

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



An Assessment of the Prediction Quality of VPIN

Antoine Bambade and Kesheng Wu

Abstract

VPIN is a tool designed to predict extreme events like flash crashes. Some concerns have been raised about its reliability. In this chapter we assess VPIN prediction quality (precision and recall rates) of extreme volatility events including its sensitivity to the starting point of computation in a given data set. We benchmark the results with the ones of a “naive classifier.” The test data used in this study contains 5.6 year’s worth of trading data of the five most liquid futures contracts of this time period. We found that VPIN has poor “flash crash” prediction power with the traditional 0.99 decision threshold. Increasing the decision threshold does not significantly improve overall prediction quality. Nevertheless we found VPIN has a more interesting predictive power for flash events of lower amplitude. Finally, we found that, for practice, the last bar price structure is the least sensitive to the starting point of computation.

Keywords: high-frequency data, probability of informed trading, VPIN, liquidity, flow toxicity, volume imbalance, flash crash

JEL codes: C02, D52, D53, G14, G23

1. Introduction

1.1 Main study purpose

Easley et al. [1] designed a tool, nicknamed volume-synchronized probability of informed trading (VPIN), with the aim to predict flash crashes. It appeared it could predict the “flash crash” of May 6, 2010, a few hours before it happened [2]. A lot of papers were published [3–5], and it was proposed to use it for regulation through a VPIN contract [2, 6]. However, critics pointed out some flaws, questioning its reliability [7–11] but without providing a quantitative evaluation of the prediction quality (e.g., in terms of precision and recall rates). In this study, we design a framework to detect flash crashes and thereby assess the behavior of the VPIN tool enabling as well as comparing and benchmarking with other predictive algorithms.

1.2 Motivation

The amount of trading data has exploded in finance thanks to the continuing progress of high-frequency techniques. It constrains practitioners to use more and more state-of-the-art algorithms to deal with this overwhelming amount of information. Computers and algorithms are more and more efficient, but still

decision-making is highly dependent on both the quantity and the quality of information. Thus, errors and speculations that can make the financial market toxic, i.e., conducive to crashes, are possible. Examples in the past, such as the “flash crash” of May 6, 2010, have shown that this new paradigm in finance has made it possible to introduce a new kind of crashes characterized by their suddenness. Such quick crashes seem dangerous because of a kind of inherent unpredictability. However, predictive models to model this new framework do exist.

1.3 Model

Easley et al. [12] designed a model of the high-frequency financial market based on flows of informed and uninformed traders. They showed that information is a key parameter of the spread between ask and bid of prices. The model works as follows. Each day, general conditions and circumstances may or may not result in events that can help predict the evolution of the price of a future. More precisely, for each day, nature decides whether or not there is an event that can help predict the evolution of the price of a future. This is modeled with a Bernoulli law of parameter α . If an event occurs, nature also decides with a Bernoulli law of parameter δ if this event is a low signal. With these conditions, buys and sells for this future come then from flows of informed and uninformed traders. They are modeled by Poisson processes of respective parameters μ and ϵ . This framework can be summarized by the following tree in **Figure 1** [13].

The whole trading process studied is thus a mixture of Poisson processes. It enabled authors to compute ask and bid and then the spread. They showed that for reasonable cases the spread is linearly linked with the following probability they named probability of informed trading (PIN) [12]:

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\epsilon}. \tag{1}$$

Later, Easley et al. [2] designed a new framework to easily compute this probability. Indeed PIN numbers come from a parametrized framework, and one does not have access to all these parameters. They showed however that PIN can be well approximated through a volume-clock paradigm [14], thanks to data of futures

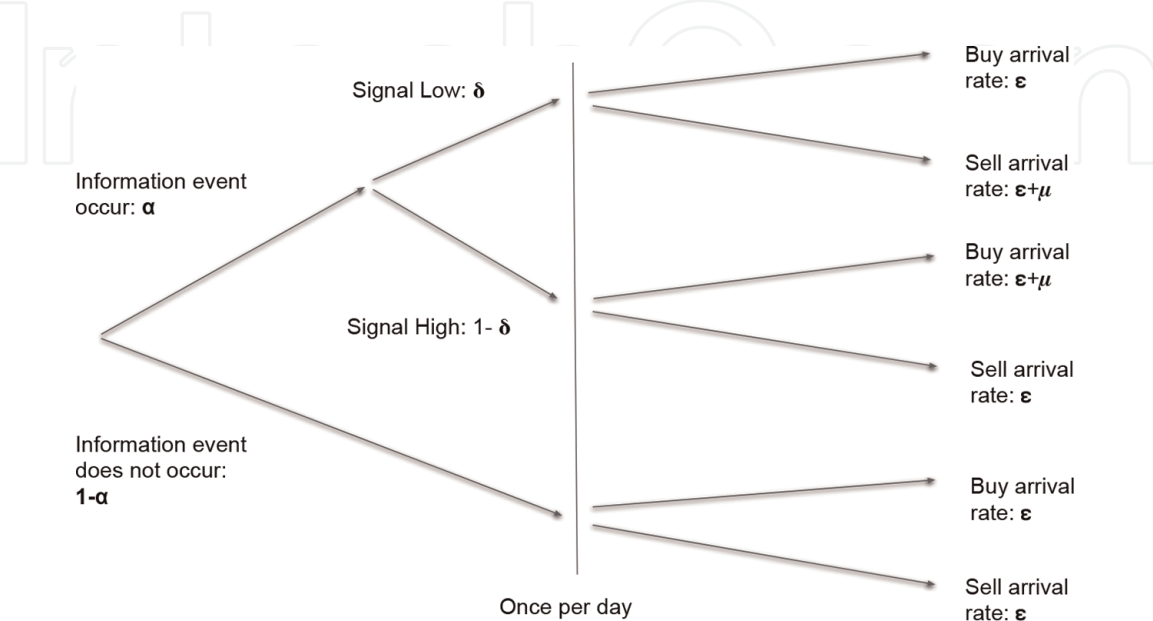


Figure 1.
A tree summarizing the trading process.

with a new formula. The approximated version of PIN was then called the volume-synchronized probability of informed trading (VPIN). It appeared that this new tool could predict the “flash crash” of May 6, 2010, a few hours before it happened [2].

Nevertheless, the model has received a lot of critics. For example, Andersen and Bondarenko have shown [7] that VPIN is quite sensitive to the starting point of when one starts computing VPIN on a data set. It indeed questions VPIN prediction quality. Moreover, they have also shown that VPIN is sensitive to other parameters, such as the trade classification rule used [8] or how one defines the average daily volume of trades [9]. Changing the classification rule may drastically change VPIN behavior [9]. Pöppe et al. have reached the same conclusions with a different approach. Using a different classification rule can change VPIN prediction power toward a crash (in their paper a German blue-chip stock [11]). Besides, controlling ex ante parameters seems to give poorer prediction quality [8, 9]. This point has also been checked by Abad et al. [10]. Controlling ex ante realized volatility, and trading intensity, as did Andersen and Bondarenko [9], prediction quality seems to vanish. More deeply, they have also underlined that it is not obvious how one should define a VPIN prediction, analyzing more precisely toxic and nontoxic halts, as well as toxic events. Furthermore, Torben G. Andersen and Oleg Bondarenko interpret VPIN as being too sensitive to trading intensity. They have also explained that VPIN metric is sometimes unexpectedly correlated with other usual ones (such as VIX or RV) [7, 8]. More recently, it has been shown theoretically that the volume-clock paradigm of VPIN framework does not enable to really approximate fully the PIN value, although the proposed formula is close [15, 16].

More generally, all these critics have pointed out that:

- First, it is not obvious how one should use VPIN.
- Second, prediction quality has not been studied sufficiently to assess it as being reliable.
- Third, the study lacks objective benchmark.

1.4 Goal

The purpose of this chapter is to quantify the prediction quality of VPIN in order to enable practitioners to assess whether or not it can be used in the real world (e.g., for trading or regulation). That's why:

- First, we want to design a proper framework to compute precision and recall rates as well as prediction length of VPIN. This will be possible by providing a formal definition of flash crashes. To be more precise, we will use the maximum of intermediate return (MIR) [5] to define it.
- Second, we want to study through this framework how sensitive VPIN is to the starting point of the data set.

1.5 Plan

In the following, we first recall VPIN model and propose a definition for flash crashes (Section 2). Second, we assess within this framework VPIN prediction quality (Section 3). Finally, we assess VPIN sensitivity to the starting point of the data set (Section 4).

2. VPIN software and formal flash crash definition

In this section, we first recall the VPIN model. Second, we propose a definition of flash crashes used to compute precision and recall rates. Finally, we present the data used in our tests.

2.1 VPIN software

Easley et al. [12] designed a model of the high-frequency financial market based on informed and uninformed traders. It is then possible to compute a probability of informed trading (PIN). Easley et al. [1] use these results and define an easy way to compute PIN only through the data of trades. We describe briefly VPIN model used in previous literature. The theoretic study of the model is treated in another research study.

2.1.1 Bars

Following Easley et al. [1], a bar is a fixed volume of trades that are successive in time. With such a definition, one can associate the following quantities with each bar:

- A nominal price, computed according to a given technique (mean price, median price, closing price, opening price, etc.)
- A nominal time (first trade time, last trade time)
- Local maximum and minimum values of trades

In practice, the last few trades that do not fill up a bar are dropped to the next bar.

2.1.2 Bulk volume classification

The computation of VPIN requires to determine directions of trades, i.e., classifying each trade as a buy or a sell. The method used here is the following: bulk volume classification (BVC) [1, 5]. Let us note V_b the volume of a bar and j the label of bar number j ($j > 0$) and P_j its price (closing, opening, median, mean). Then the number of buys V_j^b within bar j is determined according to this formula:

$$V_j^b = V_b \mathcal{Z} \left(\frac{P_j - P_{j-1}}{\sigma} \right) \quad (2)$$

where \mathcal{Z} is the cumulative function of a given law (usually student or normal distribution) and σ is the standard deviation of the numerator on successive number of bars. In our test, σ is computed once on all successive values of the data set, and the student law is of parameter one. Within bar j the number of sells V_j^s is obviously

$$V_j^s = V_b \left(1 - \mathcal{Z} \left(\frac{P_j - P_{j-1}}{\sigma} \right) \right) \quad (3)$$

2.1.3 Buckets

A bucket is defined to be a fixed number of successive trades. Here to simplify, as bars are defined also as a fixed number of trades, a bucket will be m successive bars. Let us note V_{bucket} the fixed volume of a bucket. We naturally have $V_{bucket} = mV_b$.

2.1.4 VPIN formula

VPIN formula is computed on n successive buckets, where n is VPIN support. A buffer is defined as n successive buckets. Here is VPIN formula, approximating (1) upon bucket number j ($j \geq n$):

$$VPIN_j = \frac{\sum_{i=j-n+1}^j |V_{bucket,i}^b - V_{bucket,i}^s|}{nV_{bucket}} \quad (4)$$

For a given bucket i :

- $V_{bucket,i}^s = \sum_{j \in bucket_i} V_j^s$
- $V_{bucket,i}^b = \sum_{j \in bucket_i} V_j^b$

In order to distribute all VPIN values between 0 and 1, in practice, VPIN is normalized through a normal law. We thus consider $VPIN_{normalized}$ in the following:

2.1.5 VPIN event

A VPIN event is declared when the following occurs:

$$VPIN_{normalized} \geq \theta_{VPIN} \quad (5)$$

where θ_{VPIN} is a given decision threshold. In practice [5] $\theta_{VPIN} = 0.99$.

2.2 Defining flash crashes with MIR

2.2.1 Formal definition

Let $(p_t)_t$ be a time series (e.g., of prices). Here is the definition of MIR:

$$MIR_{t,\eta} = \max_{i \neq j, i,j \in [t,t+\eta]} \frac{|p_i - p_j|}{p_i} \quad (6)$$

A flash crash will depend on two things here:

- The amplitude of the crash, which means extreme MIR values (e.g., 10%)
- The shortness of the fall, which means the shortness of the time window within η that computes $MIR_{t,\eta}$ (e.g., 10 minutes), more precisely, noting $i^*, j^* = \operatorname{argmax}_{i \neq j, i,j \in [t,t+\eta]} \frac{|p_i - p_j|}{p_i}$, the fall has length $|j^* - i^*|$

2.2.2 Empiric definition

We reported in this data set only one flash crash, i.e., on May 6, 2010, which lasted approximately 10 minutes according to media and financial institutions. Our definition of flash crash will obviously take into account this event.

2.3 The data

2.3.1 Futures used

In this work, we use a comprehensive set of liquid futures trading data to illustrate the techniques to be introduced. More specifically, we will use 67 months' worth of tick data of the five most liquid futures traded on all asset classes. The data comes to us in the form of 5 CSV files, one for each futures contract traded. The source of our data is TickWrite, a data vendor that normalizes the data into a common structure after acquiring it directly from the relevant exchanges. The total size of the comma-separated value (CSV) files is about 45.1 GB. They contain about millions of trades spanning from the beginning of January 2007 to the end of July 2012. The data set contains five of the most heavily traded futures contracts. Each has more than 100 million trades during this 67-month period. The most heavily traded futures, the file containing E-mini SP500 futures, symbol ES, has about 500 million trades involving a total number of about 3 billion contracts. The second most heavily traded futures is Euro exchange rates, symbol EC, which is 188 million trades. The next three are Nasdaq 100 (NQ), 173 million trades; light crude oil (CL), 165 million trades; and E-mini Dow Jones (YM), 110 million trades. In **Figure 2**, one can see an evolution of prices with time (here each tick corresponds to a bucket).

2.3.2 Definition of flash crash

We want to define empirically a flash crash using the tools of VPIN framework, namely, bars and buckets. As volume-clock paradigm does not allow to control filling times of fixed volume of trades, here below is a summary of the steps we have followed to manage to detect flash crashes using MIR. As it is quite long and the main purpose of study is the prediction of results of the following section, we present principles and do not go into technical details:

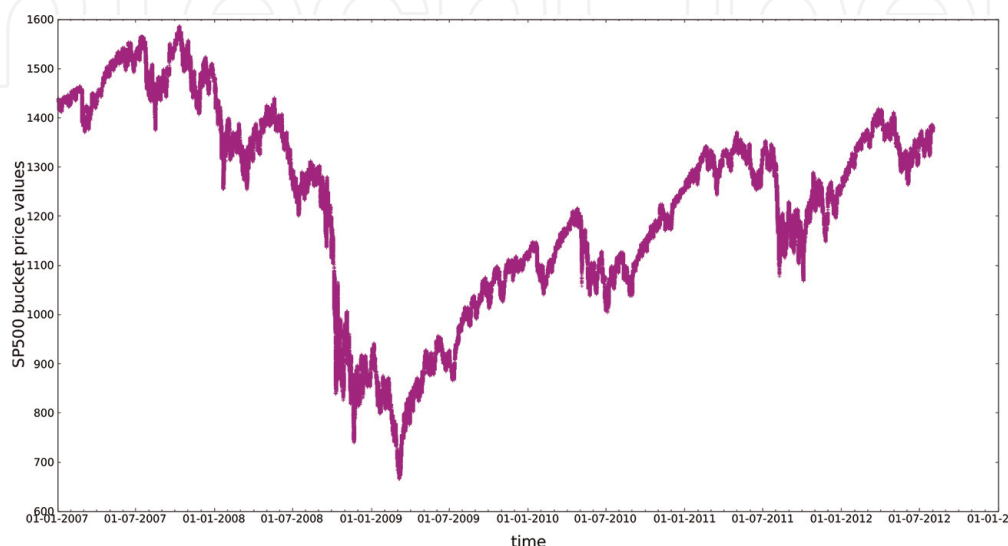


Figure 2.
Bucket S&P 500 values with time.

- To be sure not to miss a flash crash because of being too long in time bar or bucket, we have chosen a reasonable granularity level as in [5] (buckets per day, 200, and bars per bucket, 30).
- For each financial instrument, we have recorded the number of bars necessary to capture the local 10 minutes of maximum fall of May 6, 2010, known as the “flash crash”; we refer to these numbers as “window lengths” below.
- As the window lengths defined above do not have a stable distribution in time (because of the volume-clock paradigm), we have arbitrarily filtered out all events in which the time difference between minimum and maximum within a window length is longer than 20 minutes, in order to capture only quick events. Indeed, one given window length may be too big and thus allow at some date to measure a time difference between local minimum and maximum which is longer than 10 minutes whereas it would be a true flash crash with a smaller window length.¹
- For each instrument we recorded the amplitude of the “flash crash” and their respective MIR values.

The results made it possible to classify the five financial instruments into two groups:

- Data sets where the “flash crash” and other flash crashes are significantly present: ES, NQ, and YM.
- Data sets where the “flash crash” and other flash crashes are not really present. More precisely, the “flash crash” is not a rare event in the data set, and generally magnitude levels of flash crashes are low compared to other instruments.

3. Assessing VPIN prediction quality

In this section, first we present our methodology to find VPIN optimal prediction quality (for which recall and precision rates are maximal and more useful for practice). Second, we present all the results: best parameters, associated remarks, and prediction lengths.

3.1 Methodology

3.1.1 Parameters to test

Here are the parameters we will test:

¹ This is not perfect because we can still miss some crashes (whereas in this data set, it will not be that much, and it will be with a smaller probability), but first we do not want to change too much the definition in time of a flash crash (we will not increase the tolerance level to 1 day), and second this problem is inherent to the fact that fixing volume of bars and of buckets prevents us from controlling precisely filling bar and bucket times. Finding a solution for this precise data set does not guarantee at all a general solution neither for one data set nor for a financial instrument.

- Bar price: mean, median, last price, first price
- MIR decision threshold θ_{MIR} to detect a flash crash
- VPIN support n
- VPIN classifier (student, normal)
- Prediction window ω (described below)
- VPIN decision threshold θ_{VPIN} to predict a flash crash

3.1.2 Defining true positive events

Here we describe how we define true positive, false-negative, and false-positive events. For a given prediction window length ω :

- From a MIR flash crash detection (i.e., $MIR_j \geq \theta_{MIR}$) at a bucket j ($j \geq \omega$), if in the window of buckets $[j-\omega, j-1]$ there is a VPIN event (i.e., $VPIN_{Normalized, i} \geq \theta_{VPIN}$, $i \in [j-\omega, j-1]$), then we consider it as a true positive event.² Otherwise it is a false-negative event.
- From a VPIN event at a bucket j (i.e., $VPIN_{Normalized, j} \geq \theta_{VPIN}$, $j + \omega \leq \text{end Of DataSet}$), if in the window of buckets $[j + 1, j + \omega]$ there is a flash crash (i.e., $MIR_i \geq \theta_{MIR}$, $i \in [j+1, j+\omega]$), then we consider it as a true positive event.³ Otherwise it is a false-positive event.

3.1.3 Choosing the maximum value of ω

To make a useful deep search, we have computed the distribution of time difference between different amounts ω of buckets. Indeed, we want to control a temporal time window reasonable for practitioners and still sufficiently wide so that we can analyze which events VPIN can detect or not. We have focused this research to have a stable bounded distribution of time difference between ω buckets of about 1 month. Below one can see the respective distribution for the S&P500 instrument; the four other distributions of the instruments studied look the same (**Figure 3**).

In **Table 1** one can see the medians of the different distributions.

For the next step, $\omega \leq 2500$.

3.1.4 Describing deep search of flash crash prediction

Here we describe how we intend to make a first deep search of VPIN prediction quality of events close to the “Flash Crash” of May 2010. In this algorithm described below $\theta_{VPIN} = 0.99$.⁴

For each VPIN classifier (student or Gaussian), for each bar price structure (last, first, median, average) do:

- For each $\theta_{MIR} \in [5.2\%, 6.2\%]$ with step 0.1% for ES instrument, $\theta_{MIR} \in [2.2\%, 3.2\%]$ with step 0.1% for CL instrument, $\theta_{MIR} \in [0.4\%, 0.9\%]$ with

² If $j - \omega < 0$, the window of buckets considered is $[0, j-1]$.

³ If $j + \omega > \text{endOfDataSet}$, the window of buckets considered is $[j + 1, \text{endOfDataSet}]$.

⁴ Previous research, such as [5], showed that this threshold is a good one.

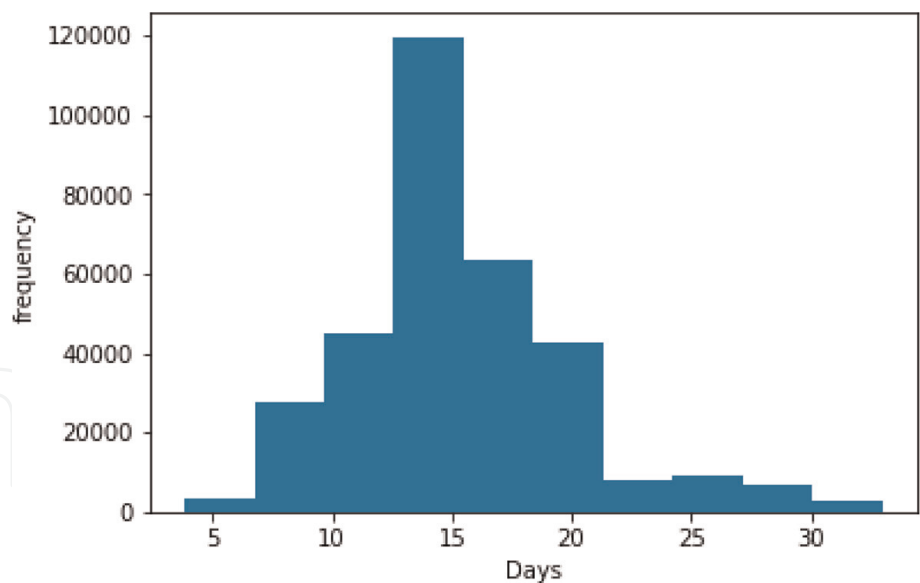


Figure 3.
Time difference distribution between 2500 S&P 500 buckets.

Futures	Days	Number of bucket chosen
ES	14.8	2500
EC	13.8	2500
CL	15.0	2500
YM	14.3	2500
NQ	15.2	2500

Table 1.
Median of time difference between 2500 buckets for the different instruments.

step 0.1% for EC instrument, $\theta_{MIR} \in [8\%, 9\%]$ with step 0.1% for NQ instrument, $\theta_{MIR} \in [5.4\%, 6.4\%]$ with step 0.1% for YM instrument⁵ do:

- For each VPIN support $n \in [30, 60]$, with step 10, do:
 - For $\omega \in [100, 2500]$ with step 100, do:
 - test prediction
 - store current parameters, precision and recall if and only if $recall + precision \geq previousLocalMaximum$
 - store prediction length (distance between VPIN event and MIR event).

Remark: we first try to maximize precision+recall rate. If the local maximum found is interesting for practice (at least superior or equal to 1.2) and more powerful than a “naive” algorithm, then it sounds worth making a more serious search of precision and recall rates separately to find a good trade-off between them (e.g., thanks to a ROC curve).

⁵ As each MIR value for the flash crash is different, one must adapt the area of deep search to be precise and have a quicker calculation time.

3.2 Results

3.2.1 Best parameters found

In **Tables 2–5** one can see the best parameters that maximize precision+recall for each financial instrument and bar price structure studied.

3.2.2 Remarks and first interpretation

We remark overall the following:

- The choice of bar structure does not really affect the optimal choice of other parameters; nevertheless mean and median bar price structures have best precision+recall rate on average.
- Recall rates are very close to 1.
- Since ES, NQ, and YM precision rates are “low”, thus precision + recall rates are “low.”
- Since EC and CL precision rates are “high,” thus precision + recall rates are “high” since recall is already “high.”
- CL and EC had on May 6, 2010, a very low flash crash threshold, which increases a lot the number of crash of same magnitude detected in the data set.
- CL and EC obtain their maximum value to the minimum bound of the deep search (respectively, a 2.2% fall and 0.6% fall). It is not the case for other

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	Bar price
ES	0.9737	0.2171	1.1908	0.062	60	2400	Gaussian	Last
EC	0.9080	0.9644	1.8724	0.006	30	2500	Gaussian	Last
CL	0.9406	0.9045	1.8451	0.022	60	2500	Student	Last
NQ	1	0.0034	1.0034	0.08	30	400	Gaussian	Last
YM	0.8421	0.1512	0.9933	0.064	60	2500	Gaussian	Last

Table 2.
Best parameters maximizing precision+recall rate for different futures and last bar price structure in the first deep search.

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	Bar price
ES	0.9737	0.2024	1.1761	0.062	60	2400	Gaussian	First
EC	0.9127	0.9681	1.8808	0.006	30	2500	Student	First
CL	0.9534	0.9012	1.8546	0.022	60	2500	Student	First
NQ	1	0.0038	1.0038	0.08	30	400	Gaussian	First
YM	0.8421	0.1449	0.9870	0.064	60	2500	Gaussian	First

Table 3.
Best parameters maximizing precision+recall rate for different futures and first bar price structure in the first deep search.

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	Bar price
ES	0.9737	0.1996	1.1733	0.062	60	2400	Gaussian	Median
EC	0.9037	0.9718	1.8755	0.006	30	2500	Student	Median
CL	0.9447	0.8951	1.8398	0.022	60	2500	Student	Median
NQ	1	0.0036	1.0036	0.08	30	400	Gaussian	Median
YM	1	0.1911	1.1911	0.054	30	2500	Student	Median

Table 4.
Best parameters maximizing precision+recall rate for different futures and median bar price structure in the first deep search.

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	Bar price
ES	0.9737	0.1950	1.1687	0.062	60	2400	Student	Mean
EC	0.9058	0.9691	1.8749	0.006	30	2500	Student	Mean
CL	0.9789	0.8654	1.8443	0.022	40	2500	Student	Mean
NQ	1	0.0036	1.0036	0.08	30	400	Gaussian	Mean
YM	1	0.1921	1.1921	0.055	30	2500	Student	Mean

Table 5.
Best parameters maximizing precision+recall rate for different futures and mean bar price structure in the first deep search.

instruments (in NQ cases, precision+recall optimal rate is constant from 0.8 to 0.9).

The results give two first findings:

- When the flash crash is significantly present for the instrument, i.e., of high magnitude and rare in the data set (ES, YM, and NQ cases), then recall is high, which means that VPIN makes a prediction before this happens, but precision is low: VPIN detects other events that are not flash crashes.
- When the flash crash is not significantly present for the instrument, i.e., of low magnitude and not rare (there are a lot of events of 10–20-minute length of same magnitude), then recall and precision are high.

This may suggest one of the following hypotheses:

- VPIN seems to be a poor indicator of flash crash prediction with the usual recommended threshold 0.99.
- VPIN can be a better indicator of another type of event (crashes of less important amplitude).

We will compare the results of the same deep search with the one of a naive classifier, to see whether or not the good prediction results in CL and ES cases are relevant.

3.2.3 Benchmark with a “naive classifier”

We made a comparison of VPIN prediction quality result with a “naive classifier,” which randomly chooses whether or not there will be a crash from each

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)
ES	1	0.0355	1.0355	0.052	50	2500
EC	1	0.9948	1.9948	0.004	50	2500
CL	1	0.3413	1.3413	0.022	50	2500
NQ	1	0.0076	1.0076	0.084	60	2500
YM	1	0.0174	1.0174	0.055	50	2500

Table 6.
Best parameters maximizing precision+recall rate for different futures for the naive classifier.

bucket of the data set. In **Table 6** one can see the results of the naive classifier for the first deep search set of parameters.⁶ As it is a naive classifier, results do not depend on direction of prices (bar price classifier) and bar price structure.

We remark the following:

- “Naive classifier” has poor results comparable to those of VPIN for ES, NQ, and YM instruments; although poor, VPIN predictions are better than “naive algorithm” on ES cases.
- “Naive classifier” has better results than VPIN on EC instrument.
- “Naive classifier” has worse results than VPIN on CL instrument.

We can interpret it as follows:

- EC flash crash definition is barely inconsistent, with a MIR threshold of 0.006%; it is obvious that a naive algorithm does better results as the constraint is very small to detect a “flash crash” of such a magnitude.
- On CL and ES cases though, VPIN predictions are better, and these results are obtained when θ_{MIR} threshold was on the lower bound of the deep search. It might indicate that VPIN software has a better predictive power than a “naive algorithm” not on a “flash crash” amplitude basis but on a lower amplitude level. Nevertheless, one may wonder whether or not this level of amplitude is useful for practitioners.

Anyway, previous results may conclude that for “flash crash” prediction, VPIN has overall equivalent poor power prediction with the traditional threshold $\theta_{VPIN} = 0.99$, as a “naive” algorithm.

That’s why in the next paragraph, we benchmark predictive power of “naive” and VPIN algorithms:

- First on higher θ_{VPIN} constraints
- Second on lower bounds of crash amplitude θ_{MIR} while $\theta_{VPIN} = 0.99$
- Third on higher θ_{VPIN} constraints and at the same time lower bounds on θ_{MIR}

⁶ First tests conducted with EC instrument have been realized with an average to get more robust results. They are really close to the one obtained here with a single realization of randomness.

Indeed, the first hypothesis is that there are too many false VPIN predictions, i.e., false-positive events, as precision rate is too low and recall rate is too high. That’s why one may hope that making θ_{VPIN} constraints higher may reduce the number of VPIN “useless” predictions while not reducing too much recall rate.

3.2.4 Deep search allowing higher bounds for θ_{VPIN}

In the following we have looked to higher bounds for θ_{VPIN} from 0.99 to 0.99999. All other parameters of the deep search are the same. Below, one can see the results in **Tables 7–10**. The results for the naive algorithm are indeed the same.

We remark the following:

- Precision rate has increased for each bar price structure for ES instrument, maintaining recall rate constant to $\theta_{VPIN} = 0.99$ case.
- Precision + recall rate has increased for YM instrument only with a last or first bar price structure, but recall decreased a bit compared to $\theta_{VPIN} = 0.99$ case.

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	θ_{VPIN}
ES	0.9737	0.4677	1.4414	0.062	60	1600	Gaussian	0.99999
EC	0.9080	0.9644	1.8724	0.006	30	2500	Gaussian	0.99
CL	0.9406	0.9045	1.8451	0.022	60	2500	Student	0.99
NQ	1	0.0034	1.0034	0.08	30	400	Gaussian	0.99
YM	0.7091	0.3160	1.0251	0.054	60	2500	Student	0.9999

Table 7.
Best parameters maximizing precision+recall rate for different futures and last bar price structure allowing higher bounds for θ_{VPIN} .

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	θ_{VPIN}
ES	0.9737	0.3412	1.3149	0.062	60	1200	Gaussian	0.99999
EC	0.9127	0.9681	1.8808	0.006	30	2500	Student	0.99
CL	0.9534	0.9012	1.8546	0.022	60	2500	Student	0.99
NQ	1	0.0038	1.0038	0.08	30	2500	Gaussian	0.99
YM	0.7091	0.3545	1.0636	0.054	60	2500	Student	0.9999

Table 8.
Best parameters maximizing precision+recall rate for different futures and first bar price structure allowing higher bounds for θ_{VPIN} .

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	θ_{VPIN}
ES	0.9737	0.3306	1.3043	0.062	30	1700	Gaussian	0.99999
EC	0.9037	0.9718	1.8755	0.006	30	2500	Student	0.99
CL	0.9447	0.8951	1.8398	0.022	60	2500	Student	0.99
NQ	1	0.0036	1.0036	0.08	30	400	Gaussian	0.99
YM	1	0.1911	1.1911	0.054	30	2500	Student	0.99

Table 9.
Best parameters maximizing precision+recall rate for different futures and median bar price structure allowing higher bounds for θ_{VPIN} .

- Compared to the “naive” algorithm, VPIN results are effectively better in ES case. In YM case we still find comparable results.
- On average, mean and median bar price structures have the best precision +recall rate.

To verify whether or not we can get at least better results than a naive algorithm in data sets with a real flash crash, we study in the following first the results allowing lower bounds on θ_{MIR} while $\theta_{VPIN} = 0.99$ and second the results allowing lower bounds on θ_{MIR} and higher constraints on θ_{VPIN} . Indeed, the intuition is that on NQ case, the “flash crash” amplitude constraints are far too high to have a good precision rate, because in this case there are too few events detected with MIR algorithm.

3.2.5 Deep search allowing lower bounds for θ_{MIR}

We remark the following in **Tables 11–14**:

- Results have changed for every instrument except the ES one which has kept the same local maximum as in the first deep search.
- Precision is far higher than before, while recall is still high. Therefore, overall precision + recall rates are “high.”
- Optimal θ_{MIR} is around 0.015 for ES, CL, NQ, and YM financial instruments, whereas for EC the previous local maximum around 0.006 remains higher.
- On average, median bar price structure has the best precision+recall rate.

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	θ_{VPIN}
ES	0.9737	0.3786	1.3523	0.062	60	1600	Gaussian	0.99999
EC	0.9058	0.9691	1.8749	0.006	30	2500	Student	0.99
CL	0.9789	0.8653	1.8442	0.022	40	2500	Student	0.99
NQ	1	0.0036	1.0036	0.08	30	400	Gaussian	0.99
YM	1	0.1921	1.1921	0.055	30	2500	Student	0.99

Table 10.
Best parameters maximizing precision+recall rate for different futures and mean bar price structure allowing higher bounds for θ_{VPIN} .

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	Bar price
ES	0.9421	0.9541	1.8962	0.015	30	2500	Student	Last
EC	0.9080	0.9644	1.8724	0.006	30	2500	Gaussian	Last
CL	0.9297	0.9806	1.9103	0.016	30	2500	Student	Last
NQ	0.9179	0.9019	1.8198	0.015	30	2500	Gaussian	Last
YM	0.9460	0.9696	1.9156	0.015	50	2500	Gaussian	Last

Table 11.
Best parameters maximizing precision+recall rate for different futures and last bar price structure allowing higher bounds for θ_{MIR} .

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	Bar price
ES	0.9404	0.9402	1.8806	0.015	30	2500	Gaussian	First
EC	0.9127	0.9681	1.8808	0.006	30	2500	Gaussian	First
CL	0.9233	0.9728	1.8961	0.016	30	2500	Student	First
NQ	0.8291	0.9833	1.8124	0.01	30	2500	Student	First
YM	0.9517	0.9673	1.9190	0.015	50	2500	Student	First

Table 12.
Best parameters maximizing precision+recall rate for different futures and first bar price structure allowing higher bounds for θ_{MIR} .

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	Bar price
ES	0.9499	0.9498	1.8997	0.015	30	2500	Student	Median
EC	0.9037	0.9717	1.8754	0.006	30	2500	Student	Median
CL	0.9265	0.9718	1.8983	0.016	30	2500	Student	Median
NQ	0.9243	0.9017	1.8260	0.015	30	2500	Gaussian	Median
YM	0.9829	0.9427	1.9256	0.015	30	2500	Gaussian	Median

Table 13.
Best parameters maximizing precision+recall rate for different futures and median bar price structure allowing higher bounds for θ_{MIR} .

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	Bar price
ES	0.9526	0.9454	1.8979	0.015	30	2500	Student	Mean
EC	0.9058	0.9691	1.8749	0.006	30	2500	Student	Mean
CL	0.9302	0.9670	1.8972	0.016	30	2500	Gaussian	Mean
NQ	0.9407	0.8796	1.8203	0.02	60	2500	Gaussian	Mean
YM	0.9446	0.9779	1.9225	0.015	60	2500	Student	Mean

Table 14.
Best parameters maximizing precision+recall rate for different futures and mean bar price structure allowing higher bounds for θ_{MIR} .

In the following, we will first compare the results to the case where we allow higher bound on θ_{VPIN} , to see if there is a difference. Second, we will benchmark both results to the one of a “naive” classifier.

3.2.6 Deep search allowing lower bounds for θ_{MIR} and higher bounds for θ_{VPIN}

We remark in **Tables 15–18** that compared to previous deep search:

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	θ_{VPIN}
ES	0.9421	0.9541	1.8962	0.015	30	2500	Student	0.99
EC	0.9080	0.9644	1.8724	0.006	30	2500	Gaussian	0.99
CL	0.9297	0.9806	1.9103	0.016	30	2500	Student	0.99
NQ	0.9076	0.9217	1.8293	0.02	50	2500	Student	0.999
YM	0.9460	0.9696	1.9156	0.015	50	2500	Gaussian	0.99

Table 15.
Best parameters maximizing precision+recall rate for different futures and last bar price structure allowing lower bounds for θ_{MIR} and higher bounds for θ_{VPIN} .

- There are changes only for NQ and YM instruments in, respectively, last, median, and mean bar price structures and first bar price structure, where θ_{VPIN} equals 0.999.
- There is no general trend for precision or recall rates with the increase of θ_{VPIN} .
- On average median bar price structure has the best precision+recall rate.

3.2.7 Benchmark with a “naive” classifier

We remark the following for the “naive” classifier (Table 19):

- It has worse results than VPIN on ES and YM cases.
- It has comparable results than VPIN on NQ case.

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	θ_{VPIN}
ES	0.9404	0.9402	1.8806	0.015	30	2500	Gaussian	0.99
EC	0.9127	0.9681	1.8808	0.006	30	2500	Student	0.99
CL	0.9232	0.9728	1.8960	0.016	30	2500	Student	0.99
NQ	0.8291	0.9833	1.8124	0.01	30	2500	Student	0.99
YM	0.9341	0.9872	1.9213	0.015	50	2500	Gaussian	0.999

Table 16.
Best parameters maximizing precision+recall rate for different futures and first bar price structure allowing lower bounds for θ_{MIR} and higher bounds for θ_{VPIN} .

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	θ_{VPIN}
ES	0.9499	0.9498	1.8997	0.015	30	2500	Student	0.99
EC	0.9037	0.9717	1.8754	0.006	30	2500	Student	0.99
CL	0.9265	0.9718	1.8983	0.016	30	2500	Student	0.99
NQ	0.8881	0.9525	1.8406	0.02	30	2500	Student	0.999
YM	0.9829	0.9427	1.9256	0.015	30	2500	Gaussian	0.99

Table 17.
Best parameters maximizing precision+recall rate for different futures and median bar price structure allowing lower bounds for θ_{MIR} and higher bounds for θ_{VPIN} .

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)	Classifier	θ_{VPIN}
ES	0.9526	0.9454	1.8980	0.015	30	2500	Student	0.99
EC	0.9058	0.9691	1.8749	0.006	30	2500	Student	0.99
CL	0.9302	0.9670	1.8972	0.016	30	2500	Gaussian	0.99
NQ	0.9188	0.9150	1.8338	0.02	40	2500	Gaussian	0.999
YM	0.9446	0.9779	1.9225	0.015	60	2500	Gaussian	0.99

Table 18.
Best parameters maximizing precision+recall rate for different futures and mean bar price structure allowing lower bounds for θ_{MIR} and higher bounds for θ_{VPIN} .

Futures	Recall	Precision	Precision+recall	θ_{MIR}	n	ω (buckets)
ES	1	0.7483	1.7483	0.01	40	2500
EC	1	0.9999	1.9999	0.001	60	2500
CL	1	0.9995	1.9995	0.01	40	2500
NQ	1	0.8465	1.8465	0.01	30	2500
YM	1	0.6892	1.6892	0.01	40	2500

Table 19.
Best parameters maximising precision+recall rate for different futures for the naive classifier allowing lower bounds for θ_{MIR} .

- It has better results than VPIN on EC and CL cases, where the flash crash is not really effective.
- It reaches obviously best local results on lowest MIR bound of the deep search.

We may partially conclude that:

- VPIN has an interesting predictive behavior on flash events of magnitude far lower (around 1.5%) than what would be considered as a crash for specific financial instrument (relatively liquid such as NQ, YM, or ES).
- But VPIN has poor results comparable to those of a “naive” classifier (precision +recall rate inferior to 1.2) on flash crash events for these financial instruments.
- For other instruments such as CL or EC, VPIN behaves worse than a naive classifier for these flash events. On flash events of higher amplitude (at least 1.5%), VPIN behaves better than a “naive” classifier for CL instrument.

4. VPIN sensitivity to the starting point of a data set

In this section, first we present the problem of VPIN’s sensitivity to the starting point of the bucketing process. Second, we present different calibrations to test its sensitivity. Third we make a summary of our results.

4.1 The problem

VPIN received among critics one which is important to precisely assess. Indeed, Bodarenko and Anderson [7] pointed out in their work that VPIN is sensitive to the starting point of the bucketing process. More precisely, if one removes the first buckets of the data set, results change. It is indeed right. We would like to know to which extent one can or cannot mitigate this effect. One idea is to test the different price bar structures. Indeed a bar structure influences trade imbalance and thus influences the appearance of VPIN events.

4.1.1 Methodology

There are at least two interesting ways of analyzing the sensitivity to the starting point of a data set:

- Study the sensitivity of best precision+recall rate to the number of trades erased and to the bar price option.
- Given one set of local optimal parameters, study the sensitivity of precision and recall rates to bar price option and data removed.

We have removed $l \in 0, 1000, 2000, 3000$ number of bars to study the sensitivity in the two previous cases, which corresponds to several hours of trading data removed. Indeed one does not want to erase first flash crash detected in the data set and erase more buckets than the average prediction length to detect it. Moreover we would like to study to which extent VPIN is locally sensitive.

4.2 Summary of results

4.2.1 Sensitivity of precision+recall rate

We summarize in **Table 20** for each bar price structure the average percentage change of local new best precision+recall rates with the number of bar erased. We remark the following:

- The sensitivity mentioned by Bodarenko and Anderson does exist.
- Its amplitude is not very big, at least for best precision+recall rate, as the maximum change is about 6%.
- Median bar price structure is far less sensitive than other price structure.

Bar price structure	1000 bars erased	2000 bars erased	3000 bars erased
Last	3.089	2.166	0.939
First	2.410	3.649	6.727
Median	0.611	0.801	0.781
Mean	1.348	2.149	3.944

Table 20.
Average absolute percentage change of local best precision+recall rates with the number of bar erased for each bar price structure.

Bar price structure	1000 bars erased	2000 bars erased	3000 bars erased
Last	1.192	1.171	1.067
First	1.612	1.725	1.049
Median	3.514	3.137	1.396
Mean	2.648	3.180	2.489

Table 21.
Average absolute percentage change of local best precision+recall rates with the number of bar erased for each bar price structure.

4.2.2 Sensitivity to local best parameter choice

In **Table 21** we summarize for each bar price structure the average percentabe change of the initial local best precision+recall rates with the number of bar erased.

We remark the following:

- Again the amplitude of the sensitivity is not very large as the maximum change is about 3.5%.
- Last bar price structure is less sensitive than other price structure to this phenomenon.

5. Conclusion

In this last section, we present first a general summary of our findings. Then we propose new suggestion of research concerning this precise subject.

5.1 Summary of results

We found that:

- VPIN has interesting predictive power (i.e., better than a naive algorithm and at least of local prediction+recall maximum higher than 1.2) for flash events of lower amplitude than flash crashes (about 1.5%) for a certain class of instruments, where flash crashes are at least present (which is not the case for currency Euro FX or Energy Light Crude NYMEX).
- VPIN is sensitive to the starting point of computation, but the amplitude of this sensitivity is not really high. For practice, which means not changing local best parameters while erasing some data, last bar price structure is the least sensitive to this phenomenon.

5.2 Suggestion for further studies

For further studies, this might be worth analyzing:

- Define a bigger constraint to capture crashes taking into account, for example, their V-shape. It would indeed filter out more events and enable analyzing more accurately which kind of crash VPIN predicts better.
- Benchmark within this framework other predictive tools between them (VIX with a naive algorithm, with VPIN, etc.).
- Analyze VPIN time-clock version predictive power.
- If previous predictive power of lower amplitude flash events is interesting for practitioners, analyze more precisely parameters that would be interesting for them.

- Describe more precisely to which class of financial instrument VPIN predictive power is most effective (if such one is worth being more studied for practitioners).
- Define a normalization of events defining crash events within a whole cluster of instruments. It is not easy to put in place as instruments are more or less correlated by crashes and response times are not trivial to analyze, but it would be also interesting indeed to assess prediction quality on common events shared by different instruments of a same cluster. It would make it possible to see whether or not VPIN predictive power is effective beyond different financial instruments embedding different aspects of the financial world to which VPIN is sensitive to.
- This area of research studies a very particular class of events: those that are potentially very rare. Taking into account this setting and that the algorithms used are fed with previous information and are sensitive to the starting point of computation, is it possible to build a consistent cross-validation approach? This aspect has not been treated yet as others needed to be first addressed, but it is still important to be studied.

Appendix

See Table 22.

Symbol	Description	Exchange	Class	Volume
ES	S&P500 E-mini	CME	Equity	478,029
EC	Euro FX	CME	Currency	188,837
CL	Light Crude NYMEX	NYMEX	Energy	165,208
YM	Dow Jones E-mini	CBOT	Equity	110,122
NQ	Nasdaq 100	CME	Equity	173,211

Table 22.
List of future contracts and their total volume of trades from January 2007 to July 2012.

IntechOpen

Author details

Antoine Bambade^{1*} and Kesheng Wu²

¹ École Polytechnique (Palaiseau), France

² Lawrence Berkeley National Laboratory, Scientific Data Management (SDM) group, USA

*Address all correspondence to: antoine.bambade@polytechnique.edu

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Easley D, de Prado ML, O'Hara M. Flow toxicity and liquidity in a high frequency world. *Review of Financial Studies*. 2012;**25**(5):1457-1493
- [2] Easley D, de Prado ML, O'Hara M. The microstructure of the 'flash crash': Flow toxicity, liquidity, crashes and the probability of informed trading. *The Journal of Portfolio Management*. 2011;**37**(2):118-128
- [3] Zheng Y. VPIN and the China's circuit-breaker. *International Journal of Economics and Finance*. 2017;**9**(12)
- [4] Abad D, Yagüe J. From PIN to VPIN: An introduction to order flow toxicity. *The Spanish Review of Financial Economics*. 2011;**6**(2):8-13. Johnson School Research Paper Series No. 10-2011
- [5] Wu K, Bethel EW, Gu M, Leinweber D, Rübel O. A big data approach to analyzing market volatility. *Algorithmic Finance*. 2013;**2**(3-4):241-267
- [6] Easley D, de Prado ML, O'Hara M. The exchange of flow toxicity. *The Journal of Trading*. 2011;**6**(2):8-13. Johnson School Research Paper Series No. 10-2011
- [7] Andersen TG, Bondarenko O. VPIN and the flash crash. *The Journal of Financial Markets*. 2014;**17**:1-46
- [8] Andersen TG, Bondarenko O. Reflecting on the VPIN dispute. *Journal of Financial Markets*. 2014;**17**:53-64
- [9] Andersen TG, Bondarenko O. Assessing measures of order flow toxicity and early warning signals for market turbulence. *Review of Finance*. 2015;**19**:1-54
- [10] Abad D, Massot M, Pascual R. Evaluating VPIN as a trigger for single-stock circuit breakers. *Journal of Banking and Finance*. 2017;**86**(C):21-36
- [11] Pöppe T, Moss S, Schiereck D. The sensitivity of VPIN to the choice of trade classification algorithm. *Journal of Banking and Finance*. 2016;**73**:165-181
- [12] Easley D, Engle RF, O'Hara M, Wu L. Time-varying arrival rates of informed and uninformed trades. *Journal of Financial Econometrics*. 2008;**6**(2):171-207
- [13] Easley D, Kiefer NM, O'Hara M, Paperman JB. Liquidity, information, and infrequently traded stocks. *Journal of Finance*. 1996;**51**(4):1405-1436
- [14] Easley D, de Prado ML, O'Hara M. The volume clock: Insights into the high frequency paradigm. *Journal of Portfolio Management*. Vol. 39, No. 1, (01 Sep Fall). 11
- [15] Ke W-C, Lin H-WW. An Improved Version of the Volume-Synchronized Probability of Informed Trading (VPIN)
- [16] Easley D, de Prado ML, O'Hara M. An Improved Version of the Volume-Synchronized Probability of Informed Trading (VPIN): A Comment