N.V. Tlekhugov,
K.F. Kush,
A.I. Stolyarov
Higher School of Economics

METHODS FOR DETECTION OF NON-TYPICAL TRANSACTIONS. APPLICATION FOR RUSSIAN STOCK MARKET

Introduction

Russian stock market has quite a long history from its establishment in 90th of XX century till nowadays. Current operational and legal infrastructure of the financial market in Russia has [so far] been established as a complete system. And the market itself has become an essential part of the Russian economy.

Despite all modern developments, high market liquidity and securities turnover level, especially during the pre-crisis period of the year 2008, regulatory and controlling infrastructure has not developed sufficiently to be considered mature and complete. Allegedly, the most widespread market abuses in Russian stock market are illegal insider trading and market manipulation. It worth noting that the legal framework for that kind of control and monitoring activities on the market was not established till the end of 2010 when the federal act on prohibition of insider trading and markets manipulation was adopted (Federal Law № 224 – FL 224).

The FL № 224 introduced the notion of «insider», «insider (material) information» and declared norms and regulations for prevention of illegal activities on the financial market. It is obvious that it will take a while to fully incorporate all the innovations into the market and make the majority of the Russian stock market participants adopt «prudent» behavior. And it will definitely be accompanied by the development of comprehensive control and monitoring infrastructure, a bulk of government information letters and intra-firm bylaws concerning insider trading, material information, compliance guidelines etc. The process will be a little chaotic and step-like along with the information and experience accumulation within the market and by its participants.

The above mentioned regulatory, control and compliance documents can be viewed as a «cover» (or «shell») for its «contents» – the effective anti-fraud infrastructure in the market. Here one can refer to the anti-fraud infrastructure as specific software and hardware-based monitoring solutions for exchanges and for the regulator (Federal Department of Financial Markets, Russian stock market regulator), proper disclosure system, dedicated staff within market participants etc.

Therefore the development of an efficient control and monitoring system supported by thorough research work is highly topical. And one of the most important components of such a system is market abuse detection methods. The rest of the article includes the discussion of some aspects of this «ideal» monitoring system, a brief summary of the existing literature on the subject and the indicative example of one newly developed math approach.

Preliminary comments

Every monitoring system consists of computational (numerical) analytical methods, visual analysis and implementation (operational) blocks. This article considers numerical procedures only, but the remaining parts of the system are also of high importance and have to be thoroughly investigated. The motivation for numerical methods is straightforward – they are the most relevant field for scientific research.

First of all, let us define the subject of the paper – non-typical transactions. The preliminary definition for the purpose of the papers is: «the exact transactions that were implemented on the market or were planned for the purpose which does not agree with the common sense and market wisdom or the implementation is not legal under the legislation applied». The definition covers not only market abuse cases but also those transactions that are «technical» mistakes, e.g. wrong order implementation. At the same time transactions which are the result of wrong information interpretation should be treated as «typical».

The definition should be viewed in association with generally accepted market abuse classification because every «non-typical transaction» is the result of some kind of abusive behavior on financial market.

Comments to fig. 1:

• Information-based manipulation is spreading false data and rumors about a company;

• Action-based manipulation is non-trading covert actions for the purpose of self-enrichment of a manipulator, e.g. felonious bankruptcy etc.;

- Trade-based manipulation is a wide-range of trading strategies aimed at illegal profiting from misleading other market participants.
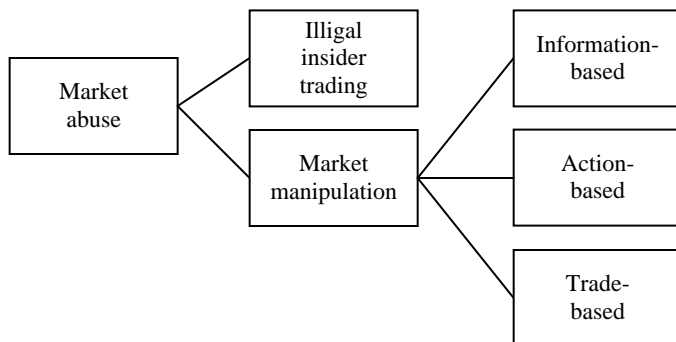

Fig. 1. Market abuse classification

As can be seen from fig. 1 a non-typical transactions phenomenon is the result of illegal insider trades, trade-based market manipulation and technical mistakes occurring on the market. This view clarifies the notion and distinguishes it from [the] broad definition of «market abusive transactions». Thus any algorithmic methodology for detection of market abusive behavior eventually comes to a method for detecting illegal insider trades and trade-based manipulative actions of market participants.

So far we have defined the notion of non-typical transactions in contrast to «typical» ones. In reality every transaction can be attributed to either class by studying some of its parameters. So the detection procedure should clearly specify the list of these parameters, a computational or other suitable procedure to obtain them, a way to generate signals, and a method for interpretation of these signals.

Existing literature survey

Insider trading phenomenon is well studied in existing empirical and theoretical research literature, and, surprisingly, it is not the case for market manipulation phenomenon.

There is a well-known debate about the consequences of insider trading for the market. Some say that it augments market efficiency by quickly introducing new information to the market, while others treat it as a serious abuse and claim that it must be prohibited in all forms. The solution for the debate is not so obvious for developed markets, like in USA or Europe. For an emerging market in Russia where there is

no sufficient market infrastructure and information allocation is highly asymmetrical the attitude towards insider trading is surely negative. As opposed to insider trading phenomenon there is no doubt about the harm of manipulators to the market: it is said to undermine the «fairness» of the market.

There are several reasons why for the purpose of construction of the detection procedure one should focus on research papers that analyze manipulative behavior. Firstly, in essence the insider behavior is covert and therefore difficult to detect. Manipulation strategies can be profitable if a lot of participants are engaged and so they can be observable.

Further, some research papers with a focus on detection of market manipulation are discussed.

There are 3 broad groups of detection methods.

The first group. Simple detection procedure that considers raw market data. The signal is generated when some predetermined indicator or coefficient deviates significantly from its «band» or authorized interval. This method is quite simple but flexible as it can be applied to almost every transaction series at every time interval. It is also model-independent.

The second group comprises procedures that utilize some statistical market models to forecast the market. The signal therefore is the statistically significant deviation from calculated «forecast» one step forward. The approach is described in [Minenna, 2003] and [Cholewiński, 2009] where authors use a time-series models to forecast stock parameters. Paper [Cholewiński, 2009] uses CAPM-like market model with autoregressive component and GARCH(1,1) errors. Paper [Minenna, 2003] utilizes the diffusion model to evaluate stock parameters like price, trade volume, market concentration.

The main advantage of this approach is that it is based on well-known statistical properties of time series and use strict criteria to generate signals. The main disadvantage is that this approach cannot be applied to classification of individual transactions within a trade session.

The third group comprises a variety of non-parametric methods and numerical algorithms. The core principle of the method is computation of a number of figures through an exactly defined algorithm, filtration of the results and graphical and numerical analysis. An example can be found at [Slama, 2008] where the authors utilize a sample entropy approach to classify transactions into «typical» and «non-typical» categories. This classification ability is the main advantage of the method. Comprehesive study of various non-parametrical algorithms can be found in [Öğüt et al., 2009]

where authors test classification power of four algorithms, namely: multiple discriminant analysis, logistic regression, artificial neural networks (ANN) and a support vector machine (SVM) approach. Results show that non-parametrical procedures (ANN and SVM) are more powerful with classification of manipulated and non-manipulated samples.

The main disadvantages of these approaches are the need to constantly and precisely calibrate the algorithm parameters and potential bias towards ambiguous signals of the system. Numerical algorithms need to be tested thoroughly before one can judge their effectiveness and put them into practice.

The bottom-line is that there is no clear answer as to what is the best numerical method. Therefore this topic is an abundant field for further research.

Assessment of the possible applications of the entropy coefficient to financial market analysis and non-typical transactions

It has been mentioned that any numerical method for non-typical transactions detection incorporates: a number of parameters that can be used to classify all market transactions to either type, calibration procedure and explanation for different parameters values. One of the examples is the chaos approach. The chaos theory approach to financial market is quite new to date, first references and empirical studies can be found in the literature of 2000th. An interesting example can be found in [Pincus et al., 2004] where the authors propose some new coefficients and parameters for the market analysis and even valuation of various assets.

Due to the disadvantages of time-series methods mentioned earlier and Russian financial market specifics (e.g. low liquidity for most securities) the entropy approach has been chosen for this article. The parameter to be discussed is the sample entropy (SampEn). It is thoroughly discussed in [Slama, 2008] and [Reddy, Sebastin]. In [Slama, 2008] the authors try to develop the method to detect manipulative transactions. It is based on a presumption that when a manipulator enters the market he brings a sort of «regularity» into it, so the entropy of the market must decrease somehow. The authors considered a number of cases of proved manipulation and assessed the characteristics of entropy parameters. They conclude that signals are too ambiguous and the method requires further investigation.

SampEn computation procedure

Entropy measures a degree of irregularity within the data. To numerically asses it several different coefficients were developed. As it is stated in [Lake et al., 2002] the SampEn is the most unbiased estimator for the entropy on small samples and will be used for this paper.

The first step is to define basic parameters of raw market data, their computation formulas and so called data «scale». For illustrative purposes of this article two parameters were selected:

- Normalized return for two consecutive transactions (price incremental in %);

- Normalized transaction volume.

Scaling procedure can be applied to «smoothen» data and eliminate seasonal effects by using non-overlapping averages instead of raw numbers. In order not to complicate the example scale level 1 has been chosen which is raw data without averaging.

One can also utilize non-overlapping sampling procedure (different trade days), consider all the transactions for a period as one sample (without considering their timing) or construct a sample on rolling basis (estimation window of predetermined length).

Define $X = \{x_1, ..., x_n\}$ as a generated sample for entropy estimation. Individual elements correspond to either normalized return or transactions volume mentioned earlier. Let $r = 20\%$ be sensitivity parameter and $m = 2$ – subsequences length. Parameters «r» and «m» are chosen according to existing literature. Further research is needed to assess different variants for them.

Define $u_m(i) = \{x_i, ..., x_{i+m-1}\}$, here $1 < i < M - m + 1$. In a case with $m = 2$ and $m = 3$ this would be two- and three-component vectors. Further for all i from $1 < i < M - m + 1$ compute $n(i, m, r)$ as a number of $u_m(j), i \neq j$ that are «similar» to $u_m(i)$. «Similarity» can be in different ways but for simplicity reasons let us consider two vectors similar if corresponding coordinates differ not more than +/– $r\%$. Percentage is used because X consists of normalized data.

In fig. 2 vector (x1, x2) is similar to (x12, x14) and (x43, x44). For $m = 3$ only vector (x1, x2, x3) is similar to (x43, x44, x45). Complete enumeration of all possible m- and $(m + 1)$-component vectors needed then and it is the main time-intensive part of the algorithm. Define $A = \sum_{i=1}^{N-m} n(i, m, r)$ and $B = \sum_{i=1}^{N-m-1} n(i, m+1, r)$ – number of all similar m- and $(m + 1)$-component vectors within sample X.
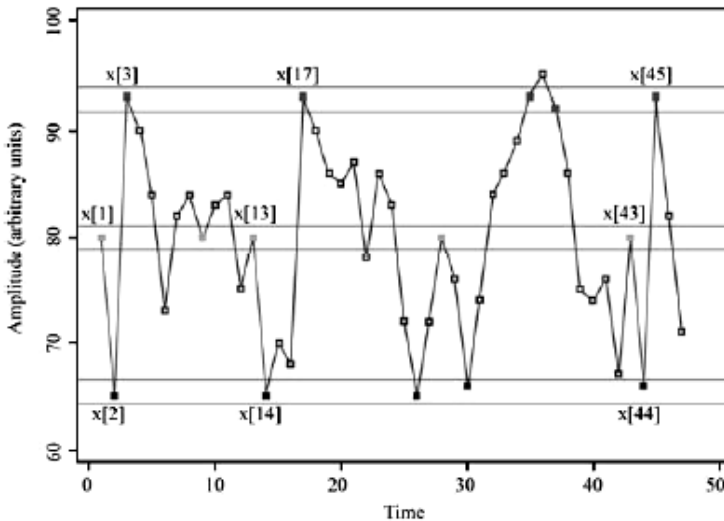
Define $SampEn(m, r, N) = \log \dfrac{A}{B}$.

Fig. 2. Illustration of vector similarity[1]

Data

It was declared on 02/12/2011 that PepsiCo acquired WimmBillDann for a bulk of cash. For the purpose of this paper WimmBillDann securities behavior around 02.12.2010 has been analyzed. Estimation period is from 22/11/2010 till 08/12/2010. It is stated by SEC that some suspicious activity from 29/11/2010 till 02/12/2010 has been detected for WBD ADRs[2]. As ADRs intraday quotes data cannot be acquired, Russian stocks data is utilized instead for transactions that occurred on MICEX for the period.

The next section considers an example of entropy coefficient computation procedure.

---

[1] The graphics is analogous to [Slama, 2008] and [Chikwasha, 2009].
[2] http://ma-journal.ru/news/77063

Results and discussion

Descriptive statistics for the sample is in Table 1.

Table 1.   Descriptive statistics for price incremental and transaction volume for WBD, 22/11/2010–08/12/2012

Date	Transactions number for the date	Turnover, pcs./day	Average price	Intraday standard deviation of price	Turnover for the date, mln. RUR	WA of price for the date	Average number of securities per transaction, pcs.	Standard deviation of the number of securities per transaction, pcs.	Average number of securities per transaction/st.dev(lag 1), %	Average price incremental, %	Standard deviation of price incremental	Average price incremental/st.dev(lag 1), %
22.11	144,0	2480	2044,8	10,2	5,1	2043,0	17,2	112,3		–0,000	4,94	
23.11	247,0	1940	2033,1	8,2	3,9	2028,2	7,9	47,0	7,0	0,001	2,78	0,03
24.11	97,0	2641	2064,5	16,5	5,5	2069,8	27,2	105,6	57,9	0,409	5,63	14,74
25.11	125,0	5941	2132,1	26,0	12,7	2143,8	47,5	126,9	45,0	0,512	6,57	9,09
26.11	332,0	21408	2287,8	54,8	49,5	2312,5	64,5	176,7	50,8	0,395	10,89	6,01
29.11	124,0	17748	2323,8	22,7	41,4	2334,1	143,1	387,4	81,0	0,443	10,39	4,07
30.11	137,0	28565	2312,4	14,5	66,1	2313,7	208,5	958,6	53,8	–0,307	7,47	–2,95
01.12	95,0	34471	2312,2	25,4	79,6	2308,4	362,9	1591,4	37,9	0,189	11,72	2,53
02.12	**4447,0**	**392605**	**3520,5**	**387,2**	**1289,9**	**3285,4**	**88,3**	**496,9**	**5,5**	**0,322**	**18,98**	**2,75**
03.12	1868,0	45015	3639,2	83,6	163,2	3624,5	24,1	93,6	4,8	–0,048	12,75	–0,25
06.12	252,0	4044	3600,8	16,9	14,6	3600,3	16,0	40,1	17,1	–0,206	7,31	–1,62
07.12	225,0	14698	3614,8	15,7	53,2	3619,5	65,3	168,1	162,9	0,053	7,78	0,73
08.12	177,0	20214	3628,8	9,9	73,4	3630,4	114,2	294,1	67,9	0,114	4,74	1,46

Proven information for the deal came to the market on 02/11/2010 and that was clearly reflected by the market in increased transactions price and volume (fig. 3). The average deal size in a number of securities traded increased too, which is the signal for increased market activity before the announcement. The main reason why this activity occurred before the announcement date is that there possibly were some market talks about the deal. The market «adapted» for this event. Also it should be noted that the difference in price between close intraday transactions diminished as can be seen in Table 1 (normalized price increment). This fact relates to market «smoothing» with increased liquidity and participants for WBD «in play».
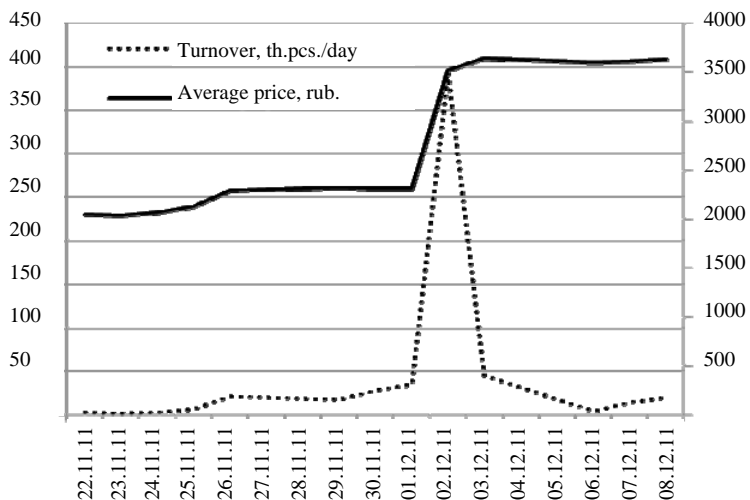
Estimation for entropy coefficient is in Table 2.

Fig. 3. Deal was announced on 02/11/2010.
Average deal size and price increased prior

Table 2.   Entropy coefficient for price incremental
                                and transaction volume for WBD,
                                22/11/2010 – 08/12/2012

Date	SampEn (price incremental), %	A	B	SampEn (volume), %	A	B
22.11	9,6	14148	14613	3,2	9826	8926
23.11	10,3	27612	28086	1,7	24195	21817
24.11	26,2	3408	4037	16,9	3263	2512
25.11	101,5	1908	3490	60,4	2219	804
26.11	70,6	3137	5996	64,8	6389	3155
29.11	92,0	2070	3300	46,6	2631	1049
30.11	65,7	4825	6363	27,7	4570	2369
01.12	82,2	1635	2506	42,7	1716	754
02.12	56,4	2020	5203	94,6	18380	10456
03.12	141,2	1506	3824	93,2	3934	959
06.12	37,7	3518	6171	56,2	20385	13980
07.12	27,2	10452	13200	23,3	12206	9297
08.12	50,9	6034	7941	27,5	6256	3762

Table 2 and fig. 4 show that entropy coefficients for both data types tend to increase prior to announcement and this can be seen as market becoming more «irregular».
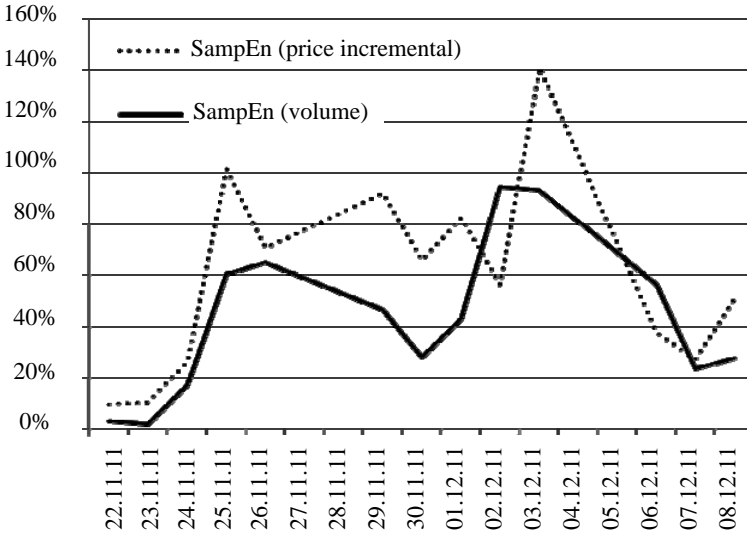

Fig. 4. SampEn signals are too ambiguous

On 02/12/2010 when information for the deal reaches the market SampEn is close to 1. Afterwards the coefficient steadily decreases. Such signals are too confusing and no conclusions can be made due to this fact. It is quite useful to compare classic event-study approach to entropy coefficient behavior. It can be seen from fig. 5 that there are no statistically significant price movements prior to 02/11/2010 and also for that period SampEn coefficients show some significant fluctuations and tend to increase long before the announcement date.
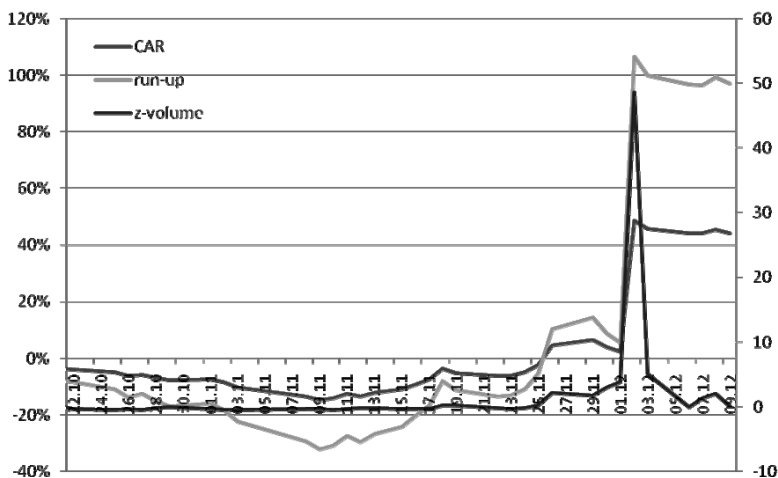
Fig. 5. Event-study basic parameters for the case discussed

Conclusion

The main purpose of this paper was to introduce sample entropy approach and the coefficient computation procedure and try to estimate it for Russian market. The results are too confusing and the approach needs to be tested further to better understand its application. The computational algorithm was described in general, which is useful for further research.

Comparison of event-study parameters and SampEn behavior revealed that the entropy coefficients are more sensitive. These findings suggest the coefficient may be a candidate for some complex non-typical transactions detection procedure.

To conclude, let us outline the main differences of entropy approach in comparison with time-series based models. Econometric models can be readily applied and estimated parameters tested for significance. There is also no need to «educate» them. The crucial disadvantage is that the data structure is incorporated in the model and therefore some non-linear structural changes are hard to detect.

Entropy approach in contrast is aimed at assessing the characteristics of the data internal structure. One particular coefficient discussed so far – SamEn, but there are also other coefficients and numerical algorithms (some of them are mentioned in this paper). Micro structural approach with learning features seems to be sufficiently interesting to be thoroughly investigated and apllied for Russian market analysis.

551

References

Chikwasha V. Time-series Analysis Using Wavelets and Entropy Analysis // BMC Bioinformatics. 2009. Vol. 10. Jan. P. 32.

Cholewiński R. Real-time Market Abuse Detection with a Stochastic Parameter Model // Central European Journal of Economic Modellig and Econometrics. 2009. Vol. 284. P. 261–284.

Lake D.E., Richman J.S., Griffin M.P. et al. Sample Entropy Analysis of Neonatal Heart Rate Variability // American Journal of Physiology. Regulatory, Integrative and Comparative Physiology. 2002. Vol. 283. Sep. P. R789–797.

Minenna M. The Detection of Market Abuse on Financial Markets: A Quantitative Approach. 2003.

Öğüt H., Doğanay M.M., Aktaş R. Detecting Stock-price Manipulation in an Emerging Market: The Case of Turkey // Expert Systems with Applications. 2009. Vol. 36. Nov. P. 11944–11949.

Pincus S., Kalman R.E. Irregularity, Volatility, Risk, and Financial Market Time Series // Proceedings of the National Academy of Sciences of the United States of America. 2004. Vol. 101. P. 13709.

Reddy Y., Sebastin A. Parameters for Estimation of Entropy to Study Price Manipulation in Stock Market.(papers.ssrn.com)

Slama M. Trade-based Stock Price Manipulation and Sample Entropy. 2008.