Physica A 462 (2016) 186–197

Contents lists available at ScienceDirect

Physica A

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

Understanding the determinants of volatility clustering in terms of stationary Markovian processes

S. Micciché

Dipartimento di Fisica e Chimica, Università degli Studi di Palermo, Viale delle Scienze, Edificio 18, I-90128 Palermo, Italy

HIGHLIGHTS

- We are interested in studying the determinants of volatility clustering.
- The proposed approach starts by diagonalizing the volatility correlation matrix.
- We then check that the diagonalized volatilities keep the usual stylized facts.
- We then present an univariate model based on a non-linear Langevin equation.
- The main ingredient of the model is a Smoluchowski potential with logarithmic tails.

ARTICLE INFO

Article history:
Received 30 March 2016
Available online 17 June 2016

Keywords:
Volatility
Econophysics
Long-range correlation
Stochastic processes
First passage time

ABSTRACT

Volatility is a key variable in the modeling of financial markets. The most striking feature of volatility is that it is a long-range correlated stochastic variable, i.e. its autocorrelation function decays like a power-law $\tau^{-\beta}$ for large time lags. In the present work we investigate the determinants of such feature, starting from the empirical observation that the exponent β of a certain stock's volatility is a linear function of the average correlation of such stock's volatility with all other volatilities.

We propose a simple approach consisting in diagonalizing the cross-correlation matrix of volatilities and investigating whether or not the diagonalized volatilities still keep some of the original volatility stylized facts. As a result, the diagonalized volatilities result to share with the original volatilities either the power-law decay of the probability density function and the power-law decay of the autocorrelation function. This would indicate that volatility clustering is already present in the diagonalized un-correlated volatilities.

We therefore present a parsimonious univariate model based on a non-linear Langevin equation that well reproduces these two stylized facts of volatility. The model helps us in understanding that the main source of volatility clustering, once volatilities have been diagonalized, is that the economic forces driving volatility can be modeled in terms of a Smoluchowski potential with logarithmic tails.

1. Introduction

Volatility is a key variable in the modeling of financial markets. In fact, it controls several risk measures associated with the dynamics of the price of financial assets as well as it affects the rational price of derivative products [1–6]. In this respect, volatility is usually assumed to be a proxy of how risky an asset is as it quantifies the un-predictability level of a given asset.

The most striking feature of volatility is that it is a long-range correlated stochastic variable, i.e. its autocorrelation function decays like a power-law $\tau^{-\beta}$ for large time lags τ [7]. This is a feature related to the fact that the volatility time

E-mail address: salvatore.micciche@unipa.it.

http://dx.doi.org/10.1016/j.physa.2016.06.081

series show persistence, i.e. the fact that when a relatively high volatility value is reached then volatility remains high for some time. In the present work we investigate the determinants of such feature, starting from the empirical observation that the exponent β of a certain stock's volatility is a linear function of the average correlation of such stock's volatility with all other volatilities [8]. This empirical observation shows that there exists a role of the multivariate volatility properties in explaining volatility clustering and therefore there exists a relationship between the level of predictability of any volatility time-series and the extent of its inter-dependence with other assets. In all considered cases, the more the asset is linked to other assets, the more its volatility retains memory of its past behavior [8]. In order to evaluate whether this is the only source of volatility clustering, we propose a simple approach consisting in diagonalizing the cross-correlation matrix of volatilities and investigating whether or not the diagonalized volatilities still keep some of the original volatility stylized facts.

As a result, the diagonalized volatilities result to share with the original volatilities either the power-law decay of the probability density function and the power-law decay of the autocorrelation function. This would indicate that volatility clustering is already present in the diagonalized un-correlated volatilities.

With the aim of understanding what are the mechanisms that give rise to volatility clustering in the diagonalized un-correlated volatilities, we present a parsimonious univariate model that well reproduces these two stylized facts of volatility. Differently from many models already known in the literature [9–17], such model is based on a simple non-linear Langevin equation. It is therefore a stationary Markovian model that, for an appropriate choice of the parameters, describes a long-range memory process. The models help us in understanding that the main source of the volatility clustering effect, once volatilities have been diagonalized, is the fact that the economic forces driving volatility can be modeled in terms of a Smoluchowski potential with logarithmic tails, as detailed below.

Such an effect can be further quantified by investigating the mean First Passage Time (mFPT) of the un-correlated volatility time-series [18]. Given a certain time-series, the mFPT $T_{x_0}(\Lambda)$ is the average time that is needed to reach for the first time position $x_0 \pm \Lambda$, where Λ is a threshold, starting from x_0. We show that the proposed model predicts a power-law functional form for the mFPT and this is also observed in real data. The predicted power-law exponents are in good agreement with the one observed in real data, thus indicating that indeed the existence of the logarithmic potential may be considered as the main source of volatility clustering.

The paper is organized as follows. In Section 2 we introduce the three datasets we will consider in this study. In Section 3 we will consider univariate properties of volatility time-series and in Section 4 we present a simple model based on a non-linear Langevin equation that accounts for the observed stylized facts relative to the volatility probability density function and autocorrelation. Section 4.2 is devoted to the investigation of the First Passage Time (FPT) of volatility time series. In Section 5 we discuss the obtained results and finally in Section 6 we will draw our conclusions.

2. Data

2.1. Datasets

In this paper we empirically investigate the univariate behavior of stocks traded in three different markets: the New York Stock Exchange (NYSE), the London Stock Exchange (LSE), and the Paris Bourse (PB). The NYSE data refer to the period from 1995 to 2003. The LSE and PB data refer to year 2002.

The NYSE data are taken from the Trades and Quotes (TAQ) dataset maintained by NYSE [19]. In particular, 100 highly capitalized stocks are selected, see Appendix A.1. For each stock and for each trading day we consider the time series of stock prices recorded transaction by transaction. Since transactions for different stocks do not happen simultaneously, we divide each trading day, lasting $6^h30'$, into intervals of length τ. For each trading day, we define N_τ intraday stock price proxies $p_i(t_k)$ of asset i, with $k = 1, \ldots, N_\tau$. The proxy is defined as the transaction price detected nearest to the end of the interval. This is one possible way to deal with high-frequency financial data [20]. By using these proxies, we compute the price returns:

$$r_i(t) = \ln p_i(t) - \ln p_i(t - \tau) \tag{1}$$

at time-horizons τ. The time-horizon used is $\tau = 5$ min. Each volatility time-series contains $N_\tau = 176\,514$ records for the whole period 1995–2003 and 19 656 records when considering year 2002 only.

The LSE data are taken from the "Rebuild Order Book" dataset, maintained by LSE [21]. In particular, we consider only the electronic transactions occurred in year 2002 for 90 highly traded stocks belonging to the SET1 segment of the LSE market, see Appendix A.2. However, most of the transactions, a mean value of 75% for the 90 stocks, are of the electronic type. This market is commonly believed to be very active and can be regarded as a realization of a "liquid" market. For each stock i and for each trading day we consider the time series of stock price recorded transaction by transaction and generate N_τ intraday stock price proxies $p_i(t_k)$ according to the procedure explained above. Also for the LSE data, the time-horizon used was $\tau = 5$ min. Each trading day lasts $8^h30'$. Each volatility time-series contains $N_\tau = 25\,602$ records.

The PB data are taken from the "Historical Market Data" dataset, maintained by EURONEXT [22]. In particular, we consider the electronic transactions of 39 highly traded stocks in year 2002, see Appendix A.3. For each stock i and for each trading day, lasting $8^h30'$, we consider the time series of stock price recorded transaction by transaction and generate N_τ intraday