# img.jpg

# Methods for Detecting Non-typical Transactions. Application for the Russian Stock Market.

Tlekhugov, N., analyst, Laboratory of Financial Markets Analysis, Department of Financial Markets, NRU-HSE

[ntlekhugov@gmail.com](mailto:ntlekhugov@gmail.com)

Kush, K., analyst, Laboratory of Financial Markets Analysis, Department of Financial Markets, NRU-HSE

[konsta@yandex.ru](mailto:konsta@yandex.ru)

*Supervisor*: Stolyarov, A., associate professor, Laboratory of Financial Markets Analysis, Department of Financial Markets, NRU-HSE

[a\_stol@inbox.ru](mailto:a_stol@inbox.ru)

# Introduction

The Russian stock market has quite а long history, from its appearance in the 1990s to the present day. Current operational and legal settings of the financial market in Russia have so far been established as a complete system. 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 2008, regulatory infrastructure has not developed sufficiently to be considered mature and complete. The most widespread market abuses in the Russian stock market are said to be illegal insider trading and market manipulation. It is worth noting that the legal framework for monitoring that kind of activities had not established till the end of 2010 when the federal act on prohibition of insider trading and market manipulation was adopted (Federal Law №224 – FL №224).

FL №224 introduced the notion of “insider”, “insider (material) information” and declared norms and regulations for preventing illegal activities on the financial market. Obviously, 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. Also, it will definitely be accompanied by the development of comprehensive controlling and monitoring infrastructure, a bunch 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 with market participants gaining information and experience.

The regulatory, controlling, and compliance documents mentioned above can be viewed as a “cover” (or “shell”) for its “contents” that is, an effective anti-fraud infrastructure in the market. One could refer to the anti-fraud infrastructure as specific software- and hardware-based monitoring solutions for exchanges and for the regulatory authority (the Federal Department of Financial Markets), 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 a pressing matter. And one of the most important components of such a system is market abuse detection methods. The rest of the article is structured as follows: a discussion of some aspects of an “ideal” monitoring system, a brief summary of the existing literature on the subject, and an indicative example of a recently developed math approach.

# Preliminary comments

Every monitoring system consists of computational (numerical) analytical methods, visual analysis, and implementation (operational) blocks. The 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 that is, non-typical transactions. The preliminary definition for the purpose of the paper 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.

Figure 1 Market abuse classification

Comments to Figure 1:

* Information-based manipulation consists in spreading false info and rumors about a company
* Action-based manipulation refers to non-trading covert actions for the purpose of self-enrichment of a manipulator, e.g. felonious bankruptcy etc.
* Trade-based manipulation refers to a wide range of trading strategies aimed at illegal profiting from misleading other market participants

As displayed in Figure 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 a broader definition of “market abusive transactions”. Thus, any algorithmic methodology for detecting market abusive behavior eventually comes to a method for detecting illegal insider trades and trade-based manipulative actions by market participants.

So far we have defined the notion of non-typical transactions as apposed to “typical” ones. In reality every transaction can be referred to either class by studying some of its parameters. Thus, the detection procedure should clearly specify the list of these parameters, a computational or other suitable procedures to obtain them, a technique to generate signals, and a method for interpreting these signals.

# Existing literature review

Insider trading is well studied in existing empirical and theoretical research literature but, surprisingly, the same can not be said regardig market manipulation phenomenon.

There is a well-known debate about the consequences of insider trading for the market. Some argue that it augments market efficiency by quickly introducing new information to the market, while others consider it to be 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 the US or European markets. For the Russian emerging market where there is no sufficient market infrastructure and information allocation is highly asymmetrical the attitude towards insider trading is surely negative. Contrary to insider trading, there is no doubt about the harm of manipulators to the market: they are 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 discernible.

Further on some research papers focussing on detecting market manipulation are discussed.

***There are three broad groups of detection methods***.

1. Simple detection procedure that deals with 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 for every time interval. It is also model-independent.
2. 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 the authors employ time-series models to forecast stock parameters. Paper [Cholewiński, 2009] employs 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.

1. 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 in [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. Comprehensive 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 a 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 2000. 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 evaluation of various assets.

Due to the disadvantages of time-series methods mentioned earlier and specific features of the Russian financial market (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 for detecting manipulative transactions. It is based on an assumption 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 revealed 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 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 (percentage price incremental)
* Normalized transaction volume

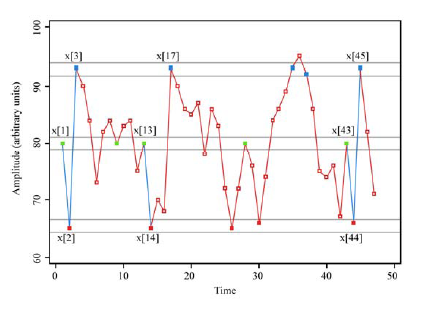
Scaling procedure can be applied to “smoothen” data and eliminate seasonal effects by using non-overlapping averages instead of raw numbers. In order not to complicate the example scale level 1 has been chosen which is raw data without averaging.

One can also utilize non-overlapping sampling procedure (different trading 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 a predetermined length).

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

Define , here *1<i<M-m+1.* In the 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 that are “similar” to . “Similarity” can be in different ways but for simplicity let us consider two vectors to be similar if corresponding coordinates differ by no more than +/- r%. Percentage is used because X consists of normalized data.

Figure 2 Illustration of vector similarity[[1]](#footnote-2)



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 and – number of all similar m- and (m+1)-component vectors within sample X.

Define .

*Data*

It was declared on 02 December 2011 that PepsiCo acquired WimmBillDann for a certain amount of cash. For the purpose of this paper WimmBillDann securities behavior around 02 December 2010 is analyzed. Estimation period is from 22 November to 08 December 2010. It is stated by SEC that some suspicious activity in period from 29 November – 02 December 2010 has been detected for WBD ADRs[[2]](#footnote-3). Since 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.

# Results and discussion

Descriptive statistics for the sample is given in Table 1.

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

Proven information for the deal came to the market on 02 November 2010 and that was clearly reflected by the market in increased transactions price and volume (Figure 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 “ajusted” for this event. Also, it should be noted that the difference in price between close intraday transactions decreased 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 given in Table 2.

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 | 2 480 | 2 044,8 | 10,2 | 5,1 | 2 043,0 | 17,2 | 112,3 |  | - 0,000 | 4,94 |  |
| 23.11 | 247,0 | 1 940 | 2 033,1 | 8,2 | 3,9 | 2 028,2 | 7,9 | 47,0 | 7,0% | 0,001 | 2,78 | 0,03% |
| 24.11 | 97,0 | 2 641 | 2 064,5 | 16,5 | 5,5 | 2 069,8 | 27,2 | 105,6 | 57,9% | 0,409 | 5,63 | 14,74% |
| 25.11 | 125,0 | 5 941 | 2 132,1 | 26,0 | 12,7 | 2 143,8 | 47,5 | 126,9 | 45,0% | 0,512 | 6,57 | 9,09% |
| 26.11 | 332,0 | 21 408 | 2 287,8 | 54,8 | 49,5 | 2 312,5 | 64,5 | 176,7 | 50,8% | 0,395 | 10,89 | 6,01% |
| 29.11 | 124,0 | 17 748 | 2 323,8 | 22,7 | 41,4 | 2 334,1 | 143,1 | 387,4 | 81,0% | 0,443 | 10,39 | 4,07% |
| 30.11 | 137,0 | 28 565 | 2 312,4 | 14,5 | 66,1 | 2 313,7 | 208,5 | 958,6 | 53,8% | -0,307 | 7,47 | -2,95% |
| 01.12 | 95,0 | 34 471 | 2 312,2 | 25,4 | 79,6 | 2 308,4 | 362,9 | 1 591,4 | 37,9% | 0,189 | 11,72 | 2,53% |
| **02.12** | **4 447,0** | **392 605** | **3 520,5** | **387,2** | **1 289,9** | **3 285,4** | **88,3** | **496,9** | **5,5%** | **0,322** | **18,98** | **2,75%** |
| 03.12 | 1 868,0 | 45 015 | 3 639,2 | 83,6 | 163,2 | 3 624,5 | 24,1 | 93,6 | 4,8% | - 0,048 | 12,75 | -0,25% |
| 06.12 | 252,0 | 4 044 | 3 600,8 | 16,9 | 14,6 | 3 600,3 | 16,0 | 40,1 | 17,1% | - 0,206 | 7,31 | -1,62% |
| 07.12 | 225,0 | 14 698 | 3 614,8 | 15,7 | 53,2 | 3 619,5 | 65,3 | 168,1 | 162,9% | 0,053 | 7,78 | 0,73% |
| 08.12 | 177,0 | 20 214 | 3 628,8 | 9,9 | 73,4 | 3 630,4 | 114,2 | 294,1 | 67,9% | 0,114 | 4,74 | 1,46% |

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

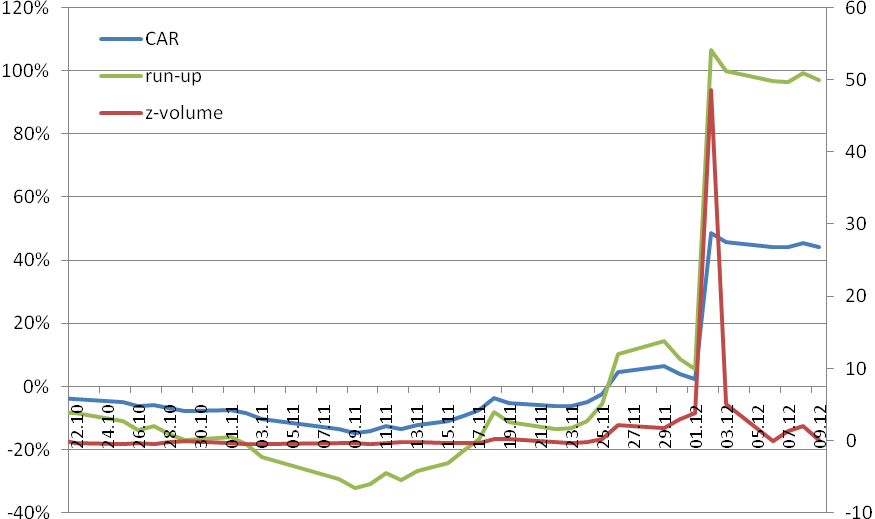
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Date** | **SampEn**  **(price incremental)** | **A** | **B** | **SampEn**  **(volume)** | **A** | **B** |
| 22.11 | 9,6% | 14 148 | 14 613 | 3,2% | 9 826 | 8 926 |
| 23.11 | 10,3% | 27 612 | 28 086 | 1,7% | 24 195 | 21 817 |
| 24.11 | 26,2% | 3 408 | 4 037 | 16,9% | 3 263 | 2 512 |
| 25.11 | 101,5% | 1 908 | 3 490 | 60,4% | 2 219 | 804 |
| 26.11 | 70,6% | 3 137 | 5 996 | 64,8% | 6 389 | 3 155 |
| 29.11 | 92,0% | 2 070 | 3 300 | 46,6% | 2 631 | 1 049 |
| 30.11 | 65,7% | 4 825 | 6 363 | 27,7% | 4 570 | 2 369 |
| 01.12 | 82,2% | 1 635 | 2 506 | 42,7% | 1 716 | 754 |
| 02.12 | 56,4% | 2 020 | 5 203 | 94,6% | 18 380 | 10 456 |
| 03.12 | 141,2% | 1 506 | 3 824 | 93,2% | 3 934 | 959 |
| 06.12 | 37,7% | 3 518 | 6 171 | 56,2% | 20 385 | 13 980 |
| 07.12 | 27,2% | 10 452 | 13 200 | 23,3% | 12 206 | 9 297 |
| 08.12 | 50,9% | 6 034 | 7 941 | 27,5% | 6 256 | 3 762 |

Table 2 and Figure 4 reveal that entropy coefficients for both data types tend to increase prior to announcement and this can be seen as market becoming more “irregular”.

Figure 4 SampEn signals are too ambiguous

On 02 December 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 Figure 5 that there are no statistically significant price movements prior to 02 November 2010 and also for that period SampEn coefficients show some significant fluctuations and tend to increase long before the announcement date.

Figure 5 Event-study basic parameters for the case discussed



# Conclusion

The main purpose of this paper is to introduce sample entropy approach and the coefficient computation procedure and try to estimate it for the Russian financial 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 that has been discussed so far is SamEn but there are also other coefficients and numerical algorithms (some of them are mentioned in this paper). Microstructural approach with learning features seems to be quite interesting to be thoroughly investigated and apllied for the analysis of the Russian financial market.

# References

Chikwasha V., “Time-series analysis using wavelets and entropy analysis,” *BMC bioinformatics*, vol. 10, Jan. 2009, p. 32.

Cholewiński, R.“Real-Time Market Abuse Detection with a Stochastic Parameter Model,” vol. 284, 2009, pp. 261-284.

Lake D.E., Richman J.S., Griffin M.P., and Moorman J.R., “Sample entropy analysis of neonatal heart rate variability.,” *American journal of physiology. Regulatory, integrative and comparative physiology*, vol. 283, Sep. 2002, pp. R789-97.

Minenna, M. “The detection of market abuse on financial markets: a quantitative approach,” 2003.

Öğüt H. Doğanay M. Mete, and Aktaş R., “Detecting stock-price manipulation in an emerging market: The case of Turkey,” *Expert Systems with Applications*, vol. 36, Nov. 2009, pp. 11944-11949.

Pincus S. and Kalman R.E., “Irregularity, volatility, risk, and financial market time series,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, 2004, p. 13709.

Reddy, Y. and Sebastin A. , “Parameters for Estimation of Entropy to Study Price Manipulation in Stock Market,” *papers.ssrn.com*, 1867.

Slama, M., “Trade-Based Stock Price Manipulation and Sample Entropy,” 2008.

1. The graphics is analogous to [Slama, 2008] and [Chikwasha, 2009] [↑](#footnote-ref-2)
2. <http://ma-journal.ru/news/77063> [↑](#footnote-ref-3)