Physica A 444 (2016) 205–212

Contents lists available at ScienceDirect

Physica A

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

Long-range correlation analysis of economic news flow intensity

S.P. Sidorov *, A.R. Faizliev, V.A. Balash, E.A. Korobov

Risk Institute, Saratov State University, Russian Federation

HIGHLIGHTS

- Auto-correlation properties for time series of the news flow intensity are examined.
- News analytics data shows the presence of long-range correlations for the time series of news intensity data.
- The detrended fluctuation analysis and the detrending moving average analysis are run both on the original and on the deseasonalized data.

ARTICLE INFO

Article history:
Received 27 April 2015
Received in revised form 4 September 2015
Available online 14 October 2015

Keywords:
Long-range correlation
Detrended fluctuation analysis
Time series
Auto-correlation

ABSTRACT

The goal of the paper is to examine the auto-correlation properties for time series of the news flow intensity using different methods, such as the fluctuation analysis, the detrended fluctuation analysis and the detrending moving average analysis. Empirical findings for news analytics data show the presence of long-range correlations for the time series of news intensity data.

1. Introduction

The output signals generated by a complex system often exhibit long-range correlations. To have a better understanding for the dynamics of the underlying complex systems it is important to estimate such long-range correlations. Different methods were developed to detect long-range correlations in time series, among them the rescaled range analysis [1], the wavelet transform module maxima approach [2], the fluctuation analysis (FA) [3], the detrended fluctuation analysis (DFA) [4], the detrending moving average analysis (DMA) [5], and many others.

DFA is one of the well-known and used method for indirect scaling the long-range auto-correlation in non-stationary time series. The method was proposed in the paper [4] extending the ideas of the work [3]. Since then DFA proved its effectiveness for many scientific and engineering problems such as DNA analysis [6–8], human gait [9,10], analysis of dynamics for daily internet traffic [11], biomedical signal processing [12–16], finance and economical time series [17–21], and many others.

It should be noted that the DFA has some drawbacks. The paper [22] states that DFA

1. can lead to uncontrolled bias;
2. is more expensive than unbiased estimator;
3. cannot provide protection against nonlinear nonstationaries.

* Corresponding author.
E-mail address: sidorovsp@info.sgu.ru (S.P. Sidorov).

http://dx.doi.org/10.1016/j.physa.2015.10.025
0378-4371/

In order to avoid one-sided results, we will compare the performances of four estimators (Fluctuation Analysis (FA), Detrended Fluctuation Analysis of orders 1, 2 and 3 (DFA1, DFA2 and DFA3), and Detrending Moving Average (DMA)) in their capability to study long-range auto-correlation in time series of news flow intensity. The work [23] points out that DMA and DFA remain "The Methods of Choice" in determining the Hurst index of time series.

Our empirical analytics used the time series of news analytics data from Raven Pack. News analytics is a relatively new tool designed to improve the trading strategies of investors. Knowing the characteristics of news in numerical indices one can use them in mathematical and statistical models [24] and automated trading systems [25]. Currently, the tools of the news analytics have been increasingly used by traders in the US and Europe.

There are more than 50 providers of economic news in the world. Bloomberg, Dow Jones and Thomson Reuters are the three largest of them. About 200 agencies are involved in providing financial analytics. In our research we use the Raven Pack data, one of the most well-known providers of news analytics data.

Two-month period was divided into non-overlapping consecutive segments of equal (small) length. Using news analytics data from Raven Pack we calculated the amount of economical and finance news reported in the world during each interval of time. Obtained in such a way time series of news flow intensity were used for the long-range auto-correlation analysis.

2. News analytics data

News analytics is a new approach to the analysis of news flow based on methods of artificial intelligence. The process of news analysis in information systems is automated and usually includes the following steps:

1. collecting news from different sources;
2. preliminary analysis of news;
3. analysis of news-related expectations (sentiments) taking into account the current market situation;
4. design and use of quantitative models.

The review of methods and tools of news analytics can be found in book [25].

Providers of news analytics extract and use data from various sources: news agencies, pre-news (SEC reports, court documents, reports of various government agencies, business resources, company reports, announcements, industrial and macroeconomic statistics) and social media (blogs, social networks, etc.).

For each news wire, Raven Pack generates the following fields for sources of news analytics data: time stamp, company name, company id, relevance of the news, event category, event sentiment, novelty of the news, novelty id, composite sentiment score of the news, word/phrase level score, projections by company, editorials & commentary, reports corp actions, news impact projection, story ID.

Our sample covers the period ranging from September 1, 2010 to October 29, 2010 (i.e. 43 trading days). It is composed of the all news that were released during this period of time. First we selected only those news data which we think would be useful for our analysis. We carried out data selection and cleaning by the following criteria:

1. There is a clear presence of weekly seasonality in the data (the average amount of a company's news announcements released during the week-end is much lower than the one of the other days of the week). The same picture is held for all companies indeed. That is why we excluded all weekend news from our analysis.
2. If a piece of news is related to a group of companies then the news will be multiplicated in the Raven Pack data for every company in the group (for example, if a news is related to a certain industry, Raven Pack will repeat the news for every company in the industry). Therefore, we selected only unique news for our final sample.
3. Finally, we removed all the news regarding the imbalance of supply and demand before both the opening and the closing of trading time of different stock exchanges which occurs every trading day at the same time (usually we can see several hundreds of news coming out in a short time at the beginning and at the end of the trading sessions).

We constructed time series of news flow intensity as follows. Two-months period Δ was divided into n non-overlapping consecutive segments $\Delta_1, \ldots, \Delta_n$ of equal (small) length δ min, $\Delta = \Delta_1 \cup \cdots \cup \Delta_n$. Using news analytics data from Raven Pack we calculated x_t, the amount of economical and finance news reported in the world during each segment Δ_t, $t = 1, 2, \ldots, n$. The sequence $x_1, x_2, \ldots, x_n$ is the time series of news flow intensity with δ min window.

Using the final data sample, we divide the whole period into segments of equal length $\delta = 1, 2, 5$ min and calculated the amount of all news during each interval of time. The total amount of news released during the two-month period is equal to 313 598. The description statistics of time series one can find in Table 1.

There is no clear trend of the total daily amount of news wires for all companies. Fig. 1 displays an example of news flow intensity with 1-min window. It is visually apparent that fluctuations are highly volatile and demonstrate a non-stationary behavior. Some periods have the rate of news intensity below the average (e.g. holidays, Christmas time). On the other hand, one can witness the increase of the rate at the periods of the quarterly reports and releases of the intermediate figures and earnings of companies.