Detrended cross-correlations between returns, volatility, trading activity, and volume traded for the stock market companies

Rafał Rak¹, Stanisław Drożdż²,³, Jarosław Kwapień² and Paweł Oświęcimka²

¹ Faculty of Mathematics and Natural Sciences, University of Rzeszów - Rzeszów, Poland
² Institute of Nuclear Physics, Polish Academy of Sciences - ul. Radzikowskiego 152, Kraków, Poland
³ Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology - Kraków, Poland

received 16 October 2015; accepted in final form 9 November 2015
published online 30 November 2015

PACS 89.75.–k – Complex systems
PACS 89.75.Da – Systems obeying scaling laws
PACS 89.65.Gh – Economics; econophysics, financial markets, business and management

Abstract – We consider a few quantities that characterize trading on a stock market in a fixed time interval: logarithmic returns, volatility, trading activity (i.e., the number of transactions), and volume traded. We search for the power-law cross-correlations among these quantities aggregated over different time units from 1 min to 10 min. Our study is based on empirical data from the American stock market consisting of tick-by-tick recordings of 31 stocks listed in Dow Jones Industrial Average during the years 2008–2011. Since all the considered quantities except the returns show strong daily patterns related to the variable trading activity in different parts of a day, which are the most evident in the autocorrelation function, we remove these patterns by detrending before we proceed further with our study. We apply the multifractal detrended cross-correlation analysis with sign preserving (MFCCA) and show that the strongest power-law cross-correlations exist between trading activity and volume traded, while the weakest ones exist (or even do not exist) between the returns and the remaining quantities. We also show that the strongest cross-correlations are carried by those parts of the signals that are characterized by large and medium variance. Our observation that the most convincing power-law cross-correlations occur between trading activity and volume traded reveals the existence of strong fractal-like coupling between these quantities.

Introduction. – Complex systems are typically characterized by a hierarchical structure, in which the hierarchy is observed both as different levels of the system’s organization from a microscale to progressively higher macroscales (for example, from molecules to the multicell organisms and ecosystems) and as the structures of different size that coexist at the same organization level (for example, organisms with different body size or species with different population). This same-level hierarchy is often observed indirectly as the power-law distributions or the power-law temporal correlations of measurement outcomes. The latter are especially interesting from the physical point of view since such behaviour (i.e., long memory) can be related to critical phenomena in the system’s evolution [1]. Financial markets are perhaps the most complex structures in the world, which makes the related data to be a particularly interesting object of studying. Here we focus on the stock market data and the most fundamental quantities: price, trades, and volume traded.

Different studies showed that there is a relation between traded volume $V$ and a price change $\Delta p(t) = p(t + \Delta t) - p(t)$ (or a logarithmic return $r(t) = \ln(p(t + \Delta t)/p(t))$ of this stock over the same time interval $\Delta t$). This relation (which should not necessarily be causal) concerns both the signed price change $\Delta p(t)$ or the return $r(t)$ and the absolute price change (volatility): $|\Delta p(t)|$, $|r(t)|$ [2,3]. An exact form of this relation is still a subject of debate since a number of studies devoted to this issue provided us with mixed results, but its very existence seems to be well established now. As regards its origin, there is no decisive conclusion yet. It might well be a consequence of a relation between volume and trading activity (the number of
transactions in a unit time interval), between volume and transaction size, or an effect of the fluctuations in order imbalance [4–11].

A specific type of coupling between different quantities are power-law cross-correlations. They may occur between the quantities that are power-law autocorrelated, which is the case in the financial market data [1]. The cross-correlations between absolute fluctuations of the S&P500 index and the corresponding volume traded were discovered in [12]. No such correlations were found between the volume and price fluctuations, though. In contrast, an analysis of the daily price and volume fluctuations of the Chinese main stock markets located in Shanghai (SHSI) and Shenzhen (SZCI) showed that these quantities do depend on each other and that this dependence is a power-law one and multifractal [13]. Similar outcomes were obtained from an analysis of the high-frequency price and volume fluctuations of the Chinese CSI300 index futures with an additional indication that these cross-correlations were persistent [14].

Here we study cross-correlations among the following quantities: price returns, volatility, trading activity, and volume traded. We analyze the tick-by-tick data representing highly capitalized stocks that were listed in the Dow Jones Industrial Average (DJIA) during the period 2008–2011. What distinguishes our study is that we remove daily trends from all types of data before we proceed with the analysis. We investigate the power-law cross-correlations by means of the multifractal detrended cross-correlation analysis with sign preserving (MFCCA [15]) that is a consistent generalization of the detrended cross-correlation approach (DCCA [16]). According to our knowledge, this is the first analysis of this type that incorporates such a set of quantities and such an approach.

Data. – Let us assume that one records every trade in which \( \omega^{(u)}(t_i) \) shares of a stock \( u = 1, \ldots, N \) at price \( p^{(u)}(t_i) \) are transferred within a pair of investors, where \( t_i \) is the moment of the \( i \)-th trade. Let us also assume that consecutive time intervals of length \( \Delta t \) are indexed by \( j \) \((j = 1, \ldots, J)\) and that the effects of all the trades \( t_i \) \((i = 1, \ldots, T^{(u)}(j))\) that were made during each such an interval are aggregated. Thus, for each interval \( j \) we may consider the following four quantities: i) the logarithmic price returns:

\[
R^{(u)}(j) = \ln p^{(u)}(j \Delta t) - \ln p^{(u)}((j - 1) \Delta t)
\]

(where we assumed that the first interval begins at \( t = 0 \) and that the price remains unchanged between the trades), ii) the trading activity \( T^{(u)}(j) \), and iii) the total volume traded

\[
V^{(u)}(j) = \sum_{i=1}^{T^{(u)}(j)} \omega^{(u)}(t_i).
\]

In addition, we define the volatility of an interval \( j \) as \(|R^{(u)}(j)|\).

The data set under our study consists of the tick-by-tick recordings representing \( N = 31 \) stocks that were listed (at least temporarily) in DJIA over the period from January 1, 2008 to July 31, 2011. For each stock, we select three time lags \( \Delta t \) equal to 1, 5, and 10 min and for each lag we create four parallel time series of length \( L \approx 3 \times 10^5 \), \( L \approx 6 \times 10^4 \), and \( L \approx 3 \times 10^4 \), respectively, corresponding to the variables listed above. These variables, especially the returns and volume traded, are known to exhibit power-law tails of the probability distribution functions for short time lags \( \Delta t \) of the order of seconds and minutes that tend to depart from power law for longer \( \Delta t \), which was documented earlier in [17–19] and recently reconfirmed in [20] for a similar data set as in our present work. Next, we remove the daily trends from each of the unsigned time series \( T^{(u)} \), \( V^{(u)} \), \( |R^{(u)}| \) by a standard procedure [21], in which the daily pattern \( D^{(u)}_Z \) is calculated as an arithmetic average of the respective quantity \( Z \) taken at a specific interval \( j \) of each trading day \( k \) over all \( N_{\text{days}} \) trading days:

\[
D^{(u)}_Z(j) = \frac{\sum_{k=1}^{N_{\text{days}}} Z^{(u)}(k, j)}{N_{\text{days}}}. \quad (3)
\]

It is removed by division:

\[
Z^{(u)}_{\text{det}},(k, j) = Z^{(u)}(k, j)/D^{(u)}_Z(j). \quad (4)
\]

For simplicity of the notation, we will omit the superscripts \( (u) \) in \( R^{(u)} \), \( |R^{(u)}| \), \( T^{(u)} \), and \( V^{(u)} \) henceforth.

Power-law detrended autocorrelations. – Such autocorrelations can be detected by applying the multifractal detrended correlation analysis that is defined as follows [22]. Let there be a time series \( x(i), i = 1, \ldots, L \) divided into \( 2M_s \) separate boxes of length \( s \) (i.e., \( M_s \) boxes starting from the opposite ends). In each box \( \nu \) \((\nu = 0, \ldots, 2M_s - 1)\) a local trend of an integrated signal is approximated by an \( m \)-th–degree polynomial \( P^{(m)} \) and removed by subtraction:

\[
X^{\nu,s}_\nu(i) = \sum_{j=1}^{i} x(i + j) - P^{(m)}_{X^{\nu,s}_\nu}(i)
\]

producing a detrended signal \( X^{\nu,s}_\nu \) of length \( s \) (we will use \( m = 2 \) throughout this study). Variance of this signal can be expressed by

\[
\frac{f^2(s, \nu)}{s} = \frac{1}{8} \sum_{s=1}^{s} X^{\nu,s}_\nu(i)
\]

and can be used to define a family of the fluctuation functions of order \( q \):

\[
F^q(s) = \frac{1}{2M_s} \sum_{\nu=0}^{2M_s-1} \left[ f^2(s, \nu) \right]^{q/2}. \quad (7)
\]

The fluctuation functions can be used to detect fractality of the signal \( x(i) \) if they obey a power-law

48001-p2
form: $F^q(s) \sim s^{\gamma(q)}$, where $\gamma(q)$ can either be linear or nonlinear. In the former case, $x(i)$ is called monofractal, while in the latter case it is called multifractal. If one prefers to look at a graphical representation of the fluctuation functions, it is more instructive to use their modified versions:

$$F_q(s) = [F^q(s)]^{1/q},$$

since in that case one can easily distinguish between a monofractal and a multifractal scaling: $F_q(s) \sim s^{h(q)}$, where $h(q)$ is constant for a monofractal signal and variable for a multifractal one.

Figure 1 shows the fluctuation functions $F_q(s)$ calculated for sample stocks with different types of scaling: XOM and AIG. One can see that all types of signals reveal at least approximate scaling but sometimes for different scale ranges. We obtain the same result for other considered stocks as well. The multifractal scaling is a well-known feature of the stock-market data [23–30], therefore we consider possible fractal cross-correlations among these signals, instead.

**Power-law detrended cross-correlations.** — We proceed along the line defined by the multifractal detrended cross-correlation analysis in its MFCCA variant [15]. It is distinct from MF DFA in that here one deals with two detrended signals: $X_u(s,i)$ and $Y_v(s,i)$ in each box $v$ instead of only one signal in eq. (5). Variance defined by eq. (6) becomes covariance:

$$f_{XY}^2(s,\nu) = \frac{1}{s} \sum_{i=1}^{s} X_u(s,i)Y_v(s,i)$$

and can have negative values now. The respective $q$-th-order fluctuation function is then expressed by the following formulae:

$$F_{XY}^q(s) = \frac{1}{2M_s} \sum_{\nu=0}^{2M_s-1} \text{sign} [f_{XY}^2(s,\nu)] [f_{XY}^q(s,\nu)]^{q/2}$$

and

$$F_{qXY}(s) = \text{sign} [F_{XY}^q(s)] [F_{XY}^q(s)]^{1/q}.$$  

The definition of the functions in eqs. (10) and (11) assures us that no imaginary parts may be obtained by raising $f_{XY}^2$ to a real power $q/2$ and $F_{XY}^q$ to a real power $1/q$, and also assures us that no information about the signs is lost by taking the moduli. Moreover, for $q = 2$ the MFCCA procedure consistently transforms itself into the basic detrended cross-correlation analysis (DCCA) [16], exactly as desired. A fractal character of the cross-correlations is related to a power-law scaling: $F_{XY}^q(s) \sim s^{h(q)}$ and $F_{XY}(s) \sim s^{\lambda(q)}$, where $\delta(q) \sim q$ and $\lambda(q) = \text{const}$ denote monofractality. (We stress the necessity of defining $F_{XY}^q$ exactly as in eq. (10), because the frequent, published attempts to avoid complex numbers by taking moduli of the detrended signals in eq. (9) lead to spurious multifractal scaling for virtually all the power-law autocorrelated signals even if they are mutually uncorrelated by construction [15].)

For each stock, we apply the MFCCA procedure to signal pairs representing 6 combinations of the 4 data types. The results show that there are significant differences in the character of the power-law cross-correlations between different pairs of signals and that these differences are consistent across all the stocks. In fig. 2 we present plots of the fluctuation functions $F_{XY}(s)$ calculated for different pairs of signals representing the stock WMT. This stock is characterized by the power-law autocorrelations existing in each of the 4 signals: $R$, $|R|$, $T$, and $V$, so it is well suited for serving as a representative example. Out of the studied data types, the returns are the only ones that are signed, so it is not surprising that the returns are the least cross-correlated with the other signals. Figure 2 (top left) displays $F_{XY}(s)$ for the pair $R-V$, while the remaining two pairs, $R-|R|$ and $R-T$, look qualitatively similar: the power-law behaviour is poor (if any) and restricted to $q > 0$ and to approximately one decade of scales (not shown here). Consequently, the scaling exponent $\lambda(q)$ cannot be defined despite the fact that the average generalized
cate that the cross-correlations might be multifractal but which \( F(\text{red/grey, insets}) \) and \( \lambda \) (black, insets) calculated for different pairs of 1 min time series representing the stock WMT: (a) returns and volume traded, (b) volatility and trading activity, (c) volatility and volume, and (d) trading activity and volume. Only the results for those values of \( q \) are shown, for which \( F_{sXY}(s) \) exhibits at least approximate scaling. The same range of \( q \) is used as in fig. 1 with \( q = 4 \) denoted by the topmost lines.

Hurst exponent,

\[
h_{xy}(q) = \frac{h_x(q) + h_y(q)}{2},
\]

is defined for the whole analyzed range of \( q \).

The next two pairs: \( |R|-T \) and \( |R|-V \) are similar to each other (fig. 2, top right and bottom left). In both cases, \( F_{sXY}(s) \) is a well-defined, two-and-a-half-decade wide power law for \( q > 0 \), while it is wildly unstable for \( q \leq 0 \) and fluctuates from positive to negative values. This means that the signals are fractal cross-correlated indeed, but these correlations are restricted mainly to large fluctuations of \( |R|, T, \) and \( V \). This result goes in parallel with the results of an earlier study [12] that also proved the existence of such cross-correlations in a market index. Both the variable slope of the fluctuation functions and the decreasing values of the scaling exponent \( \lambda(q) \) indicate that the cross-correlations might be multifractal but this multifractality is rather subtle (a small range of the \( \lambda(q) \) variability).

The cross-correlations that comprise both the small and the large fluctuations can be seen in \( T-V \) (fig. 2 (bottom right)). Here \( F_{sXY}(s) \) are power-law correlated for all the considered values of \( q \) and for more than two decades of the scales. This means that \( T-V \) is the pair that retains the fractal structure of their autocorrelations in their cross-correlations the best. Consistently, it is also the only pair, in which \( h_{xy}(q) \) is followed by \( \lambda(q) \) (inset to fig. 2 (bottom right)). Such strong power-law correlations in the properties of the trading activity and volume support the assumptions of Gabaix et al.’s model of the mechanism that produces power-law distributions in the financial data [31,32].
multifractality of the data difficult if not impossible, because such values of $\Delta \lambda(q)$ can be obtained either if the cross-correlations are actually monofractnal with only spurious multifractality due to the broad probability distribution functions of the data (see [28] for the related discussion) or if there exists actual but weak multifractality. In the case of $\lambda(q)$ being not monotonous (seen in three panels of fig. 4(b)), the type of fractality is impossible to be defined at all.

The above results were obtained for the shortest of the considered temporal scales $\Delta t = 1$ min. However, the longer time scales: $\Delta t = 5$ min and $\Delta t = 10$ min bring results that are qualitatively similar. Figure 5 shows the fluctuation functions $F_{qXY}(s)$ calculated for such longer scales in the case of two sample stocks: XOM and HPQ. These results have to be put in the context of the respective panels of fig. 4. In fact, a comparison of the behaviour of $F_{qXY}(s)$ for different $\Delta t$ indicates that there is little dependence of the quality of scaling on a time scale, at least for the considered range of $\Delta t$. Also both the exponents $h_{xy}(q)$ and $\lambda(q)$ closely resemble each other.

Finally, we join the time series representing the same quantity (either $T$ or $V$) and the same time scale $\Delta t$ for all the 31 stocks and in this way we obtain much longer signals. The above results were obtained for the shortest of the 31 stocks and in this way we obtain much longer signals. The results have to be compared with the respective panels of fig. 4. Negative values of $F_{qXY}(s)$ were removed.

---

Fig. 4: (Colour online) The cross-correlation fluctuation function $F_{qXY}(s)$ (main plots) and the scaling exponents: $h_{xy}(q)$ (red/grey, insets) and $\lambda(q)$ (black, insets) calculated for time series of trading activity and volume traded with $\Delta t = 1$ min and with $-4 \leq q \leq 4$ for different stocks: (a) sample stocks with well-defined scaling behaviour of the cross-correlations and (b) sample stocks with less clear power-law cross-correlations. Negative values of $F_{qXY}(s)$, which occur for a few scales $s$ if $q$ is negative, were removed in order to make the plots more readable. In such a case, the respective exponents $\lambda(q)$ are undefined and not shown here.

Fig. 5: (Colour online) The cross-correlation fluctuation functions $F_{qXY}(s)$ (main plots) and the scaling exponents: $h_{xy}(q)$ (red/grey, insets) and $\lambda(q)$ (black, insets) calculated for time series of trading activity and volume with $\Delta t = 5$ min and $\Delta t = 10$ min and with $-4 \leq q \leq 4$ for two representative stocks: XOM (well-defined scaling, top) and HPQ (worse scaling, bottom). These results have to be compared with the respective panels of fig. 4. Negative values of $F_{qXY}(s)$ were removed.

Fig. 6: (Colour online) The cross-correlation fluctuation functions $F_{qXY}(s)$ (main plots) and the scaling exponents: $h_{xy}(q)$ (red/grey, insets) and $\lambda(q)$ (black, insets) calculated for the joined time series of trading activity and volume traded obtained from $N = 31$ stocks with $\Delta t = 1$ min (length: $L^{(N)} \approx 10^7$, $5$ min ($L^{(N)} \approx 2 \times 10^6$), and $10$ min ($L^{(N)} \approx 10^6$) calculated for $-4 \leq q \leq 4$. Negative values of $F_{qXY}(s)$ were removed as in fig. 4.
obtained for the individual stocks. The functions $F_{qXY}(s)$ show also that for $s > L$ their plots tend to flatten and the scaling exponents $\lambda(q)$ decrease to zero. Such behaviour is triggered by the existence of seasonal components [33,34], which here are related to the long-term trends in the market evolution that influence the majority of the stocks. Such trends cannot be completely removed by the $m$-th-degree polynomial $P(m)$ in eq. (5) and they thus survive the detrending procedure and the more strongly they manifest themselves, the stronger the inequality $s > L$ is. As expected, there is no difference between $h_{2\nu}(q)$ and $\lambda(q)$.

Summary. – We studied high-frequency data from the American stock market comprising 31 stocks listed in DJIA over a 3.5-years-long period between 2008 and 2011. Four quantities were of interest to us: logarithmic returns $R$, volatility $|R|$, trading activity $T$, and volume traded $V$. In the preliminary phase, the time series were detrended in order to remove daily seasonalities. Then, in the proper phase of our study, we confirmed by means of the MFDFA method that the time series representing all these quantities are at least approximately power-law autocorrelated with the Hurst exponent $h(q) > 0.5$ for $q < 2$. Next we applied the MFCCA method to pairs of the signals representing the same stock but different quantities and found that the strongest power-law cross-correlations are seen for trading activity and volume, while the weakest ones for returns and any of the other three quantities. The pairs comprising volatility and trading activity or volatility and volume presented the behaviour that situated them somewhere in between. These results were typical from the qualitative perspective, but there were also quantitative differences among the stocks in the quality of the power-law scaling or in the fractal or multifractal properties. We also found that these results were roughly invariant under a change of time unit $\Delta t$ from 1 min to 10 min.

Our observation that the highest power-law cross-correlations occur between trading activity $T$ and volume traded $V$ reveals the existence of strong statistical coupling between these quantities. In particular, it also supports a view that high volumes are produced by a high number of small trades rather than by large individual ones, as one of the assumptions of the model by Gabaix et al. states [31,32].

***

This work was partially supported by the Centre for Innovation and Transfer of Natural Sciences and Engineering Knowledge (University of Rzeszów).

REFERENCES

[1] Kwapień J. and Drożdż S., Phys. Rep., 515 (2012) 115.
[2] Ying C. C., Econometrica, 34 (1966) 676.
[3] Crouch R. L., Am. Econ. Rev., 60 (1970) 199.
[4] Jones C. M., Kaul G. and Lipson M. L., Rev. Financ. Stud., 7 (1994) 631.
[5] Chan K. and Fong W.-M., Financ. Econ., 57 (2000) 247.
[6] Lillo F., Farmer J. D. and Mantegna R. N., Nature, 421 (2003) 129.
[7] Potters M. and Bouchaud J.-P., Physica A, 324 (2003) 133.
[8] Bouchaud J.-P., Gefen Y., Potters M. and Wyart M., Quant. Finance, 4 (2004) 176.
[9] Lim M. and Coggsins R., Quant. Finance, 5 (2005) 365.
[10] Zhou W.-X., Quant. Finance, 12 (2012) 1253.
[11] Rak R., Drożdż S., Kwapień J. and Oświęcimka P., Acta Phys. Pol. B, 44 (2013) 2035.
[12] Podobnik B., Horvatic D., Petersen A. M. and Stanley H. E., Proc. Natl. Acad. Sci. U.S.A., 106 (2009) 22079.
[13] Yuán Y., Zhiuang X.-T. and Liu Z.-Y., Physica A, 391 (2012) 3484.
[14] Wang D.-H., Suo Y.-Y., Yu X.-W. and Lei M., Physica A, 392 (2013) 1172.
[15] Oświęcimka P., Drożdż S., Forczek M., Jadach S. and Kwapień J., Phys. Rev. E, 89 (2014) 023305.
[16] Podobnik B. and Stanley H. E., Phys. Rev. Lett., 100 (2008) 084102.
[17] Drożdż S., Kwapień J., Gräumer F., Ruf F. and Speth J., Acta Phys. Pol. B, 34 (2003) 4293.
[18] Drożdż S., Forczek M., Kwapień J., Oświęcimka P. and Rak R., Physica A, 383 (2007) 59.
[19] Rak R., Drożdż S. and Kwapień J., Physica A, 374 (2007) 315.
[20] Botta F., Moat H. S., Stanley H. E. and Preis T., PLoS ONE, 10 (2015) e0135600.
[21] Liu Y., Gopikrishnan P., Cizeau P., Meyer M., Peng C.-K. and Stanley H. E., Phys. Rev. E, 60 (1999) 1390.
[22] Kantelhardt J. W., Zschiegner S. A., Bunde A., Havlin S., Koscielny-Bunde E. and Stanley H. E., Phys. A, 316 (2002) 87.
[23] Ivanova K. and Ausloos M., Eur. Phys. J. B, 8 (1999) 665.
[24] Di Matteo T., Aste T. and Dacorogna M. M., Physica A, 324 (2003) 183.
[25] Oświęcimka P., Drożdż J. and Kwapień J., Physica A, 347 (2005) 626.
[26] Muzy J. F., Bacry E., Barle R. and Poggi P., EPL, 82 (2008) 60007.
[27] Caiwet L. and Fisher A., Rev. Econ. Stat., 84 (2002) 381.
[28] Drożdż S., Kwapień J., Oświęcimka P. and Rak R., Physica A, 391 (2012) 115.
[29] Di Matteo T., Aste T. and Dacorogna M. M., Physica A, 324 (2003) 183.
[30] Podobnik B., Grosse I., Horvatic D., Ilic S., Ivanov P. Ch. and Stanley H. E., Eur. Phys. J. B, 71 (2009) 243.
[31] Drożdż J., Bogachev M. I., Kantelhardt J. W., Schumann A. Y. and Bunde A., Physica A, 390 (2011) 2480.