Physica A 451 (2016) 502–506

Contents lists available at ScienceDirect

Physica A

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

How long is the memory of the US stock market?

Paulo Ferreira [a,b,*], Andreia Dionísio [a]

[a] CEFAGE-UE, IIFA, Universidade de Évora, Largo dos Colegiais 2, 7000 Évora, Portugal

[b] Departamento de Ciência e Tecnologia Animal, Escola Superior Agrária de Elvas, Instituto Politécnico de Portalegre, Portugal

HIGHLIGHTS

- We apply DCCA and its correlation coefficient to study the efficient market hypothesis.
- We show that the correlation coefficient is significant till the 149th lag.
- It means that the US stock market has long memory for about seven months.
- It could be seen as a possible contradiction of the EMH.

ARTICLE INFO

Article history:
Received 27 August 2015
Received in revised form 23 December 2015
Available online 9 February 2016

Keywords:
Efficient Market Hypothesis
Long-range correlation coefficient
Lag
Detrended cross-correlation analysis

ABSTRACT

The Efficient Market Hypothesis (EMH), one of the most important hypothesis in financial economics, argues that return rates have no memory (correlation) which implies that agents cannot make abnormal profits in financial markets, due to the possibility of arbitrage operations. With return rates for the US stock market, we corroborate the fact that with a linear approach, return rates do not show evidence of correlation. However, linear approaches might not be complete or global, since return rates could suffer from nonlinearities. Using detrended cross-correlation analysis and its correlation coefficient, a methodology which analyzes long-range behavior between series, we show that the long-range correlation of return rates only ends in the 149th lag, which corresponds to about seven months. Does this result undermine the EMH?

1. Introduction

One of the most important hypotheses in financial economics is the Efficient Market Hypothesis (EMH). Accordingly, a financial market is considered efficient in its weak form if it is not possible to identify any deterministic pattern in its time series behavior. In other words, EMH means that through arbitrage, agents could not obtain systematic abnormal profits using past information [1]. Several studies analyze this hypothesis, checking for some windows of profit opportunities. In this paper, we do not carry out a thorough literature review, since our main objective is to use a different approach to EMH. For a more complete literature review on EMH see, for example, the work of Sewell [2].

One important conclusion of EMH, especially in its weak form, is that financial returns have no memory and are independent in time. This is an issue with many years of study, started at the turn of the 20th century by Bachelier [3] and corroborated by other studies, such as those of Kendall [4], Osborne [5], Granger and Morgenstein [6] or Fama [7]. In fact, this is the idea that originated the EMH. Several studies proved that, when linear autocorrelation exists between return rates, it quickly disappears. However, some authors argue that there may be long-range dependence in return rates. This

* Corresponding author at: CEFAGE-UE, IIFA, Universidade de Évora, Largo dos Colegiais 2, 7000 Évora, Portugal.
E-mail addresses: pjsf@uevora.pt (P. Ferreira), andreia@uevora.pt (A. Dionísio).

http://dx.doi.org/10.1016/j.physa.2016.01.080
0378-4371/

particular issue, when related with other stylized facts recognized in the literature, could mean the existence of temporal or sectional dependence in returns leading to some capacity to predict financial series, which may violate the EMH. The main stylized facts found are the existence of fat tails, asymmetries in gains and losses, volatility clustering behavior, leverage effect and the existence of autocorrelation in variance (see, for example, Ref. [8]).

Financial markets are considered to be complex systems due to their unexpected behavior. Besides, these are markets normally characterized by a large amount of available data, conditions that made some physicists interested in the behavior of financial markets. In recent years, these particular aspects have originated a new research field called Econophysics. In this new field, joint multidisciplinary teams are able to study economic issues, especially in financial markets. Studies such as those by Podobnik et al. [9], Analyali et al. [10] or Morales et al. [11] are just a few examples of this trend of joint research. It is not our objective to carry out an extensive literature review on Econophysics. Some studies containing such reviews can be consulted, for example, that of Jovanovic and Schinckus [12].

The objective of this paper is to analyze the behavior of the US stock market, using the Morgan Stanley Capital International (MSCI) stock index for the US, between 02-01-1995 and 21-02-2014, with a total of 4995 observations. We propose to analyze the behavior of return rates with detrended cross-correlation analysis and its correlation coefficient, to check for the existence of long-range dependence in the time series. This analysis remains interesting because studies are not conclusive about the existence or not of long memory in stock return rates.

If this long-range dependence exists, it could be argued that EMH could be violated. In fact, we find evidence that the DCCA correlation coefficient is significant until the 149th lag.

2. Methodology

Studies of financial markets are quite frequent, and the main goal is to analyze the possibility of EMH verification. When return rates are analyzed, in most cases results show evidence of autocorrelation which, in general, quickly vanishes. This means that returns have no memory, considering a short range analysis. However, some authors show there may be a long-range dependence in financial markets (see, for example, Ref. [13]). This result could be related to the existence of nonlinear dependence in return rates which may not be detected with the corresponding linear tests (see, for example, Ref. [14] or [15]).

Applying methodologies that can analyze nonlinear dependence in financial markets could lead us to a better understanding of these markets. These methodologies could be more informative since they follow general models and could detect not only linear dependence. This is the case of mutual information, a measure of nonlinear dependence in time series (see, for example, Ref. [16]). The analysis of long-range dependence in time series is also another example of methodologies that could give us more information.

One of these methods is detrended fluctuation analysis (DFA). Created by Peng et al. [17], this methodology studies the behavior of individual series and has several applications to financial markets (see, for example, the work of Cizeau et al. [18], Ausloos et al. [19] or Ferreira and Dionísio [20], among others). The great advantage of DFA is the fact that it can be used in both stationary and non-stationary series, while linear approaches can only be used in the former.

Besides DFA, detrended cross-correlation analysis (DCCA) can also be used to study long-range dependence. However, DCCA is used not to study the long-range behavior of one time series but the behavior between time series, namely its long-range cross-correlation. Created by Podobnik and Stanley [21], it has the advantage of also being used in non-stationary time series.

Considering the data given by x_k and y_k with $k = 1, \ldots, t$ equidistant observations. The first step of DCCA is obtained by integrating both series and calculating the values: $x(t) = \sum_{k=1}^{t} x_k$ and $y(t) = \sum_{k=1}^{t} y_k$. After the integration of those series (with N samples), they are divided into boxes of equal length, n. Afterwards, we divide them into $N - n$ overlapping boxes, defining for each box the local trend ($\tilde{x}_k$ and $\tilde{y}_k$), using ordinary least squares. After this, the detrended series is calculated: the difference between the original values and its trend. Then, we calculate the covariance of the residuals in each box given by $f_{DCCA}^2 = \frac{1}{n-1}\sum_{k=1}^{i+n}\left(x_k - \tilde{x}_k\right)\left(y_k - \tilde{y}_k\right)$. Finally, the detrended covariance is calculated summing all $N - n$ boxes of size n, given by $F_{DCCA}^2(n) = \frac{1}{N-n}\sum_{i=1}^{N-n} f_{DCCA}^2$. This process should be repeated for boxes of different length in order to find the relationship between the DCCA fluctuation function and n, which allows us to find the long-range cross correlation $F_{DCCA}(n)$ given by the power law $F_{DCCA}(n) \sim n^{\lambda}$. Interpretation of λ is as follows: if λ is equal to 0.5 the series have no long range cross-correlation; λ greater than 0.5 means persistent long-range cross-correlations while values lower than 0.5 mean anti-persistent cross-correlation (a large value in one variable is likely to be followed by a small value in another variable, and vice versa).

DCCA gives us information about cross correlation between series but does not quantify that relation. In order to make that quantification, from the results of DCCA between x and y and DFA for each series, Zebende [22] created the correlation coefficient given by $\rho_{DCCA} = \frac{F_{DCCA}^2}{F_{DFA\{xi\}}F_{DFA\{yi\}}}$. This coefficient has the general properties of one correlation coefficient, namely $-1 \leq \rho_{DCCA} \leq 1$. A value of $\rho_{DCCA} = 0$ means that there is no cross-correlation between series, while a positive or negative value means, respectively, evidence of cross-correlation or anti cross-correlation between series.

According to Podobnik et al. [23], we can test the significance of this correlation coefficient. We use that methodology to estimate critical points of our test.