Asymmetry of cross-correlations between intra-day and overnight volatilities

RUBINA ZADOURIAN\(^1\) and PETER GRASSBERGER\(^2\)

\(^1\) Max-Planck-Institut für Physik Komplexer Systeme - Nöthnitzer Strasse 38, D-01187 Dresden, Germany
\(^2\) JSC, FZ Jülich - D-52425 Jülich, Germany

received 30 January 2017; accepted in final form 15 May 2017
published online 30 May 2017

PACS 89.65.Gh – Economics; econophysics, financial markets, business and management

Abstract – We point out a stunning time asymmetry in the short-time cross-correlations between intra-day and overnight volatilities (absolute values of log-returns of stock prices). While overnight volatility is significantly (and positively) correlated with the intra-day volatility during the following day (allowing thus non-trivial predictions), it is much less correlated with the intra-day volatility during the preceding day. While the effect is not unexpected in view of previous observations, its robustness and extreme simplicity are remarkable.

Copyright © EPLA, 2017

Understanding the volatile feature of financial markets and their complex behaviour intrigues many researchers. The recent financial crises and risk portfolios require a better understanding of financial market dynamics. The risks are measured by the variance of the asset returns and the square root of the variance is called volatility. It is well known that fluctuations in equity prices traded on any stock exchange are weakly correlated, and hardly allow any non-trivial forecasting. This is different for the amplitudes of these fluctuations, namely for volatilities. If the market is hectic, volatilities are large, and usually it will take some time until the market has become calm again. This fact is related to the volatility clustering phenomenon: according to Mandelbrot [1], large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes. The fact of volatility clustering as an aggregation has inspired researchers and some works are devoted to explain the origin of the volatility clustering in terms of reflection and the behavior of traders, see, e.g., [2].

One of the challenges is to forecast the financial market volatility, which has an important meaning for portfolio management. For example if the volatilities are predicted to be high for a given asset, then the investors may reduce their commitments to the asset, in order to avoid risk. The observation of volatility clustering gives evidence of predictability of volatility. The stronger influence of negative returns than of positive returns on the future volatility and possible predictability have been studied in, e.g., [3,4]. With some modern options it is possible to make profits with forecasts for volatilities (although forecasts of signed fluctuations would be more easy to turn into money, if they were possible), whence volatilities have been studied extensively in the econometric literature [3,5].

As first shown in [6,7], the statistics of volatilities is not uniform. Rather, there is a marked daily structure, with high volatility during the opening hour of the market and a more calm period around noon. Also, equity prices at the opening of the trading session are in general different from the closing prices on the previous trading day, showing that there is a non-trivial overnight dynamics.

Finally, it has been repeatedly shown that the overnight dynamics is qualitatively different from that during the day [8–14]. Various reasons have been proposed for this:

- A foreign equity which is mainly traded on some foreign market (that is open during the night hours of the market studied) reflects mostly its activity in their overnight volatility, and this activity might be very different [8] from the market under consideration.

- The majority of news relevant for fundamental stock price evaluation (company profits, employment rates, general econometric forecasts, wars and natural disasters, ...) are released overnight [15], and there exists a correlation between frequency of news releases and volatilities [9,14].

- While the market can react during the day to any outside perturbation, it cannot do so during the night,
which might also explain the higher volatility immediately after the market opening [14].

One of the interesting and attractive areas in economy is predictability. The general consensus seems that overnight volatility is useful for predicting subsequent intra-day volatility [16–18]. This is an important result. But prediction involves a model (GARCH [12,14], SEMI-FAR [12], or different versions of the stochastic volatility model (SVM) [10,11,13]), and none of the papers cited above report model-independent analyses of the raw data themselves. This is so in spite of the fact that data analyses not involving any model and using only elementary methods and minimal assumptions would be most useful for understanding the basic mechanism(s) underlying the phenomena. It is the purpose of the present letter to provide just such an elementary analysis. The methods used here are based on correlation coefficients like Spearman [19]. Yet the result is striking and completely unexpected, as far as significance and robustness are concerned. We should point out that an extensive statistical study of intra-day and overnight returns and volatilities was recently made in [20], but since that analysis was not guided by any theoretical considerations, the effect described below was missed.

Let us use the index \( k \) to count trading days (i.e., skipping weekend and other non-trading days), and denote by \( o_k \) and \( c_k \) the opening and closing prices of one particular equity. Intra-day log-returns of this equity are defined as

\[
d_k = \ln \frac{c_k}{o_k},
\]

while overnight log-returns are

\[
n_k = \ln \frac{o_k}{c_{k-1}}.
\]

Thus overnight returns are indexed by the index of the following day. In case of weekends and holidays the overnight returns include all changes during the entire non-trading period. Volatilities are in principle defined through the variances of log-returns as observed over an extended time span. But when discussing them on a fine grained temporal scale, they are usually replaced by the absolute values of the log-returns (see, e.g., footnote 11 in [11]). We will follow this usage.

The data we studied consists of 21 individual stocks traded at various stock exchanges (Exxon, Shell, General Electric, Ford, Goldmann-Sachs, Bank of America, Citigroup, IBM, Microsoft, Cisco, AIG, BP, Caterpillar and Ford all traded at NYSE; Siemens, Deutsche Bank, Lufthansa, VW and Bayer traded in Frankfurt; and Sony & Mitsubishi traded in Tokyo) and 10 market indices and exchange-traded funds (TecDax, MDax, DAX, Dow Jones, S&P 100, Nasdaq, EuroSTOXX 50, SIM, S&P/ASX and PowerShares QQQ). They were mostly downloaded from Yahoo (https://finance.yahoo.com/) the rest from finanzen.net (http://www.finanzen.net). The time sequences cover between 10.4 and 45 years, with between 2612 and 13478 data points. Before using them, we cleaned them from some of their artifacts (missing data, wrong data, . . . ), but not of all. For instance, we did not remove jumps due to stock splitting. After cleaning, they show the typical features well known from previous analyses, such as fat tails, short-time correlations in the returns, and long-time correlations in the volatilities. For typical examples, see fig. 1. Notice that these data still have outliers (mostly negative, due to crashes, bad annual reports, . . . ). The negative outliers occur mostly for the overnight returns, consistent with the previous observation that negative news are disseminated mostly when the markets are closed. The long autocorrelations of the volatilities are seen both for daytime and overnight.

Our main concern is with cross-correlations between intra-day and overnight volatilities. Due to the artifacts, irregularities, and strong non-stationarity in the data, we did not use simple Pearson coefficients. Instead we mostly used Spearman coefficients [19]. Spearman correlation coefficients show a dependence between two variables, which are based on rank statistics. These are known to be much more robust. This implies monotonic relationships including non-linearity. To a certain extent the Spearman correlation coefficient is insensitive to the difference between raw data. For these reasons it shows more robustness for our finding.

Our main results are shown in figs. 2 and 3. In fig. 2 we show for each equity two cross-correlations between the
Asymmetry of cross-correlations between intra-day and overnight volatilities

Fig. 2: (Color online) Data for 31 equities. Each dot corresponds to one equity. The Spearman correlation between intra-day volatilities and overnight volatilities during the subsequent night are plotted on the x-axis, while the correlations with the preceding night are on the y-axis.

Fig. 3: (Color online) Ratios $C_{nd}/C_{dn}$ for the 31 equities and indices shown also in fig. 2. For stocks, the colors indicate the stock exchanges where they are traded.

ranks $r_{dk}$ and $r_{nk}$ of the two volatilities $|d_k|$ and $|n_k|$

$$C_{nd} = \frac{\langle r_{dk} r_{nk} \rangle - \langle r_{dk} \rangle \cdot \langle r_{nk} \rangle}{\sigma_d \sigma_n}$$

(3)

is the rank correlation between the intra-day volatility and the volatility during the preceding night ($\sigma_d$ and $\sigma_n$ are the square roots of the rank variances), while

$$C_{dn} = \frac{\langle r_{dk} r_{nk+1} \rangle - \langle r_{dk} \rangle \cdot \langle r_{nk} \rangle}{\sigma_d \sigma_n}$$

(4)

gives the analogous correlation with the following night.

We see that in all cases

$$C_{nd} > C_{dn}.$$  

(5)

For some equities the difference is small, but for others it can be more than a factor of two. In only one case the inequality was violated. Thus, the overnight volatility is much stronger correlated with the volatility during the following day than during the preceding day. Otherwise said, overnight volatilities seem to influence strongly what goes on during the following trading day, but do not seem to be strongly influenced by what was going on during the day before\(^1\). The ratios $C_{nd}/C_{dn}$ for the equities used in fig. 2 are plotted also in fig. 3, where we have also specified the equities. The first 10 entries in this figure are market indices, while the others correspond to individual stocks. We see no big differences, except that aggregated indices show a somewhat stronger effect. There are also no noticeable differences related to the place where the equity is traded, to the length of the time series, and — in case of individual stocks — to the type of company.

Next, we would like to demonstrate that the found asymmetry is not an artefact of the chosen correlation coefficient. It is most clearly seen by the use of the Spearman correlation coefficient, but still visible by the use of Pearson coefficients, see fig. 4. In comparison to fig. 3, however, we observe in fig. 4 a sizable fraction of data points lying below 1.

Alternatives to the Spearman correlation coefficient are Kendall’s $\tau$ [21] or mutual information [22], both of which are known to be similarly robust. Let us finally consider here mutual information for proving our results. In information theory, quantities such as entropy and mutual information are introduced, which are functions of the probability distributions of the underlying process. Mutual information is a reduction of uncertainty knowing another random variable. It tells us how much two random variables share information among each other.

Consider two random variables $X$ and $Y$ with a joint probability function $p(x, y)$ and marginal probability functions $p(x)$ and $p(y)$. Mutual information is defined by the following formula:

$$I(X, Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}.$$  

(6)

\(^1\)Notice in this context that the title of [10] is wrong. That paper is concerned with the usefulness of overnight information for prediction, not with its predictability.
It is clear that the mutual information is symmetric in $X$ and $Y$ and always non-negative and is equal to zero if and only if $X$ and $Y$ are independent, where in our case $X$ and $Y$ are the intra-day and overnight volatilities. $I_{nd}$ is the mutual information for $(x_i, y_i) = (d_k, n_k)$ and $I_{dn}$ is the mutual information for $(x_i, y_i) = (d_k, n_{k+1})$.

For the mutual information estimation we use two approaches, histogram and K-nearest neighbor statistics. While histogram is easy to comprehend, it has several disadvantages: it depends on the bin width and different initial choices can change the outcome. In contrast to this conventional method based on binnings, the K-nearest neighbour statistics are based on the entropy estimation of K-nearest neighbour distances. This means that they are data efficient, adaptive and have minimal bias [23]. The results are shown in fig. 5, which shows the robustness of our finding.

Another interesting question is about the relation between stocks, conditioned on a third one. To answer this question, we investigate in an upcoming work conditional mutual information analyses. Motivated by such analyses, the Pearson partial correlation coefficient has been also performed in [24]. However, these analyses are beyond the scope of this paper.

At first sight this strong asymmetry looks very strange, in particular since time asymmetry is usually considered to be very weak in financial data. Many popular models (most noticeably all models of the ARCH family) are time symmetric by construction, and where time asymmetry is seen [16,25] it is only seen in very special observables. But our findings are indeed compatible with previous analyses [7–14]: While the intra-day price dynamics is largely influenced by “chartist” behavior, the overnight dynamics is mostly influenced by facts exogenous to the stock market (or at least not directly related to the day-to-day price evolution of the considered equity) and thus of “fundamentalist” nature. What our results suggest is that “fundamentalist” information is more useful in prediction than “chartist” information.

The present analysis cannot of course specify which of the possible external influences (foreign stock markets, company performance reports, news about general economic indicators such as employment rates and forecasted economic growth, wars, economic crises, natural disasters, ...) is of greatest importance for the overnight dynamics, but such information could possibly be obtained by performing a larger study similar to the present one in which equities are grouped according to business sectors, stock exchanges, trading volume, bull vs. bear markets, etc. Another improvement suggested by our analysis could consist in replacing the simple cross-correlations by partial correlations or by transfer entropies [26], testing in this way for linear or nonlinear Granger causality [27]. It would be of interest to see whether the asymmetry found in the present paper is also present at larger time scales, by comparing day/night to night/day results between more distant nights and days.

Finally, with the hindsight gained from this analysis, we might also turn to signed returns (in contrast to volatilities) and test whether some parts of a full 24 hours day have more influence on periods than others. The very fact that different regions in the phase space of a recurrent system can have different powers of predictability has been known for long time [28].

Furthermore, we have done investigations by taking into account foreign markets in different time zones. Our analyses show that European markets at their opening time are largely driven by Asian and American markets. On the other side, we see that American markets are largely driven by both European and Asian market on the same day, whereas the analyses show stronger correlation between American and Asian markets for the closing prices. There is also evidence that the Asian market is affected by American markets on the previous day.

Intra-day volatilities can be considered with higher time resolution (even on the scale of seconds) than overnight volatilities, which consists of just a simple jump. The intra-day price fluctuations seem to show a kind of step response pattern. The fact that overnight and next-day intra-day volatilities are strongly coupled may indicate that intra-day volatilities are rather driven by the overnight jumps than by any short-term intra-day perturbations.

It is known that there is a strong nonlinear correlation between price returns. Therefore they can be used for prediction of bursts in time series and associated risk assessment, by using nonlinear approaches. This is mostly relevant in cases of localized breakdown of the symmetry between gain and losses, see [29,30]. Accordingly the observed strong correlations between overnight and intra-day volatilities could lead to an earlier predictability of gains/losses in financial markets, and further improve the risk assessment, giving an earlier Value-at-Risk estimate. It is also challenging to formulate a mechanistic model that allows for an investigation of the distribution function and its volatility by analytic techniques as in [31,32].
REFERENCES

[1] Mandelbrot B. B., J. Bus., XXXVI (1963) 392.
[2] Cont Rama, Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models (Springer) 2007.
[3] Bollerslev T., Engle R. F. and Nelson D. B., ARCH Models, in Handbook of Econometrics, Vol. 4 (Elsevier North Holland) 1994.
[4] Patton Andrew J. and Sheppard Kevin, Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility, Economic Research Initiatives at Duke (ERID), Working Paper No. 168 (November 16, 2013), available at SSRN: https://ssrn.com/abstract=1943825.
[5] Bouchaud J.-P. and Potters M., Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management (Cambridge University Press, Cambridge, UK) 2003.
[6] French K. R. and Roll R., J. Financ. Econ., 17 (1986) 5.
[7] Lockwood L. J. and Linn S. C., J. Financ., 45 (1990) 591.
[8] Chan K., Chockalingam M. and Lai K. W. L., J. Multinatl. Financ. Manag., 10 (2000) 495.
[9] Edmonds R. G. and Kutan A. M., Econ. Lett., 76 (2002) 223.
[10] Zhong Z., The Predictability of Overnight Information, Dissertations and Theses Collection, Paper 46, http://ink.library.smu.edu.sg/etd_coll/46.
[11] Tsiakas I., J. Bank. Finance, 32 (2008) 251.
[12] Chen C.-H., Yu W.-C. and Zivot E., Predicting Stock Volatility Using After-Hour Information, unpublished (2009).
[13] Lee C.-C. and Wang C.-S., Overnight information and stochastic volatility: A study of Asia stock markets (in Chinese), Master Thesis, Ming Chuan University (2009).
[14] Gallo G. M. and Pacini B., Stud. Nonlinear Dyn. Econ., 2 (1998) 115.
[15] Del Corral W., Colwell D., Michayluk D. and Woo L. A., News releases when the markets are closed, unpublished, University of Technology, Sidney (2003).
[16] Blanc P., Chicheportiche R. and Bouchaud J.-P., Physica A, 402 (2014) 58.
[17] Taylor N., J. Forecast., 27 (2008) 721.
[18] Chicheportiche R., Non-linear Dependencies in Finance, PhD Thesis, Ecole Centrale des Arts et Manufactures, Paris (2013).
[19] Press W. H. et al., Numerical Recipes: The Art of Scientific Computing, 3rd edition (Cambridge University Press) 2007.
[20] Wang F., Shieh S.-J., Havlin S. and Stanley H. E., Phys. Rev. E, 79 (2009) 056109.
[21] Kendall M., Biometrika, 30 (1938) 8.
[22] Cover T. M. and Thomas J. A., Elements of Information Theory (John Wiley & Sons) 2012.
[23] Kraskov Alexander, Stögbauer Harald and Grassberger Peter, Phys. Rev. E, 69 (2004) 066138.
[24] Kenett Dror Y., Huang Xuqing, Vodenska Irena, Havlin Shlomo and Stanley H. Eugene, Quant. Finance, 15 (2015) 569.
[25] Zumbach G. O., Quant. Finance, 9 (2009) 505.
[26] Schreiber T., Phys. Rev. Lett., 85 (2000) 461464.
[27] Barnett Lionel, Barrett Adam B. and Seth Anil K., Phys. Rev. Lett., 103 (2009) 238701.
[28] Deissler R. J. and Farmer J. D., Physica D, 55 (1992) 155.
[29] Bogachev Mikhail I. and Bunde Armin, Phys. Rev. E, 80 (2009) 026131.
[30] Bogachev Mikhail I. and Bunde Armin, Physica A, 390 (2011) 2240.
[31] Zadourian Rubina, Saakian David B. and Klümper Andreas, Phys. Rev. E, 94 (2016) 060102(R).
[32] Lafond François, Gotway Bailey Aimee, Bakker Jan David, Rebois Dylan, Zadourian Rubina, McSharry Patrick, Farmer J. Doyn, How well do experience curves predict technological progress? A method for making distributional forecasts, arXiv:1703.05979 (2017).