Relating Group Size and Posting Activity of an Online Community of Financial Investors: Regularities and Seasonal Patterns

P. Racca
Department of Economics and Statistics, University of Torino, Italy

R. Casarin
Department of Economics, University Ca' Foscari of Venice, Italy

Pierpaolo Dondio
Technological University Dublin, pierpaolo.dondio@tudublin.ie

F. Squazzoni
Department of Economics and Management, University of Brescia, Italy

Recommended Citation

Racca, P., Casarin, R. & Dondio, P. Relating group size and posting activity of an online community of financial investors: Regularities and seasonal patterns. *Physica A: Statistical Mechanics and Its Applications*, vol. 493, March, 458-466. doi:10.1016/j.physa.2017.11.143

Physica A 493 (2018) 458–466

Contents lists available at ScienceDirect

Physica A

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

Relating group size and posting activity of an online community of financial investors: Regularities and seasonal patterns

P. Racca [a], R. Casarin [b], P. Dondio [c], F. Squazzoni [d,*]

[a] Department of Economics and Statistics, University of Torino, Italy
[b] Department of Economics, University Ca' Foscari of Venice, Italy
[c] School of Computing, Dublin Institute of Technology, Ireland
[d] Department of Economics and Management, University of Brescia, Italy

HIGHLIGHTS

- We analysed the relation between the number of active users and posts shared on an online forum for financial investors.
- We found that the relation follows a power law, with exponent greater than 1.
- We found day-of-the-week and hour-of-the-day patterns in the fluctuations of the exponent.

ARTICLE INFO

Article history:
Received 16 June 2016
Received in revised form 2 November 2017
Available online 7 December 2017

Keywords:
Financial forum
Group size
Collective learning
Post contribution
Densification law

ABSTRACT

Group size can potentially affect collective activity and individual propensity to contribute to collective goods. Mancur Olson, in his *Logic of Collective Action*, argued that individual contribution to a collective good tends to be lower in groups of large size. Today, online communication platforms represent an interesting ground to study such collaborative dynamics under possibly different conditions (e.g., lower costs related to gather and share information). This paper examines the relationship between group size and activity in an online financial forum, where users invest time in sharing news, analysis and comments with other investors. We looked at about 24 million messages shared in more than ten years in the *finanzaonline.com* online forum. We found that the relationship between the number of active users and the number of posts shared by those users is of the power type (with exponent $\alpha > 1$) and is subject to periodic fluctuations, mostly driven by hour-of-the-day and day-of-the-week effects. The daily patterns of the exponent showed a divergence between working week and weekend days. In general, the exponent was lower before noon, where investors are typically interested in market news, higher in the late afternoon, where markets are closing and investors need better understanding of the situation. Further research is needed, especially at the micro level, to dissect the mechanisms behind these regularities.

* Corresponding author.
E-mail addresses: paolo.racca@unito.it (P. Racca), r.casarin@unive.it (R. Casarin), pierpaolo.dondio@dit.ie (P. Dondio), flaminio.squazzoni@unibs.it (F. Squazzoni).

https://doi.org/10.1016/j.physa.2017.11.143
0378-4371/

1. Introduction

Understanding determinants of collective learning is one of the most important ambitions of scientists from several disciplines, both in social and hard sciences: biologists, philosophers, sociologists, psychologists and mathematicians, among others, have dealt with this issue, either theoretically or empirically [1–4]. The study of knowledge sharing mechanisms and information transfer within a group is central in this respect: psychologists and sociologists look at knowledge sharing dynamics with regard to relational and personal cognitive aspects [5]. In managerial and organizational settings these issues are key for the maximization of value creation [6]. Opinion dynamics, conflicts, consensus formation and, more in general, human interactions in collaborative environments have also been studied in statistical physics and network science [7–12].

From an economic perspective, the potential value associated to shared information implies that collective learning depends on some forms of public-good dilemma, where an individual has to balance *self*-interest against *common* interest at the group level. Several studies have addressed the dependence of cooperation and collective learning on a variety of factors, such as group size, members' heterogeneity, the presence and role of leaders, the access to relevant information, monitoring effects and costs and benefits related to participation [1,13–18]. Mancur Olson's milestone book *The Logic of Collective Action* [19], published in 1965, proposed a paramount hypothesis that is commonly referred to as *group-size paradox* and can be summarized as follows: individual contribution to a collective good tends to be lower, the larger the size of the group. This hypothesis and the effect of group heterogeneity have been extensively examined in a variety of fields (see [16] for a recent review). More specifically, group-size paradox has been related to the nature of the collective good, in particular to its degree of privateness/publicness [20]. Oliver and Marwell pointed out that "Olson's group size argument is clearly correct *only* when the good has zero jointness of supply, i.e., when the cost of providing the good is proportional to the number who share it" [21]. Conversely, they argued that complete jointness of supply (the good costs the same no matter how many individuals enjoy it), "translates into a positive effect of group size on (1) the probability that *someone* in a group will provide the good, and (2) the total amount of contributions from the group".

Recently, the disruptive evolution and diffusion of new communication technologies has created a new context to understand these topics better. [22], for instance, contended that evolving technologies and electronic communication have substantially lowered organizational costs, introducing possibilities that Olson could not have imagined in the Sixties. On the same line, [23] noticed that, in general, self-organizing online groups, forums and meetups could hardly fit the paradigm implied by Olson's theory. In fact, they proposed an extension of the traditional collective action theory that could also account for circumstances of low coordination, information and communication costs. As an example of these lowered costs, many online platforms stimulated an increase of information sharing among users: *Wikipedia* and the open-source philosophy are examples of this. It is worth noting that in many cases these communication systems make even valuable information progressively less *excludable* and, consequently, more distant from the ideal *private good*.

In this paper, we studied the relationship between group size and overall post production in an online forum for financial investors. This is an online platform where any registered user can read and share posts about financial assets, stocks, news and any other topic related to finance. Given that registration and participation are totally free, there is no way to prevent anyone from registering and accessing the information shared by other members without contributing in turn. Although it is actually possible to create private groups where only selected users are allowed to read and write messages, this is the exception to the rule. Therefore, we can consider the content shared in this forum as non-excludable goods. Non-excludability is, indeed, a fundamental characteristic of a public good; the other one is non-rivalry: a good is non-rivalrous if the consumption by a consumer does not prevent other consumers from simultaneously consuming it. This point is not trivial, given the type of content. On the one hand, recent studies in empirical finance have shown that messages shared on online financial forums are far from being just noise; on the contrary, they contain valuable information, which might be collected and analysed to predict some market dynamics [24–26]. On the other hand, this kind of forum could provide even an ordinary user with good opportunities for gains. Let us assume that *user 1* has reasons to believe with a high degree of confidence that a stock will perform very well the next day. After taking a large long position on that stock, he/she might see no reasons to hide his/her motivations. In fact, if other users were convinced by his/her argument and took themselves a long position, they would just contribute to push the price of the stock in the direction that *user 1* expected. The point is: would that piece of information be non-rivalrous? If one assumes that all members of the forum are price takers, it would probably be the case and this is probably a good approximation for most stocks and futures. However, it cannot be excluded that stocks with low market capitalization might be sensitive to rumours [27]. The fact that a post might disclose a good trading opportunity suggests that there might be some rivalry in the exploitation of that information. As a consequence, the information shared on a financial forum cannot be classified as a pure public good, nor as a private one.

To understand these problems better, we focused on the *finanzaonline.com* forum, the leading Italian financial online community. Established at the end of 1999 by *Brown Editore*, an independent and highly influential publishing company specializing in high quality economic and financial information, *finanzaonline.com* immediately became the main information and communication online platform for Italian investors. The forum grew impressively in size and activity, from less than a thousand monthly posts shared in 2002 to hundreds of thousands in 2008. Moreover, the time span we analysed included periods of high tension and uncertainty on the markets, such as the 2008 financial crisis and the European sovereign debt crisis. This allowed us to observe the relationship of the variables across orders of magnitude and to check for its robustness.

We found robust regularities consistent with a power functional form between the number of active users and the number of posts. We then focused on the value of the parameters in order to understand to which extent an increase in size of the

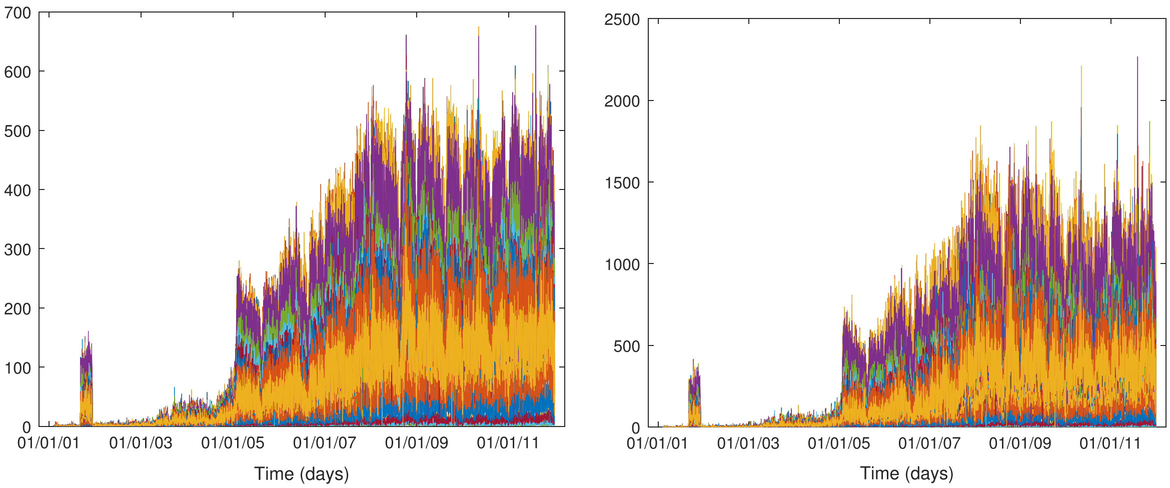

Fig. 1. Time series (synchronized) of the number of hourly active users (left) and shared posts (right): different lines correspond to different hours of the day. Horizontal axis indicates daily time.

group of contributors to the forum was related to the total post contribution. Specifically, we wanted to understand whether the change in total contribution was proportional to the active group size. We investigated the robustness of this regularity through time and its seasonality, controlling for the degree of market uncertainty.

The paper is structured as follows. The second section describes the dataset and the econometric model, the third one presents our findings. The last section discusses our findings, some limitations of this work and prospects for future investigation.

2. Data and methods

2.1. Data overview

Data included more than 24 million messages posted on *finanzaonline.com* forum from 2001 to 2011 by more than 50 thousands users. Fig. 1 shows the hourly time series of the number of active users and messages shared on the forum. We measured the number of *active* users in a given hour as the number of users who posted during that hour. The hourly activity on the forum was characterized by circadian patterns and also showed inter-day noise, partly due to a day-of-the-week effect and to exogenous market forces. In this regard, given the role of uncertainty in affecting users' behaviour [28], we used the daily Euro Stoxx 50 Volatility Index time series to take into account the level of volatility perceived by investors on the European markets.

Fig. 2 shows a scatterplot of the daily number of posts against the daily number of active users (data from 2001 to 2011): dots seem to fall on a straight line across orders of magnitude, which suggests a power relation between the two quantities. In this regard, it is worth noting that the availability of large datasets has recently led to the discovery of a similar kind of relationship in several social networks. In particular, [29] described a *densification law* between nodes and edges in the evolution of social and technological graphs: they found that the number of edges tends to increase over time as a power of the number of nodes, with a power exponent that ranges between 1 and 2. This was observed in different real graphs such as academic paper citation networks, patent citation networks, rooter networks in the Internet, email networks and networks of blogs [29–31]. These networks show a *densification* in that the number of links among nodes increases superlinearly with respect to the number of nodes, so that the average number of edges per node tends to increase over time.

Differently from platforms such as Facebook or Twitter, *finanzaonline.com* forum does not require users to establish *friendship* or *follow* links to interact. This means that we could not rely on an explicit architecture of social ties between users. More specifically, in our context, the only connections among users were post quotations. Given that it is rather common to quote the *last* message to post a reply, *quotation links* represent – at least at a short time scale – a loose proxy for ties between pairs of users. For this reason, our analysis did not look at the growth in relational links compared to the growth in the number of nodes. Conversely, we studied the dependence of the contribution of messages per unit of time on the size of the group of active users. In the following paragraphs, we presented the model that we estimated to investigate seasonality effects and robustness of this relationship.

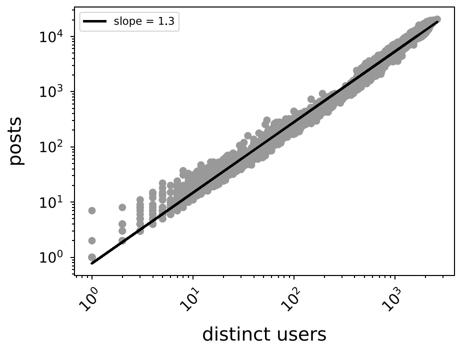

Fig. 2. Daily number of posts *vs* active users (period 2001–2011), double logarithmic scale.

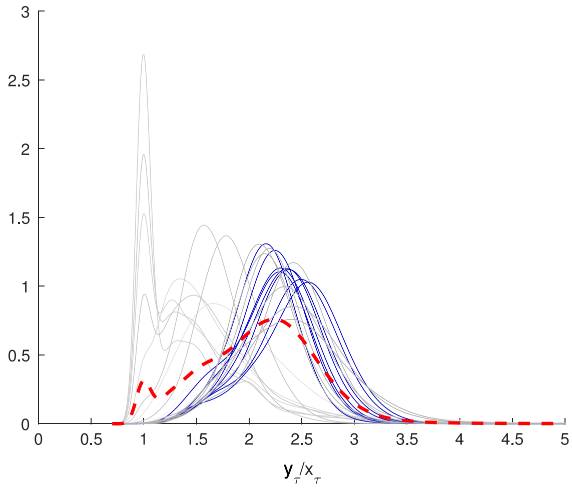

Fig. 3. Kernel density estimate of the distribution of the ratio y_τ/x_τ on all data sampled at hourly frequency (red dashed line) and on data sampled every day from 5 PM to 10 AM (closed market, grey lines) and on data sampled every day from 10 AM to 5 PM (open market, blue lines). The sample period is from 1st January 2006 to 31st December 2011. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Mathematical framework

Let y_τ be the number of messages at time τ and let x_τ be the number of active users at time τ. We considered the following model

$$y_\tau = \eta(\tau)x_\tau^{\alpha(\tau)} \tag{1}$$

where $\eta(\tau)$ was a proportionality factor which captured potential driving forces of the forum and $\alpha(\tau)$ was a power coefficient. The functional form of α and η is specified in Section 2.3.

Fig. 3 shows the kernel density estimates of the distribution of the ratio y_τ/x_τ, sampled at hourly frequency. Distributions appear clustered around an average of two/three posts *per capita* during daytime while a shift towards the left side of the graph is visible during night time, when post contribution tends to be more and more sporadic and the probability of a debate becomes extremely low. To study the long-run relationship between number of messages and number of active users, we needed to account for the part of the variability related to regular human activity and periodic behaviour of the forum (seasonal effects). More specifically, the functional form of the power coefficient should account for both a time-invariant component and a periodic deterministic component associated to the seasonal effects with intra-year periodicity.

2.3. Econometric specification

We had a set of observations (x_τ, y_τ), $\tau = 1, \ldots, T$, available at a hourly sampling frequency. We considered 24 hour-specific equations. Let $y_{t,h} = y_\tau$ be the number of messages at the hour h of the day t with $\tau = 24(t-1)+h$, $x_{t,h}$ the number of active users and $\mathbf{z}_{t,h} = (z_{1t,h}, \ldots, z_{kt,h})'$ a vector of covariates. Let $\alpha_{t,h} = \alpha_\tau$ be the time-varying power coefficient and $\kappa_{t,h}\exp\{\boldsymbol{\beta}_h'\mathbf{z}_{t,h}\} = \eta(\tau)$ the proportionality factor.

The econometric specification we used was the following:

$$\log y_{t,h} = \log \kappa_{t,h} + \alpha_{t,h} \log x_{t,h} + \boldsymbol{\beta}_h' \mathbf{z}_{t,h} + \varepsilon_{t,h}, \quad (2)$$

$$\varepsilon_{t,h} \sim \mathcal{N}(0, \sigma_{t,h}^2) \;\; t = 1, \ldots, T_h \text{ and } h = 1, \ldots, 24 \quad (3)$$

with $Cov_t(\varepsilon_{t,h}, \varepsilon_{s,h'}) = 0$ for $h \neq h'$ and $t \geq s$, where $\boldsymbol{\beta}_h = (\beta_{1,h}, \ldots, \beta_{k,h})'$ is a vector of coefficients.

As argued in many empirical studies (e.g. see [32]) strong oscillatory trends, such as seasonal effects, have to be distinguished from the intrinsic fluctuations of the variables of interest to avoid biased estimates of the quantities of interest, such as the power coefficient. Thus, we assumed the following specification of the $\kappa_{t,h}$ and $\alpha_{t,h}$ dynamics:

$$\log \kappa_{t,h} = \log c_h + \sum_{j=2}^{12} \log c_{h,j}^{(m)} D_j^{(m)}(t) + \sum_{j=1}^{6} \log c_{h,j}^{(d)} D_j^{(d)}(t) \quad (4)$$

$$\alpha_{t,h} = a_h + \sum_{j=2}^{12} a_{h,j}^{(m)} D_j^{(m)}(t) + \sum_{j=1}^{6} a_{h,j}^{(d)} D_j^{(d)}(t) \quad (5)$$

where $D_j^{(m)}(t), j = 2, \ldots, 12$ and $D_j^{(d)}(t), j = 1, \ldots, 6$ are two sets of seasonal dummies[1] :

$$D_j^{(m)}(t) = \begin{cases} 1 & \text{if } t \text{ is in the month } j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$D_j^{(d)}(t) = \begin{cases} 1 & \text{if } t \text{ is in the day } j \\ 0 & \text{otherwise} \end{cases} . \quad (7)$$

Coefficients $c_{h,j}^{(d)}$ and $a_{h,j}^{(d)}$, with $j = 1, \ldots, 6$, captured the day-of-the-week periodicity in the proportionality factor and in the power exponent, respectively; analogously, $c_{h,j}^{(m)}$ and $a_{h,j}^{(m)}$, with $j = 2, \ldots, 12$ captured the month-of-the-year periodicity.

Removing fluctuations in the dependent variable that were not of interest for the analysis since they were due to external sources motivated the use of control variables in modelling the proportionality factor dynamics. Our control variables included market volatility and lagged values of the dependent variable (to capture autocorrelation). In particular, we used:

$$\mathbf{z}_{t,h} = (\log v_t, \log v_{t-1}, \ldots, \log v_{t-p}, \quad \log y_{t-1,h}, \log y_{t-2,h}, \log y_{t-3,h}) \quad (8)$$

where v_t was the VSTOXX index at day t and p was set to 20. The VSTOXX index measures the implied volatility across all EURO STOXX 50 realtime options prices and reflects the agents' expectations about future market volatility. As recently showed in the financial literature, models which include realized or implied volatility measures produce substantial improvements in the empirical fit over models that only use daily returns and GARCH (e.g., see [33] and [34]).

Finally, in order to capture the time-variations in the fluctuation range of the error terms we specified a dynamic model for the conditional variance $\sigma_{t,h}^2 = V(\varepsilon_{t,h}|\mathcal{F}_{t-1})$, where $\mathcal{F}_t$ represents the σ-algebra generated by $y_{s,h}, s = 0, \ldots, t, h = 1, \ldots, 24$. We assumed GARCH(1,1) process (see [35]), that is:

$$\sigma_{t,h}^2 = \omega_{0,h} + \omega_{1,h}\varepsilon_{t-1,h}^2 + \omega_{2,h}\sigma_{t-1,h}^2 \quad (9)$$

where $\omega_{0,h}, \omega_{1,h}$ and $\omega_{2,h}$ were parameters to be estimated. $\omega_{1,h}$ measured the influence of random deviations in the previous period on $\sigma_{t,h}^2$ and $\omega_{2,h}$ reflected the persistence of the volatility process. E.g., see [36] and [37] for an analysis of the dynamic properties of the volatility on financial and commodity markets.

Our seasonal-effect model with GARCH innovations was estimated with a quasi maximum likelihood procedure (QMLE). For a review of the estimation methods and the asymptotic properties of the QMLE estimator for univariate GARCH models see [38].

3. Results

We focused the analysis on the period from 1st January 2006 to 31st December 2011 in order to exclude the initial explosive growth of the forum.[2] Fig. 4 shows the results of the estimation of the model for the hours from 5 PM to 6 PM (left side) and from 6 PM to 7 PM (right side). Different lines correspond to plots of the relation evaluated in January and May and on different days. It is worth noting that we used close values of the VSTOXX; as a consequence, for the hours considered in Fig. 4, our daily VSTOXX value v_t was a good proxy of the *current* uncertainty perceived on the European markets. The value of the power coefficient was significantly above 1 in both graphs.

[1] We excluded Saturday from the set of day-of-the-week dummies and January from the month-of-the-year dummies in order to avoid perfect collinearity in the columns of the regression covariates. Therefore, the omitted dummies represented the baseline case (see, for instance, the dashed line in Fig. 4) and all seasonal dummy coefficients represented deviations from the baseline.

[2] In fact, in the early phase of its development, participation and activity might have been very sensitive to exogenous events (e.g., advertising campaigns aimed at attracting new users) for which it would be hard to control in the absence of a precise track record. The initial transient was therefore disregarded here, as in [39]. However, results for the whole period from the 9*th* of January, 2001 to the 31st of December, 2011 did not deviate significantly and are reported in the *Supplementary Material* file.

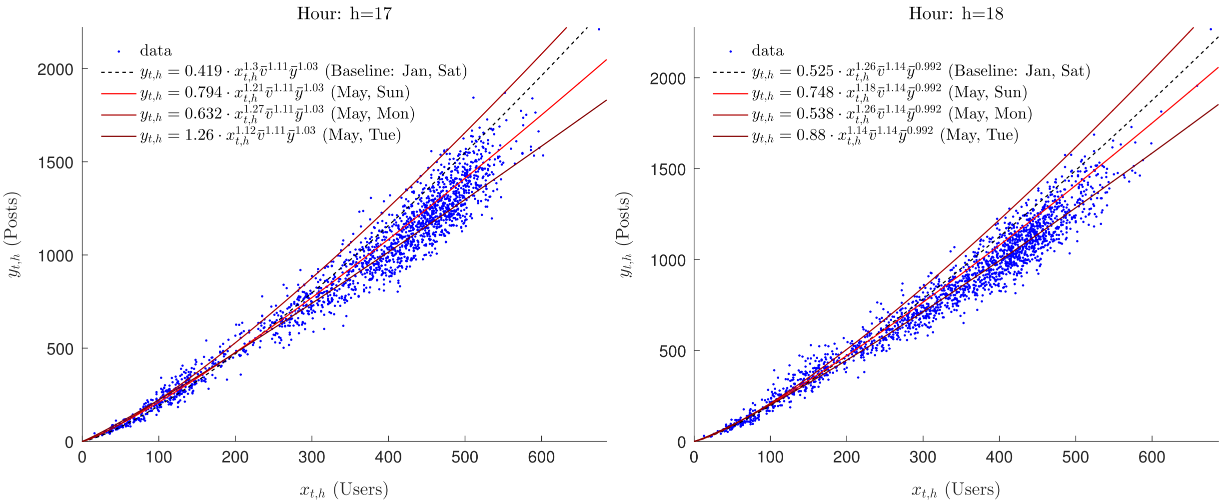

Fig. 4. Model in Eq. (2) fitted from 5 PM to 6 PM (left) and from 6 PM to 7 PM (right), evaluated in different months and on different days ($\bar{y}$ and $\bar{v}$ are whole sample averages of $y_{t,h}$ and v_t).

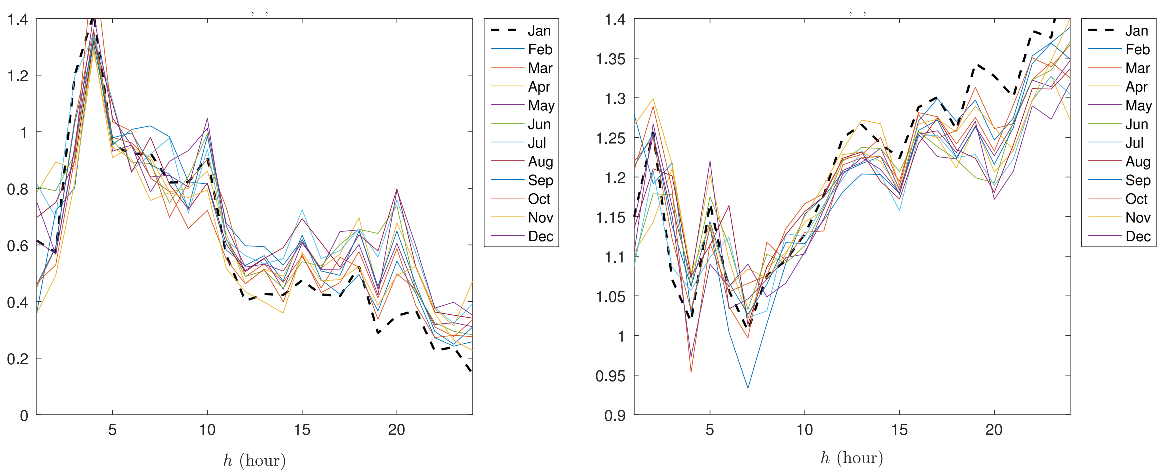

Fig. 5. Month-of-the-year effect: $\kappa_{h,m} = c_h c_{h,m}^{(m)}$ (left) and $\alpha_{h,m} = a_h + a_{h,m}^{(m)}$ (right), for different hours of the day, $h = 1, \ldots, 24$ (horizontal axis), and different months of the year, $m = 2, \ldots, 12$ (coloured lines). Baseline: Saturdays ($d = 7$) of the month of January ($m = 1$) (dashed line).

We then estimated the same model for the 24 h (assuming the value of the control variable constant during the whole day) and we found some variability. However, the main result (α significantly greater than one) was generally confirmed. Note that we also performed checks using lagged values of x in the regressions.

Figs. 5 and 6 show month-of-the-year and day-of-the-week patterns of the estimated coefficients κ and α for the 24 h. On the one hand, we found rather negligible month-of-the-year and day-of-the-week effects, the only exception being a "weekend effect": on Saturdays and Sundays, especially in the afternoon, α tended to be higher and κ tended to be lower.

On the other hand, we found remarkable intraday seasonality patterns for both κ and α coefficients. Specifically, the multiplicative factor κ tended to remain above 1 from the first hours after midnight up to 10/11 AM; then it fell below 1, reaching the minimum values after 10 PM. This could suggest that a group of a fixed size, under the same external conditions, would tend to post comparatively more at 5 AM than at 5 PM. A likely explanation is that, in the early morning, a few users posted messages (even less than ten per hour); furthermore, before the opening of the markets, users typically shared reports about the analysis of the days before, the latest news and potential market movers of the day. This is not the ordinary type of communication of a forum and it can likely be characterized by a higher post contribution per capita. After the opening of the markets, around 10 AM, the need to discuss the first market movements of the trading day attracted more active users on the forum and the ordinary communication patterns were gradually restored. The α coefficient showed an approximately reversed pattern: it remained lower (close to 1) in the early morning until 10/11 AM, then it increased around noon and remained rather stable during the afternoon. It is worth noting that weekend patterns, although more pronounced, were qualitatively similar.

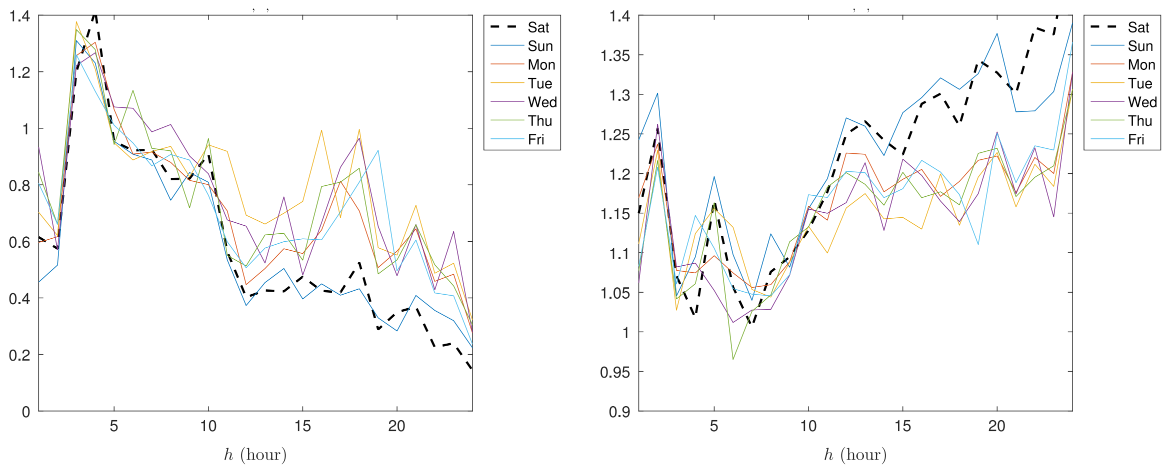

Fig. 6. Day-of-the-week effect: $\kappa_{h,d} = c_h c_{h,d}^{(d)}$ (left) and $\alpha_{h,d} = a_h + a_{h,d}^{(d)}$ (right), for different hours of the day, $h = 1, \ldots, 24$ (horizontal axis), and different days of the week, $d = 1, \ldots, 6$ (coloured lines). Baseline: Saturdays ($d = 7$) of the month of January ($m = 1$) (dashed line).

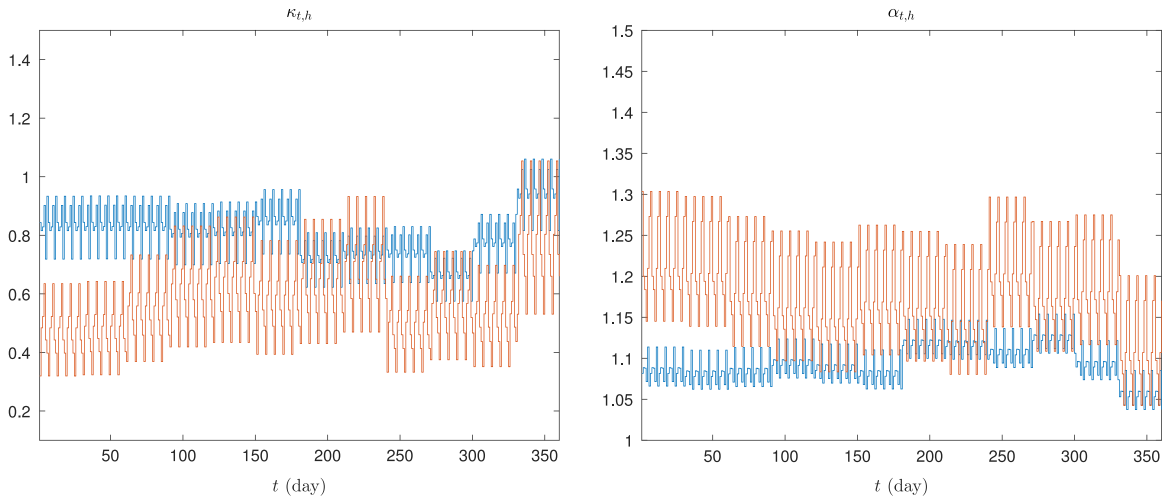

Fig. 7. Estimated evolution (daily frequency) of $\kappa_{t,h}$ and $\alpha_{t,h}$ during one year, at specific hours of the day, i.e. h = 9 AM (blue) and h = 11 PM (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 7 aggregates hourly, day-of-the-week and monthly effects to show the complete one-year estimated periodicity of κ and α, keeping the hour-of-the-day fixed. In particular, each plot compared the hour from 9 AM to 10 AM with the hour from 11 PM to midnight. First, coherently with previous comments, κ tended to be lower at night than in the morning, while the opposite happened for α. Secondly, the variability was higher in both cases for the red line (night) than for the blue line (morning). As shown in Fig. 6, the divergence between working week and weekend days was higher in the second half of the day. Thirdly, coherently with data in Fig. 5, Fig. 7 shows a negative correlation between the monthly levels of κ and α. The multiplicative and power-law effects were concurrent in explaining the variance in the data and their balance was subject to change. However, both parameters were relatively stable and α was greater than 1. Lastly, results suggest that the estimated magnitude of the day-of-the-week effect was higher than the one of the monthly effect.

Finally, we studied the impact of the VSTOXX control variables on the hourly number of posts $y_{t,h}$. Fig. 8 shows the impact of the lagged values of the volatility for different hours of the day. First, the effect of market uncertainty on the number of posts was considerable. Furthermore, the more recent values of volatility had a stronger and positive impact on the number of posts. We found no significant difference between the effect of current day volatility and previous day volatility as a control. On the contrary, values recorded more than two days before day t showed mixed effects.

Looking at the dashed line, which is our best proxy of the uncertainty perceived by forum users, we found a clear dependence of the activity upon the hour of the day, in that the effect was low around noon and increased in the late afternoon and towards the evening. This suggests that investors tend to control their *social* reaction to market tension when markets are open while they tend to engage more in discussions at the end of the trading day.

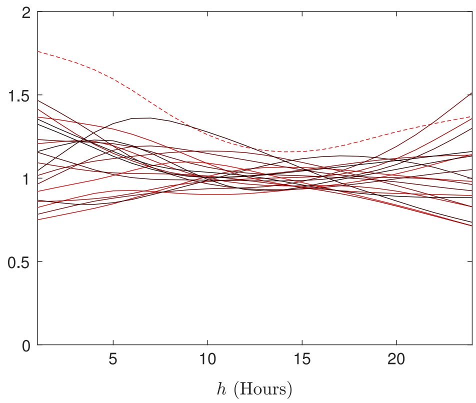

Fig. 8. Impact of the lagged market volatility terms. The dashed line corresponds to log v_{t-1}. This effect was computed for the different lags (solid lines) as $\bar{v}^{\beta_{jh}}$, where $\bar{v} = 26.632$ was the average VSTOXX value over the estimation sample period, and β_{jh} was the element of vector $\boldsymbol{\beta}_h$ associated to lag j, $j = 1 \ldots, p$. A smoothing filter was applied to the lines.

4. Discussion

Our econometric analysis looked at the relationship between active users' group size and their total post contribution across ten years of users' activity in *finanzaonline.com* forum. Results showed that power laws with $\alpha > 1$ provided a good fit: this means that the number of messages was more than proportional to the group size, in that a doubling of group size corresponds to a more than doubled number of messages posted by the group (note that this statement is implicitly *scale-free*). Furthermore, α showed rather regular patterns of intraday variability, with recognizable differences between working week and weekend days. Note that these findings are in line with other studies reporting seasonal patterns in online social behaviour and interaction [40,41] and were robust across orders of magnitude of the variables.

The main empirical finding here, namely the power law with $\alpha > 1$, carries an intuitive appeal: why do we observe a superlinear relation between the size of the active group and the number of posts produced? Especially when considering the nature of the interaction between forum members, which is the exchange of potentially valuable market information, it seems that the balance between costs and benefits in engaging in a discussion should not be overlooked. On the one hand, an active participation can allow users to show their ability, establish relations, acquire trustworthiness and receive support after losses. When the number of active users is low, some users might not see any convenience in sharing information and their tendency to engage in discussions might be lower. On the other hand, platforms such as *finanzaonline.com* likely play a role in lowering information and communication costs with respect to traditional forms of communication. For instance, thanks to search engines, the costs related to gather specific information might vary little or might even decrease with the number of contributors to the forum. Communication costs might vary little as well with the amount of active users (the cost to write a message should not depend on the expected number of people who will read it). In this sense, users might see an advantage in interacting more when there are more active users around.

However, although our results are robust, they cannot help us estimate the influence of user's behaviour and decision making. First, as we did not have the number of *online* users over time, we computed the number of *active* users only by counting those who wrote at least one post in the specified time interval. Secondly, we controlled for the effect of market volatility – which was found to be pivotal in driving online investors' behaviour [28] – while in principle we could have added other financial variables (e.g., financial stability and liquidity indices) as controls. Finally, we looked at the overall number of posts disregarding the quality of each contribution.

More importantly, while our work focused on a macro-level relation, clarifying the nature of the observed phenomenon would require a micro-level analysis. In this regard, it is worth noting that an interesting micro-level analysis has been provided by [10] to explain similar macro-level power relations found in other online platforms. In particular, [10] found a scaling law for the maximum number of contributors to a project of a given size both in Wikipedia and Github. Very interestingly, their explanation was based only on the high heterogeneity in contributors' activity and did not consider agent interactions. This suggests that the widespread presence of high variability in human commitment, which has been documented in a variety of social networks [42,43], could help explaining also the patterns we observed in our case. We leave this microscopic analysis to further investigation.

Acknowledgements

We would like to thank Antonio Pugliese (*Brown Editore*) for sharing *finanzaonline.com* data, Luca Brognara for information on the forum, Claudio Gandelli for help on the database, Niccoló Casnici for help on forum data and FIDA – Finanza Dati Analisi Srl for financial time series.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physa.2017.11.143.

References

[1] A.B. Kao, N. Miller, C. Torney, A. Hartnett, I.D. Couzin, Collective learning and optimal consensus decisions in social animal groups, PLoS Comput. Biol. 10 (2014) e1003762.
[2] A. Cabrera, E.F. Cabrera, Knowledge-sharing dilemmas, Organ. Stud. 23 (2002) 687–710.
[3] B. Maciejovsky, D.V. Budescu, Markets as a structural solution to knowledge-sharing dilemmas, Organ. Behav. Hum. Decis. Process. 120 (2013) 154–167.
[4] S. Gächter, B. Herrmann, Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment, Philos. Trans. R. Soc. B: Biological Sciences 364 (2009) 791–806.
[5] A. Bandura, Social Foundations of Thought and Action: A Social Cognitive Theory, Prentice-Hall, Inc., 1986.
[6] W. Tsai, S. Ghoshal, Social capital and value creation: The role of intrafirm networks, Acad. Manag. J. 41 (1998) 464–476.
[7] J. Török, G. Iñiguez, T. Yasseri, M. San Miguel, K. Kaski, J. Kertész, Opinions, conflicts, and consensus: Modeling social dynamics in a collaborative environment, Phys. Rev. Lett. 110 (2013) 088701.
[8] M. Formentin, A. Lovison, A. Maritan, G. Zanzotto, Hidden scaling patterns and universality in written communication, Phys. Rev. E 90 (2014) 012817.
[9] M. Formentin, A. Lovison, A. Maritan, G. Zanzotto, New activity pattern in human interactive dynamics, J. Stat. Mech. Theory Exp. 2015 (2015) P09006.
[10] M. Gherardi, F. Bassetti, M.C. Lagomarsino, Law of corresponding states for open collaborations, Phys. Rev. E 93 (2016) 042307.
[11] T. Yasseri, J. Kertész, Value production in a collaborative environment, J. Stat. Phys. 151 (2013) 414–439.
[12] A.-L. Barabási, Network Science, Cambridge University Press, 2016.
[13] M. Perc, J.J. Jordan, D.G. Rand, Z. Wang, S. Boccaletti, A. Szolnoki, Statistical physics of human cooperation, Phys. Rep. (2017).
[14] C.-j. Zhu, S.-w. Sun, L. Wang, S. Ding, J. Wang, C.-y. Xia, Promotion of cooperation due to diversity of players in the spatial public goods game with increasing neighborhood size, Phys. A: Stat. Mech. its Appl. 406 (2014) 145–154.
[15] F.C. Santos, M.D. Santos, J.M. Pacheco, Social diversity promotes the emergence of cooperation in public goods games, Nature 454 (2008) 213.
[16] S. Gavrilets, Collective action problem in heterogeneous groups, Philos. Trans. Royal Soc. Lond. B: Biol. Sci. 370 (2015).
[17] J.N. Pruitt, C.M. Wright, C.N. Keiser, A.E. DeMarco, M.M. Grobis, N. Pinter-Wollman, The achilles' heel hypothesis: Misinformed keystone individuals impair collective learning and reduce group success, in: Proc. R. Soc. B, vol. 283, The Royal Society, 2016, p. 20152888.
[18] E. Ostrom, A behavioral approach to the rational choice theory of collective action: Presidential address, american political science association, 1997, Am. Political Sci. Rev. 92 (1998) 1–22.
[19] M. Olson, in: The Logic of Collective Action, vol. 124, Harvard University Press, 2009.
[20] T. Sandler, in: Collective Action: Theory and Applications, vol. 4, University of Michigan Press Ann Arbor, 1992.
[21] P.E. Oliver, G. Marwell, The paradox of group size in collective action: A theory of the critical mass. II, Am. Sociol. Rev. (1988) 1–8.
[22] A. Lupia, G. Sin, Which public goods are endangered?: How evolving communication technologies affect the logic of collective action, Public Choice 117 (2003) 315–331.
[23] B. Bimber, A.J. Flanagin, C. Stohl, Reconceptualizing collective action in the contemporary media environment, Commun. Theory 15 (2005) 365–388.
[24] W. Antweiler, M.Z. Frank, Is all that talk just noise? the information content of internet stock message boards, The J. Finance 59 (2004) 1259–1294.
[25] T.O. Sprenger, A. Tumasjan, P.G. Sandner, I.M. Welpe, Tweets and trades: The information content of stock microblogs, Eur. Financial Manag. 20 (2014) 926–957.
[26] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, J. Comput. Sci. 2 (2011) 1–8.
[27] P. Clarkson, D. Joyce, I. Tutticci, Market reaction to takeover rumour in internet discussion sites, Account. Finance 46 (2006) 31–52.
[28] N. Casnici, P. Dondio, R. Casarin, F. Squazzoni, Decrypting financial markets through e-joint attention efforts: On-line adaptive networks of investors in periods of market uncertainty, PLoS ONE 10 (2015) e0133712.
[29] J. Leskovec, J. Kleinberg, C. Faloutsos, Graphs over time: Densification laws, shrinking diameters and possible explanations, in: Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, ACM, 2005, pp. 177–187.
[30] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters, ACM Trans. Knowl. Discov. Data (TKDD) 1 (2007) 2.
[31] X. Shi, B. Tseng, L. Adamic, Looking at the blogosphere topology through different lenses, in: Proceedings of the International Conference on Weblogs and Social Media (ICWSM 2007), vol. 1001, 2007.
[32] J.W. Kantelhardt, E. Koscielny-Bunde, H.H. Rego, S. Havlin, A. Bunde, Detecting long-range correlations with detrended fluctuation analysis, Phys. A: Stat. Mech. Appl. 295 (2001) 441–454.
[33] P.R. Hansen, A. Lunde, V. Voev, Realized beta garch: A multivariate garch model with realized measures of volatility, J. Appl. Econometrics 29 (2014) 774–799.
[34] P.R. Hansen, Z. Huang, H.H. Shek, Realized garch: A joint model for returns and realized measures of volatility, J. Appl. Econometrics 27 (2012) 877–906.
[35] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econometrics 31 (1986) 307–327.
[36] R.N. Mantegna, H. Stanley, Modeling of financial data: Comparison of the truncated lévy flight and the arch(1) and garch (1,1) processes, Phys. A: Stat. Mech. its Appl. 254 (1998) 77–84.
[37] S.R. Bentes, Forecasting volatility in gold returns under the garch, igarch and figarch frameworks: New evidence, Phys. A: Stat. Mech. its Appl. 438 (2015) 355–364.
[38] C. Francq, J. Zakoïan, A tour in the asymptotic theory of garch estimation, in: Handbook of Financial Time Series, Springer, 2009, pp. 85–111.
[39] P. Racca, R. Casarin, F. Squazzoni, P. Dondio, Resilience of an online financial community to market uncertainty shocks during the recent financial crisis, J. Comput. Sci. 16 (2016) 190–199.
[40] S.A. Golder, M.W. Macy, Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures, Science 333 (2011) 1878–1881.
[41] S.A. Golder, D.M. Wilkinson, B.A. Huberman, Rhythms of Social Interaction: Messaging Within a Massive Online Network, Springer, London, 2007, pp. 41–66. http://dx.doi.org/10.1007/978-1-84628-905-7.3.
[42] N. Perra, B. Gonçalves, R. Pastor-Satorras, A. Vespignani, Activity driven modeling of time varying networks, Sci. Rep. 2 (2012).
[43] L. Muchnik, S. Pei, L.C. Parra, S.D. Reis, J.S. Andrade Jr, S. Havlin, H.A. Makse, Origins of power-law degree distribution in the heterogeneity of human activity in social networks, Sci. Rep. 3 (2013) (1783).