

Alma Mater Studiorum Università di Bologna  
Archivio istituzionale della ricerca

Collective synchronization and high frequency systemic instabilities in financial markets

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Lucio Maria Calcagnile, ., Bormetti, G., Michele, T., Stefano, M., Lillo, F. (2018). Collective synchronization and high frequency systemic instabilities in financial markets. QUANTITATIVE FINANCE, 18(02), 237-247 [10.1080/14697688.2017.1403141].

*Availability:*

This version is available at: <https://hdl.handle.net/11585/611016> since: 2021-06-29

*Published:*

DOI: <http://doi.org/10.1080/14697688.2017.1403141>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is an Accepted Manuscript of an article published by Taylor & Francis in Quantitative Finance on 19/12/2017, available online at <https://www.tandfonline.com/doi/full/10.1080/14697688.2017.1403141>

# Collective synchronization and high frequency systemic instabilities in financial markets

Lucio Maria Calcagnile<sup>1,2,\*</sup>, Giacomo Bormetti<sup>3,2</sup>, Michele Treccani<sup>1,2,†</sup>,  
Stefano Marmi<sup>4,5</sup>, Fabrizio Lillo<sup>4,2</sup>

April 11, 2017

<sup>1</sup> *LIST S.p.A., via Pietrasantina 123, Pisa, 56122, Italy*

<sup>2</sup> *QUANTLab<sup>1</sup>, via Pietrasantina 123, Pisa, 56122, Italy*

<sup>3</sup> *University of Bologna, Piazza di Porta San Donato 5, Bologna, 40126, Italy*

<sup>4</sup> *Scuola Normale Superiore, Piazza dei Cavalieri 7, Pisa, 56126, Italy*

<sup>5</sup> *CNRS UMI 3483 – Laboratorio Fibonacci, Piazza dei Cavalieri 7, Pisa, 56126, Italy*

## Abstract

We present a series of empirical evidences on the dynamics of price instabilities in financial markets and propose a new Hawkes modeling approach. Specifically, analyzing

---

\*Corresponding author. *E-mail address:* l.calcagnile@list-group.com

†Present address: Mediobanca S.p.A, Piazzetta E. Cuccia 1, 20121 Milano, Italy

<sup>1</sup>[www.quantlab.it](http://www.quantlab.it)

the recent high frequency dynamics of a set of US stocks, we find that since 2001 the level of synchronization of large price movements across assets has significantly increased. We find that only a minor fraction of these systemic events can be connected with the release of pre-announced macroeconomic news. Finally, the larger is the multiplicity of the event – i.e. how many assets have swung together – the larger is the probability of a new event occurring in the near future, as well as its multiplicity. To reproduce these facts, due to the self- and cross-exciting nature of the event dynamics, we propose an approach based on Hawkes processes. For each event, we directly model the multiplicity as a multivariate point process, neglecting the identity of the specific assets. This allows to introduce a parsimonious parametrization of the kernel of the process and to achieve a reliable description of the dynamics of large price movements for a high-dimensional portfolio.

# 1 Introduction

Quoting from Michael Lewis’ Flash Boys “The world clings to its old mental picture of the stock market because it’s comforting” [Lewis \(2014\)](#). But trading activity has profoundly changed from the old phone conversation or click and trade on a screen to software programming. Market statistics confirm that automated algorithms carry out a significant fraction of the trading activity on US and Europe electronic exchanges [Gomber et al. \(2011\)](#); [MacIntosh \(2013\)](#). As algos feed on financial and news data, the speed of information processing has dramatically increased and potentially allows large price movements to propagate very rapidly through dif-

ferent assets and exchanges [Gerig \(2013\)](#).

The synchronization effect had its most spectacular appearance during the May 6th, 2010 Flash Crash. The crash started from a rapid price decline in the E-Mini S&P 500 market and in a very short time the anomaly became systemic and the shock propagated towards ETFs, stock indices and their components, and derivatives [CFTC and SEC \(2010\)](#); [Kirilenko et al. \(2017\)](#). The price of the Dow Jones Industrial Average plunged by 9% in less than 5 minutes but recovered the pre-shock level in the next 15 minutes of trading. The SEC reported that such a swing was sparked by an algorithm executing a sell order placed by a large mutual fund. Then high frequency traders, even though did not ignited the event, caused a “hot potato” effect amplifying the crash. In the aftermath of the crash, several studies have focused on events, evocatively named *Mini Flash Crashes*, concerned with the emergence of large price movements of an asset in a very limited fraction of time. Mini Flash Crashes may be attributed to the interaction between several automatic algorithms [Johnson et al. \(2013\)](#) or to the unexpected product of regulation framework and market fragmentation [Golub et al. \(2012\)](#).

The Flash Crash, however, has also dramatically shown how strongly interconnected different markets and asset classes can become, especially during extreme events. In this paper, by taking a different, yet complementary approach to the above literature, (i) we identify non-parametrically one-minute extreme events as over-threshold movements, studying how the frequency of collective instabilities at high frequency has changed in the last years, and (ii) we use Hawkes processes [Hawkes \(1971\)](#) to model parametrically the dynamics of these events and their mutual cross-excitation. Our identification approach shares some similarities with previ-

ous works employing non-parametric tests to identify extreme movements, see [Andersen et al. \(2007\)](#); [Lee and Mykland \(2008\)](#); [Andersen et al. \(2010\)](#); [Dumitru and Urga \(2012\)](#); [Bormetti et al. \(2015\)](#). We perform our analysis on a yearly basis from 2001 to 2013 on a data sample of highly liquid US equities and we identify extreme events affecting a sizable fraction of the investigated assets. Remarkably, very little research has been devoted to the investigation of this kind of systemic events. Few noticeable exceptions are [Bollerslev et al. \(2013\)](#), which aims at the identification of common large movements between the market portfolio and individual stocks, and [Gilder et al. \(2014\)](#), which investigates the tendency of large movements to arrive simultaneously. A very recent non-parametric test of the occurrence of simultaneous jumps across multiple assets is discussed in [Caporin et al. \(2014\)](#). Our research provides the empirical evidence that, while the total number of extreme movements has decreased along years, the occurrence of systemic events has significantly increased. As a terminology clarification, we prefer the use of the term *systemic* rather than *systematic*, since the latter has been used in the literature (see e.g. [Gilder et al. \(2014\)](#)) to define events where assets jump together with a market index, while the events we look at not necessarily imply this.

To identify the possible causes of such events we compare their time occurrences with a database of pre-scheduled macroeconomic announcements. Since macroeconomic news can be expected to have a market-level influence, they represent a natural candidate to explain market-wide events. For instance, literature has recognized the peculiar role played by Federal Open Market Committee (FOMC) meetings deciding the interest rate level [Petersen et al. \(2010b,a\)](#). However, unexpectedly, only a minor fraction (less than 40%) of events involving a

large fraction of assets has been preceded by the release of a macro news. This evidence opens the route to the more intriguing hypothesis that a genuinely endogenous dynamics is taking place. To the best of our knowledge, the association between extreme equity price movements and the news arrival has been previously investigated in [Lee and Mykland \(2008\)](#); [Lee \(2012\)](#), finding a positive association, but the results have been challenged in [Bajgrowicz et al. \(2015\)](#). Table 11 in [Gilder et al. \(2014\)](#) suggests the existence of a particularly strong relationship between FOMC announcements and the arrival of a systematic event. However, none of the previous works performs an analysis of the association between news and extreme movements conditional on the multiplicity of the event.

Finally, we show that when an event affecting a significant fraction of assets occurs, the probability of a novel extreme event in the subsequent minutes increases. More interestingly, there is a clear evidence that the more systemic the conditioning event is, the larger the expected number of assets swinging synchronously in the immediate future will be. In order to reproduce such empirical evidences, we propose a model within the class of mutually exciting point processes, termed Hawkes processes [Hawkes \(1971\)](#). In recent years, they have experienced an increasing popularity in mathematical finance and econometrics [Bowsher \(2007\)](#); [Bauwens and Hautsch \(2009a\)](#); [Muni Toke \(2011\)](#); [Muni Toke and Pomponio \(2012\)](#); [Filimonov and Sornette \(2012\)](#); [Bacry et al. \(2013\)](#); [Hardiman et al. \(2013\)](#); [Rambaldi et al. \(2015\)](#); [Bacry et al. \(2015\)](#); [Rambaldi et al. \(2016\)](#). The main technical challenge that we face in our research is the high dimensionality of the empirical dataset. Since the number of parameters of a Hawkes process grows quadratically with the dimension of the problem, the estimation of the model

rapidly becomes computationally unfeasible. A major methodological contribution of our work is to model the multiplicity of an event – how many assets are moving together – instead of modeling directly the identity of the assets. In fact, the empirical evidence shows that events of similar multiplicity influence each other more than events of very different multiplicity. Then, working with the vector of multiplicities allows to simplify the kernel of the Hawkes process and to express it as a function of the distance among multiplicities. This leads to a parsimonious parametrization of the kernel, which is not immediate to justify if we model the assets directly. In conclusion, we show that a Hawkes process with plausible restrictions on the kernel structure is able to deal with a high-dimensional portfolio, and to reproduce with remarkable realism the cross-excitation effects of over-threshold returns.

The rest of the paper is organized as follows. In Section 2 we describe the procedure to identify non-parametrically the occurrence of over-threshold events. We present the most relevant empirical findings and comment on the emergence of a significant synchronization among extreme movements of the most liquid assets belonging to the Russell 3000 index. In the second part of the Section, the macronews dataset is described and the relation with over-threshold events is investigated. Section 3 reviews the class of exponential Hawkes processes, and motivates the introduction of a parsimonious parametrization of the kernel of the processes. The results from an estimation for the year 2013 and various Monte Carlo experiments are presented. Section 4 draws the main conclusions.



## 2 Dynamics of extreme events

The main objective of this paper is the modeling of the dynamics of synchronous large price variations at high frequency. In order to detect extreme variations of the stock prices  $P_t$ , we compare price returns (defined as  $r_t = \ln P_t / P_{t-1}$ ) with an estimate of the historical spot volatility, which sets the scale of local price fluctuations. Specifically, we calculate a volatility time series  $\sigma_t$  as an exponential-moving-average version of the bipower variation (see [Barndorff-Nielsen and Shephard \(2004\)](#); [Corsi et al. \(2010\)](#); [Bormetti et al. \(2015\)](#)) of the return time series. Finally, we say that an extreme return occurs when

$$\frac{|r_t|}{\sigma_t} > \theta, \quad (1)$$

for a certain threshold  $\theta$ . In our main analyses we take  $\theta = 4$ , but we also investigate higher values of the threshold, namely  $\theta = 6, 8, 10$ , in some of our descriptive statistics.

We say that a stock *jumps*<sup>2</sup> in a given one minute interval if condition of Equation (1) is observed for a given  $\theta$ . Here we are mostly interested in *cojumps*, i.e. the simultaneous (inside the minute) occurrence of jumps for a subset of  $M$  stocks. The quantity  $M$  is termed the *multiplicity* of the cojump, and it gives a measure of the systemic nature of the event. In the following we consider three questions: (i) How has the high frequency instability changed in the last fifteen years? (ii) What fraction of the systemic instabilities can be attributed to

---

<sup>2</sup>For a terminological clarification, we term jumps the abnormal returns which are identified non-parametrically in the same spirit of [Lee and Mykland \(2008\)](#). We refer to [Bormetti et al. \(2015\)](#) for a review of the econometric literature dealing with the jump identification problem within the framework of Itô semi-martingales.

macroeconomic news? (iii) How can we model the short term dynamics of market instabilities?

## 2.1 Financial data selection and treatment

We conduct our analysis on price time series of financial stocks belonging to the Russell 3000 Index, traded in the US equity markets (mostly NYSE and NASDAQ). Data are provided by Kibot, [www.kibot.com](http://www.kibot.com). We consider the thirteen years from 2001 to 2013 and for each year we select 140 highly liquid stocks. We take the 140 stocks with the highest percentage of minutes in which at least one trade was made. The 140th selected asset of each year has such a percentage equal to 93% for 2001, 96% for 2002, between 97% and 98.5% for the years 2003–2005, and always greater than 99% for the years from 2006 onwards. We use 1-minute closing price data during the regular US trading session, i.e. from 9:30 AM to 4:00 PM. In selecting the assets to analyze, we exclude American Depositary Receipts, which are negotiable instruments representing ownership in non-US companies, since their dynamics is heavily influenced by their primary market, and thus shows a peculiar intraday pattern.

Intraday returns are first filtered for the average intraday pattern, since price fluctuations are known to exhibit significant differences in absolute size depending on the time of the day, showing a typical U shape with larger movements at the beginning and at the end of the trading day. Returns at intraday time  $t$  are rescaled by a factor  $\zeta_t$ , which is calculated as the average, over all days, of adjusted absolute returns at time  $t$ . More precisely, if  $\tilde{r}_{d,t}$  is the raw return of

day  $d$  and intraday time  $t$ , we define the rescaled return

$$r_{d,t} = \frac{\tilde{r}_{d,t}}{\zeta_t},$$

where

$$\zeta_t = \frac{1}{N_{\text{days}}} \sum_{d'} \frac{|r_{d',t}|}{s_{d'}},$$

with  $N_{\text{days}}$  indicating the number of days in the sample and  $s_{d'}$  the standard deviation of absolute intraday returns of day  $d'$ . Scaled returns no longer possess any daily regularities. We discard early-closing days (typically, the eves of Independence Day, Thanksgiving and Christmas). Data are adjusted for splits and dividends.

## 2.2 Historical dynamics of jumps and cojumps

A visual representation of how instability of financial markets has changed in the last years is shown in Figure 1, which compares the dynamics of  $\theta = 4$  cojumps in 2001 (left panel) and 2013 (right panel). The horizontal axis represents the trading day and the vertical axis indicates the hour of the day. The presence of a circle indicates the occurrence of a cojump and the color codifies the number of stocks simultaneously cojumping (i.e. the multiplicity). In 2001 there were many cojumps with low multiplicity and the high multiplicity cojumps are concentrated mostly at specific hours of the day (10 a.m. and 2:15 p.m.) corresponding to the release of important macro announcements, such as, for example, the FOMC announcements. On the contrary, in 2013 we observe less low multiplicity cojumps and many more high multiplicity

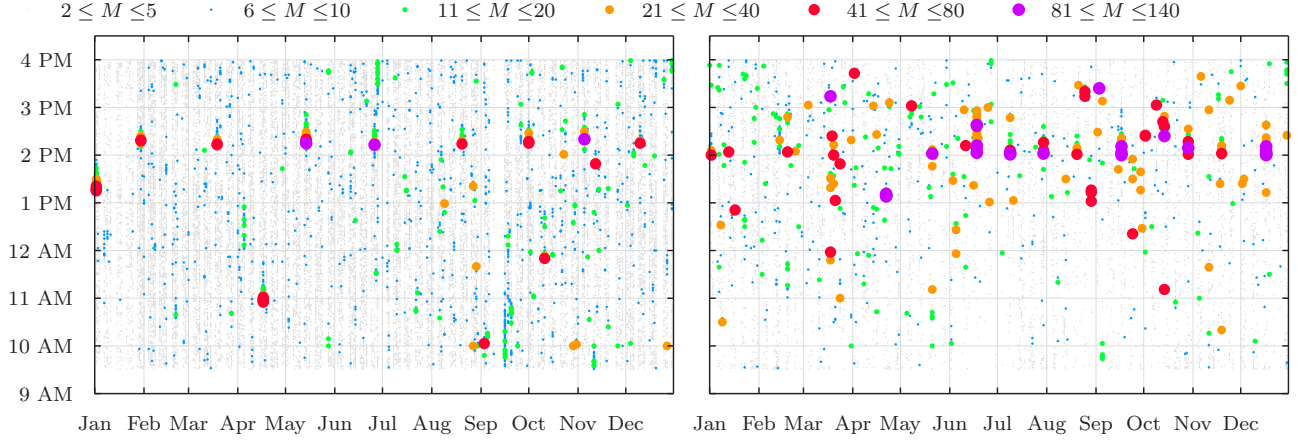


Figure 1: **Cojumps in 2001 and 2013.** Representation of the time series of cojumps detected for the dataset of 140 selected highly liquid stocks of the Russell 3000 Index during year 2001 (left panel) and 2013 (right panel). The size of the circles increases with the multiplicity of the cojump event.

cojumps, which are quite scattered during the day. This is an indication that modern financial markets have become more systemically unstable and that these instabilities are less related to macro news. In the following we show that this is the case with more quantitative analyses.

First, in the top left panel of Figure 2 we show the total number of minutes (in logarithmic scale) with at least one jump in each year, considering different values of  $\theta$ . We observe that for all  $\theta$ s the number of jumps has actually *decreased* over time. The different lines are quite parallel one to each other (especially for  $\theta \geq 6$ ) indicating that the tails of the one minute return distribution remained quite stable. A completely different pattern emerges when we consider the dynamics of cojumps. The top right panel of Figure 2 shows the frequency of cojumps of different multiplicity (normalized to its value in 2001). While the frequency of cojumps with any multiplicity ( $M \geq 2$ ) has slightly declined, the frequency of high multiplicity cojumps has become in recent years up to 10 times more frequent than its value in 2001. The result is

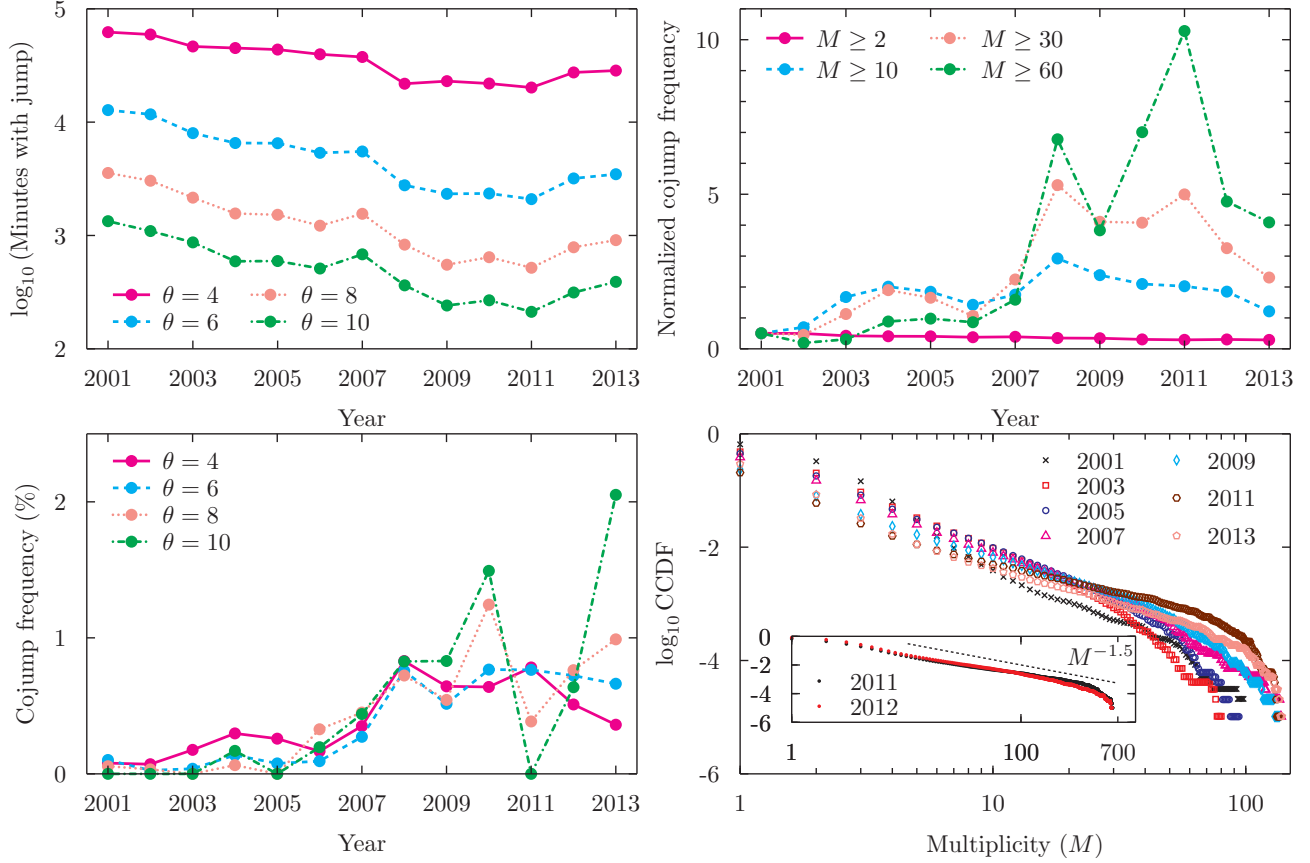


Figure 2: **Historical evolution of cojump distribution.** Top left panel: Semi-log plots of the total number of minutes where we detect at least one jump among the 140 selected assets of the Russell 3000 Index. Curves correspond to four different levels of the threshold parameter  $\theta$ . Top right panel: For  $\theta = 4$ , yearly time evolution of the fraction of minutes with at least one event of multiplicity larger than or equal to 2, 10, 30, 60. All values are normalized by the corresponding 2001 values. Bottom left panel: Yearly evolution of the percentage fraction of cojumps with multiplicity at least equal to 30 for four different values of  $\theta$ . Bottom right panel: Log-log plots of the Complementary of the Cumulative Distribution Function of the cojump multiplicity for seven different years. The panel reports the empirical evidence for a portfolio of 140 stocks, while the inset details results of the same analysis conducted with 700 liquid assets from Russell 3000 during years 2011 and 2012.

essentially unchanged when fixing the minimal multiplicity (e.g.  $M \geq 30$ ) and computing the number of cojumps for different values of  $\theta$  (bottom left panel of Figure 2). Clearly larger fluctuations are observed for larger values of  $M$ .

Finally, the bottom right panel of Figure 2 shows the distribution function of the cojump multiplicity for different years. Despite some variation is observed across the years, a clear power law tail behavior is evident. This means that, conditionally on the occurrence of a cojump event, those with a high multiplicity are frequent. Consistently with the observations above, the tail is thicker in recent years (even if in 2013 we observe a slightly thinner tail). It is important to notice that the bending of the distributions for large multiplicity is very likely due to the finite support of the distribution. Clearly for a set of  $N$  stocks the multiplicity cannot be larger than  $N$ , thus the distribution function is zero at  $M = N$ . To show the role of the finite support, in the inset we show the multiplicity distribution function for a larger set of 700 highly liquid assets. In this case the power law region extends for a wider range and close to  $M = 700$  we observe the expected bending of the function. The tail exponent of these distributions is close to 1.5 (similarly to what observed in [Joulin et al. \(2008\)](#)).

The paper mostly considers one minute (co)jumps. However one minute in 2013 is not equivalent to one minute in 2001 in terms of market activity. Hence it is important to test whether the increase in number of high multiplicity cojumps is due to the fact that in older years synchronization occurred on a time scale longer than one minute. To test this possibility we have repeated the analysis varying the time scale for jump detection from one to five minutes. The latter scale is motivated by independent analyses, not reported here, on the dynamics of the cross-correlation between stocks. Such analyses suggest that the time scale over which stocks become correlated has decreased from 2001 to 2013 by a factor approximately equal to five. Moreover, since at high sampling frequencies microstructure noise may be a source of

bias and dominate the results, we take into account the findings of Andersen et al. (2000) and, similarly to Bollerslev et al. (2008), we consider also sampling frequencies up to 15 minutes.

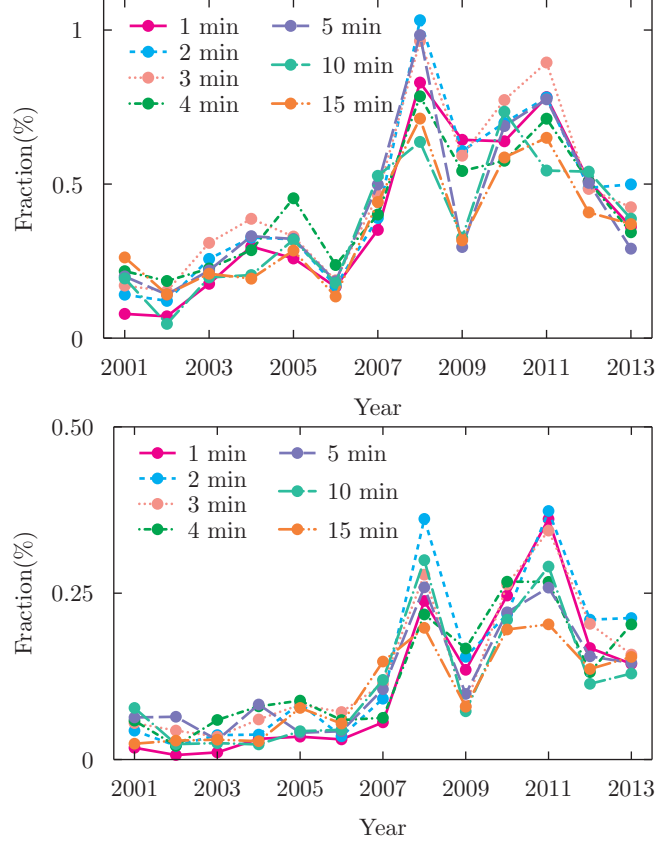


Figure 3: **Historical evolution of cojumps at different time scales.** Yearly time evolution of the fraction of cojumps with multiplicity  $M \geq 30$  (top) or  $M \geq 60$  (bottom) over the total number of cojumps ( $M \geq 1$ ) for  $\theta = 4$  and different time horizons, namely 1, 2, 3, 4, 5, 10, 15 minutes.

Figure 3 shows the yearly time evolution of the fraction of cojumps with multiplicity  $M \geq 30$  (top) or  $M \geq 60$  (bottom) over the total number of cojumps ( $M \geq 1$ ) for  $\theta = 4$  and different time scales, namely 1, 2, 3, 4, 5, 10, 15 minutes. Except for the first two years, no clear sorting of this fraction with the time scale is detectable, while the global secular trend has a much larger variability. This is particularly evident for the  $M \geq 60$  case. Hence the number

of high multiplicity one minute cojumps in 2013 is much higher than the number of high multiplicity five minute cojumps in 2001, indicating that the increased speed of market activity is a minor cause of the increase of high multiplicity systemic cojumps in recent years. We also note that this increase is evident at sampling frequencies of 10 and 15 minutes, too, which gives robustness to the results by providing the evidence that the increase of systemic cojumps is not due to microstructure noise. Finally, it is worth to point out that the fraction of cojumps with multiplicity equal to or larger than 30 reaches its maximum in 2008. When conditioning to events with even more extreme cojump events – with multiplicity equal to or larger than 60 – a second peak appears in 2011. Then, these results suggest – maybe quite expectedly – the emergence of a tighter collective market dynamics during the sub-prime mortgage crisis and the collapse of Lehman Brothers in 2008, and the rise of the sovereign debt crisis in 2011.

In conclusion, at the beginning of 2000's individual jumps were more frequent and high frequency systemic instabilities, i.e. high multiplicity jumps, were rare and mostly concentrated on macro-news announcements. In recent years, on the contrary, markets display often systemic cojumps and these are scattered across the trading day.

## **2.3 The role of macroeconomic news**

The second question we address is which fraction of systemic cojumps has an exogenous or an endogenous origin. To answer it, we study how frequently a systemic cojump is preceded by a scheduled macroeconomic news. We solely consider macronews, since it is unlikely that stock idiosyncratic news affect the whole market.



We use macroeconomic news data provided by Econoday, Inc., [www.econoday.com](http://www.econoday.com). We consider the 42 most important news categories, which are classified into two large groups according to their capacity of influencing the financial markets: the Market Moving Indicator group and the Merit Extra Attention group. Since we are concerned with matching news with market extreme events, we consider only the 27 categories whose announcement times occur during the trading session. The number of total news announcements ranges from around 150 in the first years to around 260 in the last years, for a total of 2,888 news. Table 1 in Appendix A contains the list of the news categories considered and the numbers of news announcements, organized by year and news category. We measure how frequently a systemic cojump with

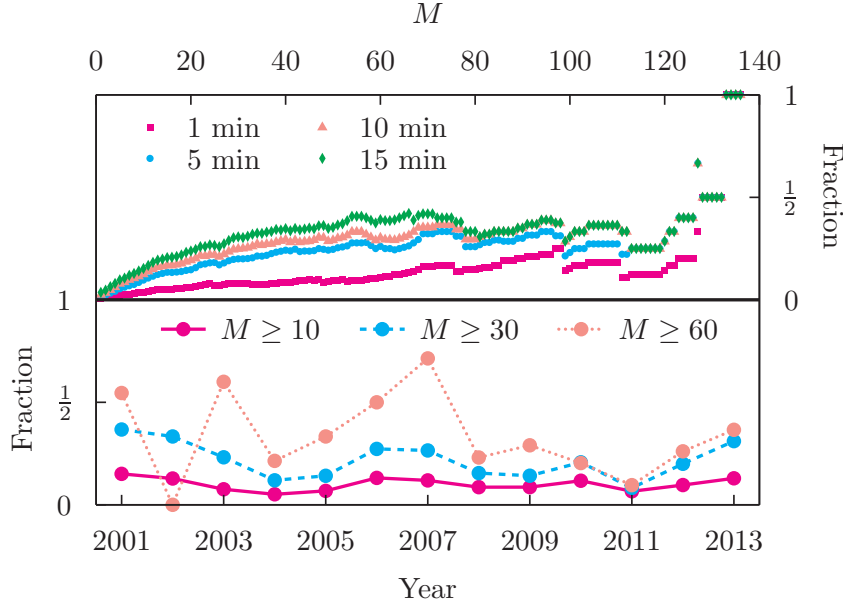


Figure 4: **Cojumps and macroeconomic news.** Top panel: Fraction of cojumps in 2012 with multiplicity larger than or equal to the value reported on the x axis for which a news occurred in the last 1, 5, 10, and 15 minutes. Bottom panel: Fraction of cojumps for different multiplicities  $M$  for which we observe at least one news in a time window of five minutes preceding the jump event.

multiplicity larger than  $M$  is preceded by a macronews in the previous  $\tau = 1, 5, 10, 15$  minutes.

The top graph of Figure 4 shows that only 40% of the high multiplicity cojumps are preceded by a macronews in the previous 15 minutes. Notice that the fractions of news-triggered systemic events in the 5, 10, and 15 minutes time windows are very close one to each other, indicating that if a macronews triggers a systemic cojump, this will typically happen within 5 minutes from the news.

For an historical perspective, the bottom graph of Figure 4 shows that the fraction of systemic cojumps triggered by macroeconomic news is quite constant across the years and, even for large  $M$ , clearly below 50%. Thus our empirical analysis shows that a relevant portion of systemic cojumps is not associated with scheduled macroeconomic announcements. Idiosyncratic company-specific news may play a role, but plausibly only for those events which involve a very limited number of assets. For high multiplicity cojump events, endogenous mechanisms are likely to play a determinant role.

### 3 Modeling with multidimensional Hawkes processes

The empirical evidence of the previous section suggests that a large fraction of the dynamics of the systemic cojumps is unrelated to macronews. Moreover, as observed for example in the 2010 Flash Crash, market instabilities tend to propagate quickly to other assets, markets, or asset classes. Thus it is important to model the self- and cross-dependence of instabilities, considering both synchronous and lagged dependencies, by studying whether and how systemic instabilities trigger other instabilities in the short run.

A class of processes which is a natural candidate to describe self- and cross-triggering effects is the family of processes dubbed Hawkes. Hawkes processes have been introduced in the early Seventies [Hawkes \(1971\)](#), and have been widely employed to model earthquake data [Vere-Jones \(1970\)](#); [Vere-Jones and Ozaki \(1982\)](#); [Ogata \(1988\)](#). For a complete overview of the properties of Hawkes processes please refer to [Daley and Vere-Jones \(2003\)](#); [Bauwens and Hautsch \(2009b\)](#), while for a review of their recent applications in a financial context see [Bacry et al. \(2015\)](#).

An  $N$ -dimensional Hawkes process is a point process characterized by the vector of intensities  $\boldsymbol{\lambda}_t := (\lambda_t^1, \dots, \lambda_t^N)^\top$ , where the  $i$ -type intensity satisfies the relation

$$\lambda_t^i = \mu_t^i + \sum_{j=1}^N \sum_{t_k^j < t} \phi_j^i(t - t_k^j),$$

where  $\mu_t^i$  and  $\phi_j^i$  are non-negative deterministic functions for all  $i, j = 1, \dots, N$ . The set  $\{t_k^j\}$  corresponds to the random sequence of increasing events associated with the  $j$ -component of the  $N$ -dimensional point process. If  $\mu_t^i = \mu^i$  is a constant and the kernel function  $\phi_j^i$  reduces identically to zero, then the Hawkes point process for the  $i$ -component reduces to a Poisson process with constant intensity  $\mu^i$ . On the contrary, if the kernel is positive, each time an event occurs for any component of the multidimensional process, the intensity  $\lambda_t^i$  increases by a positive amount.

Modeling the interaction among a set of 140, or more, assets in terms of Hawkes processes is challenging, especially from the estimation perspective. Some sort of filtering is needed. A first step in this direction is taken in [Bormetti et al. \(2015\)](#) where the authors model the

multivariate point process describing the jumps with a Hawkes factor model. Each stock is represented by a point process, each count being a jump. The coupling between the stocks is given by a one factor model structure, i.e. the intensity is the sum of the intensity of a factor and the intensity of an idiosyncratic term. In order to capture the temporal clustering of events both the factor and the idiosyncratic terms are driven by a self-exciting process. This type of modeling is very effective (and parsimonious) in describing the pairwise properties of cojumps, i.e. the probability that two stocks jump in the same time interval. However, when considering cojumps of  $M > 2$  stocks, the model shows its weaknesses. An important aspect of the empirical evidences is given by the distribution of multiplicities. From the bottom right panel of Figure 2, the power law behavior of the tail region is apparent. Unfortunately, it is readily shown that in the large  $N$  limit, the factor model predicts a distribution of multiplicities with Gaussian tails. Moreover, the multiplicity of a systemic cojump is independent from the multiplicity of previous systemic cojumps, while the right panel of Figure 1 shows clear temporal clusters of high multiplicity cojumps.

For these reasons, in this paper we propose a different modeling approach, which preserves the parsimony and is able to overcome the limitations of the factor model. The key idea is to model the vector of multiplicities directly, although losing information on the identity of the cojumping stocks. Specifically, we consider an  $N$ -dimensional point process characterized by the vector of intensities  $\boldsymbol{\lambda}_t$ . An event in the  $i$ -component at time  $t$  means that at this time a systemic cojump of multiplicity  $i$  has occurred. Within this modeling framework we know the total number of assets which have jumped, but we can no longer identify which companies

among the  $N$  possible ones have moved. To describe the self- and cross-excitation of cojumps we use an  $N$ -dimensional Hawkes process, with exponential kernels

$$\phi_j^i(t - t_k^j) := \alpha_{ij} e^{-\beta_{ij}(t - t_k^j)}, \quad (2)$$

with  $\alpha_{ij} > 0$  and  $\beta_{ij} > 0$  for all  $i, j$ . The parameter  $\alpha_{ij}$  fixes the scale of the intensity process  $\lambda^i$  and provides the deterministic amount by which the  $j$ -type event at  $t_k^j$  shocks the intensity of the  $i$ -type process. The parameter  $\beta_{ij}$  describes the inverse of the time needed by the process  $i$  to lose memory of a count of process  $j$ . In general, the model depends parametrically on the baseline intensity vector  $\boldsymbol{\mu}$ , and on the  $N \times N$  matrices  $\alpha_{ij}$  and  $\beta_{ij}$  of parameters characterizing the kernels.

As in most high-dimensional problems, the estimation of multivariate Hawkes processes might be problematic. In order to reduce the dimension of the estimation problem from  $N + 2N^2$  to a smaller number of unknowns, we impose various parameter restrictions. First, we assume that the vector  $\boldsymbol{\mu} := (\mu_t^1, \dots, \mu_t^N)^\top$  does not depend on time. Then, given the exponential form (2), the process is stationary if the spectral radius (i.e. the absolute value of the largest eigenvalue) of the matrix  $\Gamma$  of elements

$$\Gamma_{ij} = \frac{\alpha_{ij}}{\beta_{ij}}$$

is strictly smaller than one. In this case the unconditional expected intensities of the process

reads

$$\mathbb{E}[\boldsymbol{\lambda}_t] = (\mathbf{I}_N - \Gamma)^{-1} \boldsymbol{\mu},$$

where  $\mathbf{I}_N$  is the  $N$ -dimensional identity matrix.

We make the following further assumptions:

- all the  $\beta_{ij}$  are equal to a constant value  $\beta > 0$ . This means that only one time scale characterises the decay of the kernels.
- We impose the condition that  $\boldsymbol{\mu} = \eta \mathbb{E}[\boldsymbol{\lambda}_t]$ , with  $0 < \eta < 1$ . In this way, by construction, the distribution of multiplicity in the observed process will be the same as the distribution of the multiplicity in the baseline (or ancestor) process. In other words, the cross-excitation between the different components of the Hawkes process does not change the unconditional law of multiplicity. Notice that this assumption implies that

$$\Gamma \mathbb{E}[\boldsymbol{\lambda}_t] = (1 - \eta) \mathbb{E}[\boldsymbol{\lambda}_t],$$

i.e.  $\mathbb{E}[\boldsymbol{\lambda}_t]$  (or  $\boldsymbol{\mu}$ ) is the eigenvector of  $\Gamma$  with eigenvalue  $1 - \eta$ .

- The generic matrix element  $\Gamma_{ij}$  describing the intensity of the excitation of variable  $j$  on variable  $i$  is the product of a term  $D_{ii}$ , which depends on the excited variable, and a term  $\sigma(|i - j|)$ , which depends on the absolute difference of the two multiplicities. Therefore,

we rewrite  $\Gamma = D\Sigma$ , where  $D$  is a diagonal matrix of elements

$$D_{ii} := \frac{(1 - \eta)\mu^i}{\sum_{j=1}^N \mu^j \sigma(|i - j|)},$$

and  $\Sigma_{ij} = \sigma(|i - j|)$ .

- Finally, we parametrize the matrix  $\Sigma$  as

$$\Sigma_{ij} = \sigma(|i - j|) = (|i - j| + 1)^{-\gamma}$$

The hyperbolic decay describes the strong cross-excitation between two different multiplicities in terms of the sole parameter  $\gamma$ . We assess that the cross-excitation is stronger between cojumps of similar multiplicity.

To sum up, the model is completely specified in terms of  $\beta$ , and of the matrix

$$\Gamma_{ij} := \frac{(1 - \eta)\mu^i}{\sum_{k=1}^N \mu^k \sigma(|i - k|)} \sigma(|i - j|),$$

which depends on parameters  $\eta$ ,  $\gamma$ , and on the expected number of events with fixed multiplicity.

Before presenting the estimation methodology, it is worth to comment some properties of the model. Since the entries of  $\Gamma$  are strictly positive, we can apply the Perron-Frobenius Theorem.

We conclude that there exists only one eigenvector with all strictly positive components, whose associated eigenvalue is the spectral radius. Since  $\mathbb{E}[\lambda_t^i] > 0$  for all  $i = 1, \dots, N$ , the spectral

radius corresponds to  $1 - \eta$ . Incidentally, we notice that all the eigenvalues of  $\Gamma$  are real. This property readily follows from observing that  $\Gamma$  is the product of two symmetric matrices, and  $D$  is diagonal and positive definite. Denoting with  $\sqrt{D}$  the square root of the matrix  $D$ ,  $\Gamma$  is *similar* to  $\sqrt{D}^{-1}D\Sigma\sqrt{D}$ , which is by construction symmetric. Interestingly, if  $\Gamma$  is diagonal dominant, i.e. if  $|\Gamma_{ii}| > \sum_{j \neq i} |\Gamma_{ij}|$  for  $i = 1, \dots, N$ , the eigenvalues are also strictly positive.

### 3.1 Estimation and empirical results

An efficient estimation of the model parameters through likelihood maximization is certainly possible. However, we prefer to follow a methodology inspired by the weighted method of moments, which is focused on some of the empirical features that we aim at reproducing. In particular, the following two conditional expectations play a crucial role

$$f_{\tau}^{(1)}(M; J) := \mathbb{E} \left[ \mathbb{I}_{t' \in (t, t+\tau]} \text{ s.t. } M_{t'} \geq J \middle| M_t \geq M \right], \quad (3)$$

$$f_{\tau}^{(2)}(M) := \mathbb{E} \left[ M_{t'} \middle| M_t \geq M, \exists t' \in (t, t+\tau] \text{ s.t. } M_{t'} > 0 \right]. \quad (4)$$

The first quantity,  $f_{\tau}^{(1)}(M; J)$ , is the probability of observing a systemic event with multiplicity at least  $J$  inside a time interval of length  $\tau$  after a cojump of multiplicity  $M_t$  larger than or equal to  $M$ . It therefore measures the probability that a cojump of multiplicity at least  $M$  triggers a systemic cojump ( $J$  fixes the threshold for a systemic cojump). The second quantity,  $f_{\tau}^{(2)}(M)$ , is the average multiplicity of the cojumps inside a time interval of length  $\tau$  after a cojump of multiplicity  $M_t$  larger than or equal to  $M$ . It therefore measures the typical cojump



multiplicity triggered by a cojump of multiplicity at least  $M$ .

We use  $f_\tau^{(1)}(M; J)$  (for fixed  $J$  and  $\tau$ ) and  $f_\tau^{(2)}(M)$  (for fixed  $\tau$ ) to estimate the three model parameters  $\eta$ ,  $\gamma$ , and  $\beta$ . Since we are not able to compute analytically the moments of  $f_\tau^{(1)}(M; J)$  and  $f_\tau^{(2)}(M)$  from the model, we perform Monte Carlo simulations with fixed parameters. Specifically, given a multiplicity  $M$ , we indicate with  $a_d(M)$ ,  $a_m(M)$ , and  $\delta_d(M)$ ,  $\delta_m(M)$  the average values, and associated standard errors, of the conditional expectations in Equations (3) and (4) computed from the data (d) and Monte Carlo experiments (m). Then, for each expectation  $f_\tau^{(i)}$  ( $i = 1, 2$ ) we consider the loss function

$$\chi_{(i)}^2 = \sum_{M \in S} \frac{(a_d - a_m)^2}{\delta_d^2 + \delta_m^2},$$

where the sum is taken over a set of multiplicities  $S$ . The optimizer then searches for the model parameters which minimize the total loss function  $\chi_{(1)}^2 + 0.5\chi_{(2)}^2$ . Given the reduced number of parameters, we perform an exhaustive search over a large region of the three-dimensional parameter space on a 0.05-spaced grid.

As an example of the estimation procedure and to discuss the properties of the fitted model, we consider in detail the case of  $N = 140$  highly liquid assets of the Russell 3000 Index in 2013. We fix  $J = 10$  in Equation (3),  $\tau = 5$  in Equations (3) and (4),  $S = \{5, 10, 15, \dots, 65, 70\}$  and look for the parameters that minimise the total loss function. The global minimum corresponds to the values  $\eta = 0.15$ ,  $\beta = 0.6$ ,  $\gamma = 2.65$ . Thus, 85% of the cojump activity is explained by the excitation mechanism, and only 15% can be attributed to the baseline intensity. The typical

timescale of the memory is  $1/\beta \simeq 1.67$  minutes. The relatively low value of  $\gamma$  indicates a strong cross-excitation between different multiplicities. The left panel of Figure 5 reports a

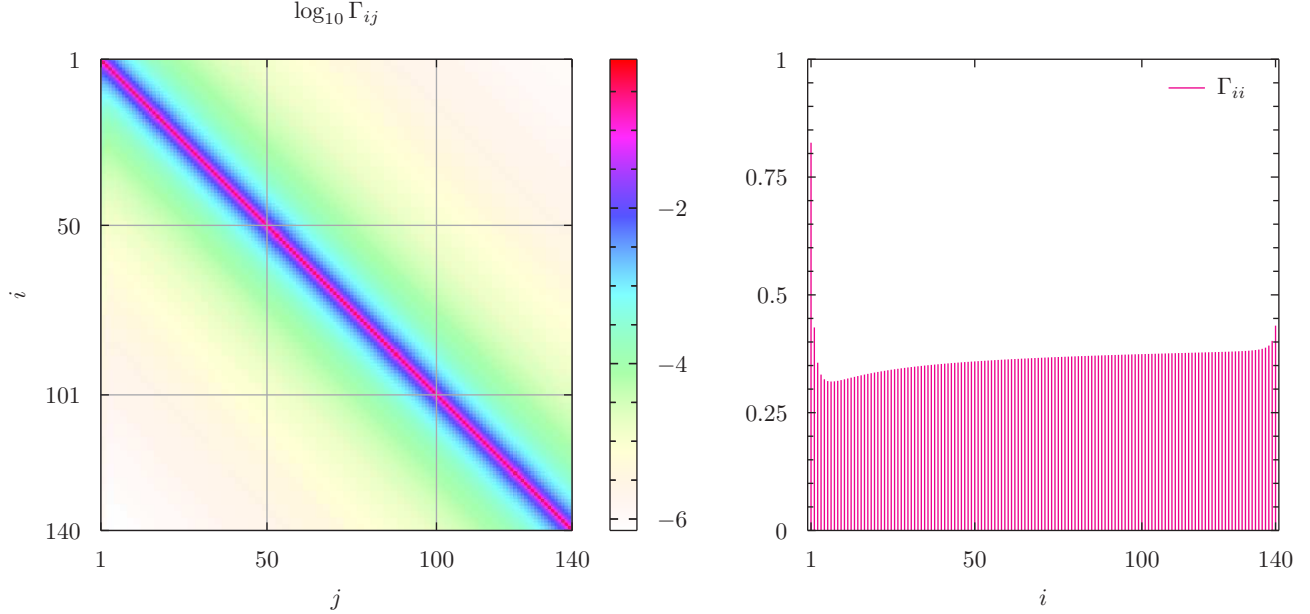


Figure 5: **Components of the  $\Gamma$  matrix.** Left panel: Contour plot of the logarithm of the matrix entries  $\Gamma_{ij} := \alpha_{ij}/\beta_{ij}$  for  $\beta_{ij} = \beta = 0.6$  for all  $i, j = 1, \dots, 140$ ,  $\eta = 0.15$ , and  $\gamma = 2.65$ . Right panel: Linear plot of the diagonal entries of  $\Gamma$  as a function of the multiplicity  $i$ .

contour plot of the logarithmic value of  $140 \times 140$  entries of the  $\Gamma$  matrix. Coherently with the definitions given above,  $\Gamma_{ij}$  for fixed  $i$ , is the impact of past events with multiplicity  $j$  on the multiplicity  $i$ . The largest value corresponds to the diagonal term  $\Gamma_{ii} = D_{ii}$  and quantifies the shock of the intensity due to a self-exciting effect. Then, moving away from the  $\Gamma_{ii}$ , the kernel matrix decreases symmetrically along the row according to a hyperbolic scaling with tail index  $\gamma = 2.65$ . The parameter  $\eta$  rescales the level of the main diagonal of the matrix  $\Gamma$ , reported in the right panel of Figure 5, and determines the degree of stationarity of the process. In Figure 6 we plot the complete spectrum of the matrix  $\Gamma$ . As expected, the largest value

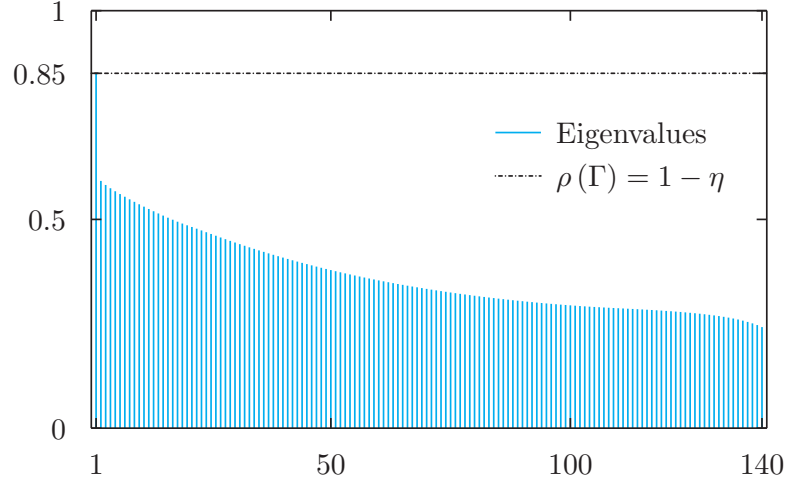


Figure 6: **Eigenvalue spectrum of the  $\Gamma$  matrix.** The spectral radius  $\rho(\Gamma)$  corresponds to  $1 - \eta$ . Since  $\eta = 0.15$ , and more generally for  $0 < \eta < 1$ , the multidimensional Hawkes process describing the stochastic evolution of the multiplicity remains stationary. For the chosen parameter values, we verified numerically that  $\Gamma$  satisfies the diagonal dominant condition and so all its eigenvalues are strictly positive.

corresponds to  $1 - \eta = 0.85$ , while the positive definiteness of all the eigenvalues follows from the evidence, verified numerically, that the matrix is diagonal dominant. To test whether the model is able to reproduce self-consistently the unconditional distribution of multiplicities, we perform a simulation experiment. The parameters  $\eta$ ,  $\beta$ , and  $\gamma$  are set equal to the empirical values, and the entries of the matrix  $\Gamma$  are specified in terms of the empirical frequencies. Then, we generate from a 140-dimensional Hawkes process a synthetic time series with length equal to the length of the real sample, i.e. 96,861. Figure 7 reports the Complementary of the Cumulative Distribution Function of the cojump multiplicities of the empirical data (bold line), and those from the numerical experiment (dashed line). The model correctly reproduces the stationary distribution of the multiplicities observed in empirical data.

Finally, Figure 8 reports the quantities  $f_{\tau}^{(1)}(M; J)$  and  $f_{\tau}^{(2)}(M)$  computed from real and

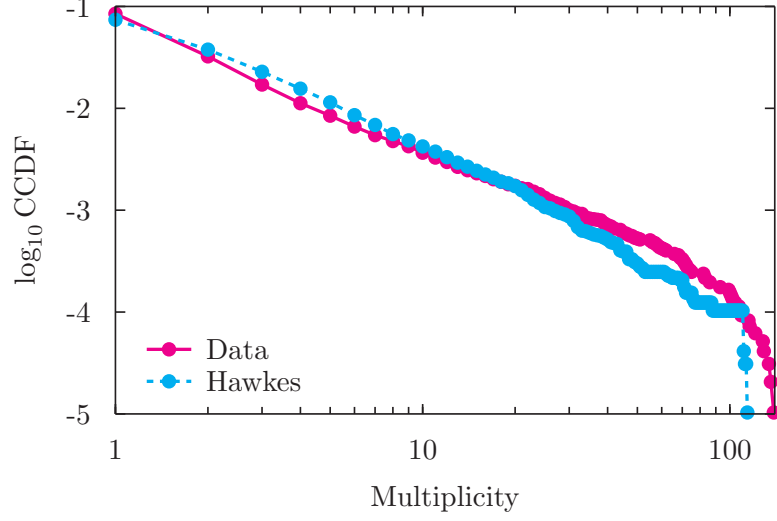


Figure 7: **Cojump multiplicity distribution in real and simulated data.** Log-log plot of the Complementary of the Cumulative Distribution Function of the cojump multiplicities. The bold line corresponds to the empirical distribution measured from the Russell 3000 data sample, 140 assets, during year 2013. The dashed line is the distribution obtained from a simulation of the multidimensional Hawkes process. The total number of minutes drawn from the simulation coincides with the length of the empirical time series and is equal to 96,861.

simulated data. The solid line corresponds to the empirical probabilities, the dotted line to the results from the Hawkes model, and as a benchmark case we also show the result from a Poisson experiment on the multiplicity time series (dashed line). It is evident that dropping the lagged correlations we obtain an unrealistic description of the multiplicity process. The Hawkes model, on the contrary, fits well the empirical data and therefore adequately describes the cross-excitation mechanism between systemic cojumps. Some discrepancies are observable for  $J = 60$ , but the general shape of the curve and its level are well reproduced and the Hawkes model is a huge improvement with respect to the benchmark case. This evidence confirms that the larger is the value of the conditioning multiplicity the greater is the probability that in the subsequent minutes an event with large multiplicity happens.

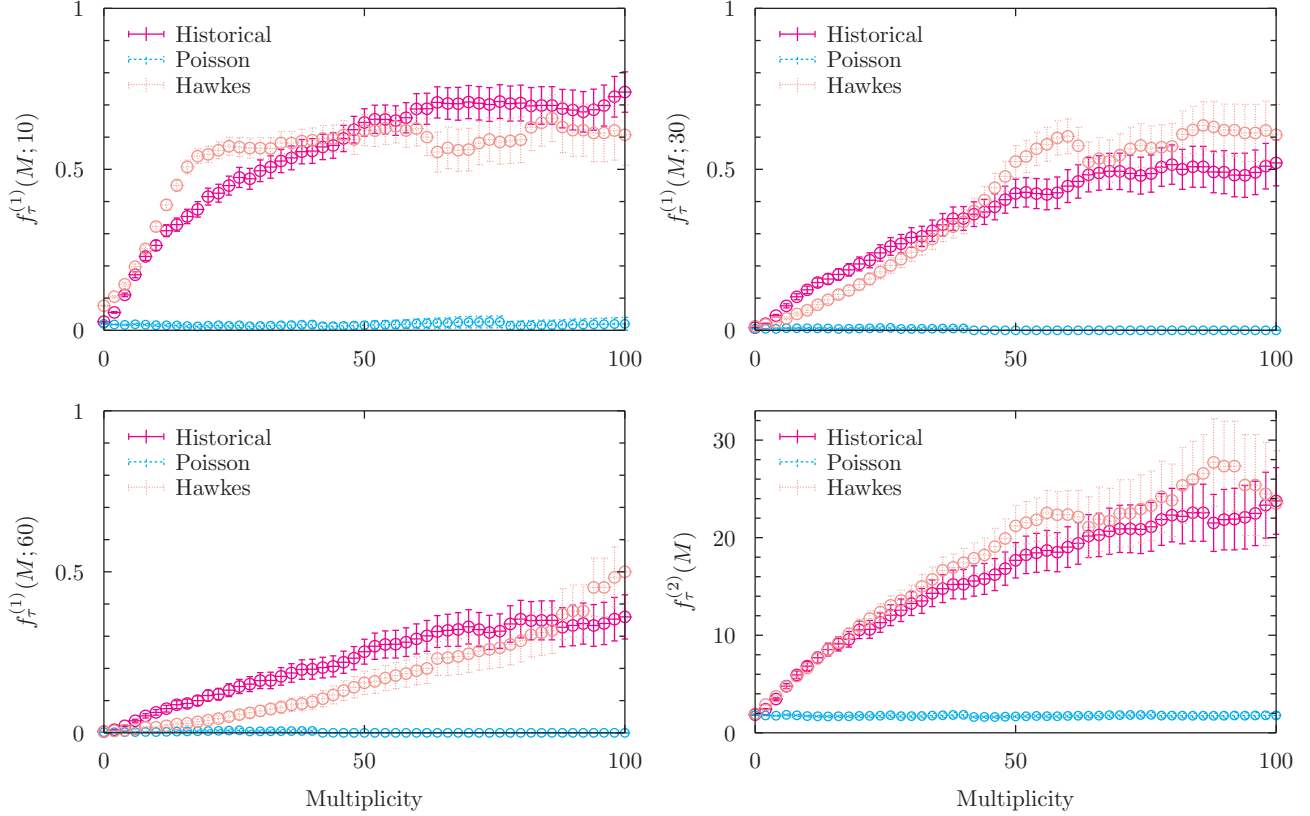


Figure 8: **Effects of cojump multiplicity on future systemic events.** Top left panel: Probability that a cojump with multiplicity larger than or equal to 10 occurs in a  $\tau = 5$  minute interval following a cojump at time  $t$  with multiplicity  $M_t \geq M$ . Plots are obtained from historical and simulated data. The error bars represent standard errors. Top right and bottom left panels: Threshold 10 replaced by 30 and 60, respectively. Bottom right panel: Expected amplitude of the cojumps in a  $\tau = 5$  minute interval following a cojump with multiplicity  $M_t \geq M$ .

## 4 Conclusion

By investigating a portfolio of highly liquid stocks, our research enlightens a clear evidence: Since 2001 the total number of extreme events has remarkably diminished, but the number of occurrences where a sizable fraction of assets jump together has increased. This trend is more and more pronounced as we consider events of higher and higher multiplicity. This evidence is a clear mark that markets are nowadays more and more interconnected and a

strong synchronization between jumps of different assets is present.

Factors responsible for the appearance of extreme movements can be either exogenous or endogenous. The former case is linked to the release of macro-economic news impacting the price dynamics, while the latter may result from unstable market conditions, such as a temporary lack of liquidity. Quite unexpectedly, only a minor fraction (up to 40%) of the cojumps involving a large number of assets can be attributed to exogenous news. The remaining 60% suggests that a more intriguing endogenous mechanism is taking place. Furthermore, it appears that the synchronization among different assets has increased through the recent years. We hypothesize that a major role is played by the dramatic and widespread diffusion of algorithmic trading. Thanks to the technological innovation, faster information processing is responsible for the more rapid propagation of large price movements through different assets. We also provide the evidence that after a highly systemic instability, as manifested by large multiplicity cojumps, (i) the probability that another systemic event takes place in the near future significantly increases, and (ii) the multiplicity of the next event is strongly correlated with the triggering one.

The low timescale of the memory of the exciting effects and the strong persistence of the cross-excitation among different multiplicities support the idea that, to achieve an accurate description of high frequency price dynamics, we should abandon conventional modeling assumptions. Coherently, we propose an innovative approach to the collective behavior of assets' prices based on the Hawkes description of the multiplicity process. Our model well describes the short term dynamics of systemic instabilities while preserving a remarkable parsimony in the number of parameters. Thus, it provides a realistic description of the market behavior which

is of prime importance from several perspectives, from trading to risk control, and market designing.

## References

- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2000. Great realizations. *Risk* 13, 105–108.
- Andersen, T. G., Bollerslev, T., Dobrev, D., 2007. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics* 138 (1), 125–180.
- Andersen, T. G., Bollerslev, T., Frederiksen, P., Ørregaard Nielsen, M., 2010. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics* 25 (2), 233–261.
- Bacry, E., Delattre, S., Hoffmann, M., Muzy, J. F., 2013. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance* 13, 65–77.
- Bacry, E., Mastromatteo, I., Muzy, J.-F., 2015. Hawkes processes in finance. *Market Microstructure and Liquidity* 1 (01), 1550005.
- Bajgrowicz, P., Scaillet, O., Treccani, A., 2015. Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science* 62 (8), 2198–2217.

- Barndorff-Nielsen, O. E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1–48.
- Bauwens, L., Hautsch, N., 2009a. Modelling financial high frequency data using point processes. In: *Handbook of financial time series*. Springer, pp. 953–979.
- Bauwens, L., Hautsch, N., 2009b. Modelling financial high frequency data using point processes. In: Mikosch, T., Kreiss, J.-P., Davis, R. A., Andersen, T. G. (Eds.), *Handbook of Financial Time Series*. Springer, Berlin, pp. 953–979.
- Bollerslev, T., Law, T. H., Tauchen, G., 2008. Risk, jumps, and diversification. *Journal of Econometrics* 144 (1), 234–256.
- Bollerslev, T., Todorov, V., Li, S. Z., 2013. Jump tails, extreme dependencies, and the distribution of stock returns. *Journal of Econometrics* 172 (2), 307–324.
- Bormetti, G., Calcagnile, L. M., Treccani, M., Corsi, F., Marmi, S., Lillo, F., 2015. Modelling systemic price cojumps with Hawkes factor models. *Quantitative Finance* 15, 1137–1156.
- Bowsher, C. G., 2007. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics* 141 (2), 876–912.
- Caporin, M., Kolokolov, A., Renò, R., 2014. Multi-jumps. Available at SSRN 2488603.
- CFTC, SEC, 2010. Findings regarding the market events of May 6, 2010. Report of the staffs



of the CFTC and SEC to the joint advisory committee on emerging regulatory issues.

Available online at: [www.sec.gov/news/studies/2010/marketevents-report.pdf](http://www.sec.gov/news/studies/2010/marketevents-report.pdf).

Corsi, F., Pirino, D., Renò, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* 159, 276–288.

Daley, D. J., Vere-Jones, D., 2003. *An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods*. Springer, Heidelberg.

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

Filimonov, D., Sornette, D., 2012. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E* 85, 056108–1–9.

Gerig, A., 2013. High-frequency trading synchronizes prices in financial markets. Available at SSRN 2173247.

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

Golub, A., Keane, J., Poon, S.-H., 2012. High frequency trading and mini flash crashes. arXiv preprint arXiv:1211.6667.

Gomber, P., Arndt, B., Lutat, M., Uhle, T., 2011. High-frequency trading. Available at SSRN 1858626.

- Hardiman, S. J., Bercot, N., Bouchaud, J.-P., 2013. Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B* 86 (10), 1–9.
- Hawkes, A. G., 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 83–90.
- Johnson, N., Zhao, G., Hunsader, E., Qi, H., Johnson, N., Meng, J., Tivnan, B., 2013. Abrupt rise of new machine ecology beyond human response time. *Scientific Reports* 3.
- Joulin, A., Lefevre, A., Grunberg, D., Bouchaud, J.-P., 2008. Stock price jumps: News and volume play a minor role. *Wilmott Magazine* Sep/Oct, 1–7.
- Kirilenko, A., Kyle, A. S., Samadi, M., Tuzun, T., 2017. The Flash Crash: High frequency trading in an electronic market. *The Journal of Finance* DOI: 10.1111/jofi.12498.
- Lee, S. S., 2012. Jumps and information flow in financial markets. *Review of Financial Studies* 25 (2), 439–479.
- Lee, S. S., Mykland, P. A., 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies* 21 (6), 2535–2563.
- Lewis, M., 2014. *Flash boys: a Wall Street revolt*. WW Norton & Company.
- MacIntosh, J. G., 2013. High frequency traders: Angels or devils? *CD Howe Institute Commentary* 391.

- Muni Toke, I., 2011. “Market making” in an order book model and its impact on the spread. In: Abergel, F., Chakrabarti, B. K., Chakraborti, A., Mitra, M. (Eds.), *Econophysics of Order-Driven Markets*. Springer-Verlag, Milan, pp. 49–64.
- Muni Toke, I., Pomponio, F., 2012. Modelling trades-through in a limit order book using Hawkes processes. *Economics: The Open-Access, Open-Assessment E-Journal* 6, 1–23.
- Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* 83, 9–27.
- Petersen, A. M., Wang, F., Havlin, S., Stanley, H. E., 2010a. Market dynamics immediately before and after financial shocks: Quantifying the Omori, productivity, and Bath laws. *Physical Review E* 82 (3), 036114.
- Petersen, A. M., Wang, F., Havlin, S., Stanley, H. E., 2010b. Quantitative law describing market dynamics before and after interest-rate change. *Physical Review E* 81 (6), 066121.
- Rambaldi, M., Bacry, E., Lillo, F., 2016. The role of volume in order book dynamics: a multivariate Hawkes process analysis. *Quantitative Finance*, 1–22.
- Rambaldi, M., Pennesi, P., Lillo, F., 2015. Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Physical Review E* 91 (1), 012819.
- Vere-Jones, D., 1970. Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society, Series B* 32, 1–62.

Vere-Jones, D., Ozaki, T., 1982. Some examples of statistical inference applied to earthquake data. *Annals of the Institute of Statistical Mathematics* 34, 189–207.

## Acknowledgments

LMC, GB, MT, SM, FL conceived and designed the study; LMC, GB, MT performed the experiments; LMC, GB, FL analyzed the data and wrote the paper. FL acknowledges partial support by the grant SNS13LILLB “Systemic risk in financial markets across time scales” and by the European Community H2020 Program under the scheme INFRAIA-1- 2014-2015: Research Infrastructures, grant agreement no. 654024 SoBigData: Social Mining & Big Data Ecosystem. GB acknowledges research support by the Scuola Normale Superiore Grant SNS14BBORMETTI. Authors thanks E. Bacry, V. Filimonov, and M. Rambaldi for useful discussions. The opinions expressed here are solely those of the authors and do not represent in any way those of their employers.

## A Macronews summary statistics

Table 1: **News dataset.** Number of news announcements, organized by year and news category.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	all years
ADP Employment Report	MEA <sup>a</sup>	0	1	0	0	0	0	0	0	0	0	0	0	1
Beige Book	MEA	8	8	7	8	8	8	8	8	8	8	8	8	103
Business Inventories	MEA	0	0	3	6	5	12	12	12	12	12	12	12	110
Chairman Press Conference	MMI <sup>b</sup>	0	0	0	0	0	0	0	0	0	3	5	4	12
Chicago PMI	MEA	12	12	12	12	12	12	12	12	12	12	12	11	155
Construction Spending	MEA	12	12	12	12	12	12	12	12	12	12	12	12	156
Consumer Confidence	MEA	12	12	12	12	12	12	12	12	13	12	12	11	156
Consumer Sentiment	MEA	12	23	24	24	24	24	24	24	24	24	24	24	298
Dallas Fed Mfg Survey	MEA	0	0	0	0	0	0	0	0	0	8	12	12	32
Durable Goods Orders	MMI	1	1	0	0	0	0	0	0	0	0	0	0	2
EIA Petroleum Status Report	MEA	0	0	0	14	51	52	52	53	52	52	52	52	482
Empire State Mfg Survey	MEA	0	0	0	0	0	1	0	0	0	0	0	0	1
Existing Home Sales	MMI	12	12	12	12	12	12	12	12	12	12	12	12	156
Factory Orders	MEA	12	12	12	12	12	12	12	12	12	12	12	12	156
FOMC Forecasts	MMI	0	0	0	0	0	0	0	0	0	0	5	4	9
FOMC Meeting Announcement	MMI	8	8	8	8	8	8	8	8	8	8	8	8	104
FOMC Minutes	MMI	8	8	8	8	8	8	8	8	8	8	8	7	103
Housing Market Index	MEA	0	0	0	0	12	12	12	12	12	12	12	12	96
ISM Mfg Index	MMI	12	12	12	12	12	12	12	12	12	12	12	12	156
ISM Non-Manufacturing Employment Index	MEA	0	0	0	0	0	0	11	12	0	0	0	0	23
Motor Vehicle Sales	MEA	0	12	0	0	0	0	0	0	0	0	0	0	12
New Home Sales	MMI	13	12	12	12	12	12	12	12	12	12	12	12	157
Pending Home Sales Index	MEA	0	0	0	0	12	12	12	12	13	12	12	12	97
Personal Income and Outlays	MMI	1	0	0	0	0	0	0	0	0	0	0	0	1
Philadelphia Fed Survey	MMI	12	12	12	12	12	12	12	12	12	12	12	12	156
Retail Sales	MMI	0	0	0	0	0	0	1	0	0	0	0	0	1
Treasury Budget	MEA	12	12	12	12	12	12	11	12	12	12	12	10	153
all categories		147	169	159	174	212	245	244	256	256	255	266	259	2888

<sup>a</sup>Merit Extra Attention according to the classification provided by Econoday, Inc.

<sup>b</sup>Market Moving Indicator according to the classification provided by Econoday, Inc.