Doctor, is that (order)flow really toxic ?

Payment and Settlement System Simulation Seminar
Bank of Finland

Olivier Brandouy (GREThA, UMR 5113, U Bordeaux)
Philippe Mathieu (SMAC, UMR 8122, U Lille 1)

28–29/08/2014

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

Outline

Outline

Context

- Increasingly complex financial markets
- Liquidity Fragmentation : competition for order execution
- Increasing role of computers (up to 80 % of activity is monitored by boxes)
- $\nearrow$ number of orders $\searrow$ in their average size
- $\nearrow$ volatility of Markets, $\nearrow$ "extreme events" or market anomalies (such as the flash crash of May 6^{th}, 2010)

The research question

– In this context, how to plot "anomalies" in the behaviour of market participants ?
– **How can we distinguish** in this huge flow of orders **"toxic"** orders coming from traders trying to benefit from (private) information ?

– Possible answer : a metric denominated "VPIN", proposed by Easley, Lòpez de Prado and O'Hara (2012), , "Flow Toxicity and Liquidity in a High-frequency World", *Review of Financial Studies*, 25(1), pp. 1457–1493.

– We propose an empirical assessment of the VPIN in an ABASM allowing for a set of controlled experiments.

Main results

- the VPIN, most of the time, actually spots informed trading but this is not systematically the case. Failures may occur when
  - the population of traders is deeply heterogeneous,
  - sophisticated splitting algorithms are implemented
- Report situations in which VPIN levels could be interpreted as revealing some toxicity in the order flow although no informed agent trade.
- Risk that competitive market makers increase their spread althought this is not consistent with the state of the market.

Outline

A classical microstructure problem

$\rightarrow$ Easley, Kiefer, O'Hara and Paperman (1996) "Liquidity, Information and *Infrequently* Traded Stocks", *Journal of Finance*, 51(4), pp. 1405–1436

Punchline : model the price intervals in which a risk neutral market makers accepts to provide liquidity.

- time is continuous
- traders (both informed or uninformed) may enter the market at any moment.
- There is a risk neutral MM proposing quotes reflecting his expectation about the value of the traded asset.
- Authors assume that MM are Bayesian updaters learning from the rate of arrival of orders if a Good or a Bad information governs the order flow and the fix quotes to avoid losses.
- These quotes determine the BAS and from the BAS one can calculate the PIN

A classical microstructure problem (Cont.)

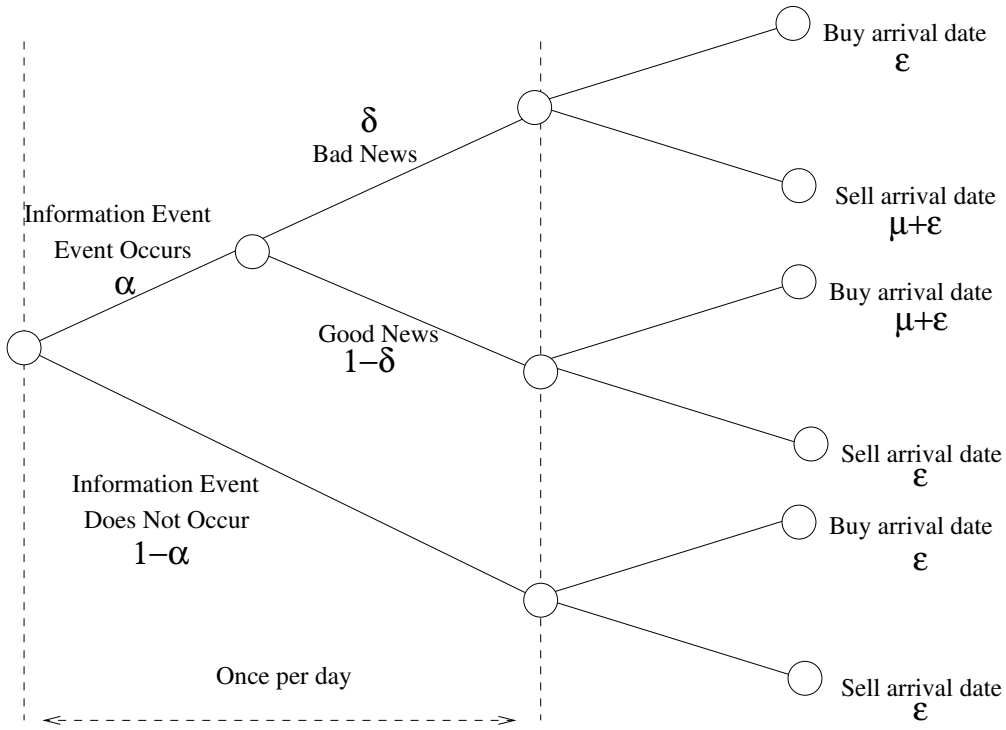


Figure : Trading process diagram Source : Easley, Kiefer, O'Hara and Paperman (1996)

Moving from infrequent trading to HFT

– Easley, Lòpez de Prado and O'Hara (2010), "The Microstructure of the 'Flash Crash': Flow Toxicity, Liquidity Crashes and the Probability of Informed Trading", Journal of Portfolio Management, 37(2), pp.118–128
– (2012), "Flow Toxicity and Liquidity in a High-frequency World", *Review of Financial Studies*, 25(1), pp. 1457–1493.

Define a new metric which is a proxy to the PIN : **Volume Synchronized Probability of Informed Trading** (VPIN)
Strength : $\Rightarrow$ easy to compute
Weakness(es) – IF ANY–: GOAL OF THE PAPER !

The VPIN

- V_θ^S and V_θ^B are volumes against Ask and Bid side of the book.
- The prob. that the flow contains toxic orders can be estimated by the ratio of :

  orders emitted by informed traders

$$E[|V_\theta^S - V_\theta^B|]$$

  volume generated by the overall activity

$$E[|V_\theta^S + V_\theta^B|]$$

- A good PIN estimator is :

$$VPIN = \frac{E[|V_\theta^S - V_\theta^B|]}{E[|V_\theta^S + V_\theta^B|]}$$

The VPIN (Cont.)

– A simple aggregation of the signed exchanged volumes in the market.
– Updated in volume-time, making it applicable to the high frequency world
– Does not require the intermediate estimation of non-observable parameters or the application of numerical methods
– Practical computation of the VPIN :

1. Define time bars over which one computes the probability
2. The time bar derives from the "Volume Bucket" that is chosen :$E[|V_\theta^S + V_\theta^B|]$
3. Calculate the imbalance $E[|V_\theta^S - V_\theta^B|]$

"How reliable is the VPIN in a HF world ?"

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

A controversial topic

- Critics
  - Andersen, T. G., Bondarenko, O., 2013. Assessing measures of order flow toxicity via perfect trade classification. CREATES Research Papers 2013-43, School of Economics and Management, University of Aarhus.
  - Andersen, T. G., Bondarenko, O., 2014. Reflecting on the VPIN dispute. Journal of Financial Markets 17, 53 – 64.
  - Andersen, T. G., Bondarenko, O., 2014. VPIN and the flash crash. Journal of Financial Markets 17, 1–46.
- Defenders
  - Easley, D., de Prado, M. M. L., O'Hara, M., 2014. VPIN and the flash crash: A rejoinder. Journal of Financial Markets 17, 47 – 52.
  - Wu, K., Bethel, W., Gu, M., Leinweber, D., Ruebel, O., 2013. Testing vpin on big data response to "reflecting on the vpin dispute". Tech. rep., SSRN e-library.

Outline

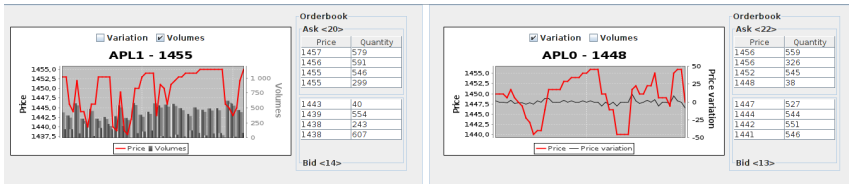
An Agent-based computational approach

- Signed volumes are hard to observe and in databases like TAQ, only quotes and prices are displayed (not the counterpart initiating the price)
- Solution consists in using a Lee-Ready algo to reconstruct these "signatures" (risk of errors)
- We want to overcome these limitations

$\Rightarrow$ Use a multi-agent system, archive quotes, orders and counterparts and develop an experimentation plan

$\Rightarrow$ The experiments run on the ABASM are inspired by the initial paper of Easley and al. (1996)

ATOM



`http://atom.univ-lille1.fr`

ATOM is a robust, flexible, and reliable platform, on which researchers can run experiments encompassing thousand of sophisticated evolving agents.

- Price-driven and order-driven systems
- Multi-asset order book
- Fast simulation engine: executes 400 000 orders in $<$ 4 seconds
- Many kinds of agents can co-exist : ZIT, Chartists, mean-variance optimizer...
- Agents heterogeneity: beliefs, memory span, risk aversion, trading frequency.

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

Outline

Agents

– Virtual autonomous entities (softwares) interacting through the market using their information and decision rules. – 3 types of agents

1. ZIT standing for uninformed traders wrt the reference paper.
2. ITRs standing for informed traders.
   - Small fishes (send small orders, weak potential impact on the market)
   - Big fishes (opposite). They can use splitting algorithms (depending upon the experiment)
3. MM posting quotes passively (no sophisticated strategy)

$\rightarrow$ Details follow (but will be skipped due to time constraints ! Back to the details if necessary)

Agents population : ZITs

Behavior based on randomness

- Only send Ask or Bid limit orders.
- Buy and Sell orders arise with equal probability ($p = 0.5$).
- Price chosen within $U \simeq [P_{min}, P_{max}]$
- Quantities : draw this value in the range $[10, 100]$

For all the experiments $P_{min} = 22.00$ and $P_{max} = 23.00$

Agents population : ITRs

Behavior based on information

- Send Ask or Bid limit orders depending upon the meaning of the signal
  1. Good $\rightarrow$ Buy : Bid between $[^{Best}Ask, S_G]$
  2. Bad $\rightarrow$ Sell : Ask between $[S_B, ^{Best}Bid]$

  For all the experiments $S_B = 22.10$ and $S_G = 22.9$
- ITRs may have different market impacts:
  1. "Small Fishes" : quantities are drawn in the range $[10, 100]$
  2. "Big Fishes" : quantities are drawn in the range $[50, 500]$

Splitting Algorithms

Depending upon the experimental treatment, we use on of the following splitting algorithm :

1. Naive Splitting : split in equal quantities and forwarded to the order book at each round.
2. Volume Weighted Average Price splitting strategy : adapt the quantities sent to the market wrt the current and expected volumes executed in the market.

Procedures are presented in the paper.

Agents population : MMs

Behavior based on best limits set by the pending orders and their volumes

- Ask quote : Best Ask from traders - 1
- Bid quote : Best Bid from traders + 1
- Done after an opening fixing (the book is full)
- Volumes proposed : $1.1 \times \sum_{i=1}^{n} V_{\theta}^{A_i}$ and $1.1 \times \sum_{i=1}^{n} V_{\theta}^{B_i}$
- In the experiments, $n = 2$
- Note that they do not adapt their spread to the VPIN. Their strategy remains unchanged during all the experiments.
- Further investigations will be made using a more sophisticated MM algorithm

Outline

Empirical strategy

Two related and complementary investigation lines

- Line 1: We want to understand whether the VPIN is accurate in spotting the presence of informed traders : (diagnosis failure).
- Line 2: We want to determine whether the metric is consistent and only detects the informed trading behaviour but no other types of uninformed trading : (erroneous diagnosis).

$\rightarrow$ We build an experimental protocol made of 10 families of experiments to answer our questions.

Experiments (2)

	Agent type	Num of Agents	Price range	Volume range	Market Maker	splitting Algorithm
Exp. 1	ZIT	100	[22.00, 23.00]	[10, 100]	No	No
Exp. 2	ZIT	100	[22.00, 23.00]	[10, 100]	Yes	No
Exp. 3	ZIT	95	[22.00, 23.00]	[10, 100]	Yes	No
	ZIT	5	[22.00, 23.00]	[50, 500]		No
Exp. 4	ZIT	95	[22.00, 23.00]	[10, 100]	Yes	No
	ZIT	5	[22.00, 23.00]	[50, 500]		Naive splitting
Exp. 5	ZIT	95	[22.00, 23.00]	[10, 100]	Yes	No
	ZIT	5	[22.00, 23.00]	[50, 500]		VWAP splitting
Exp. 6	ZIT	95	[22.00, 23.00]	[10, 100]	Yes	No
	ITR	5	[Best Ask, 22.80] if info "Good"	[10, 100]		
	ITR	5	[22.20, Best Bid] if info "Bad"			
Exp. 7	ZIT	95	[22.00, 23.00]	[10, 100]	Yes	No
	ITR	5	[Best Ask, 22.80] if info "Good"	[50, 500]		No
	ITR	5	[22.20, Best Bid] if info "Bad"			
Exp. 8	ZIT	95	[22.00, 23.00]	[10, 100]	Yes	No
	ITR	5	[Best Ask, 22.80] if info "Good"	[50, 500]		Naive splitting
	ITR	5	[22.20, Best Bid] if info "Bad"			
Exp. 9	ZIT	95	[22.00, 23.00]	[10, 100]	Yes	No
	ITR	5	[Best Ask, 22.80] if info "Good"	[50, 500]		VWAP splitting
	ITR	5	[22.20, Best Bid] if info "Bad"			
Exp. 10	ZIT	100→1, step = 1	[22.00, 23.00]	[10, 100]	Yes	No
	ITR	0→99, step = 1	[Best Ask, 22.80] if info "Good"	[50, 500]		VWAP splitting
	ITR		[22.20, Best Bid] if info "Bad"			

– Each experiment consists in 100 replications of one of these protocols

– We collect prices, volumes, orders, the make-up of the trading panel, and the VPIN over the experiments (distribution of VPINs, or dist of Mean VPIN, or dist of dispersion of VPINs).

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

Outline

VPIN and potential diagnosis failure

Is the VPIN accurate in spotting the presence of informed traders ?

– The baseline for this analysis consists in the distribution of VPIN obtained in Exp.2, where only ZIT co-evolve.
– We compare to this distribution of VPIN those of Exp. 6 to Exp. 9 where informed trading is implemented.

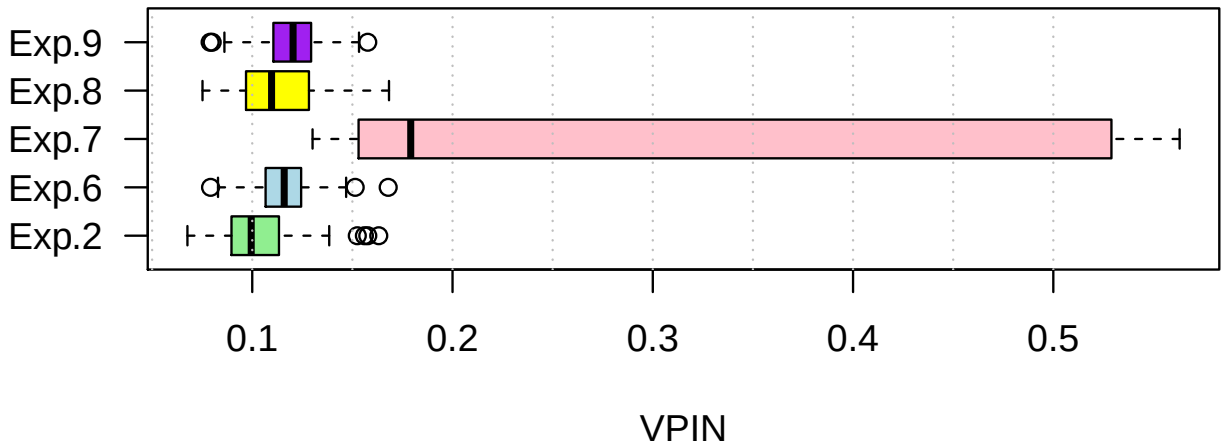


Figure : Levels of VPIN with (Exp. 6 to 9) and without ITRs (Exp. 2)

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

Statistical tests

Moving from the graphical representation to some statistics we :

- test if the VPIN samples come from the same distribution with a Kolmogorov-Smirnov test.
  For two samples D1 and D2, the null is : *"H_0 : D1 and D2 come from the same distribution"*.
- test if the distribution of each sample is symmetric around the median of the population with a Fligner-Policello test (equivalent to the Mann-Whitney test but without assuming equality in variance).

Statistical tests on Mean VPIN distributions

		Exp. 2	Exp. 6	Exp. 7	Exp. 8	Exp. 9
KS.Test	Exp. 2	0.00000	0.44000	0.93000	0.28000	0.55000
	p.value	1.00000	0.00000	0.00000	0.00079	0.00000
FP.Test	Exp. 2	0.00000	-6.34906	-69.02280	-3.88000	-9.22364
	p.value	1.00000	0.00000	0.00000	0.00010	0.00000

Statistical tests on VPIN SD distributions

		Exp. 2	Exp. 6	Exp. 7	Exp. 8	Exp. 9
KS.Test	Exp. 2	0.00000	0.42000	0.83000	0.35000	0.41000
	p.value	1.00000	0.00000	0.00000	0.00001	0.00000
FP.Test	Exp. 2	0.00000	-6.51239	-36.37792	-5.75280	-5.95439
	p.value	1.00000	0.00000	0.00000	0.00000	0.00000

Statistical tests

$\rightarrow$ First intuition :

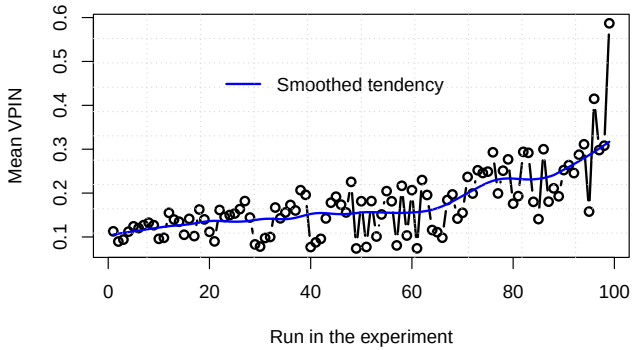
– H_0 can be rejected for all the cases

– The VPIN, on average, actually appears to detect situations where Informed Trading is implemented since one can reject the Null

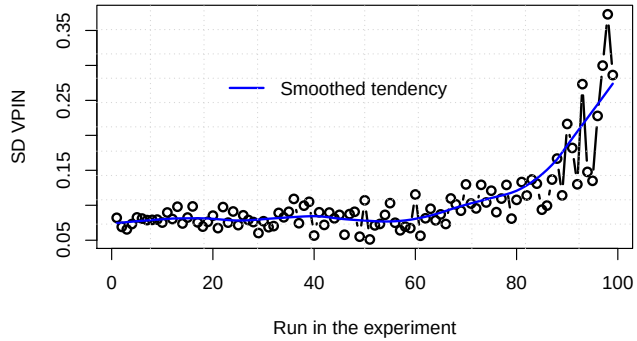
Is it always the case ? (1)

– Analysis of Exp. 10
– We increase the potential market impact of ITRs in substituting at each step, and from a population of ZIT, one ZIT with one ITR and observe how the VPIN evolves along this treatment.
– ITRs remain out of market prior the disclosure of information at round 50 within an experiment.

Is it always the case ? (2)



(a) Mean of the VPIN



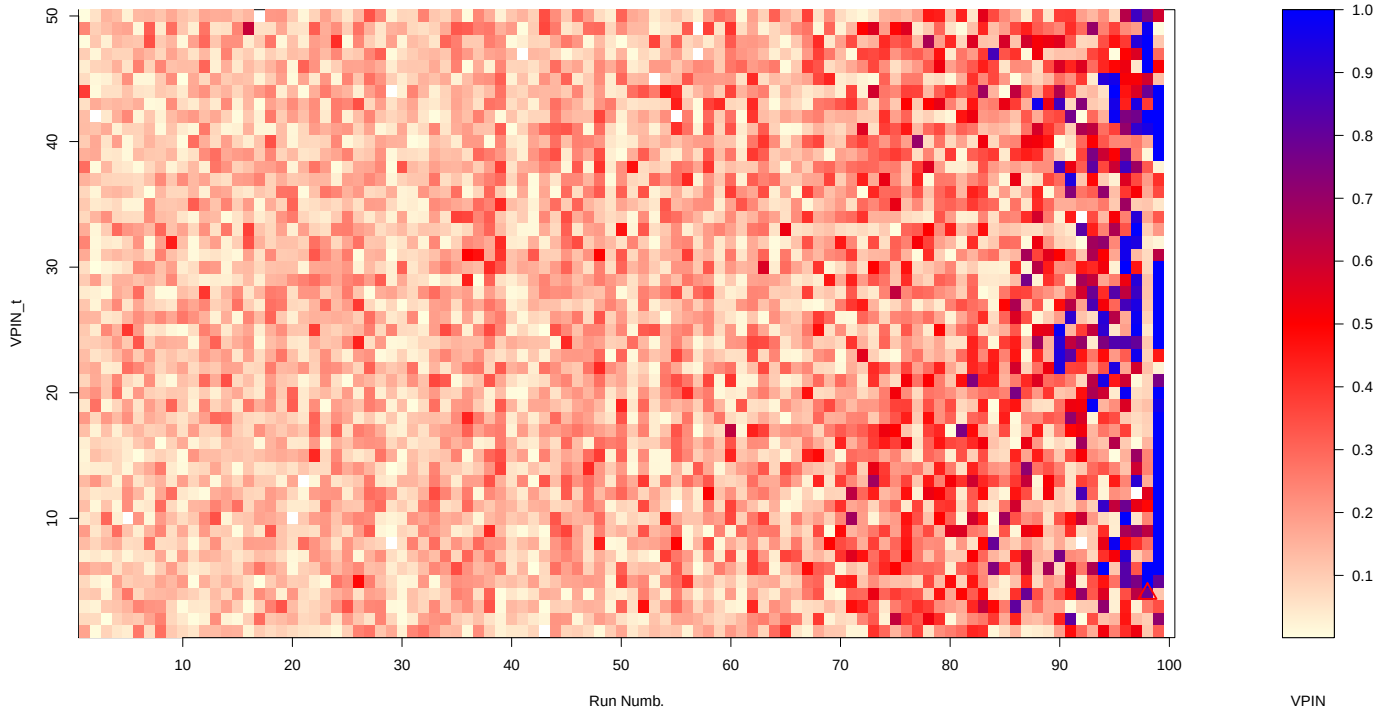
(b) Standard deviation of the VPIN

Ambiguous results:

- global evolution of the VPIN is clearly linked to the weight of ITRs
- but frequently low levels of VPIN can be observed (around runs 40 and 60 for example) within a specific run, although a large number of ITRs trade.

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

Is it always the case ? (3)



The X-axis indexes the different runs. Y-axis indexes the observations of the VPIN for each run

Idiosyncratic evolution of the VPIN for experiment 10. Visually speaking, the obviousness of the impact of ITRs only appears around the round 70.

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

Synthesis

1. VPIN globally captures informed trading even when ITRs do use splitting algorithms (COARSE GRAIN analysis)
2. However, in several situations where ITRs do trade in the market, the VPIN fails at detecting the toxic orders (FINE GRAIN analysis)

Not a full-proof tool !

Outline

Logic of the analysis

Does the VPIN only spot informed trading and no other type of uninformed trading ?

- Based on experiments 3 to 9 :
  - Exp. 3, 4 and 5: no insider trading is implemented
  - Exp. 7,8, and 9: insiders evolve in the market
- we want to compare the levels of VPIN in cunjunction with the mean VPIN or standard deviation distributions of the VPIN in different experimental treatments
- we use KS tests and FP as previously

Levels of VPIN and exp./exp. comparisons

		Subset 1			Subset 2		
		Exp.3	Exp.4	Exp.5	Exp.7	Exp.8	Exp.9
Subset 1	Exp.3	0.177	+	+	-	+	+
	Exp.4	-	0.129	+	-	+	+
	Exp.5	-	-	0.119	-	+	-
Subset 2	Exp.7	+	+	+	0.338	+	+
	Exp.8	-	-	-	-	0.114	-
	Exp.9	-	-	+	-	+	0.121

Table : Relations between the grand means of VPINs by pairs of experiments

Rows must be read before columns; for example, the grey cell has to be interpreted : "The VPIN grand mean of Exp. 7 is larger than the one of Exp. 8. The Grand mean of each Experiment is reported in the diagonal of the matrix.

Same has been done for VPIN sd. See paper.

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

Statistical tests (1)

KS tests (on mean VPIN)

		Subset 1			Subset 2		
		Exp. 3	Exp. 4	Exp. 5	Exp. 7	Exp. 8	Exp. 9
Subset 1	Exp. 3	0.00	35.32	65.50	-1.98	51.21	72.93
	p.value	1.00	0.00	0.00	0.05	0.00	0.00
	Exp. 4		0.00	4.01	-30.16	5.72	3.38
	p.value		1.00	0.00	0.00	0.00	0.00
	Exp. 5			0.00	-56.32	2.35	-0.78
	p.value			1.00	0.00	0.02	0.44
Subset 2	Exp. 7				0.00	39.51	67.03
	p.value				1.00	0.00	0.00
	Exp. 8					0.00	-2.90
	p.value					1.00	0.00
	Exp. 9						0.00
	p.value						1.00

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

Statistical tests (2)

FP Tests (on mean VPIN)

		Subset 1			Subset 2		
		Exp. 3	Exp. 4	Exp. 5	Exp. 7	Exp. 8	Exp. 9
Subset 1	Exp. 3	0.00	0.74	0.87	0.49	0.84	0.86
	p.value	1.00	0.00	0.00	0.00	0.00	0.00
	Exp. 4		0.00	0.26	0.73	0.41	0.23
	p.value		1.00	0.00	0.00	0.00	0.01
	Exp. 5			0.00	0.85	0.24	0.12
	p.value			1.00	0.00	0.01	0.47
Subset 2	Exp. 7				0.00	0.82	0.84
	p.value				1.00	0.00	0.00
	Exp. 8					0.00	0.32
	p.value					1.00	0.00
	Exp. 9						0.00
	p.value						1.00

GREThA
Groupe de Recherche en Économie Théorique et Appliquée

Other tests and overall result

– The same tests are also run on the VPIN dispersion.

- KS and FP tests lead to a rejection of the null within each group of experiments
- The levels of VPIN are sufficiently different to do so
- However KS test fails at distinguishing one situations where ITRs are compared to ZIT when both use a sophisticated VWAP splitting algorithm (Exp. 5 vs. Exp 9)
- Similar situations are reported when one studies the VPIN dispersion

Risk of erroneous diagnosis *wrt* the presence of ITRs (for example, conclude, in a situation close to Exp. 5 that ITRs actually populate the market) –

Conclusion (1)

- We investigate the VPIN pas a coherent and consistent measure of the "order flow toxicity" in High Frequency markets
- We use a price-driven, asynchronous, agent-based artificial market
- We run massive experimentations and run a set of (non-parametric) statistical tests over them.
- Our results suggest that VPIN spots, most of the time, informed trading

Conclusion (2)

- However :
  - this is not necessarily always the case notably when the population is deeply heterogeneous and when sophisticated splitting algorithms are implemented.
  - We also document situations in which VPIN levels could be interpreted as revealing some toxicity in the order flow, although no informed agent trade.

These results suggest that competitive market makers need to fine tune their estimate of the VPIN, not to be mislead in their decisions about the spread.

Thank you for your attention !

olivier.brandouy@u-bordeaux.fr

GREThA
Groupe de Recherche en Économie Théorique et Appliquée