Finance Research Letters 22 (2017) 42–48

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Identifying events in financial time series – A new approach with bipower variation

György Andor, András Bohák*

Department of Finance, Budapest University of Technology and Economics, 1117 Budapest, Magyar tudosok korutja 2, Hungary

ARTICLE INFO

Article history:
Received 13 October 2016
Accepted 3 November 2016
Available online 4 November 2016

Keywords:
Jump detection
Event detection
Realized volatility
Bipower variation

ABSTRACT

We present a statistical test to identify significant events in financial price time series. In contrast to "jumps," we define "events" as non-instantaneous, but nevertheless unusually fast and large, price changes. We show that non-parametric tests perform badly in detecting events so defined. We propose a new approach to explore the dependence of jump detection statistics on the sampling method used and find that our method improves the event detection rate of the standard test by a factor of three.

1. Introduction

The detection of discontinuities or jumps in the evolution of various financial variables (asset prices, interest rates, etc.) has attracted substantial interest in recent econometric literature. The reason behind the interest in ex-post identification of jumps is driven by multiple research fields. The first tries to extend the standard, purely continuous semi-martingale models with jump processes to better understand volatility and potentially improve the accuracy of option pricing (Aït-Sahalia, 2004) and realized volatility forecasting (Xingguo et al., 2016). The second area of research is using event studies as a tool for analyzing a broad range of financial phenomena like short-sale constraints (Yeh and Chen, 2014), informed trades around possible private information events (Ormos and Timotity, 2016) and intraday liquidity dynamics (Boudt and Petitjean, 2014). (De Moor and Sercu, 2015) show that extreme observations can bias the average return calculation and this bias affects small stocks more in CAPM alpha estimation.

All jump detection algorithms belong to one of two categories: parametric methods (e.g. Zawadowski et al. (2004)), which use user-defined parameters to define what qualifies as an "event" and non-parametric methods, where parameter definition is not necessary (e.g. the bipower variation (BPV) test proposed in Barndorff-Nielsen and Shephard (2006)). Parametric methods are not optimal for event studies, since the set of events identified and thus the results of the whole analysis are frequently very sensitive to the input parameters and it is hard to justify any particular parametrization.

In our paper, we extend the standard definition of "jumps" to include "events": unusually large price changes that occur unusually quickly, but not in an instant. To illustrate this point, in Fig. 1, we plot the price (sampled every minute) of Wal-Mart Stores Inc. (WMT) on the three business days from the 21st to 23rd of October 2006.

Identifying price jumps can have many objectives, but whatever the purpose, the event seen in Fig. 1 should definitely be detected. This jump, however, is clearly not instantaneous. This observation leads to a very loose definition of the term event in contrast to the precise mathematical definition of jumps. We call an event any window of several minutes (even up

* Corresponding author. Fax: +36 1 463 2745.
E-mail addresses: andor@finance.bme.hu (G. Andor), bohak@finance.bme.hu (A. Bohák).

http://dx.doi.org/10.1016/j.frl.2016.11.003
1544-6123/

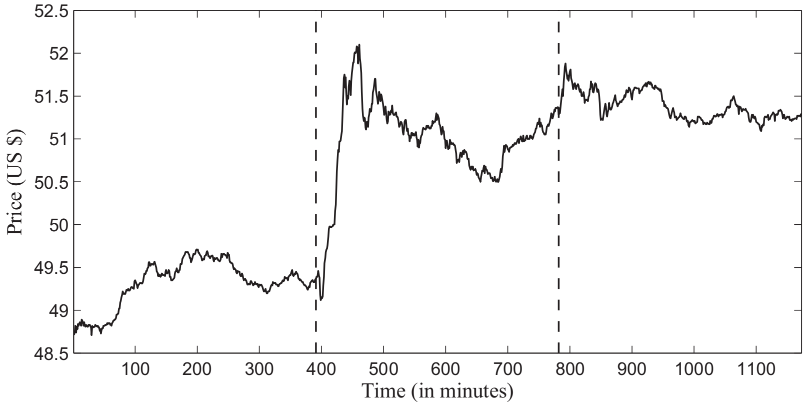

Fig. 1. Price of WMT stock on three business days between the 21st and 23rd of October 2006, clearly showing an event at the beginning of the second trading day. Black dashed lines signal day boundaries.

to an hour) where the change in price is "unusually high". We believe that this extension is crucial for event studies, since in the real world, even very fast price changes have some time duration. The main contribution of this paper is to modify non-parametric tests designed to identify jumps in a way which significantly improves their performance in identifying non-instantaneous events.

The main challenge we had to overcome was to prove that our proposed test is indeed superior to the ones already used in the literature. An obvious methodology to do this is a Monte Carlo analysis, where the time and size of the artificial price change are known and we can test if this jump is found by the algorithm or not (see Huang and Tauchen (2005) and the references therein). In one of the sections below, we show that our test outperforms the classic BPV test in terms of both size and power.

While Monte Carlo analysis is abundant in the literature, true empirical studies are limited at best and non-existent at worst. The problem with real data is that we do not know if an event occurred on a particular day, so to assess the power of an event detection algorithm, we have to find a benchmark method to compare to. However, if we accept that the benchmark method is correct, there is no point to test any algorithm against it: we could just use the benchmark method to identify events.

In our empirical analysis, we decided to start with an extensive dataset of almost a million trading days and use a parametric method with extreme parameters to define our benchmark event set. We do not believe that this benchmark method is ideal – or even acceptable: the parameters are chosen in a way that most of what we would like to call an event is probably not picked up by the algorithm. The events identified, however, are almost surely significant enough that any event detection algorithm should identify them (just like the one in Fig. 1). We show that our jump detection test significantly outperforms the original in this setup.

2. Theoretical framework and application problems

We use the framework of Barndorff-Nielsen and Shephard (2006) for our new event identification methodology. Their original test is based on the difference between the realized variance and the realized bipower variance of the returns within a trading day. The idea behind the test is that realized variance is increased by the jump component of the price movements, while bipower variation is not. The days with a significant jump component can thus be identified with a test statistic based on the appropriately scaled observed difference between the two. The family of test statistics proposed in the original article was extensively analyzed by Huang and Tauchen (2005). We use only the best performing member of the family, calculating a Z-score based on the realized variance (RV), the realized BPV and the realized tripower quarticity ($RTPQ$):

$$Z_t = \frac{(RV_t - BPV_t)/RV_t}{\sqrt{c\frac{1}{M}\max\left(1, \frac{RTPQ_t}{BPV_t^2}\right)}}$$

where c is a scaling constant.

The test statistic Z_t follows $N(0;1)$ in the case where there is no jump in the interval $[0,t]$. In the following empirical analysis, we will calculate the test statistic for each trading day of each stock separately. We refer to this method as the bipower variation test, or BPV test and we show that the choice of sampling frequency has a dramatic effect on its results.

The upper panel of Fig. 2 shows the price evolution of an event, constructed using a simple driftless geometric Brownian motion (GBM) with one-minute $\sigma = 0.11\%$ for one trading day (390 minutes) plus a non-instantaneous jump, i.e., a significant increase in the price from minute 189 to minute 199 at a rate of 0.44%/min. This could correspond to a surprising

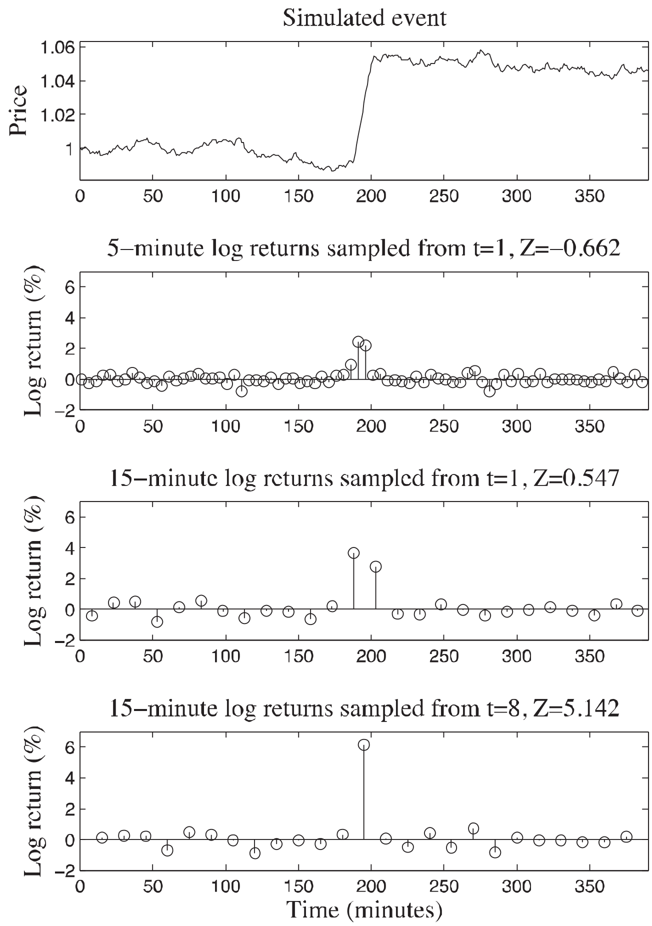

Fig. 2. A simulated event (top) and the corresponding log returns for 5-minute and 15-minute sampling frequencies, the latter for two different sampling start points.

news announcement at minute 188, with the market requiring 10 minutes to fully absorb the new information. The second panel of Fig. 2 shows the realized 5-minute log-returns associated with the same event. A Z-score of -0.806 means that the BPV test would not identify this event as a jump. The result is not much better using the 15-minute sampling interval, even though in this case the sampling interval exceeds the time required for the market to fully respond to the event. If we calculate the first log return for the first 15 minutes of the day (as is the natural way to sample), the sampling cuts the event into two consecutive jumps, which undermines the performance of the BPV test, rendering it again unable to identify the event. In the final panel, we use the same sampling interval (15-minute) but begin the sampling at minute 8 (instead of 1). This ensures that the sampling does not cut the event in half and leaves only one large outlier log return in the series. As we expected, the BPV test produces a very significant Z-score of 5.351, which clearly identifies the event. We can intuitively conclude from Fig. 2 that the Z-score of the BPV statistic might strongly depend on the start time of the sampling. Return staggering and nearest neighbor truncation suggested in (Andersen et al., 2012) would not have helped much at the 5-minute sampling interval, but could have corrected the bias at the 15-minute interval. Nevertheless, if the event lasts just a bit longer than the sampling interval, staggering will not fix the bias of the bipower variation.

To measure the dependence of the BPV test on the start time of the sampling, using actual data, we modify the volatility signature plot of Andersen et al. (2000) and calculate the RBPV statistic at different sampling intervals for all the different possible starting points in each interval, thus creating a "RBPV signature plot." Fig. 3 shows the results for a randomly chosen trading day without an event (top) and for WMT on 23rd October 2006 (clearly an event-day) (see Fig. 1 too, bottom). The plot is more than unsettling: the value of the RBPV statistic not only depends on the sampling interval but also very significantly on the start minute of the subsampling. This suggests that if the event is cut into two or more pieces (as is often the case), the BPV test will not pick it up.

3. The multi-sample BPV test

Our new event detection method is based on a natural desire to overcome the start-time dependency of the BPV test. We achieve this by calculating the maximum of the Z statistics for each sampling interval in a day. To avoid microstructure bias,

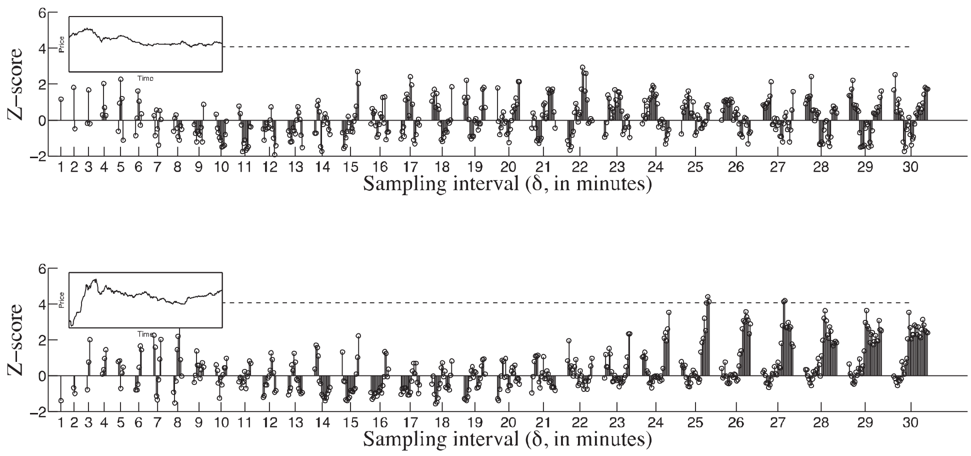

Fig. 3. RBPV signature plot of a random trading day without an event (top) and with a visually apparent event (bottom). The small figures within each plot show the price evolution for the given trading day. Note that there are δ different z-scores for each sampling frequency, each corresponding to a different sampling start point.

we do not use sampling intervals of less than 10 minutes, but calculate the maximum RBPV statistics for all other sampling intervals up to 30 minutes. If the Z-scores are independent, we get

$$\alpha = P(max(Z) > Z_{crit}) = 1 - [P(Z_i > Z_{crit})]^n$$

The appropriate critical value for each sampling interval is then

$$Z_{crit} = \Phi^{-1}(\sqrt[n]{1-\alpha})$$

where n is the number of different Z-scores (i.e. one for every starting point-frequency pair) in minutes and $_\alpha$ is the usual level of significance. Since these Z-scores are calculated for the same trading day, there is some dependence structure between them, but accounting for this would drive the critical value only lower, making the test less conservative. We show in our Results section, that the performance of our test is very good even with this limitation. In Fig. 3, we indicated the critical value with a dashed line.

Returning to the day with an event and looking at the corresponding RBPV signature plot for that day in the bottom part of Fig. 3, it is clear that if we used a sampling frequency of anything between 10 and 23 minutes, the original BPV test would not have identified this day as a jump-day (neither of the calculated statistics is even close to the critical value). Even if we sample less often, the rejection of the no-jump null depends on the start minute of the sampling: only a few calculated Z-scores exceed the critical value. Nevertheless, the multi-sampling BPV test does (correctly) identify this as a jump-day, since some of the Z-scores are above the corresponding critical value.

The bottom part of Fig. 3 supports all our conclusions from Section 2. It is clear that for the shorter sampling intervals, the sampling cuts the event into pieces leading to small test statistic values. As the sampling interval increases, the Z-score also increases and exceeds the critical value at the 25-minute interval for 3 different starting points. As the sampling interval increases further, the "large" Z-scores become increasingly pronounced. This suggests that the event lasted around 24-25 minutes. This is a new result inferred from the RBPV signature plot: we now have an estimate of the length of the identified event.

Overall, the multi-sample BPV test works as follows:

1. To determine whether there was an event on a particular trading day for a particular stock, we calculate the RBPV statistic for every sampling interval between 10 and 30 minutes (in 1 minute steps) and each possible sampling starting point.
2. We take the maximum of the calculated Z-scores for each sampling interval and compare it to the appropriate critical value.
3. If the maximal Z-score is greater than the critical value for any sampling interval (as seen, e.g., for the 25 and 27 sampling interval in the bottom part of Fig. 3), we reject the null hypothesis of no-event.

If the null hypothesis is rejected at any sampling interval, we assume that the event was not cut into pieces for at least one of the sampling starting points, meaning that we can expect that all longer sampling intervals will yield the same result. However, analysis of the RBPV signature plots for many jump days shows that this is usually not the case: for the vast majority of the events, there are only a few sampling intervals where the Z-score exceeds the threshold. This behavior is caused by the statistical nature of the BPV test and the inherent noise in the return data.

Table 1
Detection rates of the standard and the multi-sample statistic for jumps and events in a Monte Carlo simulation. The first column indicates the size of the artificial jump/event, expressed in the number of daily standard deviations of the underlying Brownian motion.

	Instantaneous jumps		Non-instantaneous events	
Relative size	Standard BPV test	Multi-sample BPV test	Standard BPV test	Multi-sample BPV test
0.2	2.7%	1.1%	1.5%	1.4%
0.4	7.3%	2.3%	1.9%	1.8%
0.6	27.9%	11.3%	3.6%	5.8%
0.8	68.6%	31.6%	4.3%	21.1%
1	89.3%	55.9%	7.9%	40.3%
1.2	97.8%	77.9%	8.9%	61.2%
1.4	99.1%	91.0%	10.0%	80.2%
1.6	99.6%	94.6%	11.9%	90.9%
1.8	99.9%	98.5%	11.9%	95.0%
2	100.0%	99.0%	12.2%	97.0%
2.2	100.0%	99.6%	13.4%	98.4%
2.4	100.0%	99.9%	13.5%	99.0%
2.6	100.0%	99.9%	13.8%	98.6%
2.8	100.0%	100.0%	14.8%	99.4%
3	100.0%	100.0%	15.5%	99.8%
3.2	100.0%	100.0%	15.3%	100.0%
3.4	100.0%	100.0%	16.8%	99.7%
3.6	100.0%	100.0%	16.5%	99.7%
3.8	100.0%	100.0%	15.3%	99.9%
4	100.0%	100.0%	18.5%	100.0%

4. Performance of the multi-sample test – a Monte Carlo analysis

We analyze the performance of the multi-sample BPV test under a simple driftless GBM model. First, we generate 10,000 paths with daily volatility $\sigma_{\text{daily}} = 1\%$, where each path contains 390 observations of the price process (this corresponds to the number of minutes in a trading day). First, we add instantaneous jumps to the generated paths. We use one jump for each trading day and choose the time of the jump from a uniform distribution across the trading day (we make sure that the jump occurs after the second and before the second-to-last price observation). We use the daily Brownian motion volatility as a numeraire for the size of jumps and the significance level at $\alpha = 1\%$. The percentages of simulated days where the non-jump null was rejected are shown in the second and third columns of Table 1, for the original and the multi-sample BPV test, respectively.

It is clear that both methods perform well in identifying large jumps, but the main advantage of the multi-sample test lies in its ability to detect non-instantaneous events. In the next part of our analysis, we use the same geometric Brownian paths, but instead of adding instantaneous jumps, we give the jumps some duration. The length of the events is drawn from a normal distribution with an 8-minute mean and a 3-minute standard deviation (our studies not reported here show that the duration distribution does not significantly affect the performance of either test). This change significantly affects the performance of the original BPV test; the detection rate remains less than 20%, even for events four times the size of the daily price variation. Our modified statistic, by contrast, is almost unaffected, showing the same detection rates as for the jump case, for events larger than two times the daily volatility. Concluding our Monte Carlo analysis, we find that the multi-sample test outperforms the standard method, in terms of both size and power, in identifying events.

5. Empirical analysis

5.1. Data

We use tick data of 816 stocks from the NYSE Trades and Quotes (TAQ) database from 1st May 2004 to 31st March 2009 to conduct our empirical analysis. We exclude trading days with missing data, typically due to reporting errors or early closing of the exchange. We exclude trading days having smaller than 0.3 bps bipower variation since these tended to be days with only a few trades, distorting our results. In the end, we are left with 1218 complete trading days for the 816 stocks, summing to a panel of almost 1 million days of 390 one-minute log returns. This means that we have the luxury of analyzing only the most significant events and still have a sample size large enough to test statistical methods.

5.2. Benchmark method

To conduct a meaningful empirical study, we need a benchmark event detection mechanism to evaluate the performance of the multi-sample BPV test. We used a modified version of the parametric method proposed by Zawadowski et al. (2004). The method has two parameters: window size and event threshold. Any return within the window size that is larger than

Table 2
Detection rate of the standard and the multi-sample test compared to different benchmarks. The benchmark event day sets were defined using the parametric event detection method of Zawadowski et al. (2004), with different threshold and window size parameters. Thresholds are expressed in terms of daily standard deviation to be comparable to the event sizes in Table 1. Window sizes are in minutes.

Window size	Threshold	Number of event days in the benchmark set	Detection rate	
			Standard BPV test	Multi-sample BPV test
10	1.12	50,602	14.1%	31.6%
10	1.28	32,967	15.5%	36.8%
10	1.44	22,279	16.8%	41.5%
10	1.60	15,823	18.1%	45.6%
20	1.59	26,751	15.0%	41.4%
20	1.81	18,010	16.3%	46.2%
20	2.04	12,503	17.8%	50.5%
20	2.26	9001	18.8%	54.2%
30	1.94	15,667	13.9%	44.3%
30	2.22	10,905	15.1%	48.5%
30	2.50	7914	16.1%	52.2%
30	2.77	5947	16.9%	55.0%

threshold times the daily volatility is considered an event. We used extreme parameters, i.e. many events we would like to find in a real event study will go undetected, but we can at least check the statistical power of our non-parametric algorithm. For examples of how the benchmark method is used in the empirical literature, see e.g. Mu et al. (2010) and Toth et al. (2009).

5.3. Results

We run the benchmark filter, the standard BPV test and the multi-sample BPV test over the entire dataset and show detection rates for both the standard and the multi-sample BPV test in Table 2.

We chose the critical value of the multi-sample test to identify approximately the same number of jump-days as the standard test. This way, it is fair to compare the percentage of events identified from our benchmark event day set. The low detection rate of the standard test suggests that the assumption that the jumps are instantaneous is too strong: real world events are rarely instantaneous and the standard non-parametric methods perform badly in finding non-instantaneous jumps. Our multi-sample test, however, performs much better in identifying events: in the case of the most prominent events (30 minutes window size with a threshold of 2.77 times the daily standard deviation), the 17% detection rate of the standard test is improved to 55% by our multi-sample test. Real life events are not monotonous: the price might drop, then increase a bit and then drop further and even visual observation of a trading day leads to many borderline cases: is what we see an “event” or just a normal random change in the price? This explains why our empirical results (while clearly better than the standard test) are still worse than what we observed in the Monte Carlo analysis, where the occurrence of the events is clear.

6. Conclusion

This paper examines the ability of a statistical test based on bipower variation to identify events i.e., non-instantaneous jumps in asset prices. Our database (5 years of NYSE TAQ data, 816 analyzed stocks) makes our analysis by far the largest-base empirical study conducted in the jump detection literature.

We used a modified version of the event detection algorithm of Zawadowski et al. (2004), running it with different parameters to define significant “events”. Our first empirical investigation shows that the original BPV statistic performs badly in identifying these events. Perhaps the most important contribution of our study is the introduction of the BPV signature plot and the demonstration that the BPV test statistic strongly depends on the starting point of a subsample. Based on this result, we propose the multi-sample BPV test that includes the calculation of the test statistic for many different subsamples of the same underlying price path. With careful selection of critical values, we were able to significantly increase the number of identified events compared to the standard test.

One of the prominent new features of the multi-sample statistic is that it not only tells us whether there was an event on a particular day, but also provides an estimate of the length of the event, i.e., the sampling interval in which a significantly high Z-score occurs. Use of this information could provide a new direction for future research in this field.

Funding

This research was partially supported by Pallas Athene Domus Scientiae Foundation. The views expressed are those of the authors’ and do not necessarily reflect the official opinion of Pallas Athene Domus Scientiae Foundation.

References

Aït-Sahalia, Y., 2004. Disentangling diffusion from jumps. J. Financ. Econ. 487–528.

Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. J. Econ. 169 (July (1)), 75–93. doi:10.1016/j.jeconom.2012.01.011.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2000. Great realizations. Risk 13 (March), 105–108.

Barndorff-Nielsen, O.E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. J. Econ.

Boudt, K., Petitjean, M., 2014. Intraday liquidity dynamics and news releases around price jumps: evidence from the DJIA stocks. J. Financ. Markets 17 (January (C)), 121–149. doi:10.1016/j.finmar.2013.05.004.

De Moor, L., Sercu, P., 2015. Measuring the impact of extreme observations on CAPM alphas: some methodological issues. Finance Res. Lett. 15 (November (C)), 1–10. doi:10.1016/j.frl.2014.05.002.

Huang, X., Tauchen, G., 2005. The relative contribution of jumps to total price variance. J. Econ. 3 (Autumn (4)), 456–499.

Mu, G.-H., Zhou, W.-X., Chen, W., Kertesz, J., 2010. Order flow dynamics around extreme price changes on an emerging stock market. New J. Phys. 12. doi:10.1088/1367-2630/12/7/075037.

Ormos, M., Timotity, D., 2016. Market microstructure during financial crisis: dynamics of informed and heuristic-driven trading. Finance Res. Lett. (June) 1–7. doi:10.1016/j.frl.2016.06.003.

Toth, B, Kertesz, J., Farmer, J.D., 2009. Studies of the limit order book around large price changes. Eur. Phys. J. B 71 (4), 499–510. doi:10.1140/epjb/e2009-00297-9.

Xingguo, L., Shihua, Q., Ye, Z., 2016. The information content of implied volatility and jumps in forecasting volatility: evidence from the Shanghai gold futures market. Finance Res. Lett. (June) 1–7. doi:10.1016/j.frl.2016.06.012.

Yeh, J.-H., Chen, L.-C., 2014. Stabilizing the market with short sale constraint? New evidence from price jump activities. Finance Res. Lett. 11 (September (3)), 238–246. doi:10.1016/j.frl.2014.02.005.

Zawadowski, A.G, Kertesz, J., Andor, G., 2004. Large price changes on small scales. Phys. A-Stat. Mech. Appl. 344 (1-2), 221–226. doi:10.1016/j.physa.2004.06.121.