Detecting spurious jumps in high frequency data

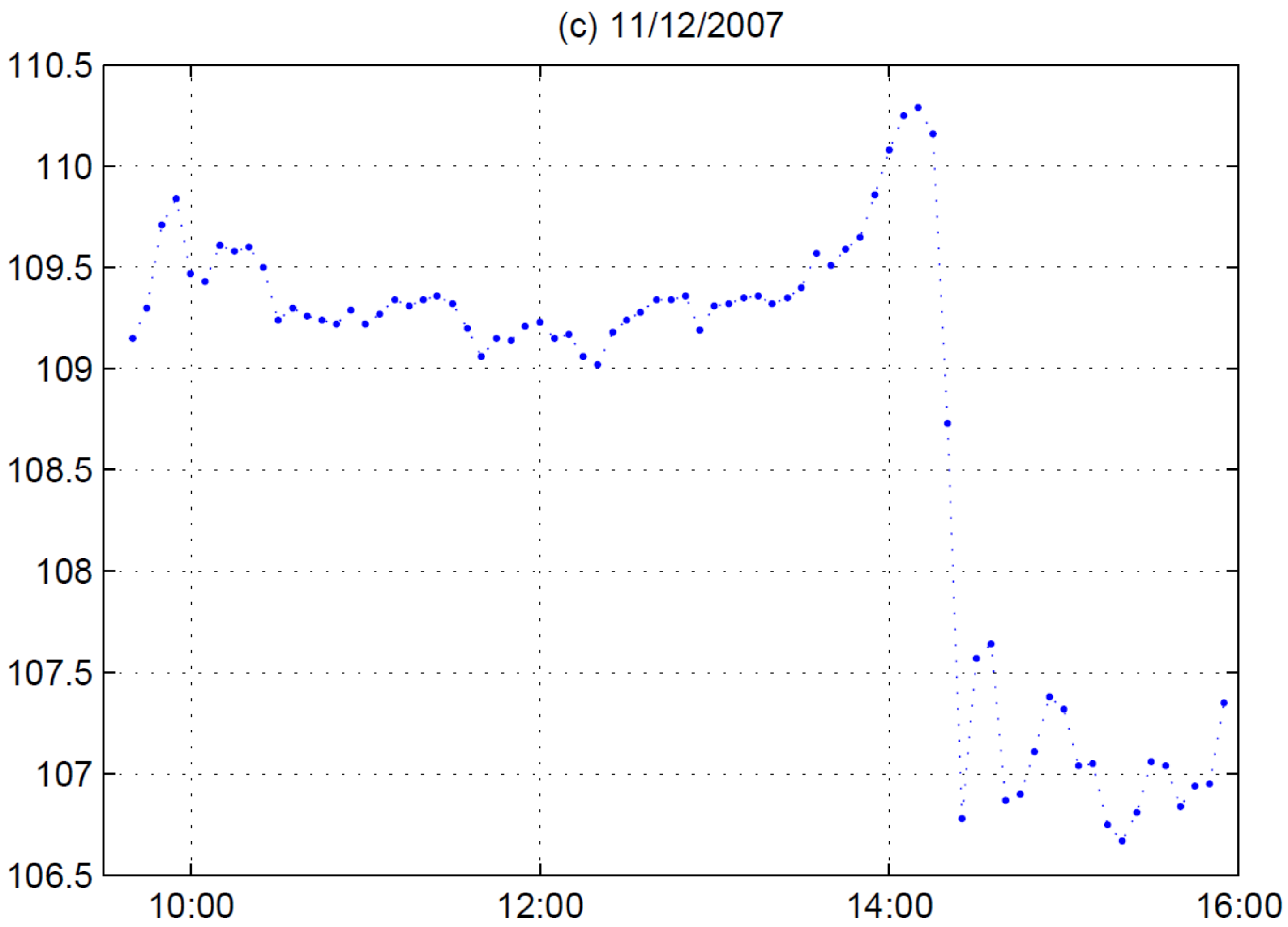
Pierre Bajgrowicz, HEC Genève
Olivier Scaillet, HEC Genève and Swiss Finance Institute

Log-price process

$$X_t = X_0 + \underbrace{\int_0^t b_s ds + \int_0^t \sigma_s dW_s}_{\text{continuous part: } X^c} + \underbrace{\sum_{j=1}^{N_t} c_j}_{\text{discrete part: } X^d}$$

Jumps

True jump (IBM)



Available jump detection tests

- Barndorff-Nielsen and Shephard (2006), (BNS)
- Aït-Sahalia and Jacod (2009), (AJ)
- Andersen, Bollerslev, and Dobrev (2007), (ABD)
- Fan and Fan (2009)
- Lee and Mykland (2008)
- Jiang and Oomen (2008)
- Fan and Wang (2007)
- Carr and Wu (2003)
- Mancini (2003)…

Multiple test

Test null hypothesis of no jumps in a particular day (e.g. at **5%** significance level) **applied over a series of days**

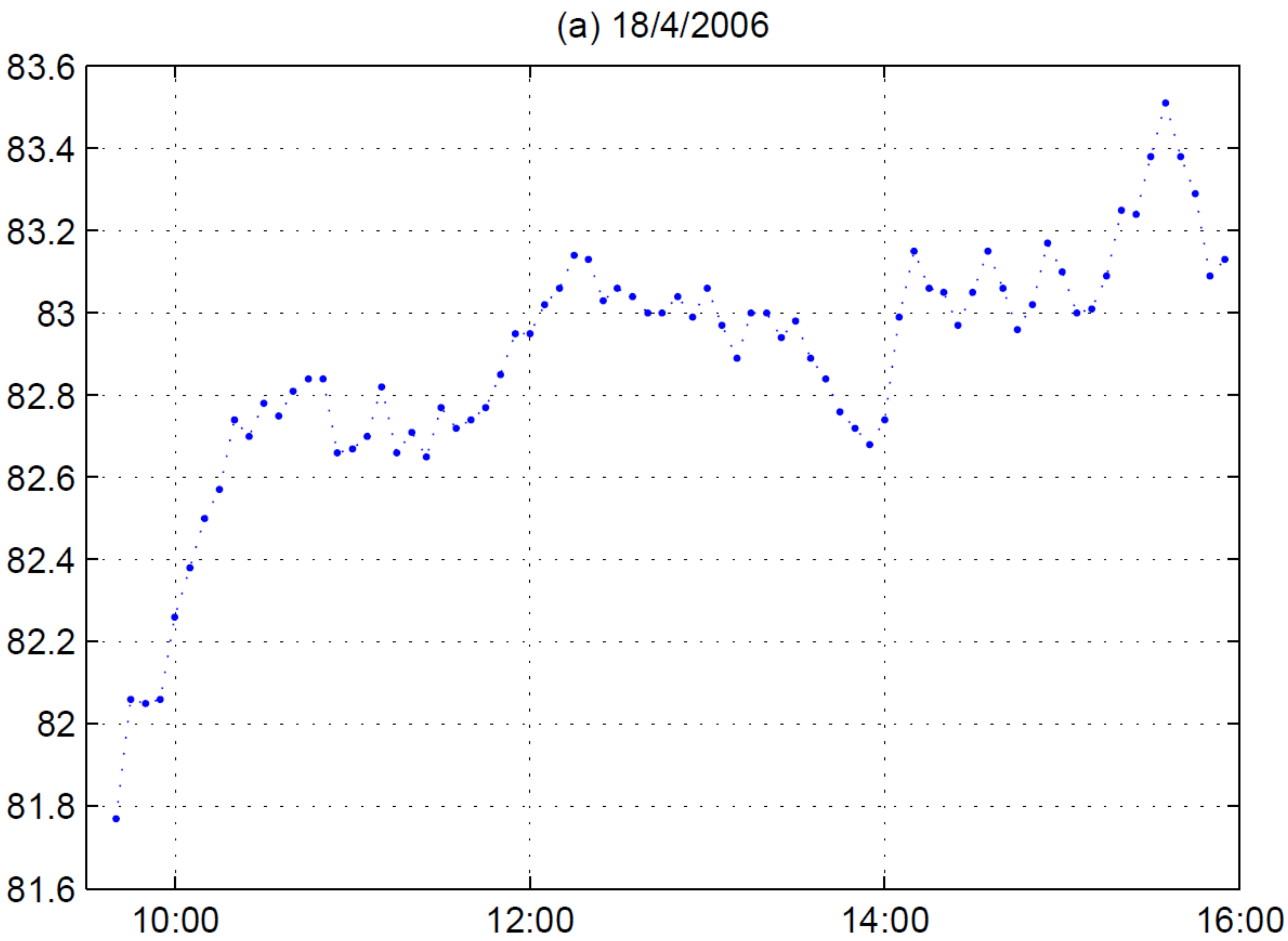
days

H_0 (no jump): 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0

Multiple test

Spurious detections of jumps!
(5% of tests in one year ≈ 13 days)

Spurious jump (IBM)



Main contributions

1. Avoid spurious detection of jumps via explicit thresholding on available test statistics:
   - Average number of jumps in U.S. equity market: 40 per year after thresholding
   - A reduction of 50% compared with before accounting for multiple testing
2. Dynamic features of irregular jump arrivals:
   Null hypothesis that jump arrival times are driven by a simple Poisson process rejected for most Dow Jones stocks.
   - Exponentiality of durations rejected for 90% of stocks
   - Serial dependence of durations for 30% of stocks

Outline

1. Methodology to eliminate spurious detections of jumps
   - Monte Carlo study
   - Empirical results
2. Dynamics of jumps arrival times
   - Are jumps arrivals driven by a simple Poisson Process?
   - Do jumps cluster in time?
3. Concluding remarks

Thresholding technique

- Number of days in the study: N
- Number of observations per day: n
- Series of daily statistics: $(S_n^1, \ldots, S_n^N)$, converging to iid N(0,1).
- **Theorem:** under the null hypothesis of no jumps:

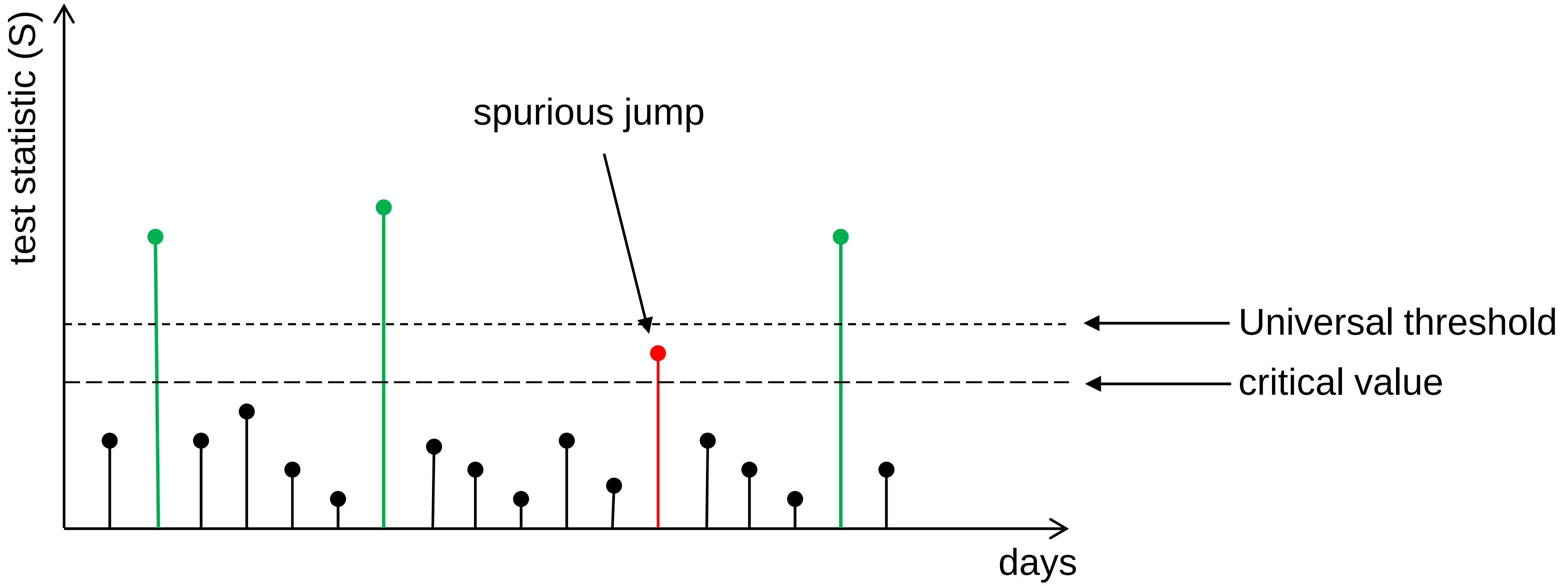
$$P\left[\sup_j |S_n^j| \leq \underbrace{\sqrt{2\log N}}_{\text{universal threshold}}\right] \to 1, \quad \text{as} \quad N, n \to \infty.$$

Thresholding technique in practice

First step:
For each day, detect jumps with existing jump detection method

→

Second step:
Apply our thresholding technique to eliminate spurious detections



Underlying detection technique

- **BNS**: Barndorff-Nielsen and Shephard (2006)
- Quadratic variation:

$$[X]_t = \lim_{n\to\infty} \sum_{j=1}^{n} \left(X_{t_j} - X_{t_{j-1}}\right)^2 = [X^c]_t + \left[X^d\right]_t$$

- Bipower variation:

$$\{X\}_t^{[1,1]} = \lim_{n\to\infty} \sum_{j=2}^{n} |X_{t_j} - X_{t_{j-1}}||X_{t_{j-1}} - X_{t_{j-2}}| = \mu_1^2 \, [X^c]_t$$

- Up to scaling factor, ratio $\dfrac{\mu_1^{-2} \{X\}_t^{[1,1]}}{[X]_t} - 1$

converges to N(0,1) under the null of no jumps.

FDR thresholding

- **Universal threshold** $\sqrt{2 \log N}$ too radical →loss of power (eliminates too many jumps)

- Less conservative threshold using **FDR thresholding** technique of Abramovich, Benjamini, Donoho, and Johnstone (2006) (keeps more jumps)

Monte Carlo study

- Same DGP as in Aït-Sahalia and Jacod (2009):

$$dX_t/X_t = \sigma_t dW_t + J_t dN_t,$$

$$v_t = \sigma_t^2, \quad dv_t = \kappa(\beta - v_t)dt + \gamma v_t^{1/2} dB_t$$

Monte Carlo study: size

		Jumps size		
		$0.1\beta^{1/2}$	$0.05\beta^{1/2}$	$0.025\beta^{1/2}$
30 sec	No thresholding	5.5	5.5	5.5
	Universal threshold	**0.2**	**0.2**	**0.2**
	FDR threshold	**0.8**	**0.8**	**0.6**
2 min	No thresholding	6.1	6.1	6.1
	Universal threshold	**0.3**	**0.3**	**0.3**
	FDR threshold	**1.1**	**1.0**	**1.7**
5 min	No thresholding	6.9	6.9	6.9
	Universal threshold	**1.2**	**1.2**	**0.4**
	FDR threshold	**1.5**	**1.5**	**3.8**

Monte Carlo study: power

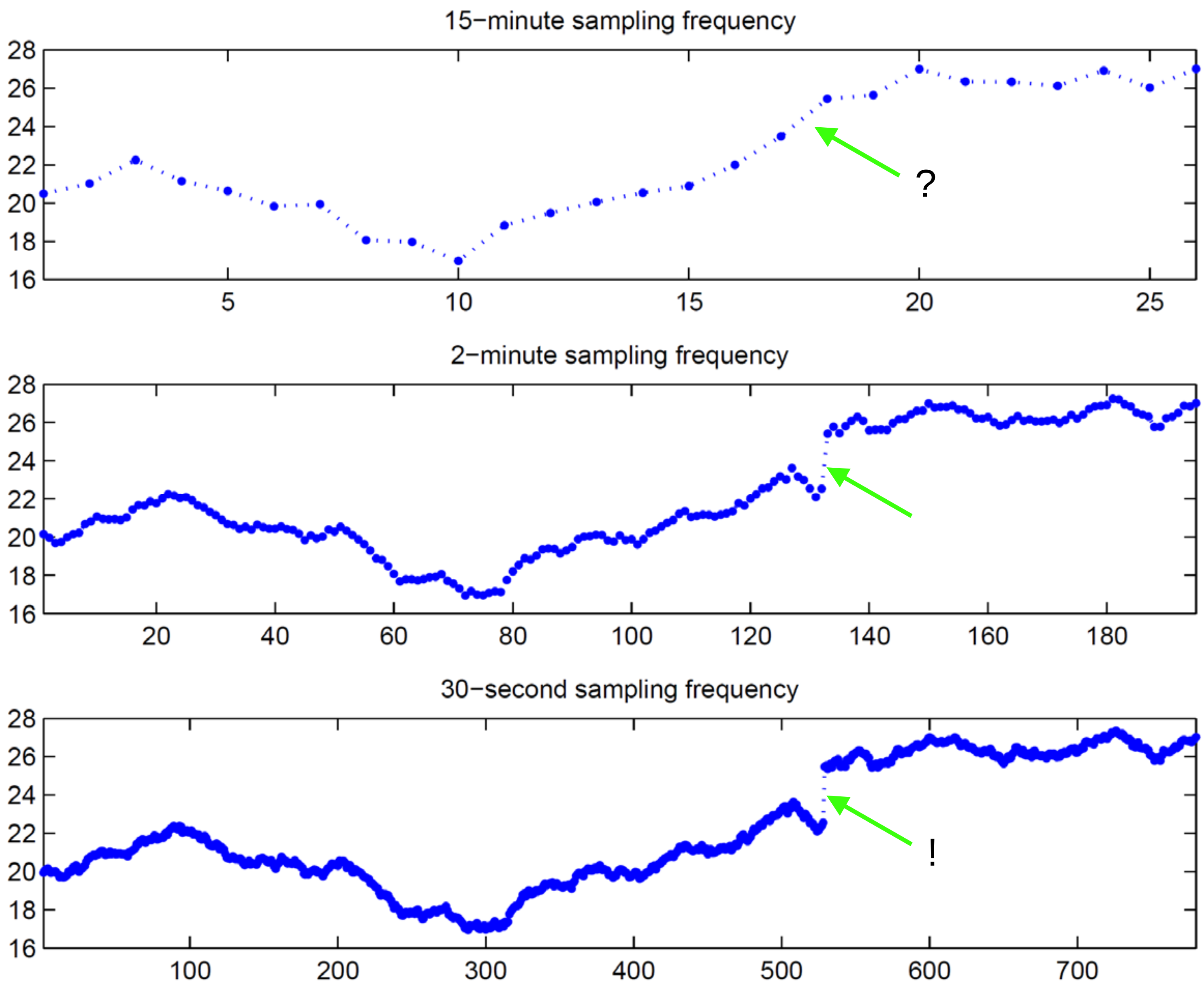
		Jumps size		
		$0.1\beta^{1/2}$	$0.05\beta^{1/2}$	$0.025\beta^{1/2}$
30 sec	No thresholding	99.9	99.9	93.0
	Universal threshold	99.9	99.9	70.0
	FDR threshold	99.9	99.9	79.7
2 min	No thresholding	99.5	98.6	46.9
	Universal threshold	99.5	91.9	13.5
	FDR threshold	99.5	95.5	23.0
5 min	No thresholding	98.8	83.2	23.3
	Universal threshold	98.6	61.4	3.7
	FDR threshold	98.6	63.2	14.4

Empirical results

- High frequency returns from the Trades and Quotes (**TAQ**) database of **Dow Jones** stocks listed on the NYSE
- **Three-year period** of January 2006 to October 2008 divided into 6-month samples
- **2-minute** sampling frequency

How often to sample?

- Avoid **microstructure noise**
- Preserve **continuous-time assumption** of underlying returns
- 15-minute frequency: only 26 observations per day
- Satisfy conditions of theorem: n>N


15−minute sampling frequency
28
26
24
22
20
18
16
5
10
15
20
25
?
2−minute sampling frequency
28
26
24
22
20
18
16
20
40
60
80
100
120
140
160
180
30−second sampling frequency
28
26
24
22
20
18
16
100
200
300
400
500
600
700
!

WMT: 01/01/2007 – 30/06/2007