Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

Q1 Global financial indices and twitter sentiment: A random matrix theory approach

Q2 A. García

Physics Research Department, Sonora University, Hermosillo, Sonora 83000, Mexico

HIGHLIGHTS

- The results from RMT analysis support the fact of the existence of true correlations between financial indices, polarities, and the mixture of them.
- The global correlation structure studied here might be preserved regardless of whether we use as an input either public tweets or financial information.
- Granger causality test hints that polarity time series have predictive information of return time series of Indonesia, Israel, South Korea, Australia, and Japan.
- The results suggest that the return and polarity indicators share a common correlation structure, at least for the selected countries and period of time studied here.

ARTICLE INFO

Article history:
Received 6 December 2015
Received in revised form 31 March 2016
Available online xxxx

Keywords:
Random matrix theory
Global financial indices
Sentiment analysis

ABSTRACT

We use Random Matrix Theory (RMT) approach to analyze the correlation matrix structure of a collection of public tweets and the corresponding return time series associated to 20 global financial indices along 7 trading months of 2014. In order to quantify the collection of tweets, we constructed daily polarity time series from public tweets via sentiment analysis. The results from RMT analysis support the fact of the existence of true correlations between financial indices, polarities, and the mixture of them. Moreover, we found a good agreement between the temporal behavior of the extreme eigenvalues of both empirical data, and similar results were found when computing the inverse participation ratio, which provides an evidence about the emergence of common factors in global financial information whether we use the return or polarity data as a source. In addition, we found a very strong presumption that polarity Granger causes returns of an Indonesian index for a long range of lag trading days, whereas for Israel, South Korea, Australia, and Japan, the predictive information of returns is also presented but with less presumption. Our results suggest that incorporating polarity as a financial indicator may open up new insights to understand the collective and even individual behavior of global financial indices.

1. Introduction

Various contributions from physics to economic systems have emerged in recent years, leading to a new field of study named *econophysics* [1–3]. Understanding the presence of correlations between different sets of stocks markets is one of the faster growing lines of research within econophysics due to its crucial importance for selecting the most efficient portfolios of financial markets [4].

E-mail address: andgarm.n@gmail.com.

http://dx.doi.org/10.1016/j.physa.2016.06.024
0378-4371/

Please cite this article in press as: A. García, Global financial indices and twitter sentiment: A random matrix theory approach, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.06.024

A novel approach to understand such correlations is Random matrix theory (RMT), which was introduced in mathematical statistics by Wishart in 1928 [5]. In the decade of 1950s Wigner used it to deal with the statistics of eigenvalues and eigenvectors of complex many-body systems in nuclear physics [6–9]. Different phenomena of physics have been resolved successfully using the formalism of RMT, but it was not until the pioneer works of Stanley et al. [10] and Bouchaud et al. [11] that a burst of studies appeared to deal with the correlation matrix structure of financial markets using RMT framework [12–16]. In the context of global financial indices, there are researches about measures of globalization and long-range correlations [17,18]. Likewise Sandoval and Franca analyzed a diversity of market indices during times of crisis [19], whereas Kumar and Deo investigated the network properties of world indices during the financial crisis of 2008 [20].

On the other hand, the social networking microblogging service Twitter allows users to broadcast short 140-character messages called *tweets*. Recent work has started to explore the extraction of real-time sentiment-tracking indicators of public mood state from twitter and relating it to economic indicators [21,22]. The work of Bollen et al. [22] is one of the first attempts to investigate correlations between mood states of public tweets and stock markets, they found a good accuracy predicting the daily changes in the closing values of the financial index DJIA. In addition, Preis et al. [23] found warning signs of stock moves by the analysis of changes in Google query volumes. The work of Alanyali et al. [24] found a positive correlation between daily transaction volume of a company and the number of mentions of that company in the *Financial Times*, and Zheludev et al. [25] point out that sentiment analysis of social media is more indicative than just message volumes.

Our approach to analyze the global financial indices relies on previous work, but here we incorporate the influence of the social media twitter through its polarity value, which is as a mood state measure. To the best of our knowledge, there are no studies that integrate polarity time series in the study of spectral deviations from RMT predictions in the financial markets. The influence of news or social media has not been extensively explored yet because the paradigm of the Efficient Market Hypothesis (EMH). According to the EMH, a share price of a given market instantly embodies all new information, and its value does not depend on past prices [26]. However, as mentioned above recent works have started to study the influence of social media in markets movements, showing that news and public sentiment extracted from twitter, google trends, and financial news can give early indications to predict changes in stock markets.

In this work, we show a methodology to construct a set of correlation matrices from public tweets and investigate its structure under RMT approach. For this purpose, the returns of 20 global financial indices are studied here, and a search keyword in the twitter database is associated to each of them. Sentiment analysis is used to quantify the collection of tweets by its polarity measure, which is taken here as a financial indicator. Thus, the goal of the present study is to compare the deviations from RMT predictions in return and polarity correlation matrices by the analysis of the eigenvalue distribution density for uncorrelated and correlated Wishart matrices, the extreme eigenvalues, and the Inverse Participation Ratio (IPR). Further, the influence of twitter mood to return values is explored through The Granger Causality Test.

The paper is organized as follow. Section 2 describes the analyzed data and the methodology to extract the collection of public tweets. Section 3 explains briefly sentiment analysis methodology, the implementation used to quantify it and how polarity time series are constructed. Section 4 contains the main results for RMT analysis, and Granger causality test. Finally, Section 5 presents the conclusions of the work.

2. Analyzed data

Our analysis was carried out for two different sets of data. The first set is composed of daily closing prices of $N = 20$ financial indices around the world, whose tickers are listed in the second column of Table 1. The second set of data was obtained by extracting public tweets associated to each of the financial indices of the first set. All the Twitter requests were made in Universal Time (UTC), while extraction times of closing prices vary in accord with the time zone where the markets trade. For both sets of data,i.e., closing prices and public tweets, the period of time under study was from February 22 to October 13 of 2014, for a total of $L = 166$ trading days. Daily closing prices were collected from Bloomberg, and have the same preprocessing as in Ref. [19].

On the other hand, tweets were extracted from Twitter database using Twitter Search API.[1] In addition, a Python wrapper code was used in order to handle Twitter API and made readily queries. The parameters and values employed to get the collection of twitter data are displayed in Table 2.

The first parameter *term* is the word or phrase we want to search, which was filled with every word listed in the third column of Table 1. The parameter *geocode* specifies a geographic radius of search around the user and is fixed as *none*. We decided to handle the parameter *until* to give us all the possible tweets until *current* time, then *since_id* and *max_id* are set as *none* because we do not want to have control on the tweet through its identifier number, but rather by its date. The parameter *count* fixes the number of tweets returned per request, and its maximum accepted value is 100. The language of search was chosen as English with the value *en* in the *lang* parameter. Moreover, *locale* is to specify the language of the sending query, but as we are searching words associated with Bloomberg symbols, it is fixed as none. Finally, with *result_type* fixed as *mixed*, we include both popular and real time results in the response.

By scheduling hourly requests, we also have increased the number of collected tweets, extracting up to 2400 tweets in a single day for each keyword. However, this number may be lower in periods of times when the requested keyword is not

[1] https://dev.twitter.com

Please cite this article in press as: A. García, Global financial indices and twitter sentiment: A random matrix theory approach, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.06.024