

Supplementary material

EMPIRICAL STUDY REGARDING THE ANALYSIS OF INVESTMENT BEHAVIOR ON HIGH-FREQUENCY TRADING AT THE ROMANIAN STOCK EXCHANGE

Florin Cornel DUMITER, Klaus Bruno SCHEBESCH, Florin TURCAȘ, Marius BOIȚĂ

Technological and Economic Development of Economy

<https://doi.org/10.3846/tede.2026.28512>

This material contains some detailed graphical representations of some data analysis steps which are pointed at but separated from the main article owing to space constraints. The contents of the article are comprehensible without this material but the full details of intermediary results are found below.

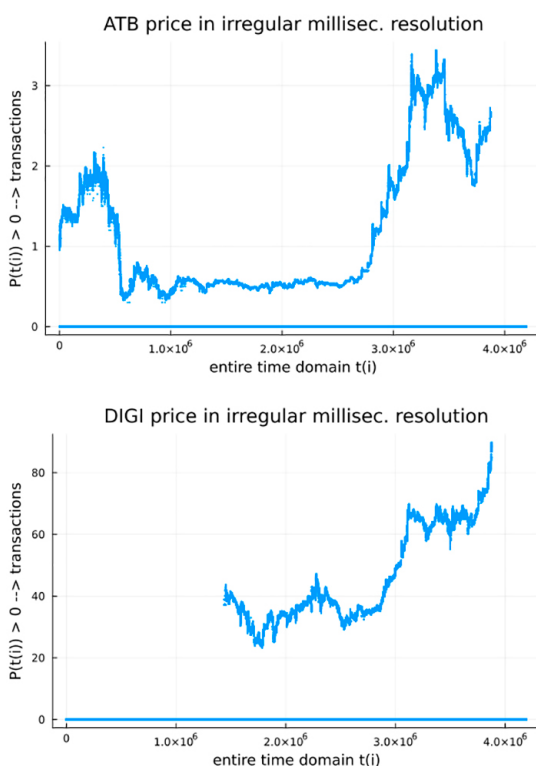


Figure S1. The stock price level evolution of the companies ATB and DIGI. These two companies have 262999 and respectively, 162881 data entries at the maximal resolution (trading event level). The fine-grained structure is not visible due to the high number of quotations. The line thickness implicitly stands for the time-local variation of the prices as a single price is plotted by a vanishingly small point

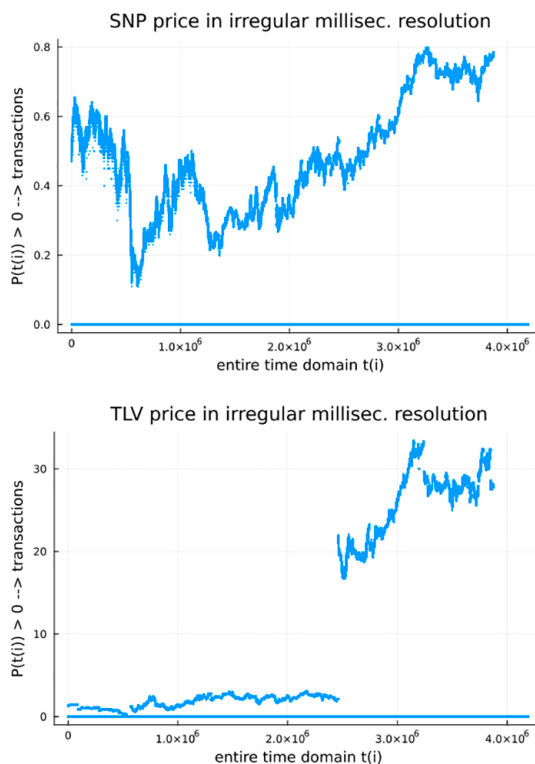


Figure S2. The stock price level evolution of the companies SNP and TLV. These two companies have 1567727 and respectively, 1883994 data entries at the maximal resolution (trading event level). The same comments apply as for Figure 2. Company TLV experienced a discontinuity caused by a merger with another major bank

The following Figures always display the data within the union over all four periods of activity. That is, the single plots will always show the entire time span (union of all activity) and will be plotted as a straight line at level zero, as stock exchange activity of single firms will start/end at different points in time.

Analyzing Figure S2, it can be seen that in the history of issuers, there have been a number of corporate events, for which it is necessary to adjust the data series to bring historical prices and volumes to the same base. In order to adjust the historical data series for the corporate events that have taken place, the prices and volumes from the date of registration of the event must be adjusted with the corresponding factors, calculated by the Bucharest Stock Exchange.

For example, at Banca Transilvania (TLV), the nominal value was consolidated on 17.08.2022 (from 1 lei/share to 10 lei/share), thus each shareholder had a number of shares 10 times lower.

However, from the point of view of price variation, it is not correct to correct all prices prior to the corporate operation, because volatility should not be recalculated retroactively. Simply, the first price change at the opening of the day in which the operation was recorded must be neglected, it is arbitrary (formal) and not determined by the investors' intention.

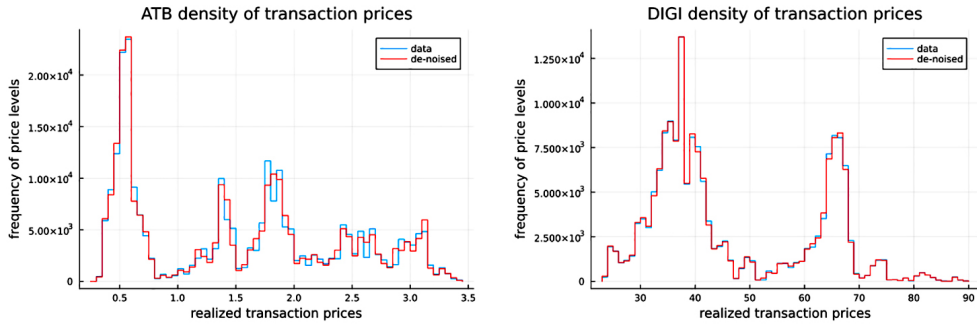


Figure S3. Histograms are used to explore the time-independent statistical properties of the stock-price behavior. Here are the cases for the companies ATB and DIGI. The original (full) data are depicted alongside a wavelet de-noised signal thereof. Owing to the long time series containing a large amount of information (variation over short time periods), de-noising of the time series and consequently showing the histogram thereof, does not lead to a strongly visible simplification of the original time series distribution for both companies. The company ATB's de-noising results in a somewhat smoother histogram than the same procedure done for company DIGI

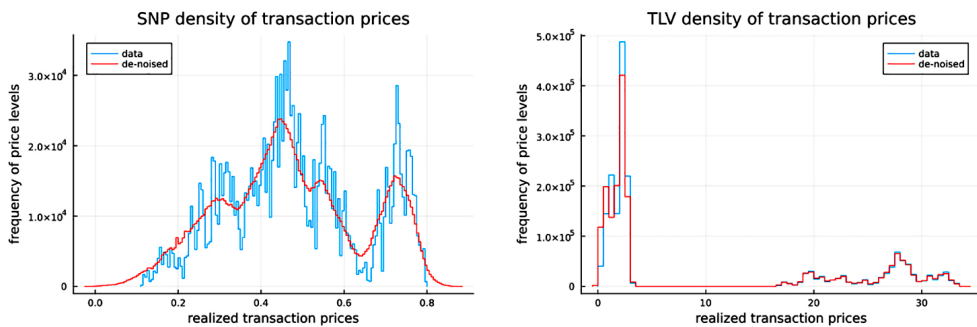


Figure S4. Histograms for the price evolution of companies SNP and TLV (in analogy to Figure S3). For the company, SNP wavelet denoising acts as a strong smoother, indicating redundancy in variation at the micro-intervals (one should not be misled by the apparent variation in the visible density). Company TLV reveals an extremely asymmetric density, which is marginally smoothed by denoising

Next, we show the time evolution of distances between consecutive transactions. We consider this to be valuable information, owing to, among others, long periods of constant prices – not visible in the above plots, which run over several hundred thousand up to more than a million data points per company. Alternatively, one may consider the time duration between actual price movements, in order to compensate that there are also lots of consecutive transactions at a constant price. Such time series referring to consecutive transactions (or, as indicated, between non-zero price movements) have the advantage that they need not be corrected over time, as is the case with the price process which may be subject to price recalibration, etc., which are convenience-based, and hence do not add the understanding of the dynamics of the trading process per se.

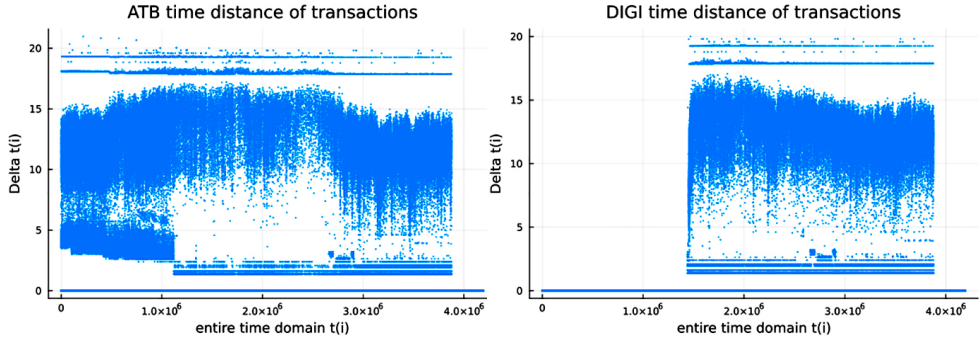


Figure S5. Log (time distance + 1) between consecutive transactions for companies ATB and DIGI. For ATB, one may distinguish several regimes. Long distinguished distances in the upper parts of the plots refer to calendric trade pauses. Distances of zero (at 1 in the specific log-plots) are multiple trades at the maximal time resolution unit (see also Figure 1 of the main article). Note two distinct trade patterns in terms of consecutive trade time distances for company ATB after passing over the 10^6 trade event mark

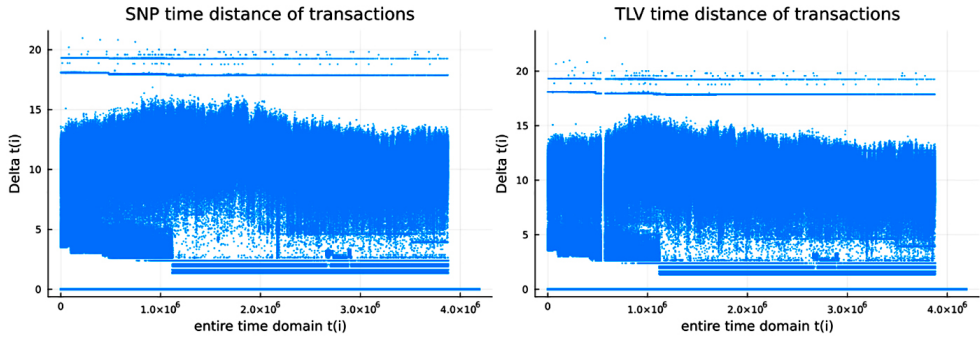


Figure S6. Log (time distance + 1) between consecutive transactions for companies SNP and TLV. These two companies had many more trading events than ATB and DIGI. Apparently, they share many qualitative features, at least as far as the limited graphics resolution permits to show

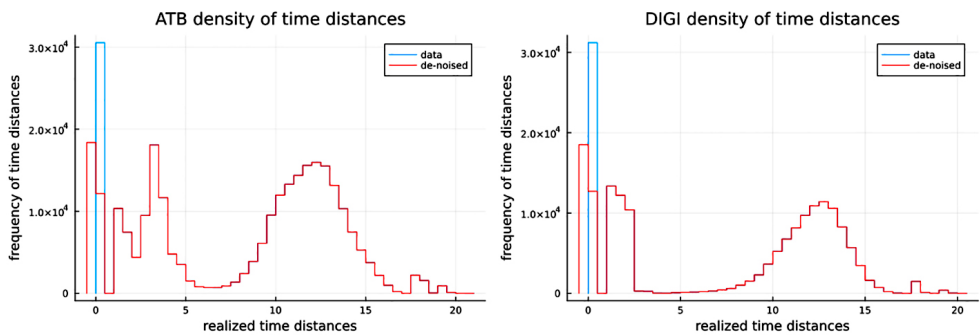


Figure S7. Histograms of the log-distance between trading events as depicted in Figure S5. Their de-noised variants essentially show a corrected regime for lower distances (towards the left end) and an unaffected regime with a bell-shaped structure

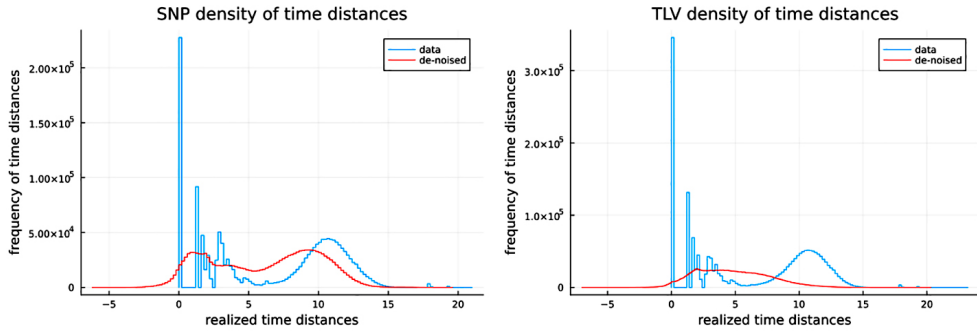


Figure S8. Histograms of the log-distance between trading events as depicted in Figure S6. These two stocks in their denoised variants show much more uncertainty (less structure) than the histograms for the companies with significantly fewer trading events

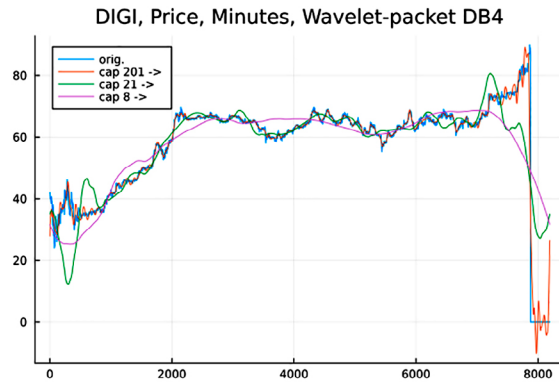


Figure S9. Wavelet smoothing of time series at the minute-resolution level for the company DIGI. This case requires more information for a higher-fidelity reproduction of the original, i.e., as cap 201 and more will do. It may indicate a more complex market, possibly owing to a predominantly more sophisticated investor type. Cap n refers to the number of wavelet coefficients used for the reproduction of the time series (see also the main article). The number n is obviously much smaller than the number of coefficients for the exact reproduction of the original time series

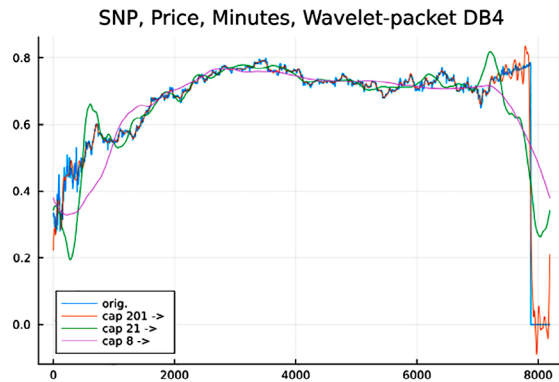


Figure S10. Wavelet smoothing of time series at the minute-resolution level for the company SNP. This situation, too, indicates that cap 201 (and above) will be needed to reach a higher-fidelity reproduction of the original. As this company is popular among investors, the increased information content may be due to a broad variety of investor types (with different time horizons and understanding). Also, the company may be popular in risk-reducing investment portfolios

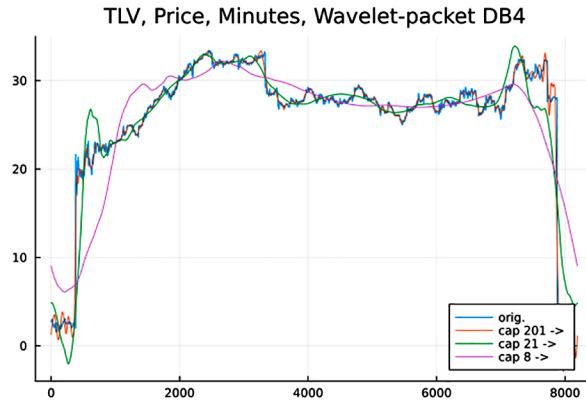


Figure S11. Wavelet smoothing of time series at the minute-resolution level for company TLV. This highly popular investment target contains information that may be reproduced with an acceptable to very good approximation by retaining cap 21 until cap 201. In fact, for Cap 201, the original can be very successfully reproduced

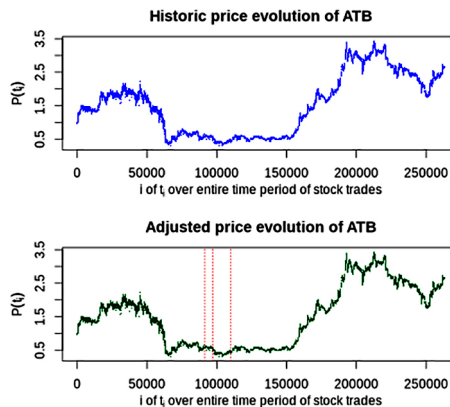


Figure S12. The historically recorded price evolution of company ATB and its adjusted variant

In the following three figures we show the historically recorded price data and the respective adjusted prices for the companies ATB, SNP, and TLV.

Analyzing Figure S12, it can be seen that vertical lines indicate the three points in (event-) time when adjustments were made. One notices that the differences from the original are minimal. Figure S13 highlights that there is just one price adjustment. As with the case of the company, the differences to the original are here minimal, too.

Figure S14 suggests that there are two major adjustments (nominal price changes by a factor of ten) and 16 much smaller ones is just one price adjustment. In this case, there is a substantial change (shift) in the visual pattern. In further time series analysis, the adjusted variant may be used. However, for many aspects of the statistical analysis, no significant difference in the outcomes may be expected.

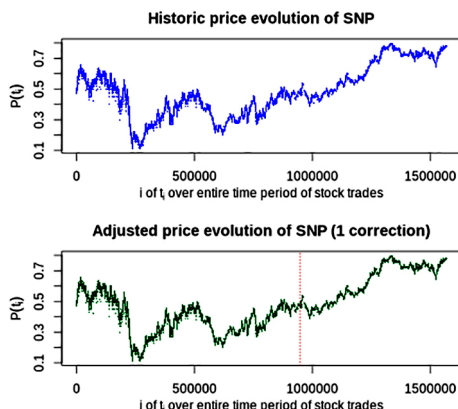


Figure S13. The historically recorded price evolution of the company SNP and its adjusted variant

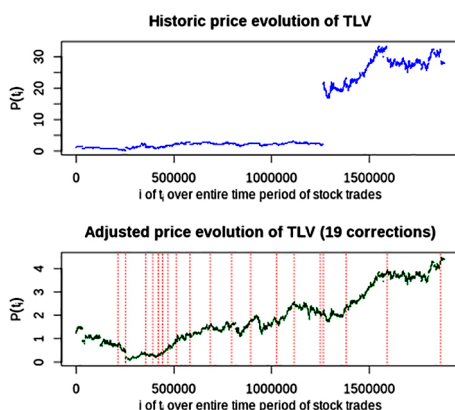


Figure S14. The historically recorded price evolution of the company TLV and its adjusted variant

Histograms. An aggregate view of company stock data

Another related view on the stock data, which might lead to confirmatory or, possibly, contradictory or confusing conclusions, is via confronting histograms with a suitable Gaussian reference.

From Figure S15, it can be established that the departure from a Gaussian density is clearly visible. Note that, besides many zero movements, small magnitude movements form a strong majority, indicating the action of higher order moments.

Figure S16 suggests that there are two concentrations in terms of price levels. Here too, a clear departure from a Gaussian density, in terms of asymmetry as well as higher symmetric moments, is visible.

Similar observations from Figure S17 concerning the departure from a Gaussian density; there are many zero price movements, asymmetry, and higher moments apply here as well.

From Figure S18, it can be seen that the departure from a Gaussian density, with many zero and small price movements, in analogy to the previous company plots, is confirmed for this stock as well.

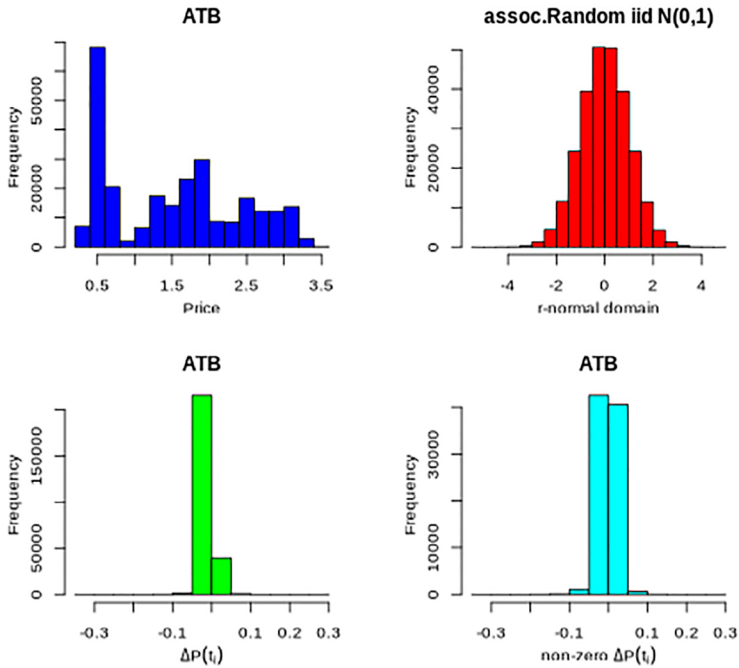


Figure S15. The histograms for the company ATB of price levels (upper left), for the Gaussian reference (upper right), and for the first differences (lower left), as well as the first differences without zero movements (lower right)

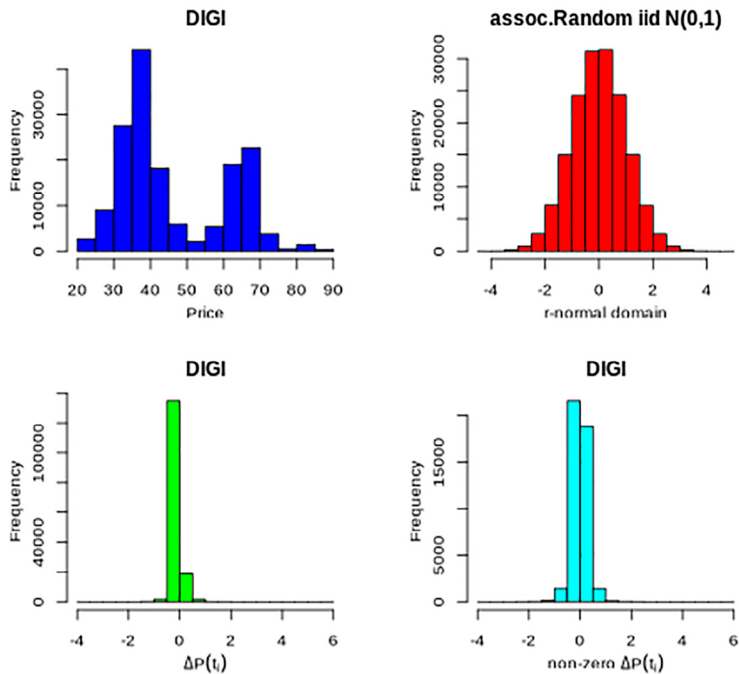


Figure S16. The histograms for the company DIGI

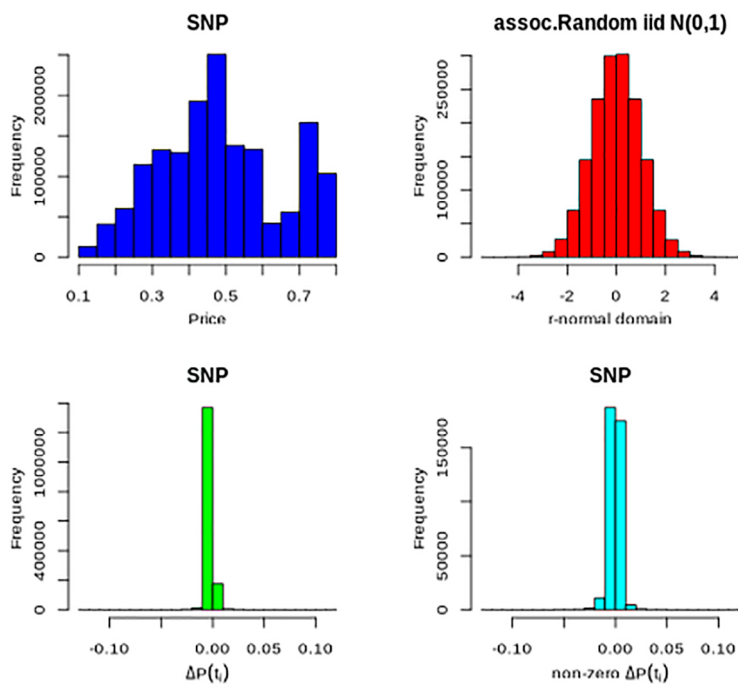


Figure S17. The histograms for the company SNP

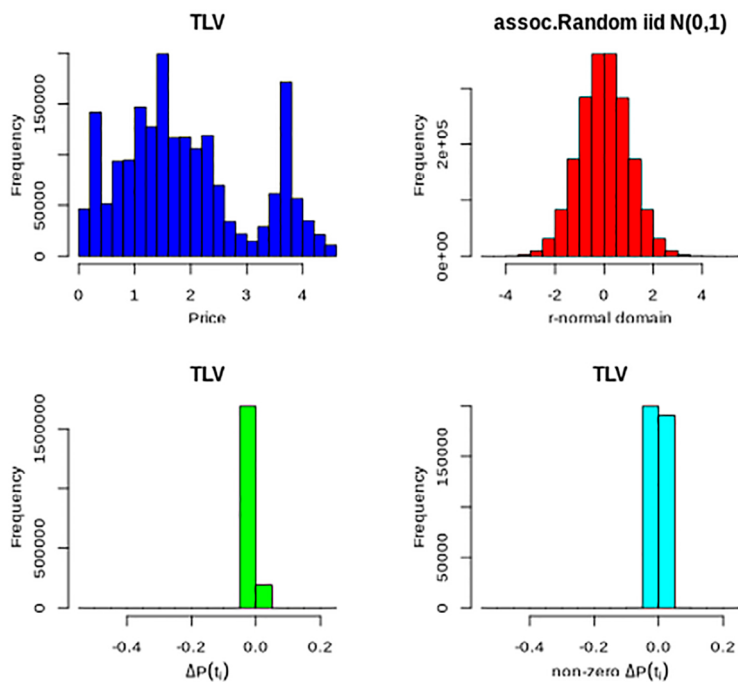


Figure S18. The histograms for the company TLV

Correlation analysis: a multivariate view

A first type of correlation analysis relates the four companies of our data set over time. Even this basic approach allows for a series of meaningful variants. Not that owing to the trading event-driven (irregular) successions of time points, down-sampling to coarser time resolutions is required. The criterion for using a time resolution is that more or all companies must have trade activity within the selected consecutive real-time intervals. While one may select arbitrary time intervals (from seconds all the way down to years) the chances that all four companies trade within a very small time interval are slim – at least for the stock exchange considered here. By down-sampling, the number of “simultaneous” transactions increases while the number of totally available aggregate data points decreases. For our data, **minutes** and **hour’s** time resolutions retain the most cases with simultaneous trades of all four firms. The order of magnitude reduces, however, sharply, from around 10^6 to 10^4 .

In the following, we choose the **minutes** time resolution to exemplify the type of correlation structure found between the company’s trading activities. This hopefully represents an approximation for each company’s overall trading activity, which must be verified by some indirect method (i.e., by comparing autocorrelations, etc.). In the following figures, the price evolution and the trading value (number of transactions times prices aggregated over all events from within each chosen minute) are shown. Hence, we correlate eight variables (i.e., P-ATB, P-DIGI, P-SNP, P-TLV, V-ATB, V-DIGI, V-SNP, V-TLV) over time. This is just a preliminary selection, as more (derived, exogenous) variables may be included.

Figure S19, left panel: the correlation matrix with numerical coefficients. Figure S19, right panel: the same correlation matrix deformed by a suitable monotonous function, **$\tanh(4x)$** , in order to highlight the differences between lower correlated variable pairs. In addition, for those pairs with insignificant coefficients (at the 0.95 level), their p-values are displayed (e.g., $p = 0.91$ for the (V-SNP, P-ATB)-entry, etc.). The variables are grouped by hierarchical clustering (**hclust**, the squares in the rhs panel), which, however, are as expected. Overall, there are surprisingly high correlations between price evolutions of companies, while all other variable pairs seem much more unrelated, but mostly significant.

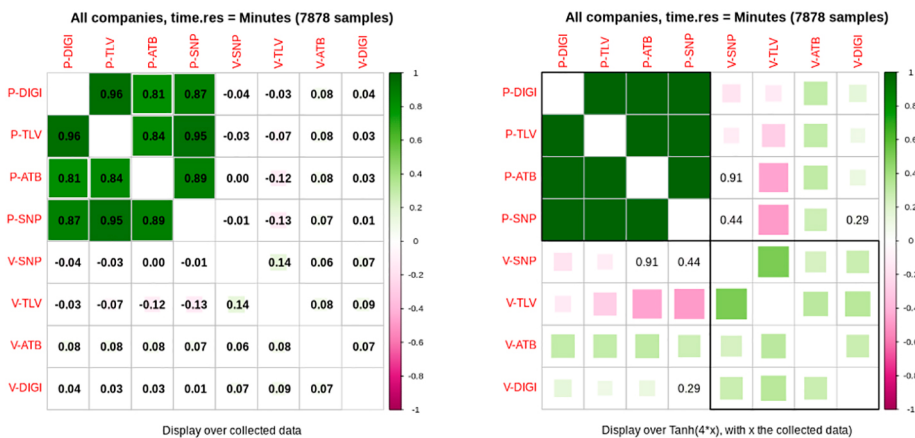


Figure S20 shows that it is intended to mimic a step in a search of sub-sequencing the data in order to reveal temporal information. Left panel: the positive correlation coefficient between the price evolution reduces, some (negative) correlation between the remaining variable pairs do increase markedly, and so does also the number of insignificant correlations (mainly due to the smaller number of samples).

Inversion of the ordering of the h-clustering implied squares is just an artifact of the software used and bears no meaning. The correlation coefficients are between value country pairs are more higher than in those from Figure S20, and those between the prices are still lower. Rather strong negative correlations between price–value pairs as found in Figure S20, also decrease in magnitude.

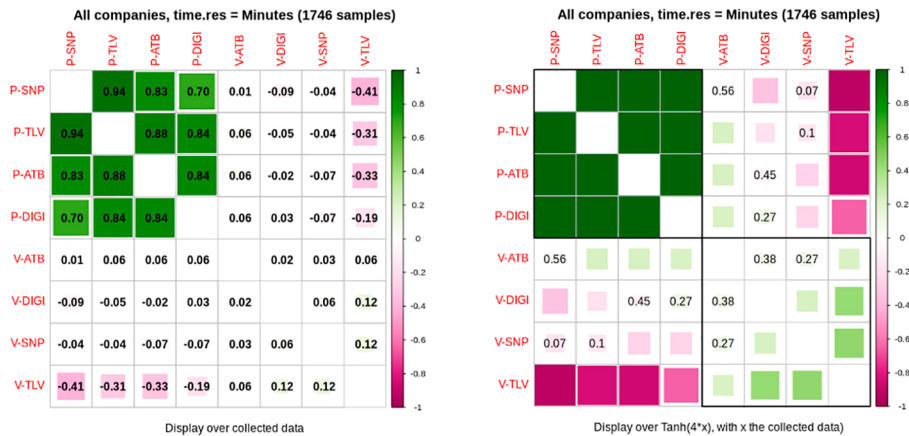


Figure S20. A pair of correlation plots with the same display logic as that from Figure S19, here for a randomly bounded subsequence – namely, the first 1746 data points from those used in Figure S19

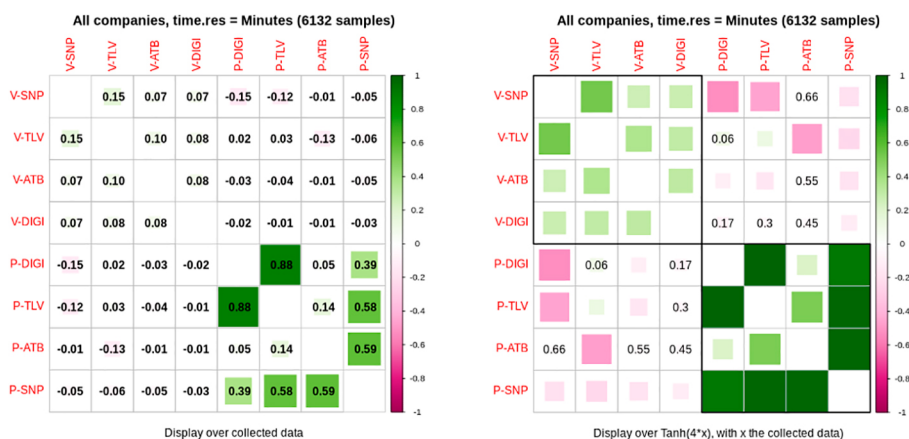


Figure S21. This pair of correlation plots should be read as a complement to Figure S20, in that here the correlations are computed on the remaining 6132 data points from those used in Figure S20 and not used in Figure S19

The lessons learned from these incipient correlation analyses are that long-term correlations hide more structured medium-term correlations, which would have different implications for trading systems and for stock portfolio design. The overall strong to very strong correlation of stock prices may indicate immaturity of a stock market or very optimistic overall expectations in an emerging economy. For further analysis concerning detecting degrees of uncertainty and finding other market features, one would seek two types of parametrizations for answering the questions of

- What time resolution of the data to use? There are a large number of potentially useful “intermediate” resolutions!
- Given the time resolution, which sub-sequencing of the data to use in order to obtain sequence informative correlation structures?

Finally, there might be utility in designing a multi-resolution approach for analysis and modeling, which would replicate different memory types and objectives of stock market agents.