



# Nonlinear intraday trading invariance in the Russian stock market

Tamara Teplova<sup>1</sup> · Sergei Gurov<sup>1</sup>

Accepted: 18 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Using high-frequency transaction-level data for liquid Russian stocks, we empirically reveal a joint nonlinear relationship between the average trade size, log-return variance per transaction, trading volume, and the asset price level described by the Intraday Trading Invariance hypothesis. The relationship is also confirmed during stock market crashes. We show that the invariance principle explains a significant fraction of the endogenous variation between market activity variables at the intraday and daily levels. Moreover, our tests strongly reject the mixture of distributions hypotheses that assume linear relationships between log-return variance and transaction intensity variables such as trading volume or the number of transactions. We demonstrate that the increase in the ruble risk transferred by one bet per unit of business time was accompanied by the rise in the average spread cost. Different aggregation schemes are used to mitigate the impact of errors-in-variables effects. Following the predictions of the Information Flow Invariance hypothesis, we also study the relationship between trading activity and the information process approximated by either the flows of news articles or Google relative search volumes of Russian stocks over the 2018–2021 period. The evidence suggests that a sharp increase in the number of retail investors who entered the Moscow Exchange in 2020 entailed a higher synchronization between trading activity and search queries in Google since February 2020, in contrast to the arrival rates of news articles. The changes are driven by the increasing influence of the trading behavior of individual investors using Google Search rather than professional news services as the main source of information.

**Keywords** Liquidity · Market microstructure · Trade size · Invariance · Non-linearity · Transaction data

**JEL Classification** G01 · G10 · G12 · G15

---

✉ Sergei Gurov  
sgurov@hse.ru

Tamara Teplova  
tteplova@hse.ru

<sup>1</sup> National Research University Higher School of Economics, Moscow, Russian Federation

# 1 Introduction

In this paper, we test the implications of the Intraday Trading Invariance (ITI) hypothesis (Andersen et al., 2020)—the extension of the Market Microstructure Invariance (MMI) hypothesis (Kyle & Obizhaeva, 2016) about the variations in microscopic and macroscopic market characteristics over long intervals. Using tick-by-tick transaction-level data for the most liquid stocks of Russian issuers included in the MOEX Russia Index from January 2014 to July 2018, we test the quantitative predictions regarding the average trade size, log-return variance per transaction, trading volume, and the asset price level over five-minute intervals. Since we use the log-linear regression specifications, we include stocks with the highest level of liquidity in the sample to minimize the number of intervals without any trading activity. In addition, we use the quantitative predictions of the Information Flow Invariance (IFI) hypothesis (Kyle & Obizhaeva, 2017b) to investigate possible differences in relations between trading activity in the Russian stock market and different flows of information used in investors' decision-making processes before and during the Covid-19 crisis.

The main contribution of this study lies in estimating potential deviations from invariance during major stock market falls from January 2014 to July 2018. In particular, we analyze the plunges in stock prices on March 3, 2014, during the Russian ruble crisis in the middle of December 2014, and during the stock market crash on April 6 and 9, 2018. To the best of our knowledge, the paper also presents the first study of the impact of the increased number of active retail investors in a stock market using the quantitative predictions of microstructure invariance theory about the relationship between information flows and trading flows. The sample period spans from the first week of August 2018 to the last week of June 2021. According to the client statistics provided on the Moscow Exchange (MOEX) website, there was a sharp increase in individual investor participation in 2020 (+ almost 5 million retail investors over 2020). Moreover, the total inflow of private investors' funds increased by 301 billion rubles (around 4.2 billion dollars according to the average U.S. dollar/Russian ruble exchange rate in 2020) from January 2020 to December 2020. The use of invariance-implied methodology leads to new insights about the changes in the relationship between the trading and the information processes during the significant price/volume events in financial markets.

The main quantitative prediction of the ITI hypothesis is as follows: log-return variation per transaction is proportional to the—2 power of the average trade size times the stock price. Andersen et al. (2020) confirm this relationship using data on the E-mini S&P 500 futures market. Kyle and Obizhaeva (2016) report the results supporting this nonlinear relationship for individual U.S. stocks. Bae et al. (2020) show that invariance principles also hold for the Korean stock market. Kyle and Obizhaeva (2019) show that the invariance-implied market impact cost model gives more accurate estimates of price declines during the bet-induced stock market crashes compared to some alternative models (e.g., Frazzini et al., 2018; Wurgler & Zhuravskaya, 2002).

Using several high-frequency measurement techniques, we test the log-linear regression specifications related to various assumptions about the relationship between trading variables. According to our results, the invariance theory explains the relations between market activity variables for Russian liquid stocks over short time intervals much better than alternative hypotheses assuming linear relationships between such variables. Applying the aggregation schemes that mitigate the errors-in-variables effects leads to an increase in the coefficients of determination while maintaining the economic closeness to the invariance-implied quantitative predictions. We also confirm the ITI relationship for the set of trading days characterized

by the largest price declines (between January 2014 and July 2018). Additionally, we briefly discuss the intraday dynamics of bid-ask spread.

Market Microstructure Invariance hypothesis allows us to look at the Russian stock market from another angle. In this paper, we examine whether there was a changepoint (an abrupt shift) in the dynamics of the estimates characterizing the relationship between public information flows (the number of news articles provided by Thomson Reuters Eikon or Google relative search volumes of Russian stocks) and trading activity. We focus on these two sources of information by making the following assumptions. First, the same business-time clock  $R$  governs both Russian retail investors' aggregate behavior and Google search activity of Russian stocks. Second, the professional market participants' aggregate activity and the arrival rate of news articles provided by Thomson Reuters Eikon unfold in the same business-time clock  $P$ . Third, the market-wide business-time clock  $M$  is a linear combination of two clocks  $M = \alpha_R \cdot R + \alpha_P \cdot P$  with positive time-varying coefficients  $\alpha_R, \alpha_P : \alpha_R + \alpha_P = 1$  depending on external factors. Intuitively, our conjecture about the possibility of a significant growth of  $\alpha_R$  relative to  $\alpha_P$  is primarily because retail investor participation in the Russian stock market was drastically increased in 2020. Since local individual investors and professional market participants (e.g., institutionals or asset managers) tend to use different sources of financial information,<sup>1</sup> as we have already noticed, we expect that market-wide trading activity started to be more synchronized with Google search activity of Russian stocks. Our regression analysis confirms this finding. We approximate the information flows by the negative binomial process and show that an abrupt shift upwards in the estimates characterizing this relationship occurred in the last week of February 2020, when the Russian stock market faced a huge inflow of private investors' funds.

We next review the literature that analyzes the link between the price and information processes and trading activity.

## 2 Literature review

Since the seminal paper by Bachelier (1900), the connection between the fluctuations of asset prices and several characteristics of the trading process has been under scrutiny. Bachelier's ideas that the movement of stock prices takes the form of a random walk and price volatility is proportional to the square root of time were the main assumptions in the initial attempts to model price-transaction intensity relations. Osborne (1962) was the first to formulate the hypothesis that the return variance is proportional to trading volume and the number of transactions. Many subsequent studies (e.g., Clark, 1973; Crouch, 1970; Epps & Epps, 1976; Jain & Joh, 1988; Tauchen & Pitts, 1983; Wood et al., 1985) also found positive correlations between trading volume and the absolute value of price changes.<sup>2</sup> One of the main theoretical explanations for this phenomenon was the assumption that price changes are selected from a set of distributions characterized by different variances. This speculation became known as the "mixture of distributions hypothesis" (MDH). As a mixing variable, researchers used trading volume  $V$  (e.g., Westerfield (1977)) or the number of transactions  $N$  (e.g., Ané & Geman, 2000; Jones et al., 1994). In the previously mentioned article by Andersen et al. (2020), the specifications were referred to as the Mixture-of-Distributions-Hypothesis in

<sup>1</sup> A review of the literature on the use of information by different types of investors is presented in Cascino et al. (2014), among others.

<sup>2</sup> A review of theoretical and empirical studies on this topic published in 1950–1980 is provided by Karpoff (1987).

Volume ( $\sigma^2 \sim V$ ) and the Mixture-of-Distributions-Hypothesis in Transactions ( $\sigma^2 \sim N$ ). Andersen et al. (2020) also first extrapolate the invariance principles to an intraday dimension. The market structure of the Moscow Exchange provides an opportunity to develop the topic introduced in Andersen et al. (2020) by turning to the panel analysis. Like the CME Group Globex platform, the Moscow Exchange is a centralized marketplace. The tick sizes and margin requirements are adjusted by exchange officials promptly, and small lot sizes prevent significant market frictions.

The relationship between macroscopic market characteristics and information available in financial markets was investigated in numerous papers. In some articles, small or even insignificant relationships between public information and market activity was found (e.g., Berry & Howe, 1994; Mitchell & Mulgerin, 1994). Several other papers documented the significant relationship between different information flows and several measures of trading intensity (e.g., Da et al., 2011; Heston & Sinha, 2017; Tetlock, 2007). Kyle and Obizhaeva (2017b) were the first to derive empirically testable predictions about the trading and information processes. The authors demonstrate that trading activity and information flow approximated by the arrival rate of news articles are highly synchronized. The researchers implemented count regression tests and showed that the number of bets per news item about a given firm is approximately constant across U.S. stocks. In other words, they proved that an invariant “amount of money changes hands on average per news article.” In this paper, we do not test this prediction using news and transaction-level data for Russian stocks; rather, we use this methodology to assess the impact of the significant rise in the number of retail investors in 2020.

In the next section, we provide a more detailed formulation of the Intraday Trading Invariance hypothesis and the two specifications of the Mixture-of-Distributions-Hypothesis: MDH-V and MDH-N.

### 3 Hypotheses about the relationship between high-frequency trading variables

#### 3.1 Basic notation and estimation methodology

Before formulating the main testable hypotheses, we introduce the necessary notations, following the methodology described in Andersen et al. (2020).

The sample starts at time 0 and contains  $D$  trading days. Each consists of  $T$  intraday intervals of length  $\Delta t = 1/T$ . Thus, the sample contains  $\tau = 1, \dots, D \cdot T$  non-overlapping intervals. For further convenience, we also introduce a double-index notation:  $d \in D = 1, \dots, D$  for a trading day and  $t \in T = 1, \dots, T$  for an intraday interval. A double-index notation can be converted to a single-index notation according to the following rule:  $\tau = (d - 1) \cdot T + t$ .

For each stock  $i \in I$  and each interval  $\tau$  from a sample, we define random realizations of some variable by placing the “tilde” sign on top of it:  $\widetilde{Q}_{i\tau}$  is the average number of shares in one transaction for the interval  $\tau$ ;  $\widetilde{\sigma}_{i\tau}^2$  is log-return variance for the interval  $\tau$  (per unit of time);  $\widetilde{\Gamma}_{i\tau}$  is the number of transactions per unit of time;  $\widetilde{V}_{i\tau}$  is the cumulative trading volume (in the number of shares per unit of time);  $\widetilde{P}_{i\tau}$  is the average trading price for the interval  $\tau$  (in rubles per share). Lower case letters, in turn, denote the logarithms of these variables:  $\widetilde{q}_{i\tau} = \ln \widetilde{Q}_{i\tau}$ ;  $\widetilde{s}_{i\tau} = \ln \widetilde{\sigma}_{i\tau}^2$ ;  $\widetilde{\gamma}_{i\tau} = \ln \widetilde{\Gamma}_{i\tau}$ ;  $\widetilde{v}_{i\tau} = \ln \widetilde{V}_{i\tau}$ ;  $\widetilde{p}_{i\tau} = \ln \widetilde{P}_{i\tau}$ .

We next assume that unobservable bets volume coincides with observable trading volume.<sup>3</sup> Given this assumption, the following equalities automatically hold:  $\widetilde{V}_{i\tau} = \widetilde{Q}_{i\tau} \cdot \widetilde{\Gamma}_{i\tau}$  and  $\widetilde{v}_{i\tau} = \widetilde{q}_{i\tau} + \widetilde{\gamma}_{i\tau}$ . When determining the expectation of a random variable on the interval  $\tau$  (conditional on the available information at time  $\tau - 1$ ), we remove the “tilde” sign. For instance,  $s_{i\tau} = E_{\tau-1}\{\widetilde{s}_{i\tau}\}$ .

The methodology for estimating these conditional expectations is based on the multiplicative error model (MEM).<sup>4</sup> The dynamics of strictly positive random variables  $\widetilde{Y}_{i\tau}$  is defined as follows:

$$\widetilde{Y}_{i\tau} = Y_{i\tau} \cdot \widetilde{U}_{i\tau}, \tag{1}$$

where  $\widetilde{U}_{i\tau}$  are strictly positive independent and identically distributed random variables with  $E\{\widetilde{U}_{i\tau}\} = 1$  and  $Var\{\widetilde{U}_{i\tau}\} = \sigma_U^2$ .

Thus  $E_{\tau-1}\{\widetilde{Y}_{i\tau}\} = Y_{i\tau}$  and  $Var_{\tau-1}\{\widetilde{Y}_{i\tau}\} = Y_{i\tau}^2 \cdot \sigma_U^2$ . This approach assumes that for each stock  $i$  and each interval  $\tau$ , the estimate  $Y_{i\tau}$  is unbiased and has finite variance. According to Andersen et al. (2020), the corresponding logarithmically transformed quantities  $\widetilde{y}_{i\tau} = \ln \widetilde{Y}_{i\tau}$  can be represented as

$$\widetilde{y}_{i\tau} = y_{i\tau} + c + \widetilde{e}_{i\tau}, \tag{2}$$

where  $y_{i\tau}$  is the conditional expectation of the logarithmic value of a variable, and  $\widetilde{e}_{i\tau}$  is a random variable with  $E\{\widetilde{e}_{i\tau}\} = 0$  and  $Var\{\widetilde{e}_{i\tau}\} < \infty$ . According to Jensen’s inequality,  $E\{\ln \widetilde{U}_{i\tau}\} < \ln E\{\widetilde{U}_{i\tau}\}$ . Therefore, a constant  $c = E\{\widetilde{y}_{i\tau} - y_{i\tau}\} < 0$ .

Equation (2) implies that significant errors-in-variables problems may arise if variance of an error term  $Var\{\widetilde{e}_{i\tau}\}$  is large. To mitigate these effects, we aggregate the variables through the aggregation scheme (3), resulting in  $D$  daily observations for each stock  $i \in I$ . We also use the aggregation technique (4) to obtain  $T$  separate time-of-day observations for each stock  $i \in I$ :

$$y_{id} = \frac{1}{T} \cdot \sum_{t=1}^T \widetilde{y}_{idt} \approx c + \frac{1}{T} \cdot \sum_{t=1}^T y_{idt} \text{ for } d = 1, \dots, D; i \in I, \tag{3}$$

$$y_{it} = \frac{1}{D} \cdot \sum_{d=1}^D \widetilde{y}_{idt} \approx c + \frac{1}{D} \cdot \sum_{d=1}^D y_{idt} \text{ for } t = 1, \dots, T; i \in I. \tag{4}$$

According to the law of large numbers, the contribution of an error term  $\widetilde{e}_{i\tau}$  should considerably decrease after temporal aggregation.

The regression-based framework for testing the Intraday Trading Invariance hypothesis and the MDH specifications is described in Appendix 8.2.

### 3.2 Data

We use tick-by-tick data on trades and bid-ask quotes (“Top of the book” specification) from January 6, 2014 to July 31, 2018 provided by the Moscow Exchange. The sample consists of only those 32 shares of Russian issuers included in the MOEX Russia Index during all

<sup>3</sup> In general, the endogeneity problem arises when a researcher constructs proxies for bets using data on executed trades since such datasets do not contain information on unexecuted trades (see Obizhaeva, 2012).

<sup>4</sup> The general structure of the MEM framework is presented in Engle (2002).

55 months. We emphasize once again that such a strict limitation is caused by testing log-linear relations over short intervals. We eliminate those observations that do not belong to the continuous trading period of the main trading session.

We aggregate the observations over five-minute intervals for each stock and each trading day. For each interval  $\tau$  and each stock  $i$ , we estimate market activity variables in the following way.  $P_{i\tau}$  is calculated as the average trading price of stock  $i$  over  $\tau$ ;  $Q_{i\tau}$  is the average number of stocks  $i$  in one transaction over  $\tau$ ;  $\Gamma_{i\tau}$  is the number of transactions of stock  $i$  over the interval  $\tau$ ;  $V_{i\tau}$  is the cumulative trading volume (in the number of stocks  $i$ ) over  $\tau$ . We use a standard unbiased high-frequency estimate of the realized return variance (see Andersen et al., 2003). We first calculate the midpoint of ask and bid quotes at the end of each minute and then sum the five consecutive squared one-minute log-returns to obtain an estimate of  $\sigma_{i\tau}^2$  for each five-minute interval. Since we use the log-linear regression specifications, we remove all stock-interval observations with zero trading volume or the realized variance estimate close to 0 (less than  $10^{-30}$  in absolute value).

Table 1 presents summary statistics for trading variables after data cleaning. Though the percentage of all omitted observations is 65.71% of the initial sample, the final sample consists of a significant number of stock-interval observations: 1,279,833.

## 4 Testing ITI and MDH over the 2014–2018 period

In this section, we test the quantitative predictions of invariance hypotheses, as well as two specifications of the mixture-of-distributions hypothesis (MDH-N and MDH-V) over the 2014–2018 period. We test the intraday trading invariance hypothesis over the 2018–2021 period in Sect. 7.

The main testable log-linear specification is as follows:

$$s_{ik} - \gamma_{ik} + 2p_{ik} = c - \beta q_{ik} + e_{ik} \text{ for } k = 1, \dots, K; i \in I \quad (5)$$

where  $k$  is an index for  $\tau$  (without aggregating five-minute observations),  $d$  (averaging the five-minute observations for each activity variable and each stock across intervals within a given trading day), or  $t$  (averaging the observations for each activity variable, each stock, and each intraday interval across trading days). To test the MDH hypotheses, we exclude the price level, so the regressand is  $s_{ik} - \gamma_{ik}$  in this case.

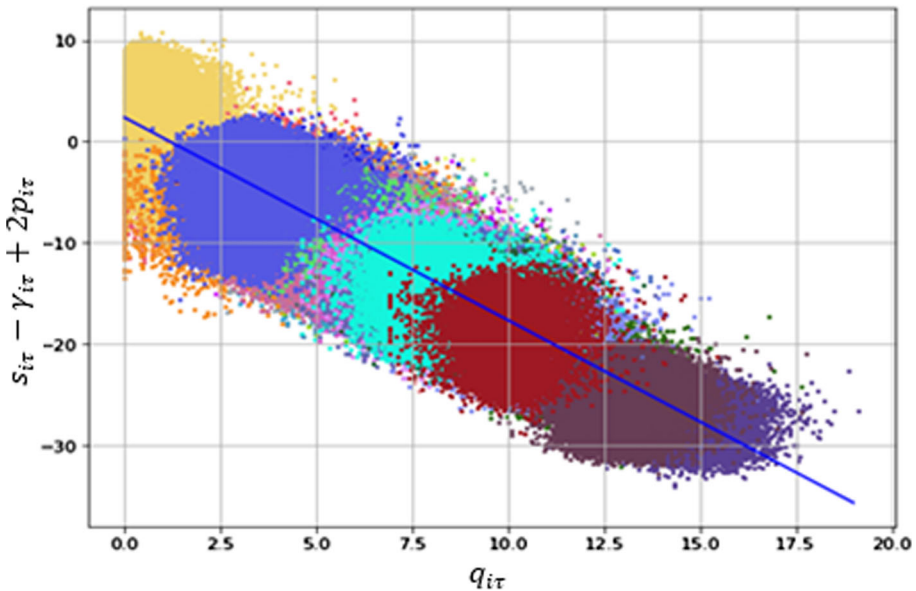
### 4.1 Tests at five-minute sampling frequency

In the first step, we perform a regression analysis exploiting five-minute observations.

#### 4.1.1 Invariance hypothesis

Figure 1 shows a cloud of 1,279,833 points representing  $\ln \sigma_{i\tau}^2 P_{i\tau}^2 / \Gamma_{i\tau}$  against  $\ln Q_{i\tau}$  for 32 stocks included in the MOEX Russia Index from January 2014 to July 2018. Different colors represent different stocks henceforth. We also add a line  $s_{i\tau} - \gamma_{i\tau} + 2p_{i\tau} = 2.364 - 2q_{i\tau}$ , where the slope is fixed at  $-2$  and the intercept is estimated by OLS regression. We can see that observations cluster along the plotted line.

The fitted line is  $s_{i\tau} - \gamma_{i\tau} + 2p_{i\tau} = 2.388 - 2.003q_{i\tau}$  with Driscoll-Kraay standard errors equal to 0.025 and 0.003, respectively. The coefficient of determination  $R^2$  is 0.779. It is worth noting that the hypothesis that the slope is equal to  $-2$  is not rejected at the 5% significance



**Fig. 1** The figure plots  $s_{i\tau} - \gamma_{i\tau} + 2p_{i\tau}$  on the vertical axis against  $q_{i\tau}$  on the horizontal axis for each of the 32 stocks of Russian issuers included in the MOEX Russia Index from January 2014 to July 2018, where  $s_{i\tau}$  is the logarithmic value of log-return variance,  $\gamma_{i\tau}$  is the logarithmic value of the transaction rate,  $p_{i\tau}$  is the logarithmic value of the average price level,  $q_{i\tau}$  is the logarithmic value of the average number of stocks per transaction. All values are computed at five-minute sampling frequency. The solid line is  $s_{i\tau} - \gamma_{i\tau} + 2p_{i\tau} = 2.364 - 2q_{i\tau}$ , where the intercept is estimated from an OLS regression with the slope fixed at  $-2$

level. The results of testing this relationship for individual stocks are the following: all 32 slope coefficients are significantly higher than  $-2$ ; they range from  $-1.105$  to  $0.207$ .<sup>5</sup> The average is  $-0.177$  and the median equals  $-0.099$ .<sup>6</sup>

As a robustness check, we test the invariance hypothesis using high-frequency data for each month in our sample separately. Figure 2 displays the slope coefficients on the  $q_{i\tau}$  variable across months. All coefficients are economically close to the theoretical value of  $-2$ .

Figure 3 shows relationships between the logarithms of log-return variance  $s_{i\tau}$  and the number of transactions  $\gamma_{i\tau}$  (a) and the logarithms of log-return variance  $s_{i\tau}$  and trading volume  $v_{i\tau}$  (b) for 32 stocks included in the MOEX Russia Index from January 2014 to July 2018. We should notice that the linear relationships between the corresponding variables predicted by the MDH-N and MDH-V hypotheses are not observed in the cross-section.

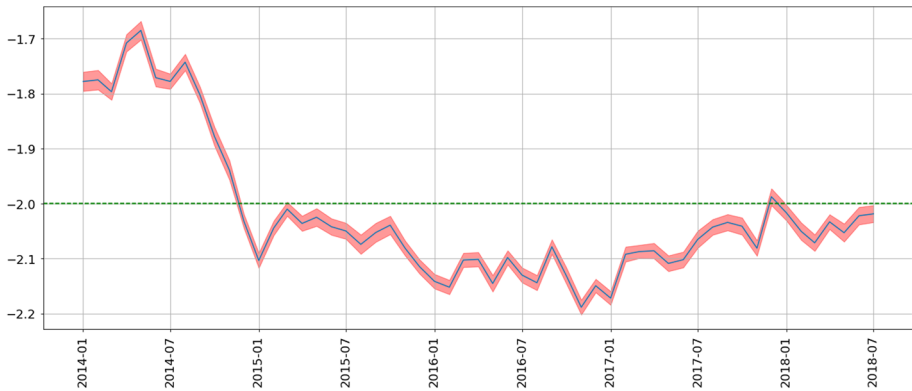
#### 4.1.2 MDH-N

For the entire sample, the fitted line is  $s_{i\tau} = -18.235 + 0.419\gamma_{i\tau}$  with Driscoll-Kraay standard errors equal to  $0.021$  и  $0.004$ , respectively. The coefficient of determination  $R^2$  equals  $0.048$ .

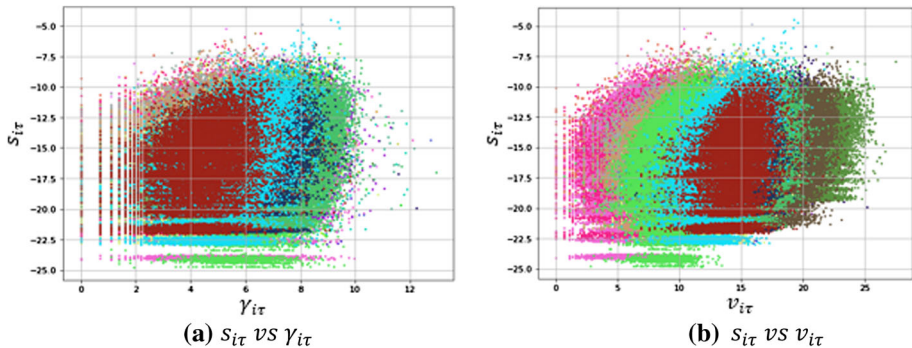
<sup>5</sup> As for testing hypotheses for individual stocks, we use White standard errors henceforth.

<sup>6</sup> The presence of significant noise in regressors for individual Russian and U.S. securities was noted in Kyle & Obizhaeva (2017a) as one of the main explanations for the difference between the fitted and invariance-implied slopes.





**Fig. 2** The figure plots the slope coefficient  $\beta$  for OLS regressions  $s_{i\tau} - \gamma_{i\tau} + 2p_{i\tau} = c + \beta \cdot q_{i\tau} + e_{i\tau}$ , estimated for each month separately. The confidence intervals are computed as  $\pm 2$  Driscoll-Kraay standard errors. The green dashed line indicates the theoretical value predicted by the invariance hypothesis



**Fig. 3** The left panel (a) depicts the logarithms of log-return variance against the logarithms of the transaction rate. The right panel (b) depicts the logarithms of log-return variance against the logarithms of trading volume. All values are computed at five-minute sampling frequency. The sample consists of the 32 stocks of Russian issuers included in the Moscow Exchange Russia Index from January 2014 to July 2018

For individual stocks, the slope coefficients range from 0.254 to 1.113. The mean value is 0.741, and the median is equal to 0.809. At the 5% significance level, the MDH-N hypothesis is not rejected for only 4 out of 32 stocks.

#### 4.1.3 MDH-V

For the entire sample, the fitted line is  $s_{i\tau} = -17.109 + 0.101v_{i\tau}$  with Driscoll-Kraay standard errors equal to 0.019 or 0.001, respectively. The coefficient of determination  $R^2$  is 0.017. The slope coefficients vary between 0.129 and 0.754 for individual stocks. The mean is 0.497, and the median equals 0.542. The MDH-V hypothesis is rejected for all 32 stocks.

As we previously mentioned, such estimates can be noisy due to the influence of error terms. Therefore, we next perform an analysis using different aggregation schemes.



## 4.2 Tests at the daily frequency

To reduce the influence of errors-in-variables effects, we next use the aggregation scheme (3).

### 4.2.1 Invariance hypothesis

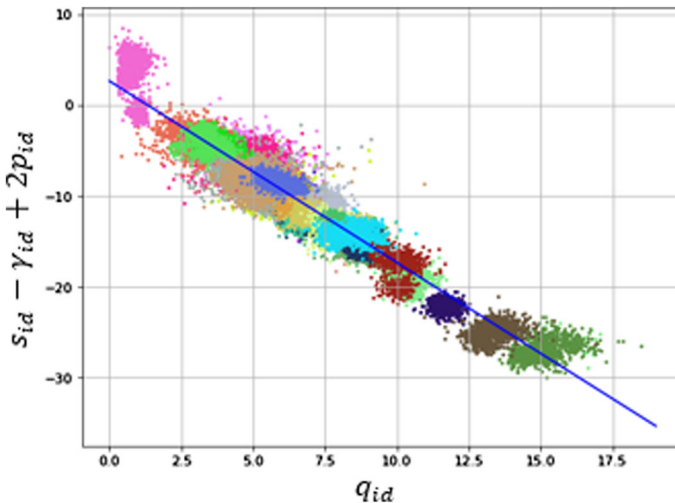
Figure 4 shows a cloud of 35,707 points, displaying  $\ln \sigma_{id}^2 P_{id}^2 / \Gamma_{id}$  versus  $\ln Q_{id}$  for 32 stocks included in the MOEX Russia Index from January 2014 to July 2018.

For comparison, we add the line  $s_{id} - \gamma_{id} + 2p_{id} = 2.075 - 2q_{id}$ . The slope is fixed at  $-2$  as predicted by the invariance hypothesis, and the intercept is estimated using OLS regression.

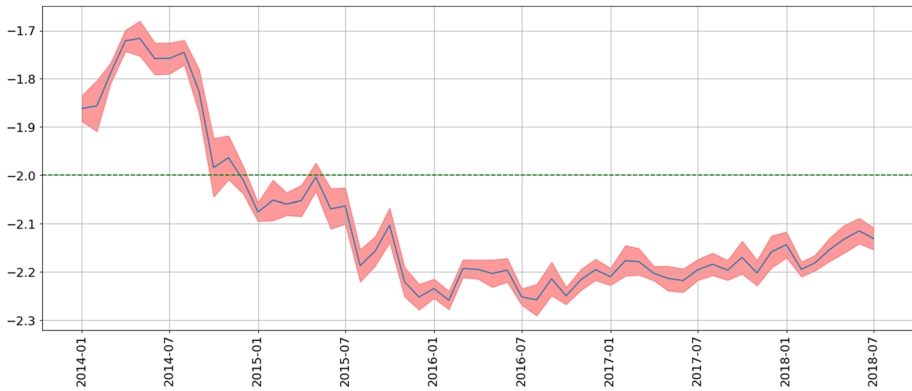
The fitted line is  $s_{id} - \gamma_{id} + 2p_{id} = 3.287 - 2.090q_{id}$  with Driscoll-Kraay standard errors equal to  $0.127$  и  $0.014$ , respectively. The coefficient of determination  $R^2$  equals  $0.933$ . It is worth noting that the hypothesis that the slope is equal to  $-2$  is rejected due to small standard errors. Nevertheless, the coefficient remains economically close to the value implied by the invariance hypothesis. For individual stocks, the slope ratios range from  $-2.027$  to  $0.582$ . The mean is  $-0.311$ , and the median is equal to  $-0.233$ . At the  $5\%$  significance level, the invariance hypothesis is not rejected for only 1 stock in our sample.

We next perform a regression analysis for each month from our first sample.

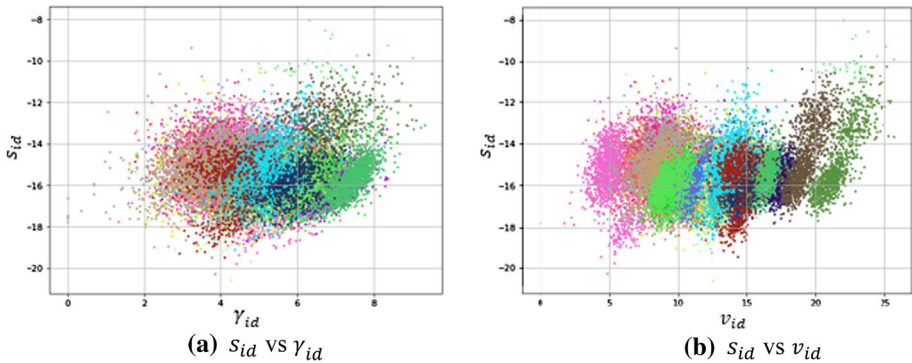
Figure 5 displays the slope coefficients on the  $q_{id}$  variable across months. We can observe similar dynamics that we found earlier in Fig. 2: a fall to the level of  $-2.3$  and subsequent fluctuations in a range from  $-2.3$  to  $-2.1$ . Figure 6 shows graphical relationships between the logarithms of log-return variance  $s_{id}$  and the number of transactions  $\gamma_{id}$  (a) and the logarithms of log-return variance  $s_{id}$  and trading volume  $v_{id}$  (b) for 32 stocks of Russian



**Fig. 4** The figure plots  $s_{id} - \gamma_{id} + 2p_{id}$  on the vertical axis against  $q_{id}$  on the horizontal axis for each of the 32 stocks of Russian issuers included in the MOEX Russia Index from January 2014 to July 2018.  $s_{id}$  is the logarithmic value of log-return variance,  $\gamma_{id}$  is the logarithmic value of the transaction rate,  $p_{id}$  is the logarithmic value of the average price,  $q_{id}$  is the logarithmic value of the average number of stocks per transaction. The values are averaged in accordance with the scheme (3). The solid line is  $s_{id} - \gamma_{id} + 2p_{id} = 2.075 - 2q_{id}$ , where the intercept is estimated from an OLS regression with the slope fixed at  $-2$



**Fig. 5** This figure plots the slope coefficient  $\beta$  for OLS regressions  $s_{id} - \gamma_{id} + 2p_{id} = c + \beta \cdot q_{id}$  estimated for each month separately. The confidence intervals are computed as  $\pm 2$  Driscoll-Kraay standard errors. The green dashed line indicates the theoretical value predicted by the invariance hypothesis



**Fig. 6** The left panel (a) depicts the logarithms of log-return variance against the logarithms of the transaction rate. The right panel (b) depicts the logarithms of log-return variance against the logarithms of trading volume. The observations are averaged across the trading day. The sample consists of the 32 stocks of Russian issuers included in the MOEX Russia Index during January 2014–July 2018

issuers included in the MOEX Russia Index from January 2014 to July 2018. Here we also do not observe linear relationships between the corresponding variables predicted by MDH-N и MDH-V in the cross-section.

#### 4.2.2 MDH-N

For the entire sample, the fitted line is  $s_{id} = -15.458 - 0.027\gamma_{id}$  with Driscoll-Kraay standard errors equal to 0.135 and 0.023, respectively. The coefficient of determination  $R^2$  is 0.001. For individual stocks, slope coefficients range from  $-0.420$  to  $1.662$ . The mean equals 0.577, and the median is 0.552. At the 5% significance level, the MDH-N hypothesis is not rejected for only 2 out of 32 stocks.

### 4.2.3 MDH-V

For the entire sample, the fitted line is  $s_{id} = -15.928 + 0.026\gamma_{id}$  with Driscoll-Kraay standard errors equal to 0.107 and 0.008, respectively. The coefficient of determination  $R^2$  is 0.007. For individual stocks, slope coefficients range from  $-0.403$  and  $1.101$ . The mean is  $0.444$ , and the median equals  $0.463$ . At the 5% significance level, the MDH-V hypothesis is not rejected for only 1 out of 32 stocks.

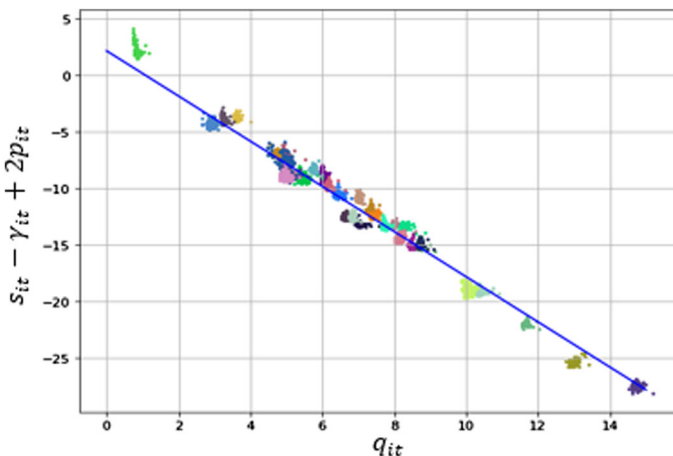
In summary, none of the three tested hypotheses show high statistical significance in the cross-section and in the case of considering individual stocks. Nevertheless, when we test the ITI relationship, we find an increase in the coefficient of determination and the retention of the economic significance of the slope coefficient, which is close to  $-2$ .

### 4.3 Tests across the intraday pattern

In this subsection, we use the second aggregation scheme, represented by formula (4). Averaging over a small number of five-minute intervals for the sample covering several years of observations undoubtedly yields fewer points compared to the aggregation scheme (3). For each stock  $i$  and each five-minute interval  $t$ , we obtain the value of each variable by taking the arithmetic average across all trading days  $d \in D$ .

#### 4.3.1 Invariance hypothesis

Figure 7 shows 3296 points representing the value of  $\ln\sigma_{it}^2 p_{it}^2 / \Gamma_{it}$  versus  $\ln Q_{it}$  for 32 stocks included in the MOEX Russia Index during January 2014–July 2018. For comparison, we add the line  $s_{it} - \gamma_{it} + 2p_{it} = 2.185 - 2q_{it}$ , where the slope is fixed at level  $-2$  and the intercept is estimated using OLS regression. We can observe that a significant part of the



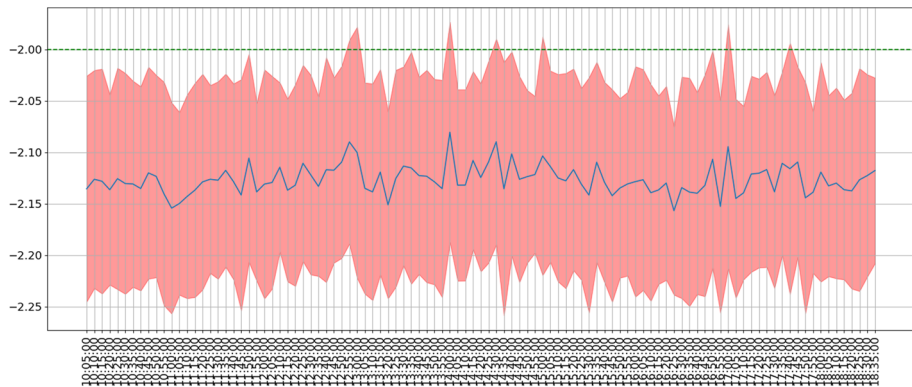
**Fig. 7** The figure plots  $s_{it} - \gamma_{it} + 2p_{it}$  on the vertical axis against  $q_{it}$  on the horizontal axis for each of the 32 stocks of Russian issuers included in the MOEX Russia Index during January 2014–July 2018.  $s_{it}$  is the logarithmic value of log-return variance,  $\gamma_{it}$  is the logarithmic value of the transaction rate,  $p_{it}$  is the logarithmic value of the average price,  $q_{it}$  is the logarithmic value of the average number of stocks per transaction. The observations are averaged in accordance with the scheme (4). The solid line is  $s_{it} - \gamma_{it} + 2p_{it} = 2.185 - 2q_{it}$ , where the intercept is estimated from an OLS regression with the slope fixed at  $-2$

points lies very close to this line. The fitted line is  $s_{it} - \gamma_{it} + 2p_{it} = 3.068 - 2.126q_{it}$  with Driscoll-Kraay standard errors equal to 0.045 and 0.002, respectively. The coefficient of determination  $R^2$  equals 0.982. It is worth noting that the hypothesis that the slope is equal to  $-2$  is rejected due to small standard errors. However, the coefficient is economically close to the value predicted by the invariance hypothesis.

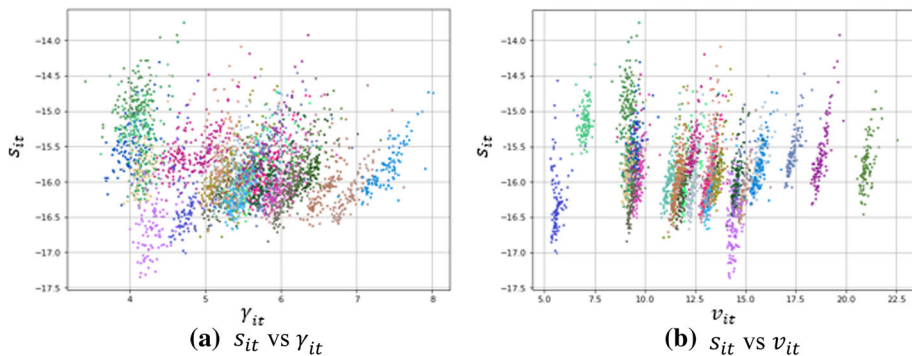
We note that the increased variation of the slope coefficients  $\beta$  for individual stocks is caused by a decline in the number of observations. The minimum (maximum) value is  $-2.887$  ( $1.295$ ). The mean equals  $-0.395$ , and the median is  $-0.355$ .

We next investigate the intraday dynamics of the slope coefficient  $\beta$  (Fig. 8) by running the OLS regression for each five-minute interval. As we can observe, there is a small variation in this parameter without a pronounced time trend: the  $\beta$  coefficient ranges from  $-2.16$  to  $-2.07$ .

Figure 9 depicts the relationship between the logarithms of log-return variance  $s_{it}$  and the number of transactions  $\gamma_{it}$  (a) and the logarithms of log-return variance  $s_{it}$  and trading



**Fig. 8** This figure plots the slope coefficient  $\beta$  for OLS regressions  $s_{it} - \gamma_{it} + 2p_{it} = c + \beta \cdot q_{it} + e_{it}$  estimated for each five-minute interval separately. The confidence intervals are computed as  $\pm 2$  White standard errors. The green dashed line indicates the theoretical value predicted by the invariance hypothesis. The timestamp is the right border of the corresponding five-minute interval



**Fig. 9** The left panel (a) depicts the logarithms of log-return variance against the logarithms of the transaction rate. The right panel (b) depicts the logarithms of log-return variance against the logarithms of trading volume. The observations are averaged across the five-minute intraday intervals. The sample consists of the 32 stocks included in the MOEX Russia Index during January 2014 – July 2018

volume  $v_{it}$  (b) for 32 stocks included in the MOEX Russia Index during January 2014 – July 2018. As we can see, there is no explicit linear relationship between the variables in the cross-section in either case.

### 4.3.2 MDH-N

For the entire sample, the fitted line is  $s_{it} = -15.352 - 0.089\gamma_{it}$  with Driscoll-Kraay standard errors equal to 0.059 and 0.013, respectively. The coefficient of determination  $R^2$  is equal to 0.026. For individual stocks, the slope coefficients vary between 0.609 and 2.351. The mean is 1.270, and the median equals 1.260. At the 5% significance level, the MDH-N hypothesis is not rejected for 21 out of 32 stocks (mainly due to large standard errors).

### 4.3.3 MDH-V

For the entire sample, the fitted line is  $s_{it} = -15.764 - 0.006v_{it}$  with Driscoll-Kraay standard errors equal to 0.050 and 0.001, respectively. The coefficient of determination  $R^2$  is 0.002. For individual stocks, the slope coefficients vary between 0.203 and 1.379. The average is 0.833, the median reaches 0.859. At the 5% significance level, the MDH-V hypothesis is not rejected for 20 out of 32 (also mainly due to large standard errors).

In summary, we have shown that the Intraday Trading Invariance hypothesis has the highest explanatory power among the three different models. As an additional check, we repeat the analysis by aggregating the variables at the twenty-minute sampling frequency. To obtain a high-frequency unbiased estimate of the realized variance, we sum consecutive squared one-minute log-returns over the corresponding twenty-minute intervals. We exclude the intervals with no trading volume and the realized variance estimate less than  $10^{-30}$ . Finally, we obtain the sample, consisting of 307,968 observations (32.96% of the initial sample). In this paper, we do not present the results of our tests using twenty-minute sampling frequency. Nevertheless, we note that no substantial differences in the estimates have been found.

## 5 Invariance during market turbulence

### 5.1 Discussion

Significant deviations from the invariant ratios can occur during various market crises.

1. A breakdown in the ITI relationship may be associated with a sharp decline in market makers' and arbitrageurs' activity.

The demonstrative example is the events in the E-mini S&P 500 market on May 6, 2010, which took place within minutes before reaching the nadir of the crash at 13:45 Central Time. As shown by Andersen et al. (2020), the collapse in the provision of liquidity was accompanied by a sharp rise in the values of  $\ln I_{dt}$  calculated at a granularity of one minute. Due to the absence of arbitrage activity, the values of  $\ln I_{dt}$  were at peak levels during those most turbulent minutes.

2. A (fast) execution of large bets can create a short-term but significant impact on asset prices.

According to the estimates presented in Obizhaeva (2016), the two-day destabilization in the Russian foreign exchange market in mid-December 2014 was an example of such a

collapse. It was triggered by the execution of a large bet to sell rubles of about \$6-\$9 billion over a short period. Due to the spillover effect, the Russian stock market was also in a crisis: the RTS index lost more than 9.3% on December 15 and more than 11.4% on December 16. A few months earlier, on March 3, 2014, there was a massive sale of Russian stocks amid political tensions in Ukraine and the decision of the Bank of Russia to raise the key rate by 1.5 percentage points. As a result, the RTS index fell by more than 10% on that trading day. The third most turbulent episode over the period under review was the collapse of stock prices on April 6 and 9, 2018, associated with the introduction of new U.S. sanctions against some Russian officials and businesspersons, as well as with the aggravated situation in Syria. In two trading days, the RTS index fell in aggregate by more than 11.9%.

In this section, we examine the degree of deviation from invariance relationships during the deepest market price falls. As we noted earlier, the invariance principle is a benchmark against which it is convenient to monitor market dynamics during stressful episodes. Such analysis may be of interest not only to market participants with short-term trading strategies but also to financial regulators.<sup>7</sup>

## 5.2 Empirical evidence

This part of the study aims to estimate the average differences in the logarithms of the invariant  $\iota \equiv \ln I = p+q+0.5s-0.5\gamma$  for turbulent days, characterized by the most significant market declines, and other periods. Earlier, we have already indicated the main reason, according to which we expect a significant rise in the value of the trading invariant during the market downturn. During high market volatility associated with external shocks, selling pressure is exacerbated, which can lead to increased liquidity demand and limited arbitrage activity. Due to the limited liquidity supply from intermediaries, the average trade size may increase. As a result, an unusually high value of the logarithm of the trading invariant  $\iota$  can be observed. Thus, the main hypothesis is formulated as follows:

**Hypothesis 1** There is a significant positive deviation of the standardized logarithm of the trading invariant during periods of market stress.

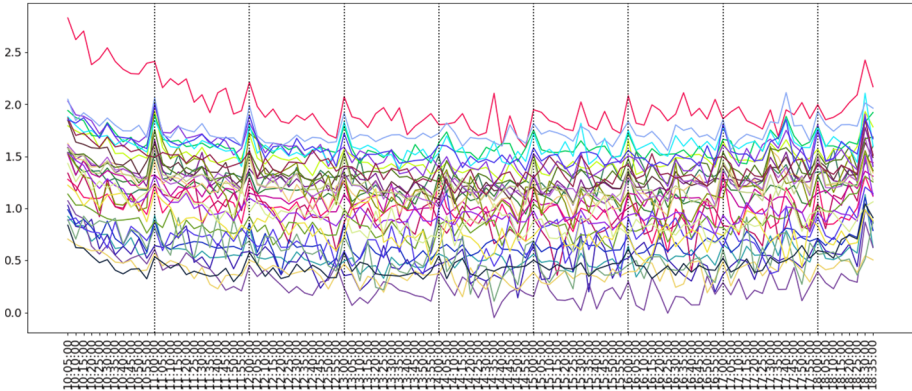
For each stock  $i$  and each five-minute interval  $\tau$ , the standardized logarithm of the trade invariant  $\tilde{\iota}_{it}$  is calculated as follows:

$$\tilde{\iota}_{it} = \frac{\iota_{it} - \overline{\iota_{it}}}{S_{\iota_{it}}}, \quad (6)$$

where  $\overline{\iota_{it}}$  is the mean of  $\iota_{it}$  for each stock  $i$  over each five-minute interval  $t$ ;  $S_{\iota_{it}}$  is the square root of unbiased sample variance of  $\iota_{it}$  calculated on the same set.

The transition to standardized values is caused by the need to take into account possible violations in the assumption about the distribution of the invariant  $I_{it}$ . Figure 10 shows the logarithms of the trading invariant  $\iota_{it}$  averaged according to scheme (4) for each specified five-minute interval and each stock from our sample. It is worth noting that all observations are concentrated in a relatively narrow interval. However, the universal distribution for all

<sup>7</sup> Obizhaeva (2016) notes that some U.S. regulators have already been using a methodology based on the principles of market microstructure invariance to assess systemic risks. Examples: Office of Financial Research 2014 Annual Report (the U.S. Treasury) and Tuzun (2013) (Board of Governors of the Federal Reserve System).



**Fig. 10** The figure shows the logarithms of trading invariant  $t_{it}$  averaged according to scheme (4) for each five-minute interval and each security from our sample. The dotted lines represent the last five-minute intervals of each calendar hour. The timestamp displays the right border of the corresponding five-minute interval

stocks is not observed: many lines have no intersection with each other. Thus, standardization is used to smooth out such heterogeneities.<sup>8,9</sup>

Based on Fig. 10, we briefly discuss the dynamics of intraday trading activity in the most liquid part of the Russian market. First, it is worth mentioning the U-shaped curve showing the dependence of the logarithm of the trading invariant on time for each stock from our sample. Similar results for dynamics of intraday bid-ask spreads and/or trading activity were reported by Wood et al. (1985), Chan et al. (1995), Madhavan et al. (1997), Andersen et al. (2020), etc.

As shown by Madhavan et al. (1997), the information asymmetry between market participants is high at the opening. As a result, liquidity suppliers are less active due to the adverse selection problem. In our opinion, the logarithms of the trading invariant have inflated values precisely because of the limited intermediary activity. A similar effect is observed during the last minutes of the trading day when liquidity suppliers increase the order-processing and/or inventory holding components of the bid-ask spread and limit the size of their positions. Second, it is necessary to pay attention to the local maxima of the logarithms of the trading invariant for various stocks, a significant part of which falls on the last five-minute intervals of each calendar hour. This empirical result may be associated with the corresponding dynamics of financial news releases or the specifics of algorithmic strategies used in the Russian stock market.<sup>10</sup>

We now test hypothesis H1. In the first step, we apply basic specifications using high-frequency trading variables.

(1.1) Baseline model

$$\tilde{v}_{i\tau} = \beta_0 + \beta_1 1\{\text{Market turbulence}\}_d + e_{i\tau}; \tag{7}$$

(1.2) Baseline model with firm fixed effects

$$\tilde{v}_{i\tau} = \beta_0 + \beta_1 1\{\text{Market turbulence}\}_d + \alpha_i + e_{i\tau}; \tag{8}$$

<sup>8</sup> When we use the aggregation scheme (4), the universal distribution for 32 stocks is also not observed.

<sup>9</sup> As a robustness check, we also consider non-standardized values of the logarithm of the trading invariant and do not find significant differences in the results.

<sup>10</sup> Further study of the causes of the intraday dynamics of the bid-ask spread takes us beyond the scope of this paper.



## (1.3) Baseline model with firm and month fixed effects

$$\tilde{v}_{i\tau} = \beta_0 + \beta_1 1\{\text{Market turbulence}\}_d + \alpha_i + \delta_m + e_{i\tau}; \quad (9)$$

where  $1\{\text{Market turbulence}\}_d$  is a dummy variable that is equal to +1 for five most turbulent trading days in the Russian market over January 2014 – July 2018: March 3, 2014; December 15 and 16, 2014; April 6 and 9, 2018 and 0. Otherwise;  $\alpha_i$  is firm fixed effects;  $\delta_m$  is month fixed effects.

The results are presented in Panel A of Table 2. The hypothesis that there are no significant positive deviations of the standardized logarithm of the trade invariant for turbulent trading days is rejected for all three specifications.

As a robustness check, we turn to an extended set, which consists of five above-mentioned most turbulent days and seven additional days: April 17, 2015; January 15, 2016; March 10, 2015; September 1, 2015; December 8, 2014; January 26, 2015; December 2, 2014. These trading days are also characterized by significant negative changes in the RTS index (from – 6.2% to – 4%). The results of our tests of the same three basic specifications are presented in Panel B of Table 2. We do not find any significant differences with the previous results: the estimated coefficients on a dummy variable are positive and statistically significant but slightly lower in absolute value.

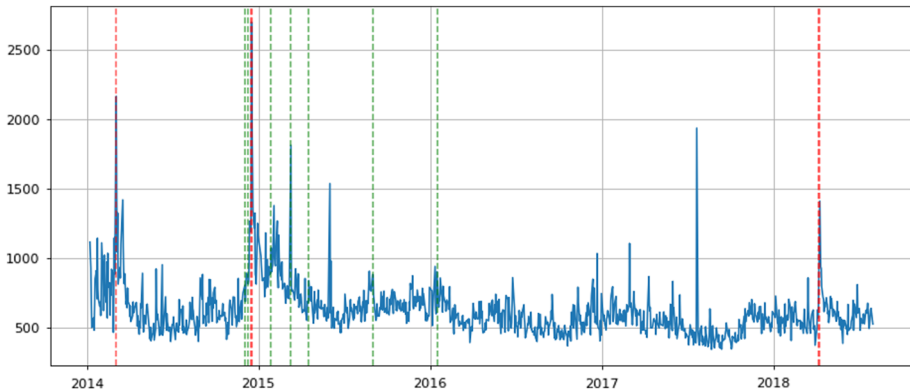
Adding fixed effects to the model does not increase the coefficient of determination. In addition, we test the hypothesis for non-standardized values of the logarithm of the trading invariant  $l_{i\tau}$  and also reveal the significance of the estimated coefficients on a dummy variable. We deliberately do not use scheme (3) to test these specifications at daily intervals. Since there were no trades at some five-minute intervals during the most turbulent trading days, some logarithms of the trading invariant are not defined. Thus, when averaging across the trading days, the previously detected differences in  $\tilde{v}_{i\tau}$  blur out. Instead, we concentrate on the top 5% of maximum values of the standardized trading invariant for each stock and each trading day ( $\tilde{v}_{i\tau}^{Top}$ ) to test the following hypothesis.

**Hypothesis 2** The market downturn is associated with extreme positive values of the standardized logarithm of the trading invariant at the daily level.

According to the results presented in Table 3, the coefficients on a dummy variable are positive and statistically significant. They are close to the estimates obtained earlier when testing specifications (1.1)–(1.3). Fixed effects do not lead to a considerable increase in the explained fraction of the variance of the dependent variable  $\tilde{v}_{it}$ .

We next test the hypothesis  $\beta = -2$  for the specification (5) when considering only those 5 or 12 days that were defined as turbulent. In both cases, this hypothesis is not rejected at the 5% significance level. We obtain the same result if the dummy variable  $1\{\text{Market turbulence}\}_d$  is added to the model (5): the coefficients on this variable are positive and statistically significant. We also apply unequal variances (Welch's) t-test for non-standardized values of the logarithm of the trading invariant. The first sample consists of observations belonging to trading days with a significant level of market drawdown (5 or 12), and the other observations belong to the second sample. According to the test results, the null hypothesis of equality of the mean values of the two samples is rejected in both cases (p-values < 0.01).

To summarize, we find that the nonlinear ITI relationship holds across 32 most liquid Russian stocks for the sample of trading days with the largest market decline. In other words, we conclude that the fundamental mechanism that determines the average trade size depending on changes in trading intensity does not change in the case of market turbulence. At the same time, we show that the  $I$  has not invariant distribution across different stocks.



**Fig. 11** The figure shows the dynamics of the average spread cost at five-minute frequency. It is defined as the average trade size computed at five-minute frequency averaged across stocks times the bid-ask spread computed at five-minute frequency for every stock and then averaged in the cross-section. The red dashed vertical lines indicate March 3, 2014; December 15 and 16, 2014; April 6 and 9, 2018. The green dashed vertical lines indicate December 2, 2014; December 8, 2014; January 26, 2015; March 10, 2015; April 17, 2015; September 1, 2015; January 15, 2016

Bucci et al. (2020) got similar results after investigating the ANcerno data on bets. The researchers found only “weak universality” when testing the invariance hypothesis. They demonstrated significant variations in the trading invariant  $I$  when examining some American stocks and futures contracts. Nonetheless, authors confirmed quantitative predictions about the relationship between trading variables.

Figure 11 demonstrates the average spread cost at five-minute frequency. It is defined as the average trade size computed at five-minute frequency averaged across stocks times the bid-ask spread computed at five-minute frequency for every stock and then averaged in the cross-section. The red dashed vertical lines indicate March 3, 2014; December 15 and 16, 2014; April 6 and 9, 2018. The green dashed lines represent seven days from the extended set of turbulent trading days defined earlier.

We can observe that the average spread cost had peak values during almost all crisis days. In our opinion, severe information asymmetry and high volatility of securities’ fundamentals were among the main causes of soaring average spread cost of trade and a sharp increase in the ruble risk transferred by one bet per unit of business time.

## 6 Trading activity and information during the period 2018–2021

### 6.1 Basic notation and estimation methodology

Similar to Kyle and Obizhaeva (2017b), we consider the extension of the invariance hypotheses about the relationship between microscopic and macroscopic market characteristics to the hypothesis about the information process. However, we apply this methodology to the Russian stock market. According to the Information Flow Invariance (IFI) hypothesis, public information is “expected to arrive at a rate proportional to the rate at which the business-time clock ticks, with a proportionality constant being the same across assets and across time.”

Quantitatively, the IFI hypothesis predicts that the rate of information flow  $\mu$  is proportional to  $W^\gamma$ , where  $W$  is trading activity, defined as the product of volatility and ruble volume,

and  $\gamma = 2/3$ . There is a simple intuition behind this coefficient. Suppose business-time clock related to professional investors' activity slows down to half speed for some reason. As a result, we start to observe two similar effects. Information flows are also retarded: providers release twice fewer news articles about firms. In addition, both ruble volume and variance go down by a factor of two. Finally, the trading activity decreases by a factor of  $2 \cdot 2^{1/2} = 2^{3/2}$ , and we obtain the following relationship between the new (marked with an asterisk) and old rates of trading activity and the arrival rates of information:

$$\mu = \mu^* \cdot \left( \frac{W}{W^*} \right)^{2/3}. \quad (10)$$

The similar principles apply in the case of retail investors' activity and Google relative search volumes of Russian stocks.

To study the time-series variation in the estimated exponent  $\gamma$  capturing contemporaneous relation between the trading and information processes, we implement negative binomial regressions:

$$\mu(W_{it}) = \exp\left(\eta + \gamma \cdot \ln\left(\frac{W_{it}}{W^*}\right)\right) \cdot \widetilde{G}_{it}(\alpha), \quad (11)$$

where  $\mu(W_{it})$  is either the number of news articles  $\mu(W_{it}^{News})$  or Google relative search volumes  $\mu(W_{it}^{Search})$  of stock  $i$  and week  $t$ ;  $\eta$  and  $W^*$  are constants corresponding to the average number of news articles (or Google relative search volumes) and the trading activity of some benchmark stock;  $\widetilde{G}_{it}(\alpha)$  is the Gamma variable with the mean of 1 and the variance of  $\alpha$ . The invariance theory predicts that  $\gamma = 2/3$ . We next discuss the advantages of using this count data regression.

## 6.2 Data

### 6.2.1 Google trends

Google Trends provided the weekly data on relative search volumes of Russian stocks  $\mu_{it}^{Search}$ . We analyze the relative popularity of search queries only within Russia during the period from the first week of August 2018 to the last week of June 2021. The search query has the following structure: "Name of a public company" + "stocks" written in Russian. The initial sample consists of only those 32 shares of Russian issuers included in the MOEX Russia Index during this period. We exclude relative search volumes of Aeroflot, MTS, and Magnit stocks since the search queries written in Russian have double meanings in these cases. Besides the meaning "stocks of the company," such queries can also be interpreted as "the company's special offers".<sup>11</sup>

It is worth noting that Google Trends imposes some limitations on downloading and analyzing data. Firstly, it is impossible to compare search volumes of more than 5 queries simultaneously. Secondly, the algorithm does not show absolute values. Instead, it divides all numbers by the sample maximum, multiplies by 100, and rounds off all of them to the nearest integer between 0 and 100. We implement the following procedure to obtain relative search volumes data. In the first step, we find a benchmark query with the highest maximum search volume at some week among all queries over August 2018–June 2021. In our case,

<sup>11</sup> Aeroflot is a Russian airline, MTS is a mobile network operator in Russia, and Magnit is Russia's food retailer.

the benchmark query is “Gazprom stocks” written in Russian. We next get data on the search volumes of all remaining words relative to the benchmark’s highest maximum. Ultimately, all relative search volumes are directly related to the benchmark’s maximum, which equals 100. It is worth mentioning that we do not get rid of zero relative search volumes using this algorithm. Kyle and Obizhaeva (2017b) note that both the negative binomial (NB) model and the Poisson model can be used to study the relationship between the information flow and the trading flow. However, unlike the Poisson model, the NB model takes into account possible over-dispersion of the news/queries data. The additional variation caused by many zero relative search volumes is the main reason why the negative binomial model is preferable in our case.

### 6.2.2 Professional news service Thomson Reuters Eikon as the source of information

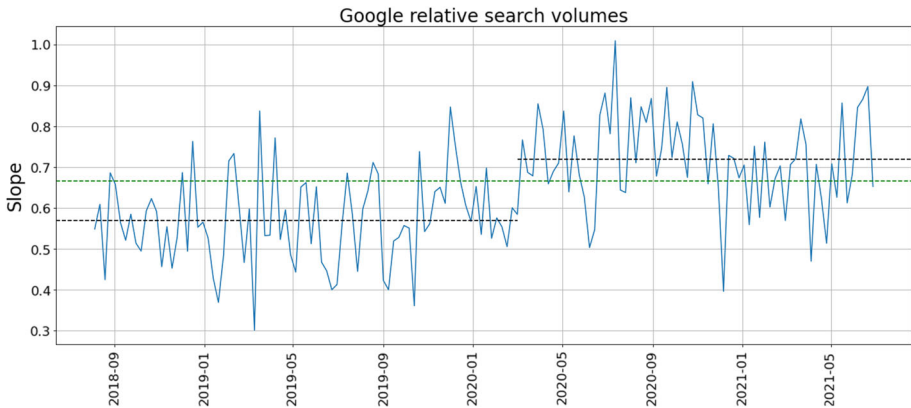
Thomson Reuters Eikon provided daily news data on 29 Russian stocks included in the Google relative search volumes sample. It also covers the period August 2018–June 2021. Each news article has information on the ticker of a firm, the time stamp, the headline, and the topic code. In the first step, we apply several filters. First, we include in the sample a news item if it is associated with at least one of the following topics: “Significant News”, “Significant Company News”, “Significant Economic News”, “Significant Equity News”, “Major News”, “Company News”. We also exclude all duplicated news articles.

Additionally, we use daily data on the open ( $O_{id}$ ), high ( $H_{id}$ ), low ( $L_{id}$ ), and closing ( $C_{id}$ ) prices and share volume ( $V_{id}$ ) from August 1, 2018 to June 30, 2021 provided by Thomson Reuters Eikon. The sample consists of 29 Russian stocks included in the Google relative search volumes dataset. For each day  $d$  and each stock  $i$ , we estimate market activity variables in the following way.  $P_{id}$  is calculated as the average of opening and closing prices of stock  $i$  and day  $d$ . We use Garman and Klass (1980) variance measure:  $\sigma_{id}^2 = \frac{1}{2}(\ln H_{id} - \ln L_{id})^2 - (2 \ln 2 - 2) \cdot (\ln C_{id} - \ln O_{id})^2$ . We next calculate trading activity  $W_{id}$  for each stock  $i$  and day  $d$  as the product of  $P_{id}$ ,  $V_{id}$ , and  $\sigma_{id}$  and convert daily data on trading activity and the number of news articles to weekly data:  $W_{id} \rightarrow W_{it}$  and  $\mu_{id}^{News} \rightarrow \mu_{it}^{News}$ . Finally, we sum weekly trading activity of ordinary and preferred stocks of the three companies (Sberbank, Tatneft, and Surgutneftegas) for each week from the sample and match this data with corresponding weekly data on the news articles and Google relative search volumes. The same procedure is implemented for the remaining companies in the sample. Table 4 presents summary statistics for the trading activity, Google relative search volumes, and news articles provided by the Thomson Reuters Eikon.

### 6.3 Empirical evidence

In this subsection, we perform the empirical tests by estimating the coefficients  $\gamma$  for the negative binomial model and testing whether it was an abrupt shift in the estimates at the beginning of 2020.

We use the CUSUM method to detect a single changepoint in the time-series data. According to the results of the OLS-based CUSUM test, the null hypothesis that there is no mean shift is rejected ( $p$ -value =  $8.6 \cdot 10^{-11}$ ), and the estimate of a changepoint is the last week of February 2020. Figure 12 shows the estimates of the slope  $\gamma$  for the NB regressions with  $\mu(W_{it}) = \mu(W_{it}^{Search})$  run separately for each week from the first week starting with the first week of August 2018 and ending with the last week of June 2021. Figure 12 shows the time-series variation in weekly estimates of parameters  $\gamma$  in the case of using the Google



**Fig. 12** The figure shows the estimates of the slope  $\gamma$  from the negative binomial regression, with the arrival of Google relative search volumes  $\mu\left(W_{it}^{Search}\right)$  for stock  $i$  and week  $t$  being modeled as,

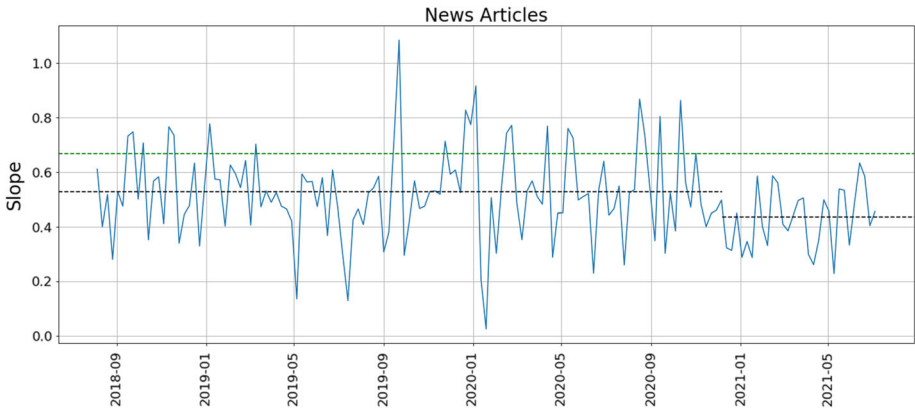
relative search volumes data. Indeed, we can observe that there is a changepoint in means around February–March 2020. The black dashed lines indicated the averages before and after the changepoint: 0.570 and 0.720, respectively. The green dashed line shows the value of  $\gamma = 2/3$ , predicted by the invariance hypothesis. It is worth noting that the average coefficient  $\gamma$  after the changepoint is closer to  $2/3$  than the average coefficient  $\gamma$  before the abrupt shift. Thus, we conclude that trading activity and the information flow approximated by Google relative search volumes became more synchronized.

We replace the Google relative search volumes data with the news articles data and repeat our analysis. In the case of news queries, the estimate of a changepoint is the last week of October 2020. However, we cannot reject the null hypothesis that there is no abrupt shift because  $p$ -value = 0.067. Moreover, according to our results, the average value of  $\gamma$  before the last week of October 2020 was closer to the theoretical value of  $2/3$  than the average value of  $\gamma$  after this week: 0.527 and 0.437, respectively. Figure 13 shows the time-series variation in weekly estimates of parameters  $\gamma$  in the case of using the news articles data. The black dashed lines indicated the averages before and after the last week of October 2020. The green dashed line shows the value of  $\gamma = 2/3$ , predicted by the invariance hypothesis.

$$\mu\left(W_{it}^{Search}\right) = \exp\left(\eta + \gamma \cdot \ln\left(\frac{W_{it}}{W^*}\right)\right) \cdot \widetilde{G}_{it}(\alpha),$$

where  $W_{it}$  is the trading activity, the product of weekly volatility and weekly ruble volume. The  $\eta$  and  $W^*$  are constants corresponding to the average number of Google relative search volumes and the trading activity of some benchmark stock. The estimates are plotted for each of 153 weeks starting with the first week of August 2018 and ending with the last week of June 2021. The black dashed lines indicated the averages before and after the changepoint (the last week of February 2020): 0.567 and 0.720, respectively. The green dashed line shows the value of  $\gamma = 2/3$ , predicted by the invariance hypothesis. The benchmark query is “Gazprom stocks” written in Russian.

We should mention that February 2020 can be considered as the beginning of the period of a significant inflow of private investors’ funds in the Russian stock market. According to the Moscow Exchange statistics, retail investors bought Russian shares listed on the MOEX in the amount of 40.2 billion rubles (around 629 million dollars according to the average

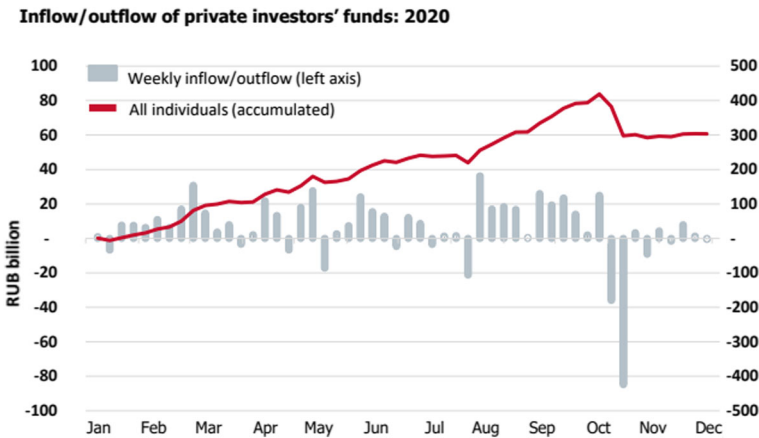


**Fig. 13** The figure shows the estimates of the slope  $\gamma$  from the negative binomial regression, with the arrival of news articles  $\mu(W_{it}^{News})$  for stock  $i$  and week  $t$  being modeled as,

U.S. dollar/Russian ruble exchange rate in February 2020). Moreover, it was the last week of February 2020 that was characterized by the most massive inflow totaling 16.7 billion rubles. Figure 14 shows the dynamics of inflow/outflow of retail investors’ funds during the period January 2020—December 2020.

$$\mu(W_{it}^{News}) = \exp\left(\eta + \gamma \cdot \ln\left(\frac{W_{it}}{W^*}\right)\right) \cdot \widetilde{G}_{it}(\alpha),$$

where  $W_{it}$  is the trading activity, the product of weekly volatility and weekly ruble volume. The  $\eta$  and  $W^*$  are constants corresponding to the average number of Google relative search volumes and the trading activity of some benchmark stock. The estimates are plotted for each of 153 weeks starting with the first week of August 2018 and ending with the last week of June 2021. The black dashed lines indicated the averages before and after the last week



**Fig. 14** The figure shows the dynamics of weekly inflow/outflow of private investors’ funds in 2020. The source is the official website of the Moscow Exchange: <https://www.moex.com/>

of October 2020: 0.516 and 0.433, respectively. The green dashed line shows the value of  $\gamma = 2/3$ , predicted by the invariance hypothesis.

Thus, it is reasonable to believe that it was the drastic increase in retail investor participation in the Russian stock market that was the reason why the trading flow and information flow approximated by Google relative search volumes better conformed to the same business-time clock  $R$ . In addition, we do not find any significant differences in the relation between the trading and information processes in the case of using the news articles data. Moreover, the tests show that the level of synchronization between these two flows is somewhat lower in 2021 compared to the 2018–2020 period.

## 7 Conclusion

In the first part of our paper, we tested the two specifications of the mixture-of-distributions-hypothesis (MDH in Volume and MDH in Transactions) and the Intraday Trading Invariance hypothesis. It should be noted that the original market microstructure hypothesis is formulated for metaorders that are usually divided by traders into many smaller orders to minimize trading costs. However, we found that the invariance principles explain much of the endogenous variation between trading variables over short time intervals. The centralized structure of the Russian stock market made it possible to correctly identify the size of transactions, a key variable in our empirical tests. After applying different techniques to mitigate the impact of errors-in-variables problems and performing several robustness checks, we showed that log-return variation per transaction is proportional to the  $-2$  power of the average trade size times the stock price. At the same time, we did not find confirmation of the invariance hypothesis for individual stocks and associate this result with significant noise in the regressors. We demonstrated that the nonlinear ITI relationship holds when considering trading days with maximum market downturns during the period January 2014–July 2018. However, we found that these days were characterized by a statistically significant increase in the trading invariant, the ruble risk transferred by one bet per unit of business time. In our opinion, the major causes were the sharp increase in the level of information asymmetry and the high volatility of securities' fundamentals since the average spread cost was soaring during these crashes. Our study also briefly touched upon the intraday dynamics of the trading invariant. We found a U-shaped pattern and drew attention to the presence of individual peaks of trading intensity at the moments of the last five-minute intervals of each hour. At the same time, we did not find the universal trading invariant for all stocks, one of the major assumptions of the invariance hypothesis. A more detailed study of the key factors of trading activity at high-frequency intervals and the use of more sophisticated econometric methodology for testing the log-linear regression specifications are among the main directions for future work.

The exploration of the relationship between trading variables over short intervals was based on the Market Microstructure Invariance hypothesis extrapolated to the intraday dimension. At the same time, to study whether the positive shock to retail investors' demand for equities changed the relationship between the trading and information processes in the Russian stock market, we used another extension of invariance principles: the Information Flow Invariance hypothesis. This invariance-implied methodology allowed us to zoom in on one of the most notable episodes in the Russian stock market over the August 2018–June 2021 period: the considerable increase in individual investor participation since the beginning of 2020. In accordance with our empirical tests, the information flow approximated by the Google search activity of Russian stocks became more synchronized with market trading activity. We showed



that a changepoint in the dynamics of estimates characterizing this relationship occurred in the last week of February 2020, when the Russian stock market faced a huge inflow of retail investors' funds. At the same time, we did not find a statistically significant change in the relationship between trading activity and news activity. To summarize, the level of synchronization between those two flows was quite stable through time but lower than the corresponding level in the case of Google search activity. In the future, we would like to examine the alternative sources of public information in the context of the Information Flow Invariance (e.g., tweets, Telegram messages, financial analytics).

**Funding** The article was prepared within the framework of the Basic Research program at HSE University.

**Data availability** The data that support the findings of this study are available from the Moscow Exchange but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. This data are however available from the authors upon reasonable request and with permission of the Moscow Exchange. Upon acceptance, news and transaction-level data provided by Thomson Reuters Eikon and Google search volumes data will be available on the official website of the Centre for Financial Research & Data Analytics.

**Code availability** Not applicable.

## Declarations

**Conflict of interest** The authors report no declarations of interest.

## Appendix

### Tables

See Table 1.

**Table 1** Statistics for trading variables aggregated by five-minute intervals from January 6, 2014 to July 31, 2018

Code	Cumulative volume ( $V \cdot 10^3$ )	Variance ( $\sigma^2 \cdot 10^6$ )	Transaction rate ( $\Gamma$ )	Stocks/transaction ( $Q$ )
HYDR	84,505 (527,243)	1.6 (10.2)	503 (1241)	141,783 (296,947)
NVTK	251 (642)	1.3 (7.2)	446 (728)	500 (665)
MFON	92 (471)	2.7 (43.0)	261 (647)	233 (687)
TATNP	52 (205)	2.4 (12.5)	127 (214)	383 (861)
ROSN	889 (1873)	0.9 (3.4)	605 (599)	1340 (1403)
MOEX	1506 (5059)	1.3 (9.2)	952 (1,983)	1481 (2043)
MTSS	484 (3793)	1.6 (10.9)	483 (589)	819 (1154)
RTKM	397 (1221)	1.2 (9.8)	329 (449)	1046 (1776)
SNGSP	5656 (16,546)	1.1 (6.7)	596 (788)	7671 (10,806)
IRAO	36,857 (1062,472)	1.8 (10.8)	339 (445)	59,047 (460,279)
NLMK	577 (2221)	1.2 (5.5)	444 (823)	1177 (2274)
LKOH	182 (399)	0.8 (3.9)	1050 (1655)	171 (166)

**Table 1** (continued)

Code	Cumulative volume ( $V \cdot 10^3$ )	Variance ( $\sigma^2 \cdot 10^6$ )	Transaction rate ( $\Gamma$ )	Stocks/transaction ( $Q$ )
AFLT	636 (1696)	2.4 (44.9)	264 (302)	2011 (2586)
PHOR	5 (15)	4.0 (31.6)	129 (174)	35 (78)
TRNFP	1 (5)	2.1 (13.5)	431 (1667)	3 (7)
POLY	56 (168)	3.8 (40.3)	136 (180)	318 (550)
PIKK	50 (228)	5.8 (37.3)	148 (235)	369 (2001)
GAZP	6408 (17,462)	0.6 (2.5)	1423 (1534)	3871 (5811)
TATN	238 (489)	1.2 (4.8)	392 (548)	559 (520)
MVID	44 (147)	(3.7) (39.7)	149 (230)	263 (1820)
GMKN	33 (90)	0.9 (5.2)	630 (893)	46 (59)
CHMF	189 (673)	1.1 (4.5)	402 (1742)	396 (424)
SBERP	1594 (6033)	(1.0) (5.1)	501 (723)	2739 (4740)
VTBR	3,734,005 (11,457,901)	1.2 (6.0)	874 (1167)	3,419,491 (4,222,240)
FEES	413,823 (1107,864)	2.0 (11.6)	522 (662)	626,922 (853,292)
ALRS	1428 (4820)	1.7 (10.1)	426 (649)	2929 (5561)
MGNT	32 (105)	1.3 (9.0)	794 (1133)	35 (64)
SBER	17,030 (61,070)	1.1 (8.1)	2265 (2738)	5874 (11,207)
MAGN	2320 (6856)	1.6 (8.2)	315 (501)	6660 (10,331)
SNGS	4443 (10,818)	1.0 (4.9)	544 (735)	7272 (7787)
AFKS	2850 (22,303)	6.6 (112.9)	363 (625)	5326 (10,121)
EONR	7891 (76,534)	1.9 (12.4)	172 (531)	33,981 (69,079)

The table contains statistics for the mean values and standard deviations (in parentheses) of trading variables, aggregated by five-minute intervals from January 6, 2014 to July 31, 2018. The cumulative trading volume ( $V$ ), the transaction rate (trades per unit time) ( $\Gamma$ ), and the average number of stocks per transaction ( $Q$ ) are computed across each five-minute interval and then averaged across all observations. The variance measure ( $\sigma^2$ ) represents realized variance and is computed from five one-minute squared log-returns across each five-minute interval. Note: On July 1, 2016 the ticker EONR changed to UPRO.

See Table 2.

Driscoll-Kraay standard errors in parentheses. In Panel A (B), the results are conditional on the number of trading days considered as turbulent: 5 (12). The sample period is from January 6, 2014 to July 31, 2018.

See Table 3.

Driscoll-Kraay standard errors in parentheses. In Panel A (B), the results are conditional on the number of trading days considered as turbulent: 5 (12). The sample period is from January 6, 2014 to July 31, 2018.

See Table 4.

**Table 2** Testing basic specifications (1.1)–(1.3) using five-minute trading variables with fixed effects. The dependent variable is the standardized logarithm of the trading invariant  $\tilde{\iota}_{it}$ 

	(1.1)	(1.2)	(1.3)
<i>Panel A: 5 turbulent trading days</i>			
Intercept	– 0.003 (0.003)	– 0.003 (0.003)	– 0.002 (0.002)
Dummy for market turbulence	0.474 (0.051)	0.474 (0.051)	0.326 (0.044)
Firm fixed effects	No	Yes	Yes
Month fixed effects	No	No	Yes
R-squared	0.0016	0.0016	0.0007
N <sup>o</sup> of obs	1,279,833	1,279,833	1,279,833
<i>Panel B: 12 turbulent trading days</i>			
Intercept	– 0.005 (0.003)	– 0.005 (0.003)	– 0.003 (0.002)
Dummy for market turbulence	0.350 (0.034)	0.350 (0.034)	0.198 (0.031)
Firm fixed effects	No	Yes	Yes
Month fixed effects	No	No	Yes
R-squared	0.0018	0.0018	0.0005
N <sup>o</sup> of obs	1,279,833	1,279,833	1,279,833

**Table 3** Testing basic specifications (1.1)–(1.3) using five-minute trading variables with fixed effects. The dependent variable is the standardized logarithm of the trading invariant $\tilde{\iota}_{it}^{Top}$ 

	(2.1)	(2.2)	(2.3)
<i>Panel A: 5 turbulent trading days</i>			
Intercept	1.605 (0.003)	1.605 (0.003)	1.606 (0.002)
Dummy for market turbulence	0.649 (0.066)	0.654 (0.068)	0.474 (0.067)
Firm fixed effects	No	Yes	Yes
Month fixed effects	No	No	Yes
R-squared	0.0098	0.0103	0.0051
N <sup>o</sup> of obs	79,862	79,862	79,862
<i>Panel B: 12 turbulent trading days</i>			
Intercept	1.603 (0.003)	1.603 (0.003)	1.606 (0.003)
Dummy for market turbulence	0.432 (0.050)	0.432 (0.052)	0.253 (0.050)
Firm fixed effects	No	Yes	Yes
Month fixed effects	No	No	Yes
R-squared	0.0091	0.0093	0.0030
N <sup>o</sup> of obs	79,862	79,862	79,862

The table contains statistics for the mean values and standard deviation (in parentheses) of the trading activity, Google relative search volumes, and news articles provided by the Thomson Reuters Eikon, aggregated by weekly intervals from the first week of August 2018 to the last week of June 2021. The benchmark query is “Gazprom stocks” written in Russian (Gazprom ordinary stock’s ticker is GAZP). For Tatneft (stocks’ tickers: TATN and TATNP), Surgutneftegaz (stocks’ tickers: SGNS and SNGSP), and Sberbank (stocks’ tickers: SBER and SBERP), we sum weekly trading activity of ordinary and preferred stocks.

**Table 4** Statistics for the trading activity, Google relative search volumes, and news articles provided by the Tomson Reuters Eikon

Code	Trading activity ( $W \cdot 10^6$ )	News articles ( $\mu^{News}$ )	Google relative search volumes ( $\mu^{Search}$ )
LKOH	1,163.3 (1,496.1)	9.6 (8.5)	9.7 (5.5)
ALRS	293.5 (303.7)	4.3 (4.4)	6.4 (3.2)
POLY	279.5 (399.3)	3.7 (4.9)	2.7 (4.1)
SBER + SBERP	2,654.8 (2,820.3)	13.1 (14.5)	15.4 (6.2)
IRAO	201.3 (240.7)	1.7 (3.1)	2.1 (2.1)
GMKN	1,070 (1,634.2)	3.8 (5.2)	8.1 (7.1)
CBOM	27.3 (32.2)	0.6 (1.2)	2.4 (1.8)
PIKK	15.7 (27.8)	0.4 (0.9)	0.6 (1.0)
PLZL	384.9 (531.7)	2.3 (3.6)	4.5 (3.4)
MAGN	184.6 (244.3)	1.3 (2.6)	3.2 (2.8)
NLMK	198.9 (206.2)	2.4 (4.1)	3.6 (2.4)
PHOR	36.1 (43.3)	2.3 (4.4)	0.9 (1.1)
VTBR	284.7 (381.5)	5.8 (5.3)	0.0 (0.0)
ROSN	552.4 (879.5)	15.2 (12.8)	8.1 (5.3)
MOEX	208.6 (221.5)	11.0 (10.7)	2.3 (1.9)
TRNFP	52.2 (62.7)	8.1 (8.8)	0.9 (1.1)
HYDR	132.9 (310.5)	1.5 (3.2)	2.9 (2.8)
GAZP	1,623.6 (1,943.3)	27.9 (24.3)	34.2 (15.0)
SNGS + SNGSP	624.5 (1,284.9)	5.1 (4.2)	6.8 (4.3)
AFKS	200.9 (324.4)	2.7 (3.9)	1.9 (1.7)
CHMF	211.8 (237.2)	2.3 (4.5)	5.5 (3.5)
FEES	65.6 (108.8)	0.3 (0.9)	0.1 (0.3)
TATN + TATNP	530.4 (700.6)	3.8 (4.8)	8.8 (5.4)
DSKY	70.4 (183.3)	1.7 (3.4)	0.0 (0.0)
RTKM	37.0 (43.3)	3.0 (4.2)	1.8 (1.6)
NVTK	320.5 (405.0)	6.3 (8.0)	2.0 (1.6)

### Details on the regression-based framework

The *Intraday Trading Invariance hypothesis* is formulated as follows: random variables  $\tilde{I}_{i\tau}$  defined as

$$\tilde{I}_{i\tau} = \frac{\tilde{P}_{i\tau}\tilde{Q}_{i\tau}\tilde{\sigma}_{i\tau}}{\Gamma_{i\tau}^{1/2}} \tag{12}$$

are independent and identically distributed for all stocks  $i$  and intervals  $\tau$ . Kyle et al. (2016) discuss the economic meaning of this variable.  $\tilde{I}$  is the monetary (e.g., dollar or ruble) risk transferred by one bet per unit of business time. The assumption about the approximate constancy of this value across assets and across time is based on the no-arbitrage condition. The authors provide the following illustrative example. Let large-cap stocks initially give more opportunities for profit than small-cap stocks. As a result, more informed traders start to take long positions in large-cap stocks. Consequently, the expected arrival of bets (the reciprocal of business time) for large-cap stocks rises. Moreover, the distance between trading prices and unobservable fundamental values tend to shrink for such stocks, and traders endogenously adjust their trade sizes to compensate for their trading costs. As a result, the monetary value of the transferred risk per unit of business time remains unchanged in equilibrium.

We can turn to an alternative formulation of Eq. (5) if we take the logarithms of both sides of the equation and then take the conditional expectations:

$$E_{\tau-1}\{\ln \tilde{I}_{i\tau}\} = p_{i\tau} + q_{i\tau} + \frac{1}{2}s_{i\tau} - \frac{1}{2}\gamma_{i\tau} = c \text{ for } \tau = 1, \dots, D \cdot T; i \in I. \tag{13}$$

Taking into account the presence of an error term, we get the following log-linear relationship between trading variables.

Intraday Trading Invariance:

$$s_{i\tau} - \gamma_{i\tau} + 2p_{i\tau} = c - 2q_{i\tau} + e_{i\tau} \text{ for } \tau = 1, \dots, D \cdot T; i \in I \tag{14}$$

where  $c$  is a constant and  $e_{i\tau}$  are residuals with zero expectation.

The inclusion of the stock price level  $p_{i\tau}$  is primarily due to the differences in stock prices in the cross-section.<sup>12</sup>

### The mixture-of-distributions-hypothesis

As basic models, we consider the following two log-linear regression specifications. We use the notations and assumptions described earlier in this section.

1. The Mixture-of-Distributions-Hypothesis in Transactions (MDH-N)

$$s_{i\tau} = c + \gamma_{i\tau} + e_{i\tau} \text{ for } \tau = 1, \dots, D \cdot T; i \in I, \tag{15}$$

where  $c$  is a constant and  $e_{i\tau}$  are residuals with zero expectation.

2. The Mixture-of-Distributions-Hypothesis in Volume (MDH-V)

$$s_{i\tau} = c + v_{i\tau} + e_{i\tau} \text{ for } \tau = 1, \dots, D \cdot T; i \in I, \tag{16}$$

where  $c$  is a constant and  $e_{i\tau}$  are residuals with zero expectation.

Let us generalize formulas (15) and (16), taking into account that  $v_{i\tau} = q_{i\tau} + \gamma_{i\tau}$ .

<sup>12</sup> Analyzing trading patterns in the E-mini S&P 500 futures data, Andersen et al. (2020) argue that variations in the price level are negligible and can be considered constant at minute time intervals since the fluctuations in other variables (volatility and trading intensity) are significantly higher.

### 3. The Mixture-of-Distributions-Hypotheses

$$s_{i\tau} - \gamma_{i\tau} = c + \beta q_{i\tau} + e_{i\tau} \text{ for } \tau = 1, \dots, D \cdot T; i \in I, \quad (17)$$

where  $c$  is a constant and  $e_{i\tau}$  are residuals with zero expectation.

Thus, when  $\beta = 0$ , Eqs. (15) and (17) become equivalent to each other; when  $\beta = 1$ , Eqs. (16) and (17) are identical. These assumptions contrast sharply with the Intraday Trading Invariance hypothesis predicting  $\beta = -2$  (with the inclusion of the term  $2p_{i\tau}$ ). It is also necessary to mention that the analysis of log-linear models is possible over intervals with positive trading activity. Therefore, we have to make the following adjustments to the original hypothesis.

*Modified Intraday Trading Invariance*  $\tilde{I}_{i\tau}$  are independent and identically distributed for all stocks  $i$  and those intervals  $\tau$  that are characterized by non-zero trading activity.

## References

- Andersen, T. G., Bondarenko, O., Kyle, A. S., & Obizhaeva, A. A. (2020). Intraday trading invariance in the E-mini S&P 500 futures market. (NES working paper № 272). New Economic School (NES). <https://www.nes.ru/files/Preprints-resh/WP272.pdf>
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579–625. <https://doi.org/10.1111/1468-0262.00418>
- Ané, T., & Geman, H. (2000). Order flow, transaction clock, and normality of asset returns. *The Journal of Finance*, 55(5), 2259–2284. <https://doi.org/10.1111/0022-1082.00286>
- Bachelier, L. (1900). *Théorie de la spéculation*. [PhD thesis, University of Paris]. Numdam. <https://doi.org/10.24033/asens.476>
- Bae, K.-h., Kyle, A., Lee, E. J., & Obizhaeva, A. A. (2020). Invariance of buy-sell switching points. (NES working paper № 273). New Economic School (NES). <https://www.nes.ru/files/Preprints-resh/WP273.pdf>
- Berry, T. D., & Howe, K. M. (1994). Public information arrival. *Journal of Finance*, 49(4), 1331–1346. <https://doi.org/10.2307/2329188>
- Bucci, F., Lillo, F., Bouchard, J.-P., & Benzaquen, M. (2020). Are trading invariants really invariant? Trading costs matter. *Quantitative Finance*, 20(7), 1059–1068. <https://doi.org/10.1080/14697688.2020.1741667>
- Cascino, S., Clatworthy, M., Osma, B. G., Gassen, J., Imam, S., & Jeanjean, T. (2014). Who uses financial reports and for what purpose? Evidence from capital providers. *Accounting in Europe*, 11(2), 185–209. <https://doi.org/10.1080/17449480.2014.940355>
- Chan, K., Chung, Y. P., & Johnson, H. (1995). The intraday behavior of bid-ask spreads for NYSE stocks and CBOE options. *The Journal of Financial and Quantitative Analysis*, 30(3), 329–346. <https://doi.org/10.2307/2331344>
- Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1), 135–155. <https://doi.org/10.2307/1913889>
- Crouch, R. L. (1970). The volume of transactions and price changes on the new york stock exchange. *Financial Analysts Journal*, 26(4), 104–109. <https://doi.org/10.2469/faj.v26.n4.104>
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *The Journal of Finance*, 66(5), 1461–1499. <https://doi.org/10.1111/j.1540-6261.2011.01679.x>
- Engle, R. (2002). New frontiers for arch models. *Journal of Applied Econometrics*, 17(5), 425–446. <https://doi.org/10.1002/jae.683>
- Epps, T. W., & Epps, M. L. (1976). The stochastic dependence of security price changes and transaction volumes: implications for the mixture-of-distributions hypothesis. *Econometrica*, 44(2), 305–321. <https://doi.org/10.2307/1912726>
- Frazzini, A., Israel, R., & Moskowitz, T. J. (2018). Trading costs. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3229719>
- Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *The Journal of Business*, 53(1), 67–78. <https://doi.org/10.1086/296072>
- Heston, S.L. & Sinha, N.R. (2017). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, 73(3), 67–83. <https://doi.org/10.2469/faj.v73.n3.3>

- Jain, P. C., & Joh, G.-H. (1988). The dependence between hourly prices and trading volume. *Journal of Financial and Quantitative Analysis*, 23(3), 269–283. <https://doi.org/10.2307/2331067>
- Jones, C. M., Kaul, G., & Lipson, M. L. (1994). Transactions, volume, and volatility. *Review of Financial Studies*, 7(4), 631–651. <https://doi.org/10.1093/rfs/7.4.631>
- Karpoff, J. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22(1), 109–126. <https://doi.org/10.2307/2330874>
- Kyle, A. S., & Obizhaeva, A. A. (2017b). News articles and equity trading. (NES working paper №233). New Economic School (NES). <https://www.nes.ru/files/Preprints-resh/WP233.pdf>
- Kyle, A. S., & Obizhaeva, A. A. (2016). Market Microstructure Invariance: Empirical hypotheses. *Econometrica*, 84(4), 1345–1404. <https://doi.org/10.3982/ECTA10486>
- Kyle, A. S., & Obizhaeva, A. A. (2017a). Dimensional analysis and market microstructure invariance. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2785559>
- Kyle, A. S., & Obizhaeva, A. A. (2019). Large bets and stock market crashes. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2023776>
- Kyle, A., Obizhaeva, A. A., & Kritzman, M. (2016). A Practitioner’s guide to market microstructure invariance. *The Journal of Portfolio Management*, 43(1), 43–53. <https://doi.org/10.3905/jpm.2016.43.1.043>
- Madhavan, A., Richardson, M., & Roomans, M. (1997). Why do security prices change? A transaction-level analysis of NYSE stocks. *Review of Financial Studies*, 10(4), 1035–1064. <https://doi.org/10.1093/rfs/10.4.1035>
- Mitchell, M. L., & Mulgerin, J. H. (1994). The impact of Public Information on the Stock Market. *The Journal of Finance*, 49(3), 923–950. <https://doi.org/10.1111/j.1540-6261.1994.tb00083.x>
- Obizhaeva, A. A. (2012) Liquidity Estimates and Selection Bias. (CEFIR/NES Working Paper №225). Centre for Economic and Financial Research at New Economic School. <https://www.nes.ru/files/Preprints-resh/WP225.pdf>
- Obizhaeva, A. A. (2016). The Russian ruble crisis of December 2014. *Voprosy Ekonomiki*, 5, 66–86. <https://doi.org/10.32609/0042-8736-2016-5-66-86>
- Osborne, M. F. M. (1962). Periodic structure in the Brownian motion of stock prices. *Operations Research*, 10(3), 345–379. <https://doi.org/10.1287/opre.10.3.345>
- Tauchen, G., & Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*, 51(2), 485–505. <https://doi.org/10.2307/1912002>
- Tetlock, P. C. (2007). Giving content to investor sentiment: the role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>
- Tuzun, T. (2013). Are leveraged and inverse ETFs the New Portfolio insurers? (Finance and Economics Discussion Series 2013–48). Board of Governors of the Federal Reserve System (U.S.). <https://www.federalreserve.gov/pubs/feds/2013/201348/201348abs.html>
- Westerfield, R. (1977). The distribution of common stock price changes: An application of transactions time and subordinated stochastic models. *Journal of Financial and Quantitative Analysis*, 12(5), 743–765. <https://doi.org/10.2307/2330254>
- Wood, R. A., McInish, T. H., & Ord, J. K. (1985). An investigation of transactions data for NYSE stocks. *The Journal of Finance*, 40(3), 723–739. <https://doi.org/10.1111/j.1540-6261.1985.tb04996.x>
- Wurgler, J., & Zhuravskaya, E. (2002). Does arbitrage flatten demand curves for stocks? *The Journal of Business*, 75(4), 583–608. <https://doi.org/10.1086/341636>