

Alternative Random Matrix Approach in Analysis of Correlations in Financial Data

M. SAWA, D. GRECH*

Institute of Theoretical Physics, University of Wrocław, Econophysics and Time Series Analysis Group (ETSA), pl. M. Borna 9, PL-50204 Wrocław, Poland

We present an alternative method based on random matrix approach that enables to distinguish the respective role of temporal autocorrelations inside given time series and cross correlations between various time series. The proposed algorithm is based on the properties of Wigner eigenspectrum of random matrices instead of commonly used Wishart eigenspectrum methodology. It is then qualitatively and quantitatively applied to financial data of stocks building WIG 30 — the main Warsaw Stock Exchange Index.

DOI: [10.12693/APhysPolA.127.A-118](https://doi.org/10.12693/APhysPolA.127.A-118)

PACS: 05.45.Tp, 02.60.-x, 89.20.-a, 89.75.-k, 89.65.Gh, 89.75.Fb

1. Introduction

One often considers cross correlations between 1-dim time series X_i^α and X_j^β , where $i, j = 1, \dots, T$ is the length of discussed data and $\alpha, \beta = 1, \dots, N$ mark different series. As a standard attempt to do so the two-point simultaneous cross correlation function $C^{\alpha\beta}$ ($-1 \leq C^{\alpha\beta} \leq 1$)

$$C^{\alpha\beta} = \frac{1}{T} \sum_{i,j=1}^T X_i^\alpha X_j^\beta \delta_{ij} \quad (1)$$

for centered and normalized data in X^α , X^β is usually evaluated. The cross correlations with some time lag can also be considered in a similar manner.

The very elegant way to look at global cross correlation properties between all considered series is based on random matrix (RM) approach. In this description, one calculates the spectrum of N eigenvalues λ_n ($n = 1, \dots, N$) of $C_{N \times N}^{\alpha\beta}$ matrix (eigenspectrum), which in turn is the subject of comparison with the corresponding eigenspectrum of independent and identically distributed Y_i^α data with finite variance. The eigenspectrum $\rho(\lambda)$ of correlation matrix $M^{\alpha\beta}$ of Y^α and Y^β series, known as Wishart-Marčenko-Pastur (WMP) spectrum [1, 2], reads for $T, N \rightarrow \infty$

$$\rho_{\text{WMP}}(\lambda) = \frac{Q}{2\pi\sigma} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \quad (2)$$

where $Q = \frac{T}{N} = \text{const}$ is kept, σ is the standard deviation of Y_i^α data and

$$\lambda_{\pm} = (1 \pm 1/\sqrt{Q})^2 \quad (3)$$

are the edge values of WMP spectrum.

Since the Wishart spectrum is limited to $\lambda_- \leq \lambda \leq \lambda_+$, any deviation from this limit may have significance of cross correlation present in data. Moreover, the spread of eigenvalue spectrum of $M^{\alpha\beta}$ with respect to

$\lambda_- \leq \lambda \leq \lambda_+$ tells us on the strength of such cross correlations in a system producing data X_i^α ($\alpha = 1, \dots, N$; $i = 1, \dots, T$).

This RM approach [3–6] was successfully used in econophysics and in finance to look for cross correlations between various one-dimensional subseries of multidimensional time series built by various stocks data (see, e.g., [7–11]). To illustrate this idea we provide an example based on stock data taken from the main Polish stock exchange index WIG 30 in the period April 1, 2010–Dec. 30, 2013 which corresponds to total $T = 936$ inputs of $N = 26$ companies[†].

Let us define the returns

$$r_i = \frac{p_i - p_{i-1}}{p_{i-1}}, \quad (4)$$

where p_i is a price of a given stock at i -th day which are thereafter organized into $W_{N \times T}$ matrix of centered and normalized returns. The top part of Fig. 1 shows the eigenspectrum of $C^{\alpha\beta}$ correlation matrix compared with the corresponding Wishart spectrum, while the bottom part of Fig. 1 reveals the results of similar analysis done for absolute returns $|r_i|$, i.e., the simplest transform of returns discussed in economics and econophysics.

The deviation from Wishart spectrum (blue plot) is usually reported in literature [7–11] as the evidence of cross correlation existing between returns of various stocks in considered time window. However, it is not difficult to notice that an approach based only on investigation of differences from WMP eigenspectrum may suffer in many real situations from several shortcomings. First, small statistics of data (particularly a small number of investigated one dimensional subseries) leads to imprecise edges of spectrum. Second, the WMP eigenspectrum method is completely robust to temporal autocorrelations in data. Thus the question arises — to what

*corresponding author; e-mail: dariusz.grech@ift.uni.wroc.pl

[†]26 companies of WIG 30 had long enough data history in the discussed period (stocks not included: PZU, TPE, JSW, ALR).

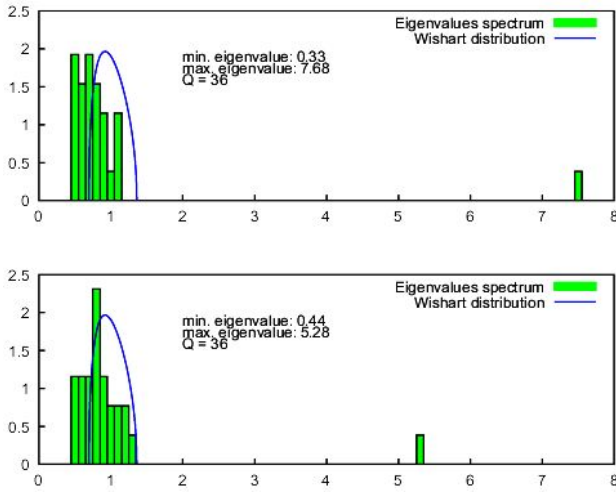


Fig. 1. Eigenspectra of correlation matrix $C(r)$ for returns (top part) and absolute returns $C(|r|)$ (bottom) of WIG 30 data in the period April 1, 2010–Dec. 30, 2013. The comparison with corresponding WMP spectrum of the same $Q = T/N$ is shown (blue plot) with numerical values from Eq. (3).

extend the particular spectrum like the one in Fig. 1 is a result of small number of considered time series N or the finite length T of series (which theoretically should be both infinite) and to what extent it is truly caused by cross correlations present in a system.

The next obstacle is that one would like to have a RM based method allowing to extract also temporal autocorrelations in data. An independent measurement of autocorrelations and cross correlations within the same RM approach would be a great advantage providing the same method of analysis to compare their strengths together. An important question about the mutual (relative) quantitative relation between autocorrelations in X^α and cross correlations between X^α and X^β ($\alpha \neq \beta$) cannot be answered within WMP analysis. Moreover, it would be nice to have a tool that investigates not only two-point correlations but also higher order contributions to correlation features.

One has to be also aware that direct calculations of cross and autocorrelation properties from given data is usually a nontrivial task. It is so because direct calculation of cross correlation (autocorrelation) function suffers from severe problems like noise present in data, possible non-stationarity and insufficient statistics (already mentioned also for WMP approach).

We therefore propose to analyze eigenvalue spectrum of square symmetric matrices constructed from entries of the primary $W_{N \times T}$ matrix built out of N time series containing T data each. For the purpose of this paper we will use WIG 30 data to do so. Then we will compare the obtained eigenspectrum with the Wigner semicircle distribution [12]. Note that the same approach can be easily extended to other data sets of similar form, even taken outside finance.

2. Description of method and data analysis

Let us recall that the eigenspectrum of the square, symmetric, real $N \times N$ matrix with independent N^2 centered entries and unit variance, known as Wigner spectrum, reads in the limit $N \rightarrow \infty$

$$\rho_W(\lambda) = \frac{1}{2\pi} \sqrt{4 - \lambda^2}. \quad (5)$$

The whole analysis is similar to the one proposed by one of us (D.G.) in [13]. In order to build square matrices from the original $W_{N \times T}$ matrix data we reshape it by splitting it into $[\frac{T}{m^2 N}]$, ($m = 1, 2, \dots$) matrices $W_{N \times m^2 N}$ with nonoverlapping entries. Then we augment these matrices one under another to form just one $W_{mN \times mN}$ matrix.

In the case of WIG 30 the original $W_{26 \times 936}$ matrix was reshaped into square matrix by splitting it into $m = 6$ horizontal/time sectors of size 26×156 , then augmenting these one under another. The final matrix obtained this way, after being symmetrized and normalized, is denoted further on as S ($S_{156 \times 156}$ in this case). The eigenspectra of $S_{156 \times 156}(r)$ and $S_{156 \times 156}(|r|)$ built respectively for returns r and absolute returns $|r|$ are shown in Fig. 2. The corresponding Wigner eigenspectrum for uncorrelated data has also been indicated in all figures as reference plot.

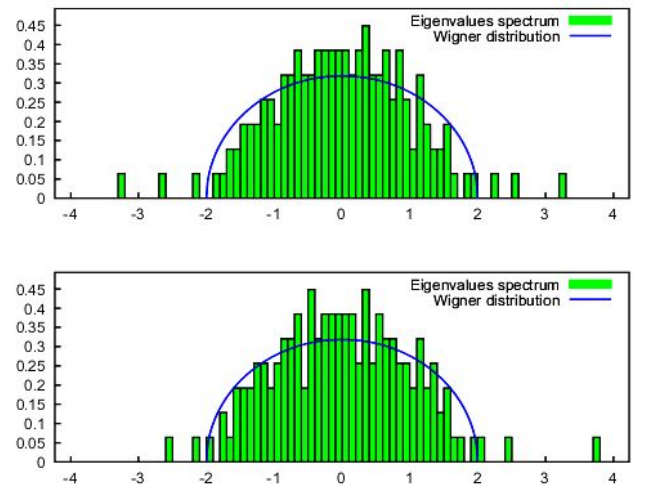


Fig. 2. Eigenspectra of just one $S_{156 \times 156}(r)$ (top) and $S_{156 \times 156}(|r|)$ (bottom) matrices constructed from $W_{26 \times 936}(r)$ and $W_{26 \times 936}(|r|)$ respective data of WIG 30 by reshaping procedure explained in the main text ($m = 6$). The comparison with corresponding Wigner eigenspectrum of uncorrelated data is made (blue curve).

The problem of small statistics seems to be evident here in a similar manner as in WMP approach because one may only speculate if tails outside the normalized eigenvalue range $\langle -2, 2 \rangle$ are the result of correlation between series or the effect of small statistics. However here, contrary to WMP approach, one may increase this statistics reshaping the primary matrix $W_{N \times T}$ in various ways. For instance, choosing $m = 2$ one finds 9 independent

matrices $S_{52 \times 52}$ instead of just one $S_{156 \times 156}$ for $m = 6$. The total number of eigenvalues in the available statistics will increase then up to 468 ($m = 2$) in opposite to the former 156 ($m = 6$). Thus, by folding up the horizontal sectors in various ways, one obtains a less or more abundant statistics from the available data as shown in Fig. 2 compared with Fig. 3. The latter figure clearly suggests that a difference from Wigner plot in Fig. 2 is not a result of small statistics because for better statistics deviation from Wigner plot becomes even more evident.

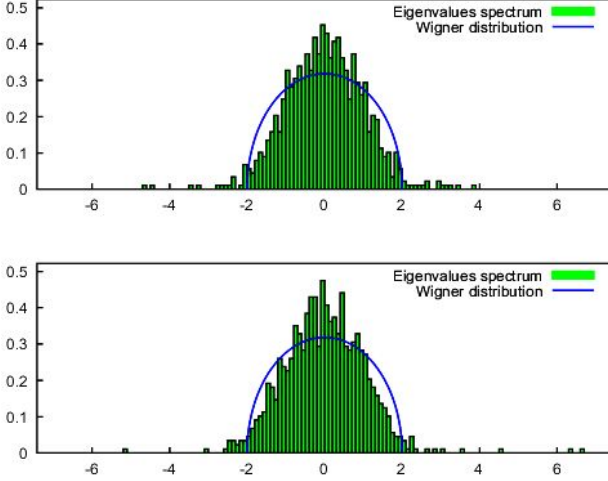


Fig. 3. Averaged eigenspectra of 9 independent matrices $S_{52 \times 52}(r)$ (top) and $S_{52 \times 52}(|r|)$ (bottom) constructed from reshaped $W_{26 \times 936}(r)$ and $W_{26 \times 936}(|r|)$ data of WIG 30 respectively ($m = 2$). The Wigner semi-circle for uncorrelated data of the same size $N \times T$ is also drawn for comparison (blue curve).

Now, we may consider the problem if within RM method one is able to distinguish the role and respective weights of cross and autocorrelations present in signals. Note that the answer to such question was negative in case of standard $C^{\alpha\beta}$ correlation approach based on WMP eigenspectrum. It turns out however that in our alternative approach it is possible to exhaust such information by applying diversified shuffling methods to all signals. In order to kill cross correlations between different companies (rows in $W_{N \times T}$ matrix), preserving however autocorrelations (of all orders), we make random “cyclic shifting” of data in rows of original matrix $W_{N \times T}$ before reshaping it into $S_{156 \times 156}$ one which has no cross correlations present. This matrix will be denoted here further on as S_{ac} .

We have chosen to perform all calculations based on $S_{156 \times 156}$ matrix ($m = 6$) since the shuffling procedure offers sufficient (as we will see) statistics, so that this statistics does not have to be additionally amplified by reshaping original WIG 30 data into very small square matrices, i.e., $S_{52 \times 52}$ ($m = 2$) or $S_{26 \times 26}$ ($m = 1$) respectively. The details of applied procedure are as follows.

A natural number n_α , ($1 \leq n_\alpha \leq T$, $\alpha = 1, \dots, N$) is chosen at random from discrete uniform distribution separately for each row of $W_{N \times T}$ matrix and

a new matrix $W_{N \times T}^{(cs)}$ is formed with cyclically shifted rows — the α -th row ($w_{\alpha 1}, w_{\alpha 2}, \dots, w_{\alpha T}$) is replaced by $(w_{\alpha n_\alpha}, w_{\alpha n_\alpha + 1}, \dots, w_{\alpha T}, w_{\alpha 1}, w_{\alpha 2}, \dots, w_{\alpha n_\alpha - 1})$. Then $W_{N \times T}^{(cs)}$ matrix is reshaped into a square matrix and symmetrized as before. We calculate its spectrum and the process is repeated 10^4 times. It results in spectra (averaged eigenspectrum) presented in Fig. 4.

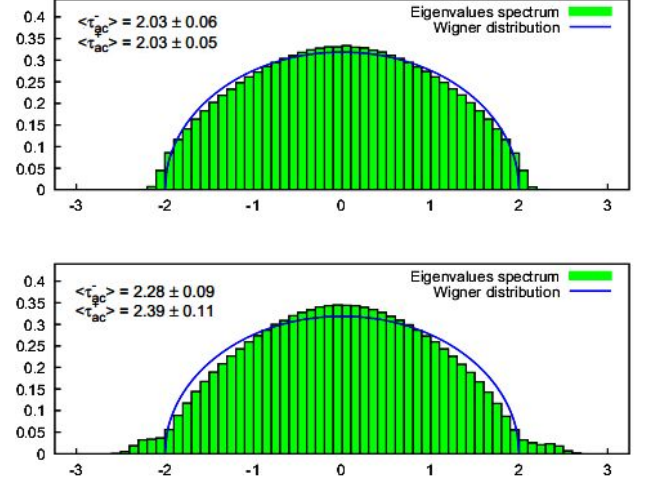


Fig. 4. Averaged eigenspectra of $S_{ac}(r)$ matrix (top) with “killed” cross correlations (only autocorrelations are left in signal) extracted from $W_{26 \times 936}(r)$ of WIG 30 data. The same for absolute returns is shown for $S_{ac}(|r|)$ matrix (bottom). Wigner spectrum is shown for comparison (blue curve).

It follows from the above description that the procedure of “cyclic shuffling” can only marginally destroy autocorrelations of length $\geq \tau$ in primary signal containing T data by respective ratio $\leq \tau/T$ (0.5% for $\tau \lesssim 5$ days, i.e., for one trading week). Thus the method is particularly effective in searching for short and medium — range autocorrelations (short or medium-term memory effects in multidimensional data). One expects that long-term memory is not very much evident in financial data so the proposed method is particularly eligible for application here.

On the other hand, to kill autocorrelations preserving however cross correlations (of any order), we perform random “shuffling” of columns of the original data matrix reshaping resultant matrices into another $S_{156 \times 156}$ ones denoted further on by S_{cc} . Due to the way the shuffling was done in this case, cross correlations are still preserved. The corresponding averaged eigenspectrum in this case after 10^4 repetitions is presented in Fig. 5. The average position of the tail edge of eigenvalue spectra is marked as $\langle \tau_{cc}^\pm \rangle$ in case of killed autocorrelations (cross correlations are left only), and $\langle \tau_{ac}^\pm \rangle$ in case of killed cross correlations (autocorrelations being left), where \pm corresponds to the right (positive) and to the left (negative) tail respectively.

The above findings should be referred to the case of entirely random uncorrelated data. The averaged

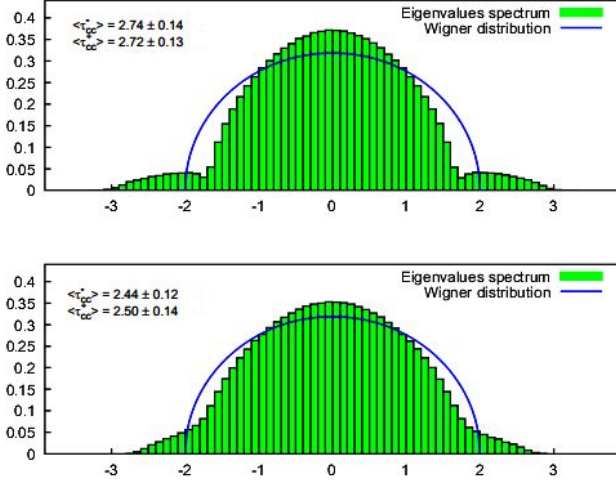


Fig. 5. Same as in Fig. 4 but for “killed” autocorrelations in signal (only cross correlations are present).

eigenspectra of matrices S_0 obtained by reshaping in a similar manner the completely shuffled $W_{N \times T}$ data (to remove all cross and autocorrelations) is presented in Fig. 6.

The eigenspectrum of $S_0(r)$ matrix has the averaged tail edges positioned at: $\langle \tau_0^+ \rangle(r) = 1.97 \pm 0.05$ and $\langle \tau_0^- \rangle(r) = -1.97 \pm 0.05$ where the similar notation as for S_{ac} and S_{cc} matrices is used. The corresponding results for eigenspectrum of $S_0(|r|)$ matrix built for absolute returns read: $\langle \tau_0^+ \rangle(|r|) = 2.00 \pm 0.07$ and $\langle \tau_0^- \rangle(|r|) = -1.99 \pm 0.05$. In all cases the mean and the standard deviation is taken from the ensemble eigenspectra of 10^4 matrices since the process of respective shuffling was repeated 10^4 times. The spectra in Figs. 4, 5 compared with Fig. 6 clearly indicate existence of correlation (autocorrelation and/or cross correlation) in analyzed signals. In particular, the comparison of tail edges between S_0 and S_{ac} indicates that the edges of spectrum between completely random signal and real financial data are at least 1σ apart for returns and 5σ apart for absolute returns.

Thus the noticed differences in average tail lengths shown in Fig. 4 (with extracted autocorrelations) and in Fig. 6 (for the shuffled signal with no correlation at all) indicate detection of very weak autocorrelations in returns and somehow more significant autocorrelations in absolute returns (both of any order). On the other hand, the comparison of results presented in Fig. 5 (extracted cross correlations) with those in Fig. 6 indicate detection of cross correlations between time series. The presented analysis is not only able to detect correlations in multi-dimensional data but offers also possibility to detect separately autocorrelations (Fig. 4) and cross correlations (Fig. 5). Moreover, the differences in mean tail lengths shown in consecutive figures, when compared with reference plot of Wigner semicircle for uncorrelated matrix entries, may be exploited to estimate quantitatively relative weights of cross correlations and autocorrelations.

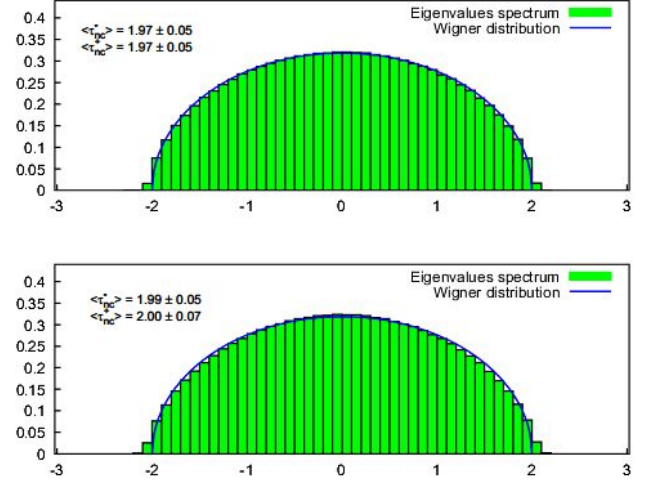


Fig. 6. Averaged eigenspectra of $S_0(r)$ (top) and $S_0(|r|)$ (bottom) matrices constructed from 10^4 shuffles of $W_{26 \times 936}(r)$ and $W_{26 \times 936}(|r|)$ data matrix of WIG 30 respectively. The very good correspondence with Wigner distribution (blue plot) is indicated.

The predominance of cross correlations over autocorrelations in returns (r) is visible as $|\langle \tau_{cc}^\pm \rangle(r)| \gg |\langle \tau_{ac}^\pm \rangle(r)|$ if one compares the average lengths of eigenspectrum distributions for $S_{ac}(r)$ and $S_{cc}(r)$. This difference does not manifest as much for absolute returns $|r|$ since here $|\langle \tau_{cc}^\pm \rangle(|r|)| \gtrsim |\langle \tau_{ac}^\pm \rangle(|r|)|$ (compare bottom parts of Fig. 4 and 5). The relative strength of cross correlations and autocorrelations can be estimated introducing the subsequent ratios defined in a straightforward manner

$$\Delta_{cc/ac}(r) =$$

$$\frac{[\langle \tau_{cc}^+ \rangle(r) - \langle \tau_{ac}^- \rangle(r)] - [\langle \tau_0^+ \rangle(r) - \langle \tau_0^- \rangle(r)]}{[\langle \tau_{ac}^+ \rangle(r) - \langle \tau_{ac}^- \rangle(r)] - [\langle \tau_0^+ \rangle(r) - \langle \tau_0^- \rangle(r)]} \quad (6)$$

in case of returns, and

$$\Delta_{cc/ac}(|r|) =$$

$$\frac{[\langle \tau_{cc}^+ \rangle(|r|) - \langle \tau_{ac}^- \rangle(|r|)] - [\langle \tau_0^+ \rangle(|r|) - \langle \tau_0^- \rangle(|r|)]}{[\langle \tau_{ac}^+ \rangle(|r|) - \langle \tau_{ac}^- \rangle(|r|)] - [\langle \tau_0^+ \rangle(|r|) - \langle \tau_0^- \rangle(|r|)]} \quad (7)$$

for absolute returns $|r|$.

Very similar analysis allows to compare quantitatively the autocorrelation levels between returns (r) and absolute returns ($|r|$):

$$\delta_{ac}(|r|/r) =$$

$$\frac{[\langle \tau_{ac}^+ \rangle(|r|) - \langle \tau_{ac}^- \rangle(|r|)] - [\langle \tau_0^+ \rangle(|r|) - \langle \tau_0^- \rangle(|r|)]}{[\langle \tau_{ac}^+ \rangle(r) - \langle \tau_{ac}^- \rangle(r)] - [\langle \tau_0^+ \rangle(r) - \langle \tau_0^- \rangle(r)]} \quad (8)$$

and analogously, between change of cross correlation levels while we transfer from returns to absolute returns:

$$\delta_{cc}(|r|/r) =$$

$$\frac{[\langle \tau_{cc}^+ \rangle(|r|) - \langle \tau_{cc}^- \rangle(|r|)] - [\langle \tau_0^+ \rangle(|r|) - \langle \tau_0^- \rangle(|r|)]}{[\langle \tau_{cc}^+ \rangle(r) - \langle \tau_{cc}^- \rangle(r)] - [\langle \tau_0^+ \rangle(r) - \langle \tau_0^- \rangle(r)]} \quad (9)$$

Substituting numerical values of tail positions for eigenspectra of the above generated matrices we find

$\Delta_{cc/ac}(r) \approx 12.67$ and $\Delta_{cc/ac}(|r|) \approx 1.40$. Simultaneously, the relative strength of autocorrelations between $|r|$ and r reaches the level $\delta_{ac}(|r|/r) \approx 5.67$, while the strength of cross correlations drops down for $|r|$ when compared with r because for the latter ratio one finds $\delta_{cc}(|r|/r) \approx 0.63$.

The latter results can be compared with the average two point Pearson correlation values $\langle C_{cc}^{(2)} \rangle$ calculated directly from Eq. (1) for each company separately, where the average is taken over all companies in the index. To be precise let us define:

$$\langle C_{cc}^{(2)} \rangle = \frac{2}{N(N-1)} \sum_{\alpha>\beta} C_{cc}^{\alpha\beta}, \quad (10)$$

where $N = 26$. On the other hand, the average two point autocorrelation function $\langle C_{ac}^{(2)} \rangle$ can be simply postulated as:

$$\langle C_{ac}^{(2)} \rangle = \frac{1}{N} \sum_{\alpha=1}^N \left[\sum_{s=1}^T \frac{1}{T} \langle X^\alpha(t) X^\alpha(t+s) \rangle_t^2 \right]^{\frac{1}{2}}, \quad (11)$$

where the maximal one month trading time lag T can be assumed to incorporate both short and medium-term autocorrelations.

The obtained numerical results $\langle C_{cc}^{(2)}(r) \rangle = 0.18$, $\langle C_{cc}^{(2)}(|r|) \rangle = 0.27$, $\langle C_{ac}^{(2)}(r) \rangle = 0.02$ and $\langle C_{ac}^{(2)}(|r|) \rangle = 0.11$ lead to corresponding quantitative ratios $\delta_{ac}^{(2)}(|r|/r) \approx 5.50$ and $\delta_{cc}^{(2)}(|r|/r) \approx 0.65$ being in very good agreement with RM outcomes shown above. Note however, that these two point correlations can only be treated as the ‘‘first order’’ approximation, while RM approach deals with higher order corrections as well.

3. Concluding remarks

Concluding, we state that RM approach based on Wigner spectrum analysis can be used to compare qualitatively and quantitatively different forms of correlations in multidimensional data. In our example based on Warsaw Stock Exchange data, the role of autocorrelations increases and the role of cross correlations decreases when one proceeds from returns (r) to absolute returns $|r|$ (thus revealing the importance of sign in returns). The approach considering the average tail lengths of probability distribution we presented in this article (instead of cumulative distributions of matrix eigenspectra) seems to be statistically more reliable since it eliminates large fluctuations from eigenspectrum. This is why we adopted it here.

In this RM based approach higher orders of correlations as well as the influence of short term memory is included. Therefore we should not be surprised that very small autocorrelation effect in primary data was observed (see Fig. 4 in comparison with Fig. 6). This effect increases about six times when one moves from r to $|r|$. Simultaneously the cross correlation level in between series is sensitive to the presence of sign in returns in an opposite direction. The cross correlations strength between absolute returns is about 40% smaller than the cross correlation strength for returns. It confirms an importance

and influence of sign in price change on the magnitude of cross correlation.

Note that in the presented analysis all quantitative results take automatically into account also higher correlation orders. Therefore they are more general than the outcomes of the standard approach when only two-point correlation function is considered. Finally, it is worth emphasizing that this kind of analysis can be easily extended to investigate mutual relationship between arbitrary data in multidimensional time series of any origin.

References

- [1] J. Wishart, *Biometrika* **A 20**, 32 (1928).
- [2] V.A. Marčenko, L.A. Pastur, *Math. USSR Sb.* **1**, 457 (1967).
- [3] M. Mehta, *Random Matrices*, Academic Press, New York 1995.
- [4] T. Guhr, A. Müller-Groeling, H.A. Weidenmüller, *Phys. Rep.* **299**, 190 (1998).
- [5] L. Laloux, P. Cizeau, J.P. Bouchaud, M. Potters, *Risk Mag.* **12**, 69 (1999).
- [6] L. Laloux, P. Cizeau, J.P. Bouchaud, M. Potters, *Phys. Rev. Lett.* **83**, 1467 (1999).
- [7] V. Plerou, P. Gopikrishnan, B. Rosenov, L.N. Amaral, H.E. Stanley, *Phys. Rev. Lett.* **83**, 1471 (1999).
- [8] V. Plerou, P. Gopikrishnan, B. Rosenov, L.N. Amaral, T. Guhr, H.E. Stanley, *Phys. Rev. E* **65**, 066126 (2002).
- [9] B. Podobnik, D. Wang, D. Horvatić, I. Grosse, H.E. Stanley, *Europhys. Lett.* **90**, 68001 (2010).
- [10] S. Drożdż, J. Kwapien, P. Oświęcimka, *Acta Phys. Pol. B* **38** 4027 (2007).
- [11] D. Wang, B. Podobnik, D. Horvatić, H.E. Stanley, *Phys. Rev. E* **83**, 046121 (2011).
- [12] E.P. Wigner, *Ann. Math.* **53**, 36 (1951).
- [13] D. Grech, J. Miśkiewicz, *Europhys. Lett.* **97**, 30005 (2012).