Article

Efficiency of the Moscow Stock Exchange before 2022

Andrey Shternshis [1,*] , **Piero Mazzarisi** [1] **and Stefano Marmi** [1]

[1] Scuola Normale Superiore, Piazza dei Cavalieri 7, Pisa, Italy, 56126
* Correspondence: andrey.shternshis@sns.it

Abstract: This paper investigates the degree of efficiency for the Moscow Stock Exchange. A market is called efficient if prices of its assets fully reflect all available information. We show that the degree of market efficiency is significantly low for most of the months from 2012 to 2021. We calculate the degree of market efficiency by (i) filtering out regularities in financial data and (ii) computing the Shannon entropy of the filtered return time series. We have developed a simple method for estimating volatility and price staleness in empirical data, in order to filter out such regularity patterns from return time series. The resulting financial time series of stocks' returns are then clustered into different groups according to some entropy measures. In particular, we use the Kullback–Leibler distance and a novel entropy metric capturing the co-movements between pairs of stocks. By using Monte Carlo simulations, we are then able to identify the time periods of market inefficiency for a group of 18 stocks. The inefficiency of the Moscow Stock Exchange that we have detected is a signal of the possibility of devising profitable strategies, net of transaction costs. The deviation from the efficient behavior for a stock strongly depends on the industrial sector it belongs.

Keywords: Shannon entropy; market efficiency; volatility estimation; price staleness; stock market clustering; Kullback-Leibler divergence

1. Introduction

When prices reflect all available information, the market is called efficient [1]. One way to claim the efficiency of a market is by testing the Efficient Market Hypothesis (EMH). In the weak form, the EMH considers that the last price incorporates all the past information about market prices [2]. If the weak form of EMH is rejected, previous prices help to predict future prices. For traders, market efficiency means that analyzing the history of previous prices does not help to design a strategy that gives an abnormal profit. For a company issuing shares, market efficiency means that the cost of its share already reflects all information about the valuation and decisions of the company. The EMH is of great interest also in research. Mathematical models of an asset price are usually based on the assumption that the price follows a martingale: the expected value of a future price is the current value of the price. If the EMH is rejected, there should be an estimation of the future price better than its current value. In such a case, new models should be thought.

The review of works confirming the EMH was presented by Fama in 1970 [2] and then in 1991 [3]. The martingale hypothesis was also tested later. It was shown that the efficiency of a market depends on the development of the country [4]. Also, the martingale hypothesis was confirmed on short time intervals, but may be violated on longer intervals [5]. In addition, there is a range of strategies designed to increase an expected profit. High-frequency and algorithmic trading strategies are discussed in [6]. Statistical and machine learning methods for high frequency trading are reviewed in [7]. The existence of such profitable strategies contradicts the Efficient Market Hypothesis. According to Grossman and Stiglitz [8], the degree of market inefficiency determines the effort investors are willing to expend to gather and trade on information.

The goal of this paper is to investigate the degree of stock market efficiency of the Moscow stock Exchange using the Shannon entropy. We quantify the degree of market

inefficiency and the degree of price randomness. We aim to distinguish between price predictability due to stylized facts of financial time series [9] and due to market inefficiency. In particular, we consider volatility clustering and price staleness as data regularities needed to be filtered out. Based on the behavior of stock prices, we group them into clusters using several measures. Combining stocks into one cluster means a common price behavior that moves prices away from complete randomness.

A range of methods is used to measure a degree of market efficiency. In particular, Cajueiro and Tabak used the Hurst exponent and R/S statistics to rank efficiency of markets [10,11]. The Hurst exponent was measured on Bitcoin data to compare it with mature markets [12]. A generalized version of the use of the Hurst exponent, multifractal detrended fluctuation analysis, was applied to investigate the efficiency of stock and credit markets [13]. The algorithmic complexity of return time series was applied to measure the efficiency of financial markets [14] and to check the Efficient Market Hypothesis [15]. Finally, the Shannon entropy as a measure of randomness is used in a range of articles, see [16–18]). The general idea of these methods is to compare the characteristic of the time series with the value corresponding to a completely random process. In our work, we use Monte Carlo simulations to determine which deviations from a completely random process are statistically significant.

Before estimating the degree of market efficiency, we need to get rid of regularities that make prices more predictable, but do not imply any profitable strategies. The method for filtering regularities was introduced in [19]. However, such a filtering has not usually been applied in other research works. In fact, deviations of price behavior from perfect randomness may be the result of some known regularity pattern, such as volatility clustering or daily seasonality, but not a signal of market inefficiency. One of the innovations of this article is a new method for filtering data regularities allowing to estimate volatility and a degree of price staleness minute by minute.

We process data by filtering regularities of financial time series including volatility clustering and price staleness. Price staleness is defined as a lack of price adjustments yielding 0-returns. Traders may trade less because of high transaction costs and so the price does not update. See [20] for more details. The price staleness produces an extra amount of 0-returns called *excess 0-returns*. The other source of 0-returns in the time series is price rounding. Estimations of volatility and degree of price staleness are mutually connected: excess 0-returns appearing due to price staleness tend to underestimate volatility. At the same time, volatility estimation is needed to calculate the expected amount of 0-returns due to rounding.

One way to estimate volatility in the presence of excess 0-returns was presented in [21]. It uses expectation-maximization algorithm [22] to estimate returns in the places of all 0-returns and uses GARCH(1,1) model to estimate volatility [23]. The maximization of the likelihood function appearing at each step of the considered algorithm requires several parameters for numerical optimization. If the estimation of volatility is sensitive to these parameters, that are user-defined, then they may affect the entropy of returns standardized by volatility and the amount of 0-returns in the time series. In this article we suggest a modification of moving average volatility estimation that requires adjusting of the only parameter that can be defined using out-of-sample testing. The idea is to adopt a simple method for volatility estimation, so that price staleness is taken into consideration. Moreover, while estimating volatility, we filter out excess 0-returns.

The degree of market efficiency has been measured for many countries. Stock indices for 20 countries were considered in [24]. The efficiency of 11 emerging markets, the US and Japan markets was measured in [10]. US stock markets were considered in a recent paper [25]. The review of articles about Baltic countries was presented in [26]. A degree of uncertainty of Chinese [27], Tunisian [28], Mexican [29], and Portuguese [30] stock markets was also considered using entropy measures. However, the efficiency of the Russian stock market has not yet been analyzed. In this paper, we present an analysis of market efficiency

based on the estimation of Shannon entropy for a group of 18 stocks of Russian companies from five industries.

Our paper introduces four original contributions in the field. First, we construct the method of filtering out heteroskedasticity and price staleness. This filtering process helps to identify a true degree of market inefficiency. Second, we calculate the degree of market inefficiency for the previous decade using monthly intervals. We conclude that the degree of market inefficiency for the Moscow Stock Exchange was greater than 80%. Third, we determine which pair of stocks exhibits the largest amount of inefficiency, as measured by estimating Shannon's entropy on their high frequency price time series. We show that months where the predictability of stock prices attains its maximum cluster together. We find out what behavior of stocks repeats most often for inefficient time periods. Finally, we estimate the closeness of price movements using two measures of entropy. Based on these results, we cluster together groups of stocks for which the efficient market hypothesis is rejected, thus pointing out how market inefficiency display some dependence from the financial sector they belong.

The article is organized as follows. Section 2 describes the dataset and the method for filtering data regularities and calculating the Shannon entropy. Section 3 presents the results on simulated and real data. Section 4 concludes the paper.

2. Materials and methods

Our main goal is to measure a degree of efficiency of the Moscow Stock Exchange. The data taken for the study are reviewed in the next section. All data processing and computing can be divided into three stages. First, we filter data regularities from financial time series. Then, we calculate the degree of efficiency of the market using the Shannon entropy. Finally, we use the resulting time series to cluster stocks using Kullback–Leibler distance discussed in Section 2.4.

2.1. Dataset

We study the Moscow Stock Exchange. We consider close prices aggregated at one-minute time scale. In particular, we select only minutes of the main trading session from 10:00 to 18:40. The time interval covers ten years from 2012 to 2021. The time period is divided into monthly time intervals. We take 18 companies, 16 of them are from five sectors: oil industry, metallurgy, banks, telecommunications, electricity. All stocks are listed in the Table 1[1]. All data are provided by Finam Holdings[2].

2.2. Apparent inefficiencies

To estimate a degree of market efficiency, we first should eliminate the known patterns of predictability, such as a daily seasonality. Financial agents operating in the market tend to trade less in the middle of a day. It is reflected in prices, but again this pattern in trading volume should be filtered out to detect genuine patterns of inefficiency. Other known regularities are volatility clustering, price staleness, and microstructure noise. See Appendix A for the guide of filtering out apparent inefficiencies. The contribution of this article is devising a simple method for filtering volatility clustering and price staleness. One of the methods to estimate volatility is the exponentially weighted moving average (EWMA). It is described in the next section.

---

[1] There are 2520 trading days. Assuming that there are 520 minutes in each trading day, there are 1310400 trading minutes in total. We use the Brownlees and Gallo's algorithm of an outlier detection [31]. See details in Appendix A.1

[2] https://www.finam.ru/

Table 1. Stocks of Russian companies traded at Moscow Exchange.

Ticker	Company	Sector	Size	Outliers
GAZP	Gazprom	Oil	1307427	50
LKOH	Lukoil	Oil	1287582	192
ROSN	Rosneft	Oil	1270592	130
SNGS	Surgutneftegaz	Oil	1211809	11
TATN	Tatneft	Oil	1191390	174
SBER	Sberbank	Bank	1309402	37
VTBR	VTB Bank	Bank	1287330	0
CHMF	Severstal	Metal	1214735	157
NLMK	Novolipetsk Steel	Metal	1194324	58
GMKN	Nornikel	Metal	1272769	197
MTLR	Mechel	Metal	1084990	161
MAGN	Magnitogorsk Iron and Steel Works	Metal	1106771	13
MTSS	Mobile TeleSystems	Telecommunications	1153527	260
RTKM	Rostelecom	Telecommunications	1140798	134
HYDR	RusHydro	Electric utility	1252584	0
RSTI	Rosseti	Electricity	1094244	0
AFLT	Aeroflot	Airline	1083552	123
MGNT	Magnit	Food retailer	1184223	544

For each company, we specify the ticker of stock, its sector, the size of data, and the amount of outliers removed. The size is given in the amount of minutes with trading activity.

2.2.1. EWMA

We define price returns as $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$, where P_t is the last price available at time t and $\ln()$ is the natural logarithm. In order to estimate volatility σ_n, we apply exponentially weighted moving average [32] of the values $\mu_1^{-1}|r_i|$, $i < n$, where $\mu_1 = \sqrt{\frac{2}{\pi}}$.

$$\bar{\sigma}_n = Sig_1(\alpha, r_{n-1}, \bar{\sigma}_{n-1}) = \alpha\mu_1^{-1}|r_{n-1}| + (1-\alpha)\bar{\sigma}_{n-1} \tag{1}$$

This form of exponential moving average was used in [19]. Here, the fact that $E[|r_n|] = \mu_1\sigma_n$ is used assuming that returns are normally distributed, $r_n \sim N(0, \sigma_n)$. More weights are given for the more recent data. An alternative formula based on the expectation $E[r_n^2] = \sigma_n^2$ is

$$\bar{\sigma}_n^2 = Sig_2(\alpha, r_{n-1}, \bar{\sigma}_{n-1}) = \alpha r_{n-1}^2 + (1-\alpha)\bar{\sigma}_{n-1}^2. \tag{2}$$

A large value of return increases the value of volatility. The current value of volatility reflects all available values of returns and changes slowly if the value of α is small.

We follow the approach suggested by [33] (p. 97) to find optimal values of α in Equations 1 and 2. The value of α is selected so that it minimizes the error $Er_\sigma(\alpha) = \sum_i(\bar{\sigma}_i^2 - r_i^2)^2$. In order to minimize $Er_\sigma(\alpha)$ as a function of the only parameter $0 < \alpha < 1$, we apply Brent's algorithm [34][3]. We modify the exponential moving average method in Section 2.2.3 so that it removes a bias due to the effect of price staleness discussed in the next Section.

2.2.2. Estimation of price staleness

Let's define an efficient price, P^e, as a continuous process following a Geometric Brownian Motion.

---

[3] The method is available in Python by using the function scipy.optimize.minimize_scalar. Alternatively, we could use the golden-section search [35] that requires the boundary of search and the only parameter for the stopping criteria.

$$P_t^e = P_0^e + \int_0^t \sigma_s P_s^e dW_s$$

An observed price moves along a discrete grid. Possible price values are multiples of the tick size, d.

$$P_t = d \cdot \left[\frac{P_t^e}{d} \right]$$

If the efficient price changes insignificantly, the return of the rounded price will be equal to 0. Analogically, if the return of rounded price is 0, the return of efficient price has a value close to 0. We use Equation 3 to estimate the probability that a return of rounded price has zero value:

$$p_i = erf(R_{i-1}) + \frac{1}{R_{i-1}\sqrt{\pi}}(\exp(-R_{i-1}^2) - 1), \qquad (3)$$

where $R_i = \frac{d}{\bar{P}_i \bar{\sigma}_i \sqrt{2\Delta}}$ and $erf(x)$ is the Gaussian error function; d is a tick size [4], Δ is a time step[5], $\bar{P}$ is a rounded price, and $\bar{\sigma}$ is an estimation of volatility [36]. It is obtained by considering the probability that a price following a Geometric Brownian Motion moves less than one tick size, assuming that price increments are normally distributed.

There is another source of getting 0-returns, namely price staleness. Price staleness represents a regularity pattern of the dynamics, namely the fundamental (efficient) price of an asset is not updated because of a number of reasons, such as no transactions because of high cost, which makes the trading as unprofitable for agents. See [20] for more details. This results in a persistence pattern of ("*excess*") 0-returns. Such a pattern, for example, tends to reduce any estimation of the volatility. Therefore, we need to filter out 0-returns due to price staleness, while keeping 0-returns due to rounding, for a genuine estimation of the volatility.

We save 0-returns in the amount of the sum of past values of the probability in Eq. 3 [36]. We set other 0-returns as missing values. We adopt this method to estimate the degree of price staleness together with volatility in the next section.

2.2.3. Modification of EWMA

In this Section, we present a modification of the EWMA that takes into consideration the effect of price staleness. Our modification of the EWMA is based on the suggestion to estimate volatility σ_n as $\bar{\sigma}_{n-1}$ (i.e., by setting $\alpha = 0$), if the value of r_{n-1} is missing because of price staleness. That is, there is no new information from returns to update the value of volatility.

Initially, the expected amount of 0-returns due to rounding is $N_{save} = 0$. Thus, each appearance of 0-returns does not affect the value of volatility. A 0-return is defined as a value due to rounding and is saved in the sequence if the sum of all p_i (Eq. 3) moves to a new integer value. Other details and the algorithm of volatility estimation can be found in Appendix B.

We update the estimation of volatility and price staleness minute by minute. This method has the clear advantage of making possible the online inference by processing data in real time.

---

[4] We estimate the tick size using 2-steps procedure for each month. First, we find the amount of significant digits in price. Then, we determine the most frequent increment in ordered prices.

[5] The time step between the end and start of the main trading session is set as 1 minute. Also, we consider any time gap without trading more than 2 hours as the closure of the market. We set the time step equal to 1 minute for these gaps.

2.3. Calculating a degree of market inefficiency

2.3.1. The Shannon entropy

A degree of randomness of price returns is assessed by the Shannon entropy. The entropy of a source is an average measure of the randomness of its outputs [37].

Definition 1. *Let $X = \{X_1, X_2, \dots \}$ be a stationary random process with a finite alphabet A and a measure μ. An n-th order entropy of X is*

$$H_n(\mu) = - \sum_{x_1^n \in A^n} \mu(x_1^n) \log \mu(x_1^n)$$

with the convention $0 \log 0 = 0$. The process entropy (entropy rate) of X is

$$h(\mu) = \lim_{n \to \infty} \frac{H_n(\mu)}{n}.$$

2.3.2. Discretization

The Shannon entropy is computed over a finite alphabet. To measure the Shannon entropy, we need to keep the length of blocks of symbols, k, sufficiently large. The predictable behavior of returns can be seen on blocks of greater length and may not be noticeable on blocks of smaller length. For this reason, we consider 3-symbols and 4-symbols discretizations using empirical quantiles.

$$s_t^{(3)} = \begin{cases} 1, r_t \le \theta_1, \\ 0, \theta_1 < r_t \le \theta_2, \\ 2, \theta_2 < r_t, \end{cases} \qquad s_t^{(4)} = \begin{cases} 0, r_t \le Q_1, \\ 1, Q_1 < r_t \le Q_2, \\ 2, Q_2 < r_t \le Q_3, \\ 3, Q_3 < r_t, \end{cases}$$

where θ_1 and θ_2 are tertiles and Q_1, Q_2, Q_3 are quartiles. The tertiles divide data into three equal parts. The quartiles divide data into four equal parts. Q_2 is also the median of the empirical distribution of returns. For the later analysis, we will need a discretization describing the behavior of a pair of stocks.

$$s_t^{(p)} = \begin{cases} 0, r_t^{(1)} \le m_1 \text{ and } r_t^{(2)} \le m_2, \\ 1, r_t^{(1)} \le m_1 \text{ and } r_t^{(2)} > m_2, \\ 2, r_t^{(1)} > m_1 \text{ and } r_t^{(2)} \le m_2, \\ 3, r_t^{(1)} > m_1 \text{ and } r_t^{(2)} > m_2, \end{cases} \tag{4}$$

where $r_t^{(1)}$ and $r_t^{(2)}$ are two time series of price returns and m_1 and m_2 are their medians.

2.3.3. The estimation of entropy

Let $x_1^n \in A^n$ be the sequence of length n generated by an ergodic source μ from the finite alphabet A, where $x_i^{i+k-1} = x_i \dots x_{i+k-1}$. There are possible missing values in the sequence generated independently from x_1^n. We consider all blocks of length k that do not contain missing values. We take

$$k = max(K : K < \lfloor \log(n_b(K)) \rfloor), \tag{5}$$

where $n_b(k)$ is the number of blocks of length k.[6] The base of the logarithm is the size of the alphabet A (3 or 4).

---

[6] The restriction on a value of k allows to have enough blocks to estimate probabilities appearing in k-th order entropy [38].

For each $a_1^k \in A^k$ empirical frequencies are defined as 203

$$f(a_1^k | x_1^n) = \#\{i \in [1, n - k + 1] : x_i^{i+k-1} = a_1^k\}.$$

Empirical frequencies are the actual amount of each block from A^k in the data. By considering an empirical k-block distribution as

$$\hat{\mu}_k(a_1^k | x_1^n) = \frac{f(a_1^k | x_1^n)}{n_b}, \tag{6}$$

an empirical *k*-entropy is defined by

$$\hat{H}_k(x_1^n) = -\sum_{a_1^k} \hat{\mu}_k(a_1^k | x_1^n) \log\left(\hat{\mu}_k(a_1^k | x_1^n)\right) = \log(n_b) - \frac{1}{n_b} \sum_{i=1}^{M} f_i \log f_i.$$

The estimation of the process entropy is 204

$$\hat{h}_k = \frac{\hat{H}_k}{k}.$$

See [38] for the proof of the consistency of this estimator and [36] for the case of missing 205 values. Since the sequence is finite, the estimation of entropy is underestimated. To remove 206 this bias, we use the correction for the entropy estimation introduced in [39,40]. 207

$$\hat{H}_k^G = \log(n_b) - \frac{1}{n_b} \sum_{i=1}^{M} f_i \log\left(\exp G(f_i)\right), \tag{7}$$

where the sequence $G(i)$ is defined recursively as 208

$$\begin{aligned} G(1) &= -\gamma - \ln(2) \\ G(2) &= 2 - \gamma - \ln(2) \\ G(2n + 1) &= G(2n) \\ G(2n + 2) &= G(2n) + \frac{2}{2n + 1}, n \geq 1 \end{aligned}$$

with the Euler's constant $\gamma = 0.577215\ldots$. 209

2.3.4. Detection of inefficiency 210

We need to do three steps to determine if the time interval is efficient or not. First, 211 we filter out apparent inefficiencies (see Appendix A). Then, we estimate the entropy of 212 the filtered return time series using Eq. 7. Finally, we determine if the value of entropy is 213 significantly low relative to the case of perfect randomness. We detect inefficiency in the 214 time interval using Monte Carlo simulations. We regard a Brownian motion as absolutely 215 unpredictable. First, we define the length of sequences as $l = n_b(k) + k - 1$. Then, we 216 simulate 10^4 realizations of Brownian motions with Gaussian increments and the length l. 217 For each realization, we calculate entropy using 3- and 4-symbols discretizations. Then, we 218 find the first percentile of the obtained entropies for each discretization. These percentiles 219 are the bounds of 99% of the Confidence Interval (CI) for testing market efficiency. Finally, 220 we define an *efficiency rate* as the ratio of the entropy of the time interval and the bound of 221 CI. If the efficiency rate is less than 1 for at least one type of discretization, *we define the time* 222 *interval as inefficient*. We provide testing for inefficiency twice using different discretizations 223 because the unique testing may not be robust. See an example in Appendix C. 224

2.4. Kullback–Leibler distance

In addition to estimating the entropy of one time series, we can also consider the difference between two time series. The Kullback–Leibler divergence [41] is used to measure similarity between two distributions. For two discrete probability distributions P and Q.

$$KL(P|Q) = \sum_i p_i \log \frac{p_i}{q_i}$$

We use p_i and q_i as empirical probabilities obtained in Eq 6. Since the Kullback–Leibler divergence is asymmetric, we consider the distance between two time series proposed in [42].

$$D(P,Q) = \frac{KL(P|Q)}{H^G(P)} + \frac{KL(Q|P)}{H^G(Q)} \tag{8}$$

The greater the distance $D(P,Q)$, the more probability distributions P and Q differ.

3. Results

3.1. Simulations

The aim of this section is to assess the accuracy of the estimation of volatility and the degree of price staleness. We will choose the method that gives the least error of the estimation for further analysis on real data. We take the following model of an observed price $\tilde{P}_t$, $t = 1 \ldots 2N$.

$$P_t = \int_0^t \sigma_s P_s dW_s^1$$
$$\tilde{P}_i = P_i(1 - B_i) + \tilde{P}_{i-1} B_i \qquad B_i = \begin{cases} 1 \text{ with probability } q_i \\ 0 \text{ with probability } 1 - q_i \end{cases}$$
$$q_t = q_0 + \int_0^t \mu_s ds + \int_0^t \nu dW_s^2$$

where W^1 and W^2 are two independent Brownian motions with the length of $2N$, $N = 10^5$, price $P_0 = 100$, and $\nu = 10^{-4}$. $B = 1$ stands for the case when price is not updated due to price staleness (see [20,43]). Prices are rounded to two digits, thus the tick size is $d = 0.01$. We consider 4 choices for q_t and σ_t listed below.

$$q_t^1 = 0$$
$$q_t^2 = 0.1 + \int_0^t \nu dW_s^2$$
$$q_t^3 = 0.2 + \int_0^t \nu dW_s^2$$
$$q_t^4 = 0.2 + \int_0^t \mu_s^4 ds + \int_0^t \nu dW_s^2$$
$$\mu_t^4 = 0.8\pi/N \cos(8t\pi/N)$$
$$\sigma_t^1 = 5 \times 10^{-4}$$
$$\sigma_t^2 \sim ARCH(1.75 \times 10^{-7}, 0.2, 0.1)$$
$$\sigma_t^3 \sim GARCH(1.25 \times 10^{-8}, 0.1, 0.85)$$
$$\sigma_t^4 \sim GARCH(1.25 \times 10^{-8}, 0.15, 0.8)$$

For price staleness, we consider 4 cases: the absence of price staleness; two stochastic probabilities with different constant means; a periodic mean. For all four cases of volatility, the unconditional expected value of σ_t is 5×10^{-4}. The first choice of volatility is a constant. Then, we consider the ARCH model [44] with two lagged values, where 0.2 and 0.1 correspond to the first and the second lags, respectively. Volatility values directly depend only on the previous returns values. The dependency on the previous return should be

reflected in the value of smoothing parameter. The third and fourth choices are GARCH(1,1) models [23], where the last parameter (0.85 or 0.8) stands for the coefficients for lagged variances. We consider two sets of parameters for a GARCH model, giving less persistence to the fourth model.

We divide data into two equal parts with the size N. The first part is a training set for finding optimal values of α from Equations 1 and 2. The second part is a testing set for calculating errors represented in Tables 2 and 3 below. We compare two methods that use Sig_1 and Sig_2 for volatility estimation. For each method we find the optimal value of α. In addition, we set a fixed value of alpha, $\alpha = 0.05$, as a benchmark for the comparison. We also apply non-modified EMWA estimation from Section 2.2.1 with selected optimal value of α to show the contribution of 0-filtering to the accuracy of volatility estimation. We simulate 10^3 prices for each model.

Table 2 represents a mean absolute percentage error (MAPE) that is $\frac{1}{N} \sum_i \left| \frac{\bar{\sigma}_i - \sigma_i}{\sigma_i} \right|$ for six different approaches. These approaches differ in the choice of a function for volatility, the value of α, and the presence of missing values. Table 3 represents three values for each of two methods using Sig_1 and Sig_2 for the volatility estimation. The first value is the optimal value of α. The second is $Er_N = \left| \frac{N_{round} N_A}{N_0 N} - 1 \right|$ where N_A is the amount of remaining non-missing returns, N_{round} is the amount 0-returns that would appear due to rounding (before adding the effect of staleness in the simulated data); N_A is the amount of non-missing returns; and N_0 is the amount of 0-returns. Er_N represents the absolute error of the proportion of 0-returns that remain in the data and are defined as 0-returns due to rounding. The third value is the proportion of data set as missing values, that is $1 - \frac{N_A}{N}$.

It can be seen from Table 2 that the method that more often gives the lowest value of MAPE is with fixed $\alpha = 0.05$ and Sig_1 used for volatility estimation. Moreover, for almost all cases, 0-filtering makes the volatility estimate more accurate. The error of the amount of 0-returns due to rounding is smaller for the function Sig_1 than for the function Sig_2 for all 16 cases.

After the comparison of the two functions of volatility estimation, we decide to use Sig_1, that uses absolute values of returns, in the next sections. For the rest of the paper, we fix the value of α as 0.05 for the simplicity of further analysis.

3.2. Moscow Stock Exchange

We calculate $18 \cdot 120 = 2160$ efficiency rates for each type of discretization, where 18 is the amount of stocks and 120 is the amount of months in 10 years. We define a degree of inefficiency as the fraction of 2160 months which are defined as inefficient according to Section 2.3.4. The degree of inefficiency for the chosen group of stocks traded at Moscow Exchange is 0.823.[7] The degree of inefficiency for each stock and discretization is presented in Table 4. We notice that the 4-symbols discretization contributes to the larger amount of inefficient months than the 3-symbols discretization. That is, the 4-symbol discretization appears to have a more predictable structure than the 3-symbols discretization.

Figure 1 shows the minimum value of efficiency rate among all months for each stock.

---

7   In our previous work, [36] we found that the degree of inefficiency for the U.S. ETF market is about 0.11 for monthly time intervals and the 3-symbols discretization only. This difference in the degrees of inefficiency can be explained by the hypothesis that developed markets have a high level of efficiency. W. A. Risso reached this conclusion in the article [24].

Table 2. Results on volatility estimation.

model	MAPE, method v_1	MAPE, v_2	MAPE with $\alpha = 0.05$, v_1	MAPE with $\alpha = 0.05$, v_2	MAPE w/o 0-filtering, v_1	MAPE w/o 0-filtering, v_2
σ_1, q^1	0.0193 (0.0007,0.0507)	**0.017** (0.0014,0.0406)	0.0975 (0.0955,0.0995)	0.0897 (0.0878,0.0915)	0.0193 (0.0007,0.0507)	**0.017** (0.0014,0.0406)
σ_1, q^2	**0.0607** (0.0245,0.1017)	0.0629 (0.0293,0.1057)	0.095 (0.093,0.0972)	0.0914 (0.0893,0.0936)	0.0862 (0.0459,0.131)	0.0674 (0.0294,0.1154)
σ_1, q^3	**0.0737** (0.0333,0.1278)	0.0756 (0.033,0.1338)	0.0948 (0.0928,0.0971)	0.0915 (0.0894,0.094)	0.138 (0.0917,0.1863)	0.0888 (0.0368,0.1592)
σ_1, q^4	**0.0716** (0.0323,0.1213)	0.0739 (0.0354,0.1268)	0.0949 (0.0926,0.0973)	0.0913 (0.089,0.0937)	0.1404 (0.1022,0.1875)	0.0873 (0.0405,0.1516)
σ_2, q^1	0.1121 (0.1082,0.121)	0.1183 (0.1146,0.1244)	0.1459 (0.1438,0.1481)	0.1446 (0.1422,0.147)	**0.1118** (0.108,0.1207)	0.1179 (0.1144,0.1243)
σ_2, q^2	0.1359 (0.1163,0.1715)	0.1411 (0.1237,0.1765)	0.1462 (0.1439,0.1487)	0.1489 (0.1457,0.1526)	**0.1341** (0.1043,0.1819)	0.1407 (0.1193,0.1832)
σ_2, q^3	**0.146** (0.1198,0.1958)	0.1519 (0.1266,0.1981)	0.1473 (0.1449,0.1499)	0.1496 (0.1464,0.1534)	0.1649 (0.1196,0.2271)	0.1589 (0.123,0.222)
σ_2, q^4	**0.146** (0.1205,0.1912)	0.15 (0.1256,0.1986)	0.1472 (0.1447,0.1498)	0.1494 (0.1463,0.1532)	0.1696 (0.1274,0.2261)	0.1571 (0.1223,0.2239)
σ_3, q^1	0.1479 (0.1446,0.1513)	**0.1473** (0.1442,0.1505)	0.1495 (0.1467,0.1522)	**0.1473** (0.1442,0.1502)	0.1479 (0.1446,0.1513)	**0.1472** (0.1441,0.1503)
σ_3, q^2	0.1592 (0.1485,0.1891)	0.1613 (0.1508,0.1857)	**0.1529** (0.149,0.1574)	0.1546 (0.1497,0.1598)	0.1622 (0.144,0.2033)	0.1628 (0.1491,0.1978)
σ_3, q^3	0.1681 (0.1528,0.2048)	0.171 (0.1556,0.2178)	**0.1567** (0.1525,0.1616)	0.1584 (0.1536,0.1639)	0.1904 (0.154,0.2477)	0.1815 (0.1546,0.2464)
σ_3, q^4	0.1668 (0.1527,0.1997)	0.1701 (0.1555,0.2178)	**0.1568** (0.1525,0.1613)	0.1583 (0.1537,0.1633)	0.192 (0.1591,0.246)	0.181 (0.1556,0.2455)
σ_4, q^1	0.1897 (0.1856,0.1952)	**0.1873** (0.1838,0.1911)	0.1881 (0.1844,0.1918)	0.1924 (0.1879,0.1968)	0.1897 (0.1856,0.1952)	0.1873 (0.1837,0.1911)
σ_4, q^2	0.2035 (0.1906,0.2454)	0.2057 (0.1921,0.2474)	**0.1954** (0.1891,0.2022)	0.2037 (0.1961,0.2119)	0.2049 (0.1836,0.2617)	0.2079 (0.1902,0.2642)
σ_4, q^3	0.2146 (0.1965,0.2623)	0.2166 (0.1996,0.2757)	**0.2015** (0.1951,0.2077)	0.2101 (0.2026,0.2177)	0.2318 (0.1912,0.307)	0.2294 (0.1988,0.3082)
σ_4, q^4	0.214 (0.1967,0.2591)	0.2155 (0.1986,0.2689)	**0.2013** (0.1951,0.2088)	0.2097 (0.2023,0.2185)	0.2338 (0.1976,0.3064)	0.2286 (0.1988,0.306)

The first column indicated a model. Columns 2 and 3 represent results for two methods described in Section 2.2.3. Columns 4 and 5 are for the same methods but with the fixed value of α. Columns 6 and 7 shows the error of the standard EMWA approach with the optimal selected value of α. 95% CI is presented below each averaged statistic. v_1 stands for using Sig_1; v_2 stands for using Sig_2.
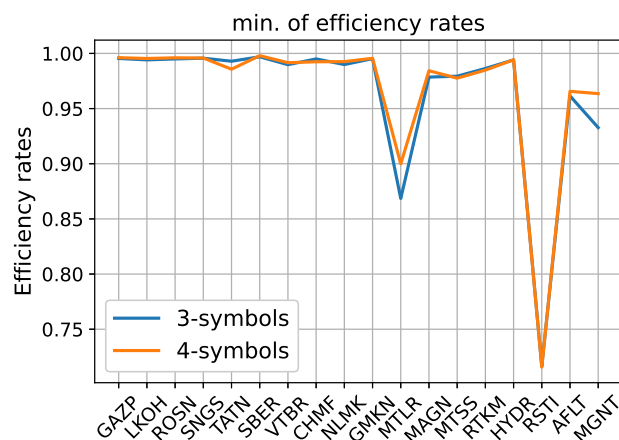
Figure 1. Minimum of efficiency rates for 18 stocks using 3- and 4-symbols discretizations.

There are two most notable deviations from 1 for stocks MLTR (Mechel, mining and metals company) and RSTI (Rosseti, power company). We investigate them in the next

Table 3. Results on filtering out 0-returns.

model	α for v_1	α for v_2	Er_N, v_1	Er_N, v_2	Fraction of data deleted, v_1	Fraction of data deleted, v_2
σ_1, q^1	0.0027 (0.0,0.0137)	0.0022 (0.0,0.0103)	0.0006 (0.0,0.0)	0.0015 (0.0,0.0259)	0.0001 (0.0,0.0)	0.0003 (0.0,0.0052)
σ_1, q^2	0.0228 (0.0033,0.0569)	0.0259 (0.0044,0.067)	0.0094 (0.0004,0.026)	0.011 (0.0005,0.0295)	0.2005 (0.0814,0.3244)	0.2008 (0.0818,0.3247)
σ_1, q^3	0.0335 (0.006,0.0902)	0.0379 (0.0058,0.1063)	0.0106 (0.0005,0.0288)	0.0121 (0.0005,0.0336)	0.3661 (0.2474,0.481)	0.3659 (0.246,0.4797)
σ_1, q^4	0.0314 (0.0056,0.0824)	0.036 (0.007,0.0966)	0.0104 (0.0004,0.0283)	0.0122 (0.0007,0.0355)	0.3628 (0.2521,0.4717)	0.3626 (0.2515,0.4713)
σ_2, q^1	0.0039 (0.0,0.0161)	0.0037 (0.0006,0.0146)	0.0149 (0.0,0.0438)	0.035 (0.0,0.0586)	0.0029 (0.0,0.0092)	0.0067 (0.0,0.0134)
σ_2, q^2	0.035 (0.0059,0.0903)	0.0367 (0.0054,0.1021)	0.0209 (0.0012,0.0448)	0.0319 (0.0039,0.0601)	0.2016 (0.0856,0.3249)	0.2032 (0.0878,0.3263)
σ_2, q^3	0.0489 (0.0079,0.1326)	0.0556 (0.0091,0.1476)	0.0217 (0.001,0.0473)	0.0275 (0.0022,0.0603)	0.3706 (0.25,0.4835)	0.371 (0.2518,0.4836)
σ_2, q^4	0.049 (0.0093,0.1268)	0.0525 (0.0082,0.1484)	0.0206 (0.0012,0.0443)	0.0274 (0.0013,0.0571)	0.3645 (0.2495,0.4695)	0.3651 (0.2491,0.4692)
σ_3, q^1	0.0424 (0.0349,0.0527)	0.048 (0.0392,0.0603)	0.0034 (0.0,0.0337)	0.0089 (0.0,0.0402)	0.0007 (0.0,0.0067)	0.0018 (0.0,0.0085)
σ_3, q^2	0.0672 (0.0267,0.1456)	0.0767 (0.0249,0.1592)	0.0155 (0.0006,0.0404)	0.02 (0.0009,0.0551)	0.1995 (0.0826,0.3248)	0.1997 (0.0826,0.3251)
σ_3, q^3	0.0825 (0.0264,0.1734)	0.0985 (0.0301,0.2371)	0.018 (0.0011,0.0477)	0.0243 (0.0008,0.0702)	0.3678 (0.2444,0.4746)	0.3671 (0.2421,0.4751)
σ_3, q^4	0.0788 (0.0266,0.163)	0.0969 (0.0325,0.2362)	0.0178 (0.001,0.0463)	0.0222 (0.0007,0.067)	0.3623 (0.2466,0.476)	0.3615 (0.2486,0.4739)
σ_4, q^1	0.0819 (0.0696,0.1037)	0.0904 (0.0757,0.119)	0.0013 (0.0,0.0248)	0.0047 (0.0,0.0329)	0.0003 (0.0,0.0052)	0.0011 (0.0,0.0075)
σ_4, q^2	0.1132 (0.0534,0.2359)	0.1339 (0.0576,0.2925)	0.0185 (0.0007,0.0564)	0.0265 (0.0008,0.087)	0.1993 (0.0765,0.3287)	0.1982 (0.077,0.3257)
σ_4, q^3	0.1338 (0.0557,0.2678)	0.1596 (0.0597,0.3734)	0.0214 (0.0009,0.0621)	0.0321 (0.001,0.1119)	0.3687 (0.2419,0.4823)	0.3669 (0.2378,0.4817)
σ_4, q^4	0.1317 (0.0571,0.263)	0.1556 (0.0613,0.3541)	0.0211 (0.0008,0.0667)	0.0315 (0.0011,0.108)	0.3641 (0.2599,0.4823)	0.3625 (0.2564,0.4822)

Values of α, errors of the number of 0-returns due to rounding, and fraction of data set as missing values. The first column indicated a model. 95% CI is presented below each averaged statistic. v_1 stands for using Sig_1; v_2 stands for using Sig_2.

section. For the other 16 stocks, the minimum value of efficiency rate is attained for the stock AFLT and is equal to 0.933 (0.964) for 3 (4) symbols.

3.2.1. Analysis of MLTR and RSTI

We plot the values of efficiency rates for monthly intervals for the MLTR and RSTI stocks. See Fig. 2 and Fig. 3.

Both types of discretization show coherent results. For MLTR, there are two notable decreases in the efficiency rates at the beginning of 2014 and in the middle of 2016. For both types of discretizations, eight months with the lowest efficiency rate (in the ascending order of time) are Jan-Feb and May-Oct of 2014. For each month we write down the most frequent block of symbols in Table 5. Note that block 1111 for the 4-symbols discretization appears as the most frequent for 6 months out of 8 for MLTR. The block denotes a slight decrease in the price for 4 minutes in a row. The meaning of the last two columns is discussed later.

For RSTI, there are two sharp decreases in 2014 and 2015. There are 11 months that have the lowest efficiency rates in common for both discretizations. These months are Apr-Sep of 2014 and Jun-Oct of 2015. Note that these inefficient months cluster together and are not distributed uniformly among the entire time period of 10 years. This is the signal of a market condition that affects the inefficiency of the stocks for more than one month.

Table 4. The degree of inefficiency for each stock.

| Ticker	Degree of inefficiency	For 3 symbols only	For 4 symbols only
GAZP	0.725	0.392	0.675
LKOH	0.65	0.342	0.542
ROSN	0.742	0.392	0.708
SNGS	0.725	0.4	0.625
TATN	0.617	0.392	0.525
SBER	0.725	0.433	0.658
VTBR	0.842	0.592	0.792
CHMF	0.858	0.55	0.692
NLMK	0.8	0.467	0.692
GMKN	0.733	0.475	0.608
MTLR	0.992	0.783	0.975
MAGN	0.833	0.65	0.758
MTSS	0.967	0.7	0.942
RTKM	0.942	0.683	0.908
HYDR	0.892	0.75	0.8
RSTI	0.917	0.742	0.875
AFLT	0.983	0.775	0.95
MGNT	0.842	0.667	0.742

Fraction of inefficient months using 3-symbols and 4-symbols discretization. Each value in the last two columns is calculated using 120 efficiency rates.

We construct a simple trading strategy on discretized returns to test the predictability of future returns. We consider blocks of length 4 obtained by the 4-symbols discretization. For each month, we divide blocks into two halves. The discretization is made using only the first half of a month. We consider the sequences of the first 3 symbols of each block. If the empirical probability of getting 0 or 1 after the sequence of 3 symbols in the first half is greater than 0.5 , this sequence is from group D (decreasing). If the empirical probability of getting 2 or 3 after the sequence of 3 symbols is greater than 0.5, this sequence is from group I (increasing). Then, for the second half of the month, we determine a success if symbols 0 or 1 follow a sequence from group D or if symbols 2 or 3 follow a sequence from group I. Then, we calculate the fraction of successes. Thus, it is the probability of making a profit: sell after group D or buy after group I. In the case of market efficiency, this probability is equal to 0.5. For example, we expect that after 111 the next symbol would be 1 according to

Table 5. The most frequent blocks appearing for the Stock MLTR and probabilities of success.

Months of 2014	The most frequent block, 3-s	The most frequent block, 4-s	prob. of success, filtered	prob. of success, original
Jan.	00000	1111	0.64	0.75
Feb.	00000	2222	0.64	0.74
May	00000	1111	0.61	0.73
June	22222	1111	0.60	0.73
July	11111	1111	0.62	0.74
Aug.	00000	1111	0.61	0.76
Sep.	00000	1111	0.63	0.74
Oct.	120120	0303	0.55	0.6

The first column represents months with the lowest efficiency rates. Columns 2 and 3 are the most frequent blocks in 3- and 4-symbols discretization. Columns 4 and 5 are the probability of the success of the simple trading strategy for filtered and original price returns.
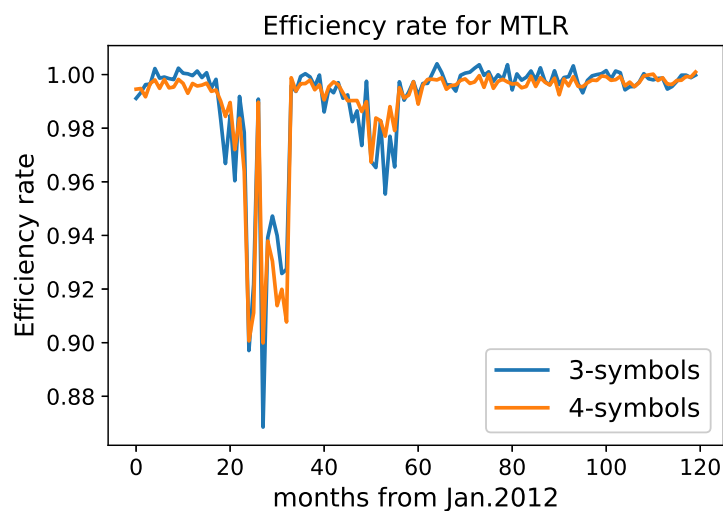
Figure 2. Efficiency rate for the MLTR stock using 3- and 4-symbols discretizations.
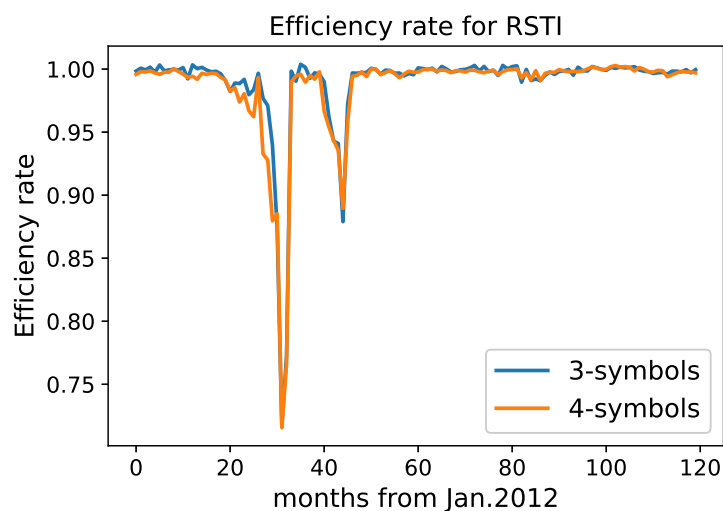
Figure 3. Efficiency rate for the RSTI stock using 3- and 4-symbols discretizations.

Table 5. That is, after this block, a trader can sell a stock. The fourth column of the Table 5 shows the results for filtered return time series. The fifth column stands for the original return time series.

For all cases the probability is greater than 0.5. Obviously, the probabilities for the original return time series are greater than those for the filtered return time series. The reason is that predictability for the original return time series follows from the sources of apparent inefficiencies.

The same analysis is done for the RSTI stock. Eleven months with the lowest efficiency rates are presented in Table 6. For the RSTI stock, the simple trading strategy gives the fraction of successes (of predicting increases and decreases of the price) greater than 0.5 for all 11 months. The frequent behavior of the price of RSTI during the chosen months is a slight increase in price for several minutes in a row denoted by symbol 2.

The simple trading strategy is an illustrative example of market inefficiency. In fact, such a strategy could result in no profit when used in practice because it does not take into account the costs of transaction and other trading frictions. Moreover, the filtering of daily seasonality pattern is made by using the whole period of analysis. That is, this method cannot be applied in real time. Finally, we consider blocks containing only observed returns, by neglecting the missing values from the analysis. Thus, the application of such a strategy

Table 6. The most frequent blocks appearing for the Stock RSTI and probabilities of success.

Months	The most frequent block, 3-s	The most frequent block, 4-s	prob. of success, filtered	prob. of success, original
Apr.2014	212121	0111	0.63	0.77
May.2014	00000	1111	0.61	0.73
June.2014	00000	1111	0.6	0.73
July.2014	00000	2222	0.62	0.74
Aug.2014	00000	2222	0.61	0.76
Sep.2014	000000	22222	0.63	0.74
June.2015	00000	2222	0.54	0.61
July.2015	00000	1111	0.55	0.6
Aug.2015	00000	2222	0.54	0.6
Sep.2015	00000	2222	0.55	0.61
Oct.2015	11111	0111	0.56	0.62

The first column represents months with the lowest efficiency rates. Columns 2 and 3 are the most frequent blocks in 3- and 4-symbols discretization. The length of a block is defined using Eq. 5. Columns 4 and 5 are the probability of the success of the simple trading strategy for filtered and original price returns.

in practice should be integrated with the case when a missing value follows a sequence of 3 symbols.

3.3. Stock market clustering

Most of the month-long time intervals are identified as inefficient. But is there some dependence between two stocks that are inefficient at the same time?

3.3.1. Kullback–Leibler distance

We measure the similarity of discretized filtered returns by using the Kullback–Leibler (KL) distance (Eq. 8). We use k, the length of blocks, as the maximum value suitable for both sequences according to Eq. 5. The 4-symbols discretization is used. The Kullback–Leibler divergence $DL(P|Q)$ is calculated using empirical frequencies. The entropy rates are calculated using Eq. 7. Using the Kullback–Leibler distance for all pairs of stocks, we cluster them in three groups using hierarchical clustering with UPGMA algorithm [45][8]. The result is in Fig. 4. Combining companies into one cluster means that their stocks have a common behavior that is not related to the value of volatility, the degree of price staleness and the structure of microstructure noise.

It can be seen that banks and oil companies are clustered together (right). There is a group of four stocks RTKM, HYDR, AFLT, MGNT, that have nothing in common at first glance. The remaining group (left) mainly consists of metallurgy companies. However, there is no visible distinction between the stocks of banks and oil companies. According to the clustering tree, two telecommunications companies differ significantly, as well as electricity companies.

Finally, two stocks with the lowest efficiency rates, RSTI and MLTR, are the furthest (in the sense of KL distance) from any other stock. That is, there are no stocks that behave similarly to these two stocks.

3.3.2. Entropy of co-movement

Now, we consider another measure of difference between two stocks, the entropy of co-movement. We calculate the Shannon entropy of the discretization describing the

---

8 This algorithm is implemented using the python function cluster.hierarchy.dendrogram with the argument distance=average.
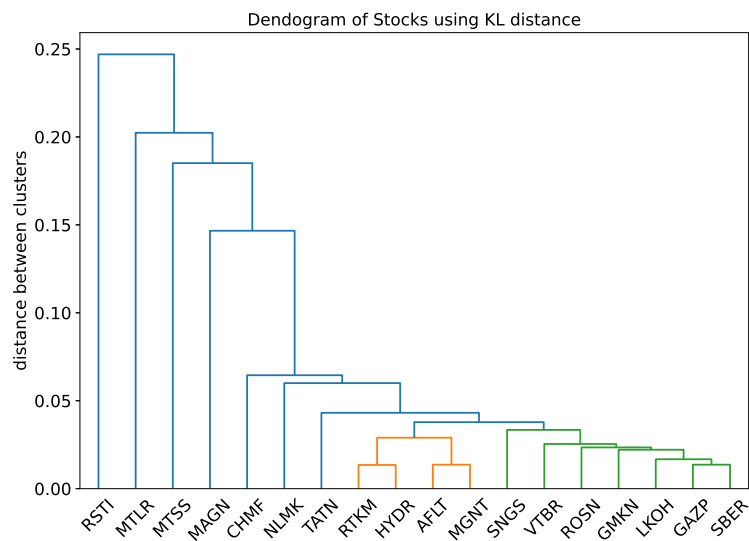
Figure 4. Hierarchical clustering tree using KL distance. The threshold for clustering into groups is 0.035.

movement of a pair of prices presented in Eq. 4. We consider only minutes that are in common for both stocks. For these minutes we consider values of residuals obtained after ARMA fitting. The result is in Fig. 5.
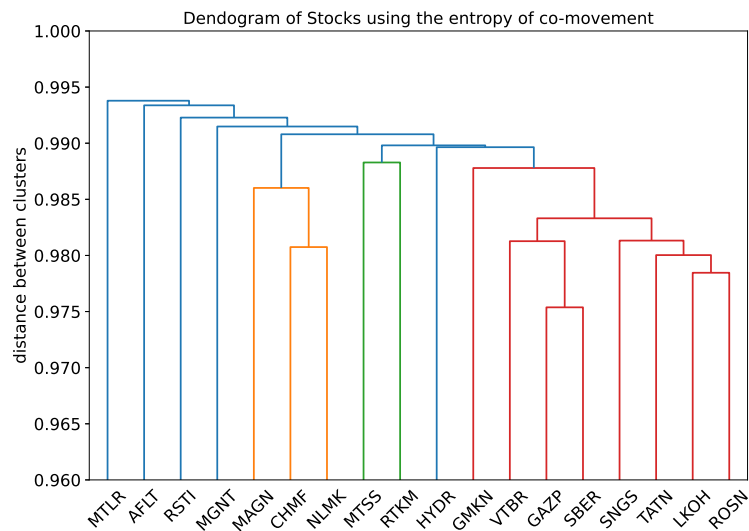
Figure 5. Hierarchical clustering tree using the entropy of co-movement. The threshold for clustering into groups is 0.989.

Two companies related to telecommunications are a separate cluster. Three metallurgy companies MAGN, CHMF, NLMK also cluster together. Stocks relating to oil and bank companies form the other cluster. The same cluster, with the exception of the TATN (oil industry), was also formed in the previous section. The "closeness" of stocks GAZP and SBER is detected either in this and in the previous section. The three stocks on the left that join other stock clusters last are the stocks with the lowest efficiency rates.

Some clusters may form on the basis that companies belong to the same industry. The division of companies into industries is noticeable from the dendrogram in Figure 5. However, this criterion does not explain all clusters. For instance, GMKN from metallurgy is in the cluster of oil companies and banks.

4. Discussion and conclusions

We have investigated the predictability of the Moscow Stock Exchange. We are interested in a measure of market inefficiency that is not related to known sources of regularity in financial time series. Usually, these sources are not filtered out and, accordingly, their impact is taken into account in the degree of price predictability (see, e.g. [16–18]).

We have focused on two sources of regularity, namely volatility clustering and price staleness [20]. Filtering of volatility clustering was made in [19] by estimating volatility using the exponentially weighted moving average. We have developed a modification of the volatility estimation by taking into consideration the effect of price staleness. Price staleness produces excess 0-returns that affect the estimation of volatility. Another approach of estimating volatility in the case of presence 0-returns was proposed in [21] where all 0-returns are reevaluated during an expectation-maximization algorithm. In our approach, we separate 0-returns that may have resulted from rounding and from price staleness. Thus, we also filter out apparent inefficiency due to price staleness. Our approach combining the estimates of volatility and the degree of staleness can be used for a real-time causal analysis, since only past observations are used.

One of the clear advantage of the proposed approach relies in its simplicity: there is only one smoothing parameter in the method which can be optimized using historical data. We fix the value of smoothing parameter equaled 0.05. In the literature, the smoothing parameter α is usually taken close to 0. Using the principle of the best one-step forecasting, the smoothing parameter is set to 0.06 for the daily data and to 0.03 for the monthly data [33]. The value of the parameter α is set to be equal 0.12 for in-sample testing and 0.22 for out-of-sample testing in [46]. Hunter [32] suggests to use $\alpha = 0.2 \pm 0.1$.

We used the Shannon entropy as a measure of randomness to infer the degree of (in)efficiency of the Moscow Stock Exchange. We used two types of the discretization of return time series to test efficiency more reliably for each month. The 4-symbols discretization helps to find more price movements that lead to market inefficiency than the 3-symbols discretization. There are 80% of months over the period from 2012 to 2021 that are defined as inefficient. According to Risso [24], a higher level of efficiency corresponds to more developed markets. Deviation from efficiency is a frequent phenomenon in various markets. For example, the authors of work [14] conclude that the Colombo Stock Exchange is only 10.5% efficient while the Pakistan Stock Exchange is 23.7% efficient. Cajueiro and Tabak [10] have shown that Asian markets are less efficient than Latin American markets. The article [28] estimates the efficiency of the Tunisian stock market as 97%. There are periods of inefficiency for some stocks traded at Tel-Aviv stock exchange as stated in [15]. Short periods of inefficiency were also detected for US stock markets in [25].

By investigating the discretized values of filtered price returns, we came to the following conclusions:

- Even after filtering out all known sources of regularity, most of the months contain signals of market inefficiency.
- The most inefficient months are grouped together for two stocks exhibiting the lowest efficiency rates.
- For such months, discretized price returns before and after filtering out apparent inefficiencies are predictable.
- We introduced the entropy of co-movement. Stock prices display common patterns that have an interpretation in terms of the sector the stocks belong to.
- The stocks of banks and oil companies cluster together in terms of co-inefficiency for the case of the Moscow stock exchange.

One possible improvement to stock clustering is to modify the entropy of co-movement such that it is possible to define a proper distance function. This is left for future research. The proposed method for measuring market efficiency using the Shannon entropy can be applied in other markets of different countries. In this work, we use monthly time intervals for entropy calculation. Our future work will be related to the optimization of the length of

return time series. One of the problems is to find a significant decrease in entropy without using Monte Carlo simulations. We also plan to switch to a higher frequency (less than one minute) to analyze the predictability of financial time series.

Author Contributions: Conceptualization, A.S., S.M. and P.M.; Data curation, A.S. and S.M.; Methodology, A.S.; Software, A.S., Formal analysis, A.S., Supervision, S.M. and P.M., Writing - original draft, A.S., Writing - review & editing, P.M. and S.M.

Funding: This research received no external funding.

Data Availability Statement: The Python codes (v. 3.9.5) used in this study are available in supplementary material. The dataset analyzed in this study can be found here: https://www.finam.ru/profile/moex-akcii/gazprom/export/

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Data cleaning and whitening

Appendix A.1. Outliers

We use the method of an outlier detection introduced in [31]. The algorithm finds price values that are too far from the mean in relation to the standard deviation. The algorithm deletes a price P_i if

$$|P_i - \bar{P}_i(k)| \geq cs_i(k) + \gamma,$$

where $\bar{P}_i(k)$ and $s_i(k)$ are respectively a δ-trimmed sample mean and the standard deviation of the k price records closest to time i. The $\delta\%$ of the lowest and the $\delta\%$ of the highest observations are discarded when the mean and the standard deviation are calculated from the sample. The parameters are $k = 20, \delta = 5, c = 5, \gamma = 0.05$.

Appendix A.2. Stock Splits

We check the condition $|r| > 0.2$ in the return series to detect unadjusted splits[9]. There are no unadjusted splits found.

Appendix A.3. Intraday Volatility Pattern

The volatility of intraday returns has periodic behavior. The volatility is higher near the opening and the closing of the market. It shows an U-shaped profile every day. The intraday volatility pattern from the return series is filtered by using the following model. We define deseasonalized returns as

$$\tilde{R}_{d,t} = \frac{R_{d,t}}{\xi_t},$$

where

$$\xi_t = \frac{1}{N_{days}} \sum_{d'} \frac{|R_{d',t}|}{s_{d'}},$$

$R_{d,t}$ is the raw return of day d and intraday time t, s_d is the standard deviation of absolute returns of day d, N_{days} is the number of days in the sample.

Appendix A.4. Heteroskedasticity

Different days have different levels of the deviation of the deseasonalized returns $\tilde{R}$. In order to remove this heteroskedasticity, we estimate the volatility $\bar{\sigma}_t$ in Appendix B. We define the standardized returns by

$$r_t = \frac{\tilde{R}_t}{\bar{\sigma}_t}.$$

---

9   A split is a change in the number of company's shares and in the price of the single share, such that a market capitalization does not change.

Appendix A.5. Price staleness

If a transaction cost is high, the price is updated less frequently, even if trading volume is not zero. This effect is called price staleness and is discussed in Section 2.2.2. We identify 0-returns appearing due to rounding (and not due to price staleness) using the Equation 3. Other 0-returns are set as missing values as shown in Appendix B.

Appendix A.6. Microstructure noise

The last step in filtering apparent inefficiencies is filtering out microstructure noise. The microstructure effects are caused by transaction costs and price rounding. We consider the residuals of an ARMA(P,Q) model of the standardized returns after filtering out 0-returns. We apply the method introduced in [47] to find the residuals of an ARMA(P,Q) model by using the Kalman filter. We select the values of P and Q that minimize the value of BIC [48], so that $P + Q < 6$. The values of P and Q are chosen for each calendar year and are used for the next year. For the year 2012 we select $P = 0$ and $Q = 1$ corresponding to an MA(1) model.

Appendix B. Algorithm

The aim of the algorithm is to estimate volatility and filter out excess 0-returns due to price staleness. Some 0-returns appear due to price rounding. These 0-returns will be saved in the data. First, we set the number of 0-returns "to save" $N_{save} = 0$ and the first value of a cumulative function $Z_1 = 0$. The cumulative function is updated $Z_t = Z_{t-1} + p_t$, if r_{t-1} is not defined as missing due to staleness. Each time when $\lfloor Z(t) \rfloor - \lfloor Z(t-1) \rfloor = 1$, N_{save} is increased by 1.

We notice that the first non-zero return after a row of 0-returns due to staleness is the sum of all missing returns generated by a hidden efficient price. This return is also set as missing. However, the value of return used for estimating volatility is calculated as its expected value: $\hat{r}_{n-1} = \frac{r_{n-1}}{\sqrt{N_0+1}}$, where N_0 is the amount of missing values strictly before the non-zero return r_{n-1}. The same is also referred to initially missing values, e.g., due to no-trading or errors in collecting the data.

Another assumption is that a 0-return appears due to staleness if the previous return had the 0-value and was defined to appear due to staleness. We include this rule, since we assume that it is more likely that two consecutive 0-returns appear due to high transaction costs than due to rounding (that is, simply speaking, two outcomes of generating Gaussian random variables are less than a tick size).

Generally, for the estimation of volatility at time t we should consider three cases: P_{t-1} was missing (or minute $t-1$ is non-trading), $r_{t-1} = 0$, $r_{t-1} \neq 0$. Thus, the algorithm is the following. We give the algorithm for the case of Sig_1 which is used in the application for the real data. We remove all 0-returns that start the sequence.

Appendix B.1. Pseudocode

Step 0: $\bar{\sigma}_1 = |r_1| / \mu_1$; $Z_1 = 0$, $N_{save} = 0$; $N_0 = 0$.
For t from 2 to N, where N is the length of time series:
Step 1:

- If r_{t-1} is missing: $\bar{\sigma}_t = \bar{\sigma}_{t-1}$; Increase N_0 by the amount of consecutive missing prices
- Else if $r_{t-1} = 0$:

  - If $N_{save} > 0$ and $N_0 = 0$: $N_{save} = N_{save} - 1$, $\bar{\sigma}_t = Sig_1(\alpha, 0, \bar{\sigma}_{t-1})$
  - Else: $\bar{\sigma}_t = \bar{\sigma}_{t-1}$, $N_0 = N_0 + 1$, $r_{t-1} = $ missing

- Else: $\bar{\sigma}_t = Sig_1(\alpha, \frac{r_{n-1}}{\sqrt{N_0+1}}, \bar{\sigma}_{t-1})$, $N_0 = 0$

Step 2:

- Calculate p_t (Eq. 3)
- If r_{t-1} is not missing, $Z_t = Z_{t-1} + p_i$
- If $\lfloor Z(t) \rfloor - \lfloor Z(t-1) \rfloor = 1$, $N_{save} = N_{save} + 1$

Finally, we check if the effect of staleness really exists in the price time series:

$$\hat{p} = \frac{\sum_i p_i}{N}$$
$$\hat{q} = 1 - \hat{p}$$
$$Var = \hat{p}\hat{q}N$$

If $N_{real} \leq \sum_i p_i + 1.96\sqrt{Var}$, we leave time series without putting any missing values, where N_{real} is the initial amount of 0-returns. The value of α can be selected using a training set. The optimal value of α minimizes the mean of $(\bar{\sigma}_t^2 - r_t^2)^2$.

Appendix C. A predictable time series with entropy at maximum

The goal of this section is to construct a price model where entropy is high because of discretization. This model shows that a high entropy value may be caused by discretization, but not because of the randomness of a return time series.

There are equal probabilities of having symbols 0, 1 and 2. 1 corresponds to log-returns, r, equal to -0.4, 2 corresponds to log-returns equal to 0.4. The structure of symbol 0 is more complicated. It covers three other symbols 3, 4, 5. They correspond to log-returns $-0.3, 0.1, 0.2$, respectively. One of the symbols 3, 4 or 5 appears with probabilities depending on the previous value of these symbols. The probabilities are presented in the Table A1. Having a symbol presented in a column, there are probabilities of getting a symbol presented in a row.

Table A1. Transition probabilities

first symbol	·3	·4	·5
3·	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
4·	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$
5·	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Rows stand for the first symbol of a block, columns stand for the second symbol.

The model implies an average zero return. However, a trading strategy that increases a profit exists. After 3 a trader should buy, and after 4 and 5 the trader should sell. However, the entropy of 3-symbols series is at maximum, that should imply absence of profitable strategies.

Considering the same example with 4-symbols discretization we get that $Q_1 = -0.4$, $Q_2 = 0.1$, and $Q_3 = 0.4$. Therefore, we have the following discretization of returns:

$$s = \begin{cases} 0, r = -0.4, \\ 1, r = -0.3 \text{ or } r = 0.1, \\ 2, r = 0.2, \\ 3, r = 0.4. \end{cases}$$

Thus, we can distinguish returns $r = 0.2$ from others the using 4-symbols discretization. Table A1 gives the following probabilities for the blocks of two symbols from the 4-symbols

discretization: $p(11) = \frac{7}{162}$, $p(12) = p(21) = \frac{5}{162}$, $p(22) = \frac{1}{162}$. Noting that $p(0) = p(3) = \frac{1}{3}$, $p(1) = \frac{2}{9}$, and $p(2) = \frac{1}{9}$, we calculate that

$$H_1 = -\frac{2}{3}\log\left(\frac{1}{3}\right) - \frac{2}{9}\log\left(\frac{2}{9}\right) - \frac{1}{9}\log\left(\frac{1}{9}\right) \approx 0.946 < 1$$

and

$$H_2 = -\frac{1}{2}\left(\frac{7}{162}\log\left(\frac{7}{162}\right) + \frac{5}{81}\log\left(\frac{5}{162}\right) + \frac{1}{162}\log\left(\frac{1}{162}\right) + \right.$$
$$\left. + \frac{4}{9}\log\left(\frac{1}{9}\right) + \frac{8}{27}\log\left(\frac{2}{27}\right) + \frac{4}{27}\log\left(\frac{1}{27}\right)\right) \approx 0.944 < H_1$$

References

1. Samuelson, P.A. Proof that properly anticipated prices fluctuate randomly. *Ind. manage. rev.* **1965**, *6*, 41–49.
2. Fama, E.F. Efficient Capital Markets: A Review of Theory and Empirical Work. *J Finance* **1970**, *25*, 383–417. https://doi.org/10.2307/2325486.
3. Fama, E.F. Efficient Capital Markets: II. *J Finance* **1991**, *46*, 1575. https://doi.org/10.2307/2328565.
4. Kim, J.H.; Shamsuddin, A. Are Asian stock markets efficient? Evidence from new multiple variance ratio tests. *J. Empir. Finance* **2008**, *15*, 518–532. https://doi.org/10.1016/j.jempfin.2007.07.001.
5. Linton, O.; Smetanina, E. Testing the martingale hypothesis for gross returns. *J Empir Finance* **2016**, *38*, 664–689. https://doi.org/10.1016/j.jempfin.2016.02.010.
6. Mandes, A. Algorithmic and High-Frequency Trading Strategies: A Literature Review. MAGKS Papers on Economics 201625, Philipps-Universität Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung), 2016.
7. Huang, B.; Huan, Y.; Xu, L.D.; Zheng, L.; Zou, Z. Automated trading systems statistical and machine learning methods and hardware implementation: a survey. *Enterp Inf Syst* **2019**, *13*, 132–144. https://doi.org/10.1080/17517575.2018.1493145.
8. Grossman, S.J.; Stiglitz, J.E. On the Impossibility of Informationally Efficient Markets. *Am Econ Rev* **1980**, *70*, 393–408.
9. Cont, R. Empirical properties of asset returns: stylized facts and statistical issues. *Quant Finance* **2001**, *1*, 223–236. https://doi.org/10.1080/713665670.
10. Cajueiro, D.O.; Tabak, B.M. Ranking efficiency for emerging markets. *Chaos Solitons Fractals* **2004**, *22*, 349–352. https://doi.org/10.1016/j.chaos.2004.02.005.
11. Cajueiro, D.; Tabak, B. Ranking efficiency for emerging markets II. *Chaos Solitons Fractals* **2005**, *23*, 671–675. https://doi.org/10.1016/j.chaos.2004.05.009.
12. Drożdż, S.; Gębarowski, R.; Minati, L.; Oświęcimka, P.; Watorek, M. Bitcoin market route to maturity? Evidence from return fluctuations, temporal correlations and multiscaling effects. *Chaos* **2018**, *28*, 071101. https://doi.org/10.1063/1.5036517.
13. Shahzad, S.J.H.; Nor, S.M.; Mensi, W.; Kumar, R.R. Examining the efficiency and interdependence of US credit and stock markets through MF-DFA and MF-DXA approaches. *Physica A* **2017**, *471*, 351–363. https://doi.org/10.1016/j.physa.2016.12.037.
14. Giglio, R.; Matsushita, R.; Figueiredo, A.; Gleria, I.; Silva, S.D. Algorithmic complexity theory and the relative efficiency of financial markets. *EPL* **2008**, *84*, 48005. https://doi.org/10.1209/0295-5075/84/48005.
15. Shmilovici, A.; Alon-Brimer, Y.; Hauser, S. Using a Stochastic Complexity Measure to Check the Efficient Market Hypothesis. *Comput Econ* **2003**, *22*, 273–284. https://doi.org/10.1023/A:1026198216929.
16. Molgedey, L.; Ebeling, W. Local order, entropy and predictability of financial time series. *Eur Phys J B* **2000**, *15*, 733–737. https://doi.org/10.1007/s100510051178.
17. Risso, W.A. The informational efficiency and the financial crashes. *J Int Bus Stud* **2008**, *22*, 396–408. https://doi.org/10.1016/j.ribaf.2008.02.005.
18. Mensi, W.; Aloui, C.; Hamdi, M.; Nguyen, D.K. Crude oil market efficiency: An empirical investigation via the Shannon entropy. *écon. intern.* **2012**, *129*, 119–137. https://doi.org/10.3917/ecoi.129.0119.
19. Calcagnile, L.M.; Corsi, F.; Marmi, S. Entropy and Efficiency of the ETF Market. *Comput. Econ.* **2020**, *55*, 143–184. https://doi.org/10.1007/s10614-019-09885-z.
20. Bandi, F.M.; Kolokolov, A.; Pirino, D.; Renò, R. Zeros. *Manage Sci* **2020**, *66*, 3466–3479. https://doi.org/10.1287/mnsc.2019.3527.
21. Sucarrat, G.; Escribano, A. Estimation of log-GARCH models in the presence of zero returns. *Eur. J. Finance* **2018**, *24*, 809–827. https://doi.org/10.1080/1351847X.2017.1336452.
22. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J R Stat Soc Series B Stat Methodol* **1977**, *39*, 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.
23. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327. https://doi.org/10.1016/0304-4076(86)90063-1.

24. Risso, W.A. The informational efficiency: the emerging markets versus the developed markets. *Appl. Econ. Lett.* **2009**, *16*, 485–487. https://doi.org/10.1080/17446540802216219.

25. Alvarez-Ramirez, J.; Rodriguez, E. A singular value decomposition entropy approach for testing stock market efficiency. *Physica A* **2021**, *583*, 126337. https://doi.org/10.1016/j.physa.2021.126337.

26. Degutis, A.; Novickytė, L. The efficient market hypothesis: a critical review of literature and methodology. *Ekonomika* **2014**, *93*, 7–23. https://doi.org/10.15388/Ekon.2014.2.3549.

27. Ahn, K.; Lee, D.; Sohn, S.; Yang, B. Stock market uncertainty and economic fundamentals: an entropy-based approach. *Quant Finance* **2019**, *19*, 1151–1163. https://doi.org/10.1080/14697688.2019.1579922.

28. Mahmoud, I.; Sebai, S.; Naoui, K.; Jemmali, H. Market Informational Efficiency of Tunisian Stock Market: The Contribution of Shannon Entropy. *J. Econ. Finance Adm. Sci.* **2014**, pp. 6–17.

29. Coronel-Brizio, H.; Hernández-Montoya, A.; Huerta-Quintanilla, R.; Rodríguez-Achach, M. Evidence of increment of efficiency of the Mexican Stock Market through the analysis of its variations. *Physica A* **2007**, *380*, 391–398. https://doi.org/10.1016/j.physa.2007.02.109.

30. Dionisio, A.; Menezes, R.; Mendes, D.A. An econophysics approach to analyse uncertainty in financial markets: an application to the Portuguese stock market. *Eur Phys J B* **2006**, *50*, 161–164. https://doi.org/10.1140/epjb/e2006-00113-2.

31. Brownlees, C.; Gallo, G. Financial econometric analysis at ultra-high frequency: Data handling concerns. *Comput Stat Data Anal* **2006**, *51*, 2232–2245. https://doi.org/10.1016/j.csda.2006.09.030.

32. Hunter, J.S. The Exponentially Weighted Moving Average. *J. Qual. Technol.* **1986**, *18*, 203–210. https://doi.org/10.1080/00224065.1986.11979014.

33. Morgan, J.; Longerstaey, J.; Limited, R.; Ltd, R.; Spencer, M.; of New York, M.G.T.C. *RiskMetrics: Technical Document*; J. P. Morgan, 1996. https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

34. Brent, R.P. An Algorithm with Guaranteed Convergence for Finding a Zero of a Function. *Comput J* **1971**, *14*, 422–425.

35. Kiefer, J. Sequential Minimax Search for a Maximum. *Proc Am Math Soc* **1953**, *4*, 502. https://doi.org/10.2307/2032161.

36. Shternshis, A.; Mazzarisi, P.; Marmi, S. Measuring market efficiency: The Shannon entropy of high-frequency financial time series. *Chaos, Solitons Fractals* **2022**, *162*, 112403. https://doi.org/10.1016/j.chaos.2022.112403.

37. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

38. Marton, K.; Shields, P.C. Entropy and the Consistent Estimation of Joint Distributions. *Ann. Probab.* **1994**, *22*, 960–977. https://doi.org/10.1214/aop/1176988736.

39. Grassberger, P. Entropy Estimates from Insufficient Samplings. *arXiv* **2003**. https://doi.org/10.48550/arXiv.physics/0307138.

40. Grassberger, P. On Generalized Schürmann Entropy Estimators. *Entropy* **2022**, *24*, 680. https://doi.org/10.3390/e24050680.

41. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat* **1951**, *22*, 79–86. https://doi.org/10.1214/aoms/1177729694.

42. Benedetto, D.; Caglioti, E.; Loreto, V. Language Trees and Zipping. *Phys. Rev. Lett.* **2002**, *88*, 048702. https://doi.org/10.1103/PhysRevLett.88.048702.

43. Kolokolov, A.; Livieri, G.; Pirino, D. Statistical inferences for price staleness. *J. Econom.* **2020**, *218*, 32–81. https://doi.org/10.1016/j.jeconom.2020.01.021.

44. Engle, R.F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **1982**, *50*, 987–1007. https://doi.org/10.2307/1912773.

45. Sokal, R.R.; Michener, C.D. A statistical method for evaluating systematic relationships. *Univ. Kansas, Sci. Bull.* **1958**, *38*, 1409–1438.

46. Bollen, B. What should the value of lambda be in the exponentially weighted moving average volatility model? *Appl Econ* **2015**, *47*, 853–860. https://doi.org/10.1080/00036846.2014.982853.

47. Jones, R.H. Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations. *Technometrics* **1980**, *22*, 389–395. https://doi.org/10.1080/00401706.1980.10486171.

48. Schwarz, G. Estimating the Dimension of a Model. *Ann Stat* **1978**, *6*, 461–464. https://doi.org/10.1214/aos/1176344136.