

Table 7: **Probability of a jump on macroeconomic news announcements.** This table presents the probability (%) of a jump on announcement days for the 30 Dow Jones stocks, the Diamonds ETF (DIA), and the price-weighted index portfolio of the 30 Dow Jones constituents (PWI). We use the BNS jump detection test at a 2-minute sampling frequency and the COP test at a 10-second frequency. Results for the latter are reported for respectively, no account for spurious detections, use of the universal threshold, and use of the FDR threshold. An asterisk indicates that the likelihood of a jump is significantly larger after a news is released than on days with no news of the same type. Numbers in parentheses correspond to standard deviations.

	Number of news		Conditional probability of jump in stocks			
			2-minute	10-second		
News subject	All	During market hours	No thresholding	No thresholding	Universal threshold	FDR threshold
Dow Jones/Reuters Top Wire News	949	196	46.6* (3.6)	12.4 (2.4)	1.0 (0.7)	1.1 (0.7)
Government Contracts	44	14	34.1 (12.7)	6.8 (6.7)	4.5 (5.6)	4.5 (5.6)
Non-governmental Contracts	125	40	6.9 (4.0)	1.3 (1.8)	0.4 (1.1)	0.4 (1.1)
Dividends	82	31	20.0 (7.2)	15.6 (6.5)	2.2 (2.6)	2.2 (2.6)
Ownership Changes	807	173	28.4 (3.4)	5.1 (1.7)	0.7 (0.7)	0.7 (0.7)
- Acquisitions/Mergers/Takeovers	660	148	30.2 (3.8)	4.3 (1.7)	1.0 (0.8)	1.0 (0.8)
- Divestitures/Asset Sales	48	9	50.0* (16.7)	28.6 (15.1)	7.1 (8.6)	7.1 (8.6)
Share Capital	53	10	33.3 (14.9)	-	-	-
- Share Buybacks	35	7	42.9* (18.7)	-	-	-
Corporate Crime/Legal/Judicial	167	57	48.4* (6.6)	11.1 (4.2)	-	-
- Insider Dealing	3	2	-	-	-	-
Corporate Credit Ratings	34	12	14.6 (10.2)	4.2 (5.8)	-	-
Management Moves	243	79	22.4 (4.7)	7.4 (3.0)	0.3 (0.6)	0.3 (0.6)
Sales Figures	58	13	30.8 (12.8)	7.7 (7.4)	-	-
Earnings Projections	693	109	25.3 (4.2)	1.9 (1.3)	0.4 (0.6)	0.5 (0.7)
Analyst Comment/Recommendation	65	11	7.1 (7.8)	-	-	-
All days	747	747	28.1 (1.6)	7.7 (1.0)	1.0 (0.4)	1.2 (0.4)

Table 8: **Probability of a jump on Reuters/DJNS news.** This table displays the probability of a jump in stocks on different types of news appearing on the Reuters and DNS news wires. We use the BNS jump detection test at a 2-minute sampling frequency and the COP test at a 10-second frequency. Results for the latter are reported for respectively, no account for spurious detections, use of the universal threshold, and use of the FDR threshold. An asterisk indicates that the likelihood of a jump is significantly larger after a news is released than on days with no news of the same type. Numbers in parentheses correspond to standard deviations.

Announcement	Nb. of ann. days	P(news if jump in stocks)	P(news if jump in DIA)	P(news if jump in PWI)
2-minute, no thresholding:				
Dividends declaration	329	1.2 (0.1)	-	-
Earnings announcements	15	0.1 (0.0)	-	-
Macroeconomic news	$30\times236^{\star}$	31.4 (0.6)	32.4 (2.8)	32.6 (3.2)
Reuters/DJNS news	629	2.8 (0.2)	-	-
All news	$1209^{\star}$	33.9 (0.6)	-	-
10-second, no thresholding:				
Dividends declaration	329	2.3 (0.4)	-	-
Earnings announcements	15	0.1 (0.1)	-	-
Macroeconomic news	236	29.1 (1.1)	25.5 (5.9)	23.9 (6.3)
Reuters/DJNS news	629	3.1 (0.4)	-	-
All news	1209	32.4 (1.1)	-	-
10-second, universal threshold:				
Dividends declaration	329	2.1 (0.9)	-	-
Earnings announcements	15	0.0 (0.0)	-	-
Macroeconomic news	236	27.5 (2.9)	25.0 (21.7)	40.0 (21.9)
Reuters/DJNS news	629	5.2 (1.4)	-	-
All news	1209	32.6 (3.1)	-	-
10-second, FDR threshold:				
Dividends declaration	329	2.6 (1.0)	-	-
Earnings announcements	15	0.0 (0.0)	-	-
Macroeconomic news	236	27.6 (2.7)	33.3 (15.7)	28.6 (12.1)
Reuters/DJNS news	629	4.9 (1.3)	-	-
All news	1209	32.5 (2.9)	-	-

Table 9: **Probability (%) of finding a news story on jump days.** This table displays the probability that a jump is caused by one of the announcements we consider. The table shows the number of announcement days, and the probability of finding a news explaining jumps in the 30 Dow Jones stocks, the Diamonds ETF (DIA), and the price-weighted index portfolio of the 30 Dow Jones constituents (PWI). We use the BNS jump detection test at a 2-minute sampling frequency and the COP test at a 10-second frequency. Results for the latter are reported for respectively, no account for spurious detections, use of the universal threshold, and use of the FDR threshold. A star highlights that, except for macroeconomic news, the number of announcements corresponds to the total over the 30 stocks. An asterisk indicates that the likelihood of a news is significantly larger on days with jump than on days with no jump. Numbers in parentheses correspond to standard deviations.

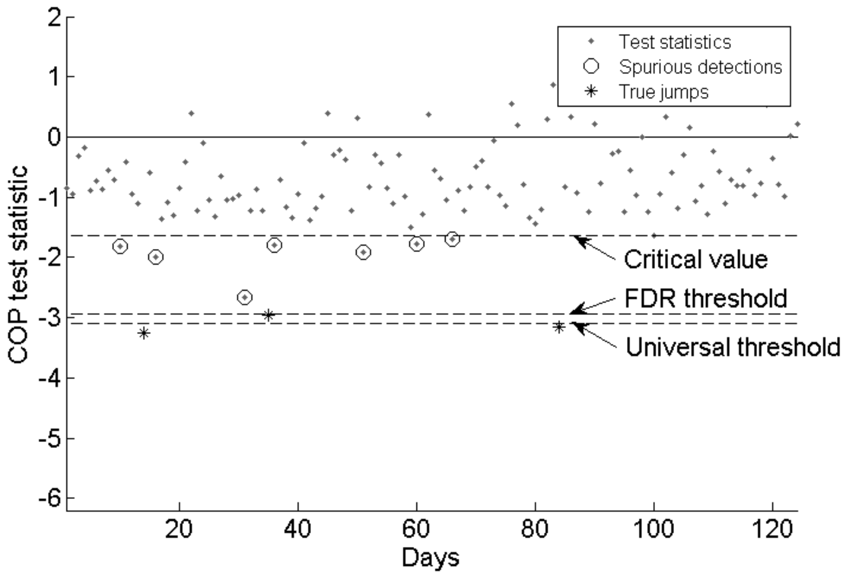

Figure 1: **Thresholding test statistics.** This figure displays daily COP test statistics (points) for Boeing, over the period between January and June 2007, using a 10-second sampling frequency. The spurious detections and the true jumps identified when applying the FDR threshold are depicted respectively by circles and asterisks.