Quantitative Finance
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rquf20

Real-time market microstructure analysis: online transaction cost analysis

R. Azencott[ab], A. Beri[c], Y. Gadhyan[d], N. Joseph[e], C.-A. Lehalle[f] & M. Rowley[g]
[a] Department of Mathematics, University of Houston, Houton, TX, USA
[b] Ecole Normale Supérieure Cachan, Cachan, France
[c] Mathematical Biosciences Institute, The Ohio State University, Columbus, OH, USA
[d] Cluster Innovation Center, University of Delhi, Delhi, India
[e] Crédit Agricole Cheuvreux, Quantitative Research, Paris, France.
[f] CFM, Quantitative Research, 23/25, rue de l'Université, Paris 75007, France
[g] Rowley Consulting LLC, Quantitative Research, Princeton, NJ, USA
Published online: 31 Mar 2014.

Quantitative Fnance, 2014
http://dx.doi.org/10.1080/14697688.2014.884283

Real-time market microstructure analysis: online transaction cost analysis

R. AZENCOTT†‡, A. BERI§, Y. GADHYAN¶, N. JOSEPH‖, C.-A. LEHALLE*†† and M. ROWLEY‡‡

†Department of Mathematics, University of Houston, Houton, TX, USA
‡Ecole Normale Supérieure Cachan, Cachan, France
§Mathematical Biosciences Institute, The Ohio State University, Columbus, OH, USA
¶Cluster Innovation Center, University of Delhi, Delhi, India
‖Crédit Agricole Cheuvreux, Quantitative Research, Paris, France
††CFM, Quantitative Research, 23/25, rue de l'Université, Paris 75007, France
‡‡Rowley Consulting LLC, Quantitative Research, Princeton, NJ, USA

(*Received 10 March 2013; accepted 6 January 2014*)

Motivated by the practical challenge in monitoring the performance of a large number of algorithmic trading orders, this paper provides a methodology that leads to automatic discovery of causes that lie behind poor trading performance. It also gives theoretical foundations to a generic framework for real-time trading analysis. The common acronym for investigating the causes of bad and good performance of trading is transaction cost analysis Rosenthal [*Performance Metrics for Algorithmic Traders*, 2009]). Automated algorithms take care of most of the traded flows on electronic markets (more than 70% in the US, 45% in Europe and 35% in Japan in 2012). Academic literature provides different ways to formalize these algorithms and show how optimal they can be from a mean-variance (like in Almgren and Chriss [*J. Risk*, 2000, **3**(2), 5–39]), a stochastic control (e.g. Guéant *et al.* [*Math. Financ. Econ.*, 2013, **7**(4), 477–507]), an impulse control (see Bouchard *et al.* [*SIAM J. Financ. Math.*, 2011, **2**(1), 404–438]) or a statistical learning (as used in Laruelle *et al*. [*Math. Financ. Econ.*, 2013, **7**(3), 359–403]) viewpoint. This paper is agnostic about the way the algorithm has been built and provides a theoretical formalism to identify in real-time the market conditions that *influenced* its efficiency or inefficiency. For a given set of characteristics describing the market context, selected by a practitioner, we first show how a set of additional derived explanatory factors, called *anomaly detectors*, can be created for each market order (following for instance Cristianini and Shawe-Taylor [*An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, 2000]). We then will present an online methodology to quantify how this extended set of factors, at any given time, predicts (i.e. have *influence*, in the sense of predictive power or information defined in Basseville and Nikiforov [*Detection of Abrupt Changes: Theory and Application*, 1993], Shannon [*Bell Syst. Tech. J.*, 1948, **27**, 379–423] and Alkoot and Kittler [*Pattern Recogn. Lett.*, 1999, **20**(11), 1361–1369]) which of the orders are underperforming while calculating the predictive power of this explanatory factor set. Armed with this information, which we call *influence analysis*, we intend to empower the order monitoring user to take appropriate action on any affected orders by re-calibrating the trading algorithms working the order through new parameters, pausing their execution or taking over more direct trading control. Also we intend that use of this method can be taken advantage of to automatically adjust their trading action in the post trade analysis of algorithms.

Keywords: Transaction costs; Performance evaluation; Learning and adaptation; Market microstructure; Empirical time series analysis

JEL Classification: C14, C81

1. Introduction

Institutional investors use optimal dynamic execution strategies to trade large quantities of stock over the course of the day. Most of these strategies have been modelled quantitatively to guarantee their optimality from a given viewpoint: it can be from an expectation (see Bertsimas and Lo (1998)), a mean-variance (like in Almgren and Chriss (2000)), a synchronized

*Corresponding author. Email: charles@lehalle.net

portfolio (e.g. Lehalle (2009)) or a stochastic control (Bouchard *et al.* 2011) perspective.

Algorithmic trading analyzes high frequency market data, viewed as actionable information, in order to automatically generate and place automated trading orders. Hence, automatic trading algorithms can be viewed as parameterized black boxes. They are initialized and monitored by human traders, who have the capability to set and adjust some high-level parameters that drive the algorithm—incorporating the trader's view and allowing reaction to unexpected events. However standard automatic trading systems, do not currently offer advanced computerized discovery of potential causes for poor trading performance. Our paper focuses on automated online monitoring of portfolio performance by real-time scanning of static (like the sector, the country, etc., of the traded stock) and dynamic (like the bid-ask spread, the volatility, the momentum, the fragmentation of the traded stock) 'explanatory' market variables quantifying their current influence on portfolio performance. The *Influence Analysis* methodology we have developed and tested can provide real-time feedback to traders by detecting the most significant explanatory factors influencing current degradations of trading performance on specific portfolios.

Transaction cost analysis (TCA) practitioners, in particular for real-time analysis, are faced with the problem of automating the understanding of how market context affects trading performance of a large number of orders. This paper is the first theoretical formalization of such a process proposing a framework to understand and improve TCA (off line or in real-time) in several aspects: (1) augmenting the description of the market context (using scoring Section 2.3 and pattern detection Section 3) to identify relationships between this enhanced description and the performance of a large basket of orders. A key feature of our influence analysis methodology is that it does not require actual knowledge of the trading algorithms' mechanisms. Online influence analysis could hence be useful in updating trade executions as well as in re-calibrating trading algorithms. Influence analysis on historical data could also improve existing trade scheduling algorithms, selecting new significant signals in their kinematics.

Our online influence analysis generates in real-time a very short list of market factors 'explaining' the current lack of performance in a collection of intra-day trades. The automatic trading algorithms can be arbitrary; it can be a result of a classical mean-variance optimization (see Almgren and Chriss (2000)), stochastic control of a brokerage benchmark (see Bouchard *et al.* (2011)), stochastic control of a market making scheme (see Guéant *et al.* (2013)), a liquidity-seeking algorithm driven by a stochastic algorithm optimization (see Laruelle *et al.* (2013)), or even purely heuristic driven ones. The analysis conducted here is supported by automatic detection of the conjunction of poor trading quality and singularity or rarity of the market context. The influence analysis can be launched as soon as a portfolio performance evaluation criterion is selected, such as the slippage with respect to the arrival price, to the VWAP (volume-weighted average price), etc. Our online influence analysis relies on extensions of classical relative entropy techniques (see for instance Brillinger (2004), Mougeot and Azencott (2011), Azencott *et al.* (2007), Billingsley (1965)) to generate in real-time optimized empirical relations between an automatically selected small set of high influence explanatory factors and any pre-assigned trading performance criterion. Our approach also quantifies at each time t the current influence of a given pool of market factors on trading performance degradation.

We first describe typical sets of market variables and trading performance criteria to which our influence analysis applies, and we outline our benchmark sets of intra-day data, used to test our approach. In Section 2 we describe the market dataset considered for our study. We then present in Section 3 the three online anomaly detectors we have developed to enrich in real-time any set of dynamic input market variables. Such detectors are a bespoke implementation of a more generic class of detectors that could be used like wavelet coefficients (see Mallat (2008)). Section 5.3 outlines our generic framework for influence analysis. We select and fix any pragmatic binary trading performance criterion Y_t detecting low trading performance at time t. This binary criterion will be deduced from a variable of performance PE_t of the trading portfolio. Given any small set G of explanatory variables $\mathbb{X}^G = \{X_t^j, j \in G\}$ (deduced from a multiscale analysis of market context descriptors $\{M_t^j, 1 \leq j \leq J\}$), we generate the current 'best' predictor $\hat{Y}_t = h_t(\{X^G\})$ of Y_t based on these explanatory variables, and we compute its current predictive power, which we call the influence coefficient $J_t(G)$ of the group G on Y_t The time dependent set $G_t^{\max}$ which maximizes $J_t(G)$ among all small groups G of explanatory factors is then determined, and if $J_t(G_t^{\max})$ is high enough, the set G_t^{max} can be exported in real-time to traders, as the set of market variables which best explains current trading performance degradations. In Section 6.2 we present the accuracy analysis of the influence computation, and obtain pragmatic conditions for robust identification of explanatory variables. We then present the steps to compute optimized parametric predictors based on single explanatory factors, and to generate hierarchical combinations of optimal predictors. Section 9 presents the test results of our influence analysis on benchmark intra-day datasets provided by Crédit Agricole Cheuvreux Quantitative Research Group. Section 10 concludes.

2. Dynamic dataset to be monitored online

2.1. The automated trading process

Only few assumptions are demanded to a trading process to be monitored by the methodology proposed here. By 'trading process' we mean the operation of buying or selling shares or any other financial instrument in more than one transaction. Since the 'last leaf' of any investment or hedging strategy is to obtain transactions, a trading process is needed if block execution cannot be obtained. With the conjunction of the financial crisis and regulation changes (mainly Reg NMS in the US in 2005 and MiFID in Europe in 2007, see Lehalle *et al.* (2013) for details), the capability to close a deal in one transaction strongly decreased. Hence most market participants are 'slicing' their large orders (see Almgren and Chriss (2000), Bouchard *et al.* (2011) or Pagès *et al.* (2011) for quantitative approaches of optimal slicing). At a δt min time scale (say $\delta t = 5$), an important variable of a trading process is its 'participation rate' $\rho_{m \cdot \delta t}$; the trader (manually or tuning parameters of some trading robots), succeeds in obtaining $\rho_{m\cdot\delta t}\, V_{m\cdot\delta t}$ shares during the mth interval of δt min when the whole market traded $V_{m\cdot\delta t}$ shares.

The adequate trading rate is difficult to achieve. First because its optimal value is a function of market conditions during the whole trading process (i.e. it is not causal). Secondly, market conditions are either price driven (like the volatility, the presence of price jumps or trends, correlations between some assets, etc.) or liquidity driven (the bid-ask spread, the market depth, etc.); that all of these are difficult to anticipate or predict. Moreover, even if the theoretically optimal trading rate would have been known in advance, the actions of the trader himself have an impact on future market conditions (see Moro *et al.* (2009)). A trader is thus continuously monitoring his trading rate in conjunction with market conditions to check: first that ex-post his trading rate has been close to the expected one, second that a change in his trading rate does not come from an unexpected change of market conditions, and last that he is not impacting the price formation process (PFP). The faster he understands what is happening to the trading process, the more efficiently he will be able to react.

The motivation to buy or sell does not change the way to apply the 'influence analysis' methodology presented in this paper. It can be driven by long term considerations (typical for orders sent by institutional investors to executing brokers' trading desks), to hedge a derivative portfolio, to implement an arbitrage strategy, or even inside a market making scheme (see Guéant *et al.* (2013)). For all of these motivations the same trading process takes place: a human being monitors a trading rate according to performance criteria (that are specific) and tries to adjust his trading rate as fast as possible to take into account changes in market conditions. The methodology proposed in this paper provides decision support to the trader.

2.1.1. Performance criteria of a trading process. In our framework, the proxy for the target of the trader is called his *performance evaluation criterion*. Let us define a few possible criteria:

- for a market-maker: a decreasing function of his inventory imbalance, and his profits are good performance criteria;
- for a brokerage trading desk: a decreasing function of the distance to a fixed participation rate (for instance 10 or 20% of the market traded volume), the obtained average price compared to the VWAP of the market, or to the arrival price, or to the close price, are meaningful criteria; and
- for an arbitrageur: a decreasing function of his tracking error, and his profits should be chosen.

2.1.2. Description of the market context. In addition to the performance evaluation criteria, we will need *market descriptors* to quantify the market context. They will be used by the proposed methodology to build an online understanding of the causes of bad performance. Typical market descriptors are:

- *Price driven market descriptors*:
  - Prices returns or price momentum (signed to be in the same direction as the side of the monitored order) are important since it is more difficult to buy when the price is going up rather than when it is going down.
  - The volatility is a common proxy for the amount of uncertainty in the price dynamics. Its influence on trading performances is not straightforward since some volatility can help to capture passively some flows, but a too high volatility level can lead to adverse selection. Moreover, market impact models link the volatility to the impact of a trade a negative way (see Almgren *et al.* (2005)), meaning that a too high level of volatility is negative for almost all trading processes.
- *Liquidity driven market descriptors*:
  - The bid-ask spread is the distance (in basis points of the current price) between the best ask price and the best bid one. It thus describes the state of the 'auction game' between market participants (see Lehalle (2013) for more details and terminology).
  - Market traded volume is an important characteristic of the market since it is easier to buy or sell when the market is active than when nothing is traded.
  - The visible size at first limits (also called *average volume on the books*) can also be used to quantify the current *market depth*.

Any other characteristic of the trading instruments can be added in the analysis. For shares, modal variables like the sector of the stock or the market capitalization of the listed firm are of importance.

2.1.3. Anomaly detectors. To be able to capture the causes of bad trading performance, it is often useful to have access to information other than the averaged values of market variables. For instance price jumps, price trend changes, volume peaks and crenels are not captured by averages and we would like to take them into account in our *influence analysis* methodology. Section 3 shows how to build such detectors that will be used in the analysis.

2.1.4. A priori and choice of variables. Our approach is non parametric, since we do not need any modelled relationship between the considered variables; nevertheless it is worthwhile to give some clues to the reader on usual models considered in optimal trading. Our selection of variables has been guided by usual models, stylized facts on market microstructure (see Lehalle *et al.* (2013) for more details) and practice.

As it is stated earlier: volatility, bid-ask spread, and traded volumes are the variable commonly used in quantitative trading. For instance in their seminal paper Almgren and Chriss (2000) provide a model dedicated to optimize the trading rate of an order thanks the following decomposition of the value of a large trade of size v^* over T discrete time intervals (buying v_t on the tth interval):

$$W = \underbrace{v\, S_0}_{\text{price without frictions}} + \underbrace{\sum_t x_t \sigma_t \xi_t}_{\text{price uncertainty}} \underbrace{- \kappa \sum_t \sigma_t (x_t - x_{t+1}) \left(\frac{x_t - x_{t+1}}{V_t}\right)^{\gamma}}_{\text{liquidity cost}}, \tag{1}$$

where $x_t = \sum_{s \geq t} v_t$ is the remaning quantity to buy at t, ξ_t the innovation of a Gaussian random walk, γ and κ parameters of the market impact and V_t the traded quantity by the whole market during the tth time slice.

Sometime the second term is named *market risk* instead of *price uncertainty* and the third one *market impact* instead of *liquidity cost*. We would like to monitor if the behaviour of these terms is *the expected one* or not, but only if it impact the quality of the trading process (i.e. the value of W).

We added a momentum component that is not often used in models since academic proposals are often agnostic to the direction of the market, leading to martingale assumptions on the price dynamics. Nevertheless, once we are monitoring in real-time the price moves, a variable is needed to capture adverse or favourable ones. Typically favourable moves (i.e. negative trends for a buy order) will have as a consequence an unexpected increase of the realized trading rate, and often lead to adverse selection (it could have been better to buy in few minutes if the price continues to fall). And adverse moves (positive trends for a buy orders), should imply an unexpected decrease of the trading rate; being late the trading algorithm will probably have to 'catch the falling knife' to cope with its planned trading rate, thus increasing its average price W.

Detecting abnormal behaviours of the monitored variables (abnormal in the sense that they are not following ex ante martingale of diffusive assumptions) can be done by likelihood computations, scoring, or pattern detection. We choose to use the two latter approaches, and provide the methodology to implement such computations, but a given practitioner could choose other ways to build these detectors. We mainly focus on how to relate the emergence of unlikely values of our variables with bad performances of the trading algorithms.

Our approach targets to capture dependencies between potential explanatory market descriptors and performances; it will nevertheless have to deal with dependencies between market descriptors themselves, as it will be observed latter.

2.2. *Trading orders*

We consider a portfolio of at most K trading orders $\mathbf{T}(k)$, $k = 1, \ldots, K$ driven by automatic trading algorithms, and supervised by one or more traders. Each trading order is defined by a few 'static' variables such as buy/sell label, order size, trading place, section, country, capitalisation, free-float, benchmark type (VWAP, arrival price, etc.), etc. In our intraday benchmark studies a portfolio typically involves $200 \leq K_t \leq 700$ active orders at any arbitrary 5 min time slice.

2.3. *Market descriptors*

Each trading order $\mathbf{T}(k)$ focuses on a specific asset whose dynamics is recorded at each time point t through a fixed number of basic 'market descriptors' $M_t^1(k)$, $M_t^2(k)$, $M_t^3(k)$, ...; in our benchmark study below, we have focused on a subset of the following market descriptors:

- M^1 = Volatility,
- M^2 = Bid-Ask spread,
- M^3 = Momentum in bid-ask spread,
- M^4 = Momentum in basis points (bp).

This list can be augmented by the *rarity scores* Score(M^i) of the market variables M^i. These scores are defined by Score $(M^i) = F^i(M^i)$ where F^i is the cunulative distribution function of M_i.

This list can be augmented by the *rarity scores* Score(M^i) of the market variables M^i. Recall that for any random variable V the rarity score is formally defined by $Score(V) = F(V)$ where F is the cumulative distribution function of V, and Score(V) always has a uniform distribution (see Borovkov (1998)).

If poor performance on a given set of stocks is due to a strong increase in the volatility level, the concrete cause may either be due to volatility reaching an 'absolute' psychological threshold, or to volatility being high *relatively to its usual levels*. In this last case the volatility score will be a better explanatory factor for poor performance. We will use here the following scores, increasing our number of market variables:

- M^5 = Volume Rarity Score,
- M^6 = Volatility Rarity Score,
- M^7 = Spread Rarity Score.

2.3.1. Empirical computation of rarity scores. At time slice n, the last $n-1$ successive observations of the market variable V, namely $[V_1, \ldots, V_{n-1}]$ are available . The best natural estimate S_n of Score(V_n) is given by $S_n = R_n/n$ where R_n is the rank of V_n in the sequence $[V_1, \ldots, V_n]$. When the V_i are independent, the accuracy of the 'quantile' estimate S_n is roughly proportional to $1/n^{1/2}$ for moderately large $n > 100$ since $n^{1/2}(S_n - Score(V_n))$ is asymptotically Gaussian as $n \to \infty$. Robust 95% confidence intervals for S_n are provided by the Woodruff formulas (see Woodruff (1952)) Indeed the easy to use Woodruff formulas have been shown (see Sitter and Wu (2001)) to perform surprisingly well even when V_n does take rare values.

When the correlations between V_i and V_{i+k} are roughly bounded by $a\rho^k$ for some positive constants $0.05 < a < 1$ and $\rho < 1$ the accuracy and confidence intervals for S_n can be handled similarly provided n is replaced by the smaller value n/r where $r = \max 1, \log(50a)/\log(1/\rho)$.

2.3.2. Short comments on market descriptor choices. For our analysis, we use the three main variables used in classical optimal trading theory (like in equation (1)) : volatility, bid-ask spread, and volumes. We added the momentum; this variable is difficult to estimate a priori, but in this analysis we just need to estimate the momentum during the last few minutes of trading, thus this very short term ex-post measurement is possible. We decided to express the volatility and the bid-ask spread in their usual units (bp per 10 min for intraday volatility, estimated using the Garman-Klass estimator (see Garman and Klass (1980)), and bp for the bid-ask spread). The momentum is expressed in two units, giving birth in two different descriptors: one in bid-ask spread (putting the emphasis on a balance between the recent price moves and the level of

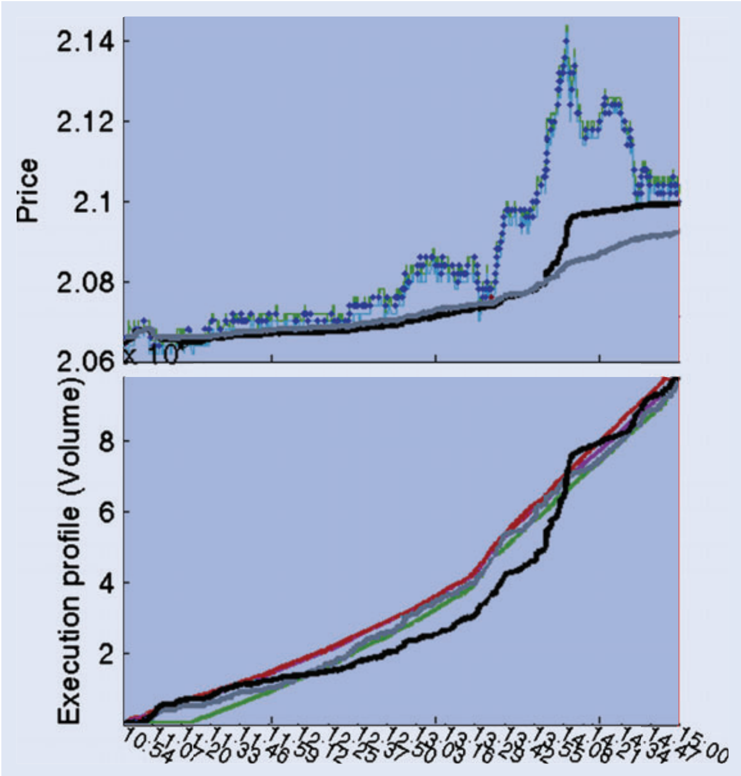

Figure 1. Example of intraday behaviour of performance variables for a given order. Top: the prices (the variations of the average obtained price is in gray, the market VWAP in dark); bottom: the cumulative traded volume (grey) and market volume (dark).

liquidity during this move), and the other in basis points (being a straightforward measurement of the move, potentially useful to identify cross sectional price moves across different traded stocks).

We use the volume not in its natural unit (i.e. a *number of traded shares* for a stock), but only as a rarity score, since it is difficult to compare the number of traded shares on two different stocks. First because the number of share is a function of the free float of the stock (i.e. its tradable capitalization) and second because the intensity of the volume seasonality (intraday, weekly, monthly, etc.) vary a lot from a security to the other. The *Volume Rarity Score* is far easier to compare from one stock to the other.

For potential ease of homogeneity, we added a *Volatility Rarity Score* and a *Bid-Ask Spread Rarity Score*.

Figure 1 displays typical intraday (14 January 2011) plots of the time series corresponding to the seven basic market variables listed above for an anonymous stock.

We also display heat map representations of rarity scores for multiple stocks in figures 2 and 3, where each row displays the time series of rarity scores for one single stock (associated to one trading order in our benchmark data), and each column represents one time slice. Clearly 'extreme' rarity scores tend to appear in clusters, and to co-occur across multiple stocks.

As will be seen below, high co-occurrence frequency of 'dynamic anomalies' such as peaks, jumps, etc. within a group of stocks tend to 'explain' simultaneous lack of performance for the corresponding trading orders (figure 4).

2.4. Trading performance evaluation

We also select a 'trading performance evaluator' PE providing at each time point t, and for each active trading order $\mathbf{T}(k)$, a quantitative evaluation $\mathrm{PE}_t(k)$ for the current performance of $\mathbf{T}(k)$.

In our study we have selected by default PE = 'slippage in bid-ask spread' (*slippage* being the average price of the order minus the benchmark—VWAP, arrival price, close price, etc.—for a sell order, and the opposite for a buy order), but there are no restrictions on the user choice for this PE variable. In particular, other examples of PE include 'slippage in bp' (basis points), 'Slippage in Dollars', 'Absolute value of slippage in bid-ask spread', etc. We are assuming that degraded performances are associated to low values of PE. For each trading order $\mathbf{T}(k)$ the performance evaluator and the seven market descriptors are volume averaged over successive time slices of arbitrary duration (set at 5 min for our benchmark study). Thus we generate eight time series $M_t^1(k), \ldots, M_t^7(k)$ and $\mathrm{PE}_t(k)$ indexed by time slices t. These time series generically have missing values since orders do not necessarily begin or end at the same time. Fix a low percentile threshold q such as $q = 3\%$ to binarize the performance evaluator.

At time slice t, call $K_t \leq K$ the number of currently active trading orders $\mathbf{T}(k)$. The $q\%$— quantile of the corresponding K_t performance evaluations $\mathrm{PE}_t(k)$ is denoted by ℓ_t. We consider ℓ_t as a PE-threshold, separating 'bad trading performances' (tagged '1') from 'normal trading performances' (tagged '0'). We then *binarize the performance evaluations*

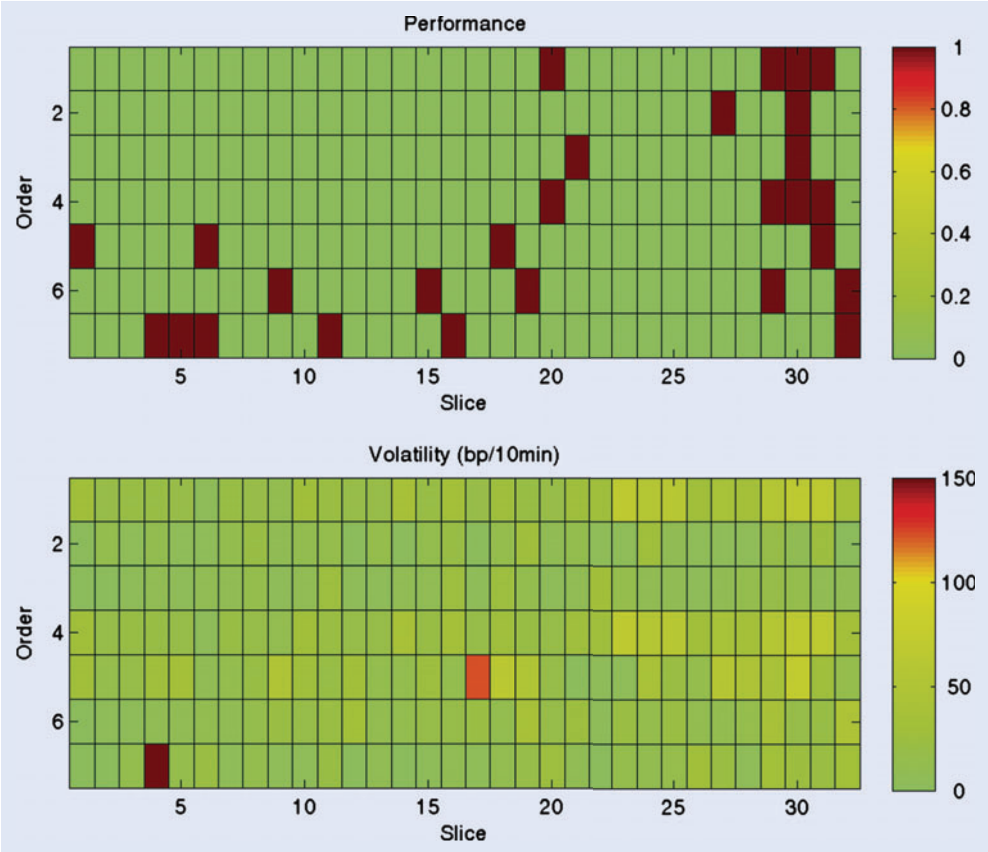

Figure 2. Heat map of online performances of a traded basket of 7 orders (top) in conjunction with values of one explanatory variable (the volatility; bottom), the correspondance between the two heat maps is not obvious.The abscissa represents the evolution over five-minute time slices.

$\mathrm{PE}_t(k)$ by setting

$$\begin{aligned} Y_t(k) &= 1 \quad \text{if} \quad \mathrm{PE}_t(k) < \ell_t, \\ Y_t(k) &= 0 \quad \text{if} \quad \mathrm{PE}_t(k) \geq \ell_t. \end{aligned} \tag{2}$$

In figure 1, we plot an example of the intraday behaviour of a trading algorithm. Its trading performance evaluation PE_t can observed within a trading day, in real-time, like some of the market context variable we used. A trading order may or may not be active at a given time slice as observed in figure 2.

Figure 3 displays synchronous intraday plots of trading performance evaluations in conjunction with the values of a few selected market variables. An essential goal of our methodology is, for each fixed time slice, to quantify on line the current influence of a market variable on trading performance degradation. Our automated online influence quantification replaces expert visual inspection of current trading orders performances, to identify critical market variables explaining trading performance degradations. For instance, visual inspection of figures 2 and 3 will naturally 'explain' the low performances observed at time slices $t = 39, 40, 41$ by the obvious trend changes simultaneously observed on rarity scores as well as by the volatility peak.

On figure 3 one can notice some conjunctions not only between the performance and the market descriptors, but between the market descriptors themselves. Around the 40th time slice the performance decreases and the volume, volatility, and bid-ask spread scores drop simultaneously. Nevertheless we can note that at the beginning of the day, two market descriptors had unexpected behaviours (traded volumes and bid-ask spread) without any performance degradation, and during the performance drop of slice 30, only one market descriptor (the bid-ask spread) show some level of abnormality.

The joint behaviour of market descriptors is very difficult to capture explicitly, since sometimes (like around slide 40 in our example), the market exhibits mutual deviations: figure 1 explains what append: the price move up probably because of another (buy) order impacting the price in few minutes: the result is an increase of traded volume, bid-ask spreads (because of the sudden consumption of liquidity), and volatility (as a measure of the up and down move of the price usually observe in market impact events).

On the contrary, the degradation of performance around slice 30 is more subtle, and probably mainly due to a sudden increase of the bid-ask spread: it may be because the monitored trading algorithm has been detected (thus other market participant moved some steps away to see how greedy the trading algorithm was).

On average some dependences between descriptors are known to be verified (like a balance between volatility, bid-ask spread and the number of trade per day, see Hastie *et al.* (2011)), but fitting a model on each stock to try to remove these common effects will not help our methodology since we need to capture explanations of the performances traders can easily understand and take means to adjust the behaviour of the trading algorithms they pilot. Correcting a performance degradation identified as an unexpected deviation from a usual common move is very difficult, but a trader knows how to correct bad performances coming from simultaneous cross sectional abnormal levels of bid-ask spread and volatility.

Our approach will typically take these dependencies into account even if there is no parametric model coding them: we will detect if there is a conjunction of common factors affecting the performance of a large significant subset of orders. The detection will deliver a clear message to the trader who will be able to fine tune in real time the parameters of the trading algorithms.

3. Online anomaly detectors

3.1. *Anomaly detection*

Online anomaly detection is a critical step in many applications, such as safety of complex systems, safety monitoring in automotive or aeronautics industries, remote health monitoring in biomedicine, real-time quality control for industrial production lines, etc. (see Aviv (1991), Basseville and Nikiforov (1993), Basseville (1988), Gustafsson (2000), Lehalle and Azencott (2007)).

In the context of trading performance online monitoring, it is also quite natural to systematically enrich raw market descriptors by automated detection of anomalies affecting their dynamics. We have thus developed algorithmics dedicated to the online implementation of this processing step.

Our approach is highly non parametric, nevertheless a parametric reading of the use of such *anomaly detectors* is simple. Since trading algorithms are usually built assuming martingality of the PFP, giving birth to price decomposition like the one of equality (1), one can expect that any abnormal behaviour of variable involved in the PFP (traded prices and volumes here) with respect to this assumption will disturb the performance of the trading process. Price jumps, price small trending periods, and volume crenels are typical abnormal under such assumptions.

The occurrence of such anomalies can be detected by algorithmic tracking of local regime changes in market descriptors dynamics, and may have potentially strong influence on performance degradation for the corresponding trading orders. We have hence developed and implemented a set of three parameterized *anomaly detectors*, dedicated to the online identification of 'significant' *Peaks or Crenels, Jumps, and Trend Changes* on generic time series (figure 5). These three detectors automatically locate emerging anomalies, quantify their intensities, and filter them through adjustable gravity thresholds.

3.2. *Building online detectors*

Consider a generic discrete time series U_t. A smoothed 'baseline' BU_t is generated as a moving local median of U_t. One then computes the local standard deviation σ_t of the 'noise' $U_t - BU_t$, and in turn, this defines 'outlier' values of U_t. Our three online anomaly detectors are based on local trend extractions at each time slice t by fitting linear or quadratic regression models on short moving time windows to the left and the right of t. The detector parameters have simple geometric interpretations for the users and are kept fixed during online influence analysis. Each anomaly detector is dedicated to a fixed type of anomaly, and generates a binary time series encoding the presence or absence of this anomaly type at successive time slices of U_t.

3.2.1. Peaks/crenels detector. A 'peak' is the sudden occurrence of a high 'outlier' value of U_t. More generally, a 'crenel' is a cluster of successive high 'outliers' with approximately equal values. Each crenel is described by three geometric 'crenel features', namely, its *time duration*, its *thickness* (i.e. absolute difference between highest and lowest crenel points), and its *height* above the baseline BU_t. Minimal threshold values are imposed on these three features, as well as a minimal *time gap* between successive crenels.

To detect peaks and/or crenels on the series U_t, one first extracts outliers with respect to the baseline BU_t; then one applies simple filters to detect local geometric configurations of outliers which satisfy the threshold constraints imposed on the three 'crenel features' described above. If a peak or crenel is detected at time t, then 'peak/crenel intensity' $Peak_t$ is set equal to 'height' of the peak/crenel above the baseline. If no peak or crenel is detected at time t, one sets $Peak_t = 0$.

3.2.2. Jumps detector. A 'jump' at time t is a sudden level change between the U_s values on finite time windows to the left and to the right of t. Bona fide jumps are described by 2 features, namely, a *duration* $2L$ and a minimal *jump size* Δ. For each t, one fits two distinct quadratic regressions to the baseline BU_s, namely Reg^- for $(t-1-L) \le s \le (t-1)$ and Reg^+ for $t \le s \le (t+L)$, where L is a fixed parameter.

A jump is detected at t if the 'jump size' $JS(t) = |Reg^+(t) - Reg^-(t)|$ is larger than Δ, provided the two regressions have small enough residuals. If a jump is detected at time t on the series U_t, then 'jump intensity' $Jump_t$ is set equal to the 'jump size' $JS(t)$. If no trend change is detected at time t, one sets $Jump_t = 0$.

3.2.3. Trend changes detector. Bona fide 'trend changes' are described by three features, a *duration* $2L$, a minimal *slope change* λ, and a *continuity modulus* ε. For each t, one fits as above two quadratic regressions to the baseline BU_s, namely Reg^- to the left of t and Reg^+ to the right of t. Call α^+, α^- the slopes of Reg^+, Reg^-, and define the 'trend change size'

$$TCS(t) = |\alpha^+ - \alpha^-|.$$

A local 'trend change' is detected at time t if Reg^+, Reg^- have sufficiently small residuals and verify

$$TCS(t) > \lambda\,; \quad \text{and } |Reg^+(t) - Reg^-(t)| < \varepsilon.$$

If a Trend Change is detected at time t on the series U_t, then 'trend change intensity' $Trend_t$ is set equal to the 'trend change size' $TCS(t)$. If no jump is detected at time t, one sets $Trend_t = 0$.

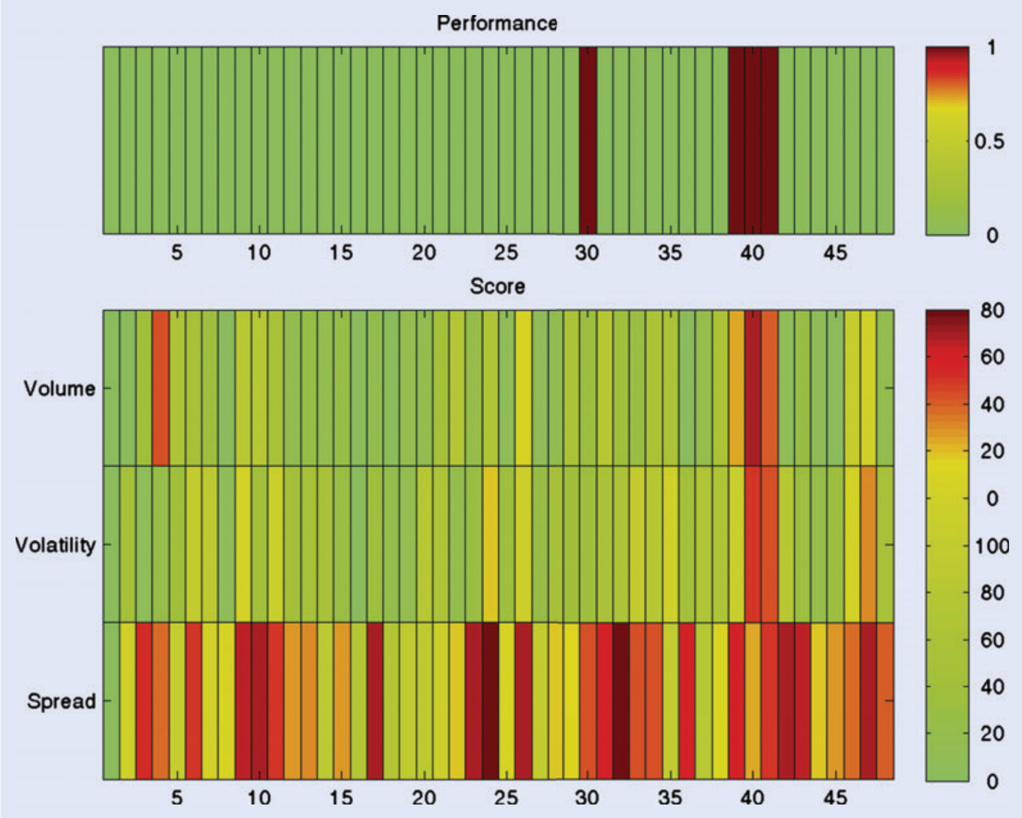

Figure 3. Conjunction of the performance (top curve) of one traded order (first line of figure 3) with the market context (bottom heatmap).

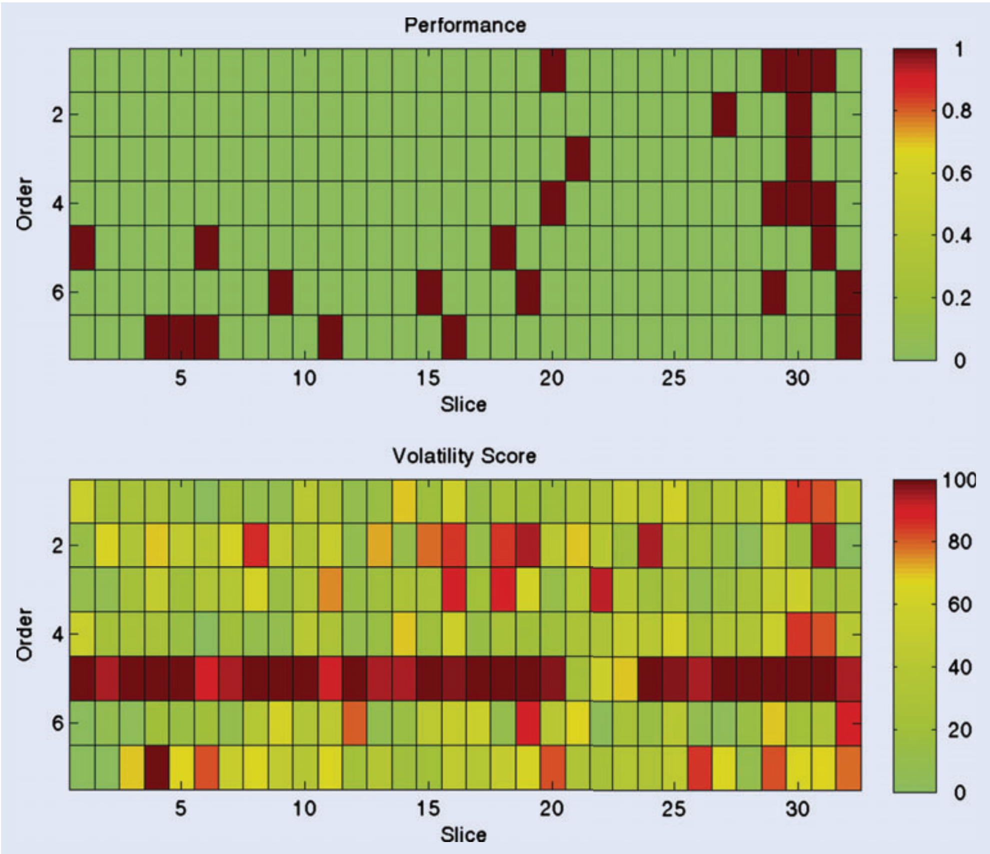

Figure 4. Heat Map of online performances of the some basket of 7 orders (top) in conjunction with the scores of the same explanatory variable (the volatility; bottom), compared to Figure 3, one can see that the scores change the high and low values of the explanatory variables, giving birth to more potential conjunctions with bad trading performances.

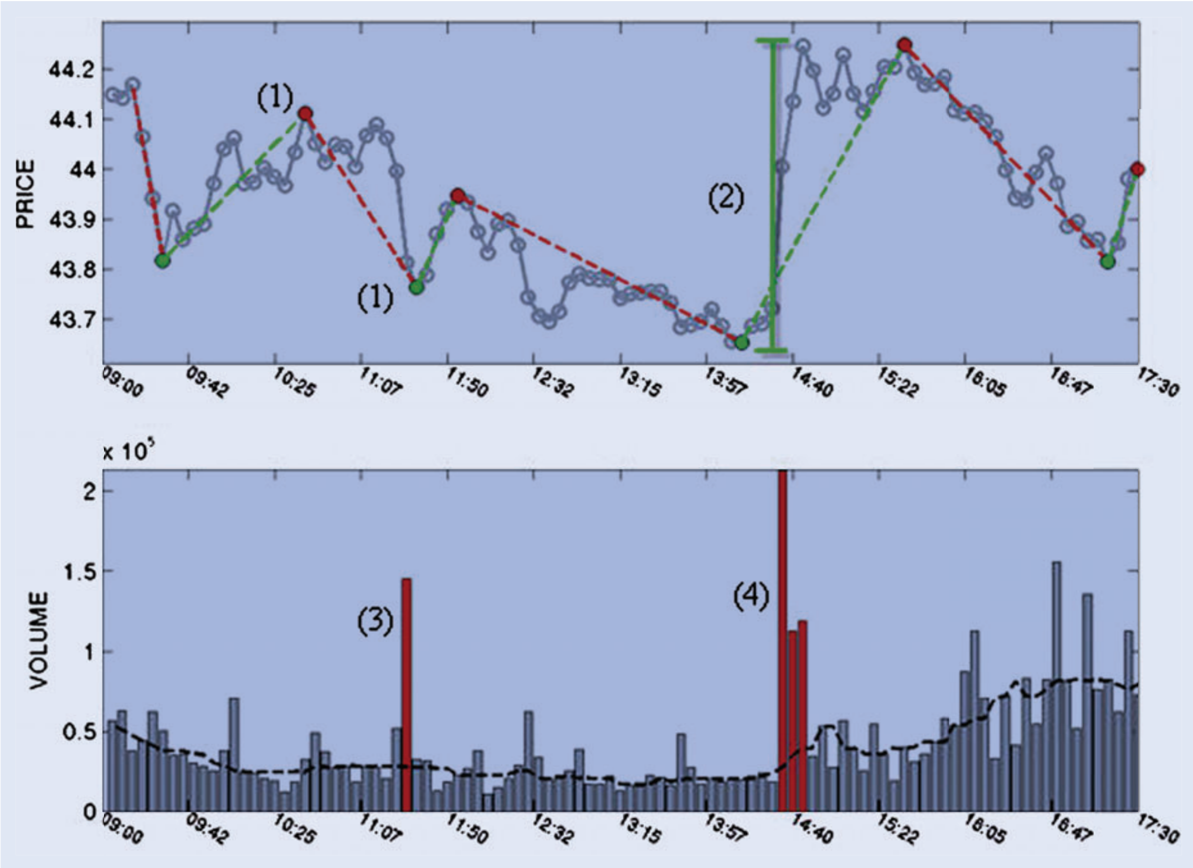

Figure 5. The three abnormal patterns targeted by our three anomaly detectors: (1) price trends, (2) price jump, (3) volume peak, (4) volume crenel.

4. Probabilistic framework for influence analysis

4.1. The enriched set of explanatory factors

For each trading order $\mathbf{T}(k)$ and each market descriptor M^j, online analysis of the time series $t \to U_t = M_t^j(k)$ by the three anomaly detectors progressively generates three time series of 'anomaly intensities' $Peak_t^j(k)$, $Jump_t^j(k)$, $Trend_t^j(k)$ which respectively encode the 'anomaly intensities' of the Peaks/Crenels, Jumps, Trend Changes detected on the time series U_t. Applying the three anomaly detectors to our seven basic market descriptors yields a set of $3 \times 7 = 21$ time series $A_t^1, \ldots, A_t^{21}$ of anomaly intensities. Let $A_t = A_t^j$ be any one of these 21 anomaly intensities. A high value of A_s detected at time s may have a degradation influence on trading performances observed not only at time slice s but also on performances observed at later time slices $t \in [s, s+\tau]$, where the short time lag τ is fixed (as a user selected parameter). When we analyze below the influence at time t of detected anomalies, we will hence take account of all the recent anomaly intensities A_s where $t - \tau \leq s \leq t$. To this end, we will replace each one our 21 anomaly intensities A_t by a *smoothed anomaly intensity at scale* τ $[[A_t|\tau]]$ defined by the following formula,

$$[[A_t|\tau]] = \max_{t-\tau \leq s \leq t} A_s.$$

From now on we will assume that a time scale τ is fixed, and note $[[A_t]]$ instead of $[[A_t|\tau]]$. Note that $[[A_t]] \geq 0$ records the maximal gravity of recent anomalies (of fixed type) affecting the dynamics of a fixed market variable, and that for 'most' time slices t one has $[[A_t]] = 0$ for most scales τ. We thus generate online $N = 28$ time series $X_t^1, \ldots, X_t^N$ of *explanatory factors*, namely the seven current market descriptors $M_t^1, \ldots, M_t^7$ and their associated 21 smoothed anomaly intensities $[[A_t^1]], \ldots, [[A_t^{21}]]$ at chosen time scales.

At each time slice t, and for each trading order $\mathbf{T}(k)$, denote by $\mathbb{X}_t(k) \in \mathbb{R}^N$ the vector

$$\mathbb{X}_t(k) = \left[X_t^1(k), \ldots, X_t^N(k) \right],$$

which regroups the current values of our $N = 28$ explanatory factors.

In the online context, the $K \times N$ time series of explanatory factors become progressively available, and provide at each time slice t the $K_t \times N$ incoming new values $X_t^j(k)$, where $1 \leq k \leq K_t \leq K$, $1 \leq j \leq N$ and K_t is the number of trading algorithms handling active orders at time t. These factors are viewed as potential 'causes' for eventual degradations affecting the current binary performance evaluation $Y_t(k)$ of trading order $\mathbf{T}(k)$ where Y_t binarizes the performance evaluations PE. The goal of our online influence analysis is to identify at each time t the explanatory factors which have the most significant degradation influence on current trading performances of the whole portfolio.

4.2. *Probabilistic framework*

To analyze functional relationships between trading performance evaluations and explanatory factors, we introduce a time dependent probabilistic framework.

Fix any *time slice* t. We consider that each one of the K_t trading order $\mathbf{T}(k)$ currently active at time t has been extracted at random from a very large finite pool Ω of 'virtually observable' trading orders $\mathbf{T}(\omega)$, with $\omega = 1, 2, \ldots, \bar{K}$, where the fixed cardinal $\bar{K}$ of Ω is much larger than K_t.

We consider Ω as a probability space endowed with the uniform probability. The binary evaluations $Y_t(\omega)$ of current trading performances defined in (2) can then be viewed as a single *binary valued random variable* Y_t defined on Ω, verifying for each $\omega \in \Omega$

$$\begin{aligned} Y_t(\omega) &= 1 \quad \text{if} \quad \mathrm{PE}_t(\omega) < \ell_t, \\ Y_t(\omega) &= 0 \quad \text{if} \quad \mathrm{PE}_t(\omega) \geq \ell_t. \end{aligned} \tag{3}$$

Similarly all our N explanatory factors X_t^j, $j = 1, 2, \ldots, N$, can be viewed as real valued random variables $X_t^j(\omega)$ defined on Ω and for which we have only observed the K_t values currently available at time t, namely the values $X_t^j(k) \in \mathbb{R}$ for $i = 1, 2, \ldots, K_t$. Then $\mathbb{X}_t = [X_t^1, \ldots, X_t^N]$ becomes a random vector defined on Ω, with values in $\mathbb{R}^N$. Our online influence analysis is performed anew at each fixed time slice t and involves only the currently available K_t joint observations of the random vector $\mathbb{X}_t$ and of the random variable Y_t. For each fixed t, we will often omit the subscript t and adopt the abbreviated notations

$$Y = Y_t\,; \quad X^j = X_t^j\,; \quad \mathbb{X} = \mathbb{X}_t = [X^1, \ldots, X^N]$$

5. Binary valued predictors and their predictive power

The time slice t is kept fixed and deliberately omitted in this whole section, where we introduce the precisely relevant notions of predictors, predictive power, and quantitative influence of explanatory factors.

5.1. *Binary predictors*

On a probability space Ω consider an arbitrary random vector $\mathbb{X} = X^1, \ldots, X^N$ of 'explanatory factors' and an arbitrary binary valued random variable Y. We call *binary valued predictor* of the binary random variable Y any random variable $\hat{Y}$ which is a deterministic function of $\mathbb{X}$. Clearly $\hat{Y}$ is then necessarily of the form $\hat{Y}_B = \mathbf{1}_B(X)$ where $\mathbf{1}_B$ is the indicator function of a fixed but arbitrary Borel subset of $\mathbb{R}^N \to \{0; 1\}$.

This class of binary predictors is naturally imbedded in the convex set of "randomized" binary predictors $\hat{Y}_\phi$ of Y, indexed by arbitrary "decision functions" $\phi \in \Phi$, where Φ is the set of all Borel functions $\phi(x)$ defined for $x \in R^N$ and such that $0 \leq \phi(x) \leq 1$. The predictor $\hat{Y}_\phi$ defined by each such ϕ verifies

$$P(\hat{Y}_\phi = 1|X) = \phi(X)\,; \qquad P(\hat{Y}_\phi = 0|X) = 1 - \phi(X)$$

Note that Φ is a closed compact convex subset of $L_\infty(R^N)$, endowed with its weak topology as dual of $L_1(R^N)$.

5.2. *Probabilities of correct predictions*

For randomized binary predictors $\hat{Y}_\phi$ of the true but yet unknown Y, the accuracy of $\hat{Y}$ is usually characterized by the two conditional *probabilities of correct prediction* p^1 and p^0, or equivalently by the two absolute probabilities of correct prediction $P^1 = P^1(\phi)$ and $P^0 = P^0(\phi)$, defined by

$$\begin{aligned} p^1 &= \mathbb{P}(\hat{Y} = 1|Y = 1) \quad \text{and} \quad P^1 = \mathbb{P}(\hat{Y} = 1; Y = 1) \\ p^0 &= \mathbb{P}(\hat{Y} = 0|Y = 0) \quad \text{and} \quad P^0 = \mathbb{P}(\hat{Y} = 0; Y = 0) \end{aligned} \tag{4}$$

The obvious expressions

$$P^1 = \mathbb{E}(\phi(X)\mathbf{1}_{Y=1})\,; \quad \text{and } P^0 = \mathbb{E}((1 - \phi(X))\mathbf{1}_{Y=0})$$

show that $P^1(\phi)$ and $P^0(\phi)$ are weakly continuous functionals of $\phi \in \Phi$. Intuitively the predictive power of a predictor should be an increasing functional of p^1 and p^0, or equivalently an increasing functional of P^1 and P^0. Indeed the classical 2×2 *confusion matrix* for the estimation of Y by the binary predictor $\hat{Y}$ is determined by p^1, p^0 as follows:

	$Y = 0$	$Y = 1$
$\hat{Y} = 0$	p^0	$1 - p^1$
$\hat{Y} = 1$	$1 - p^0$	p^1

This motivates the following definition of predictive power (figure 6).

5.3. *Predictive power*

For any random vector of 'explanatory variables' $X \in R^N$ and any binary random variable Y jointly defined on a probability space Ω, the joint probability distribution μ of (X, Y) belongs to the compact convex set $\mathcal{M}$ of all probabilities on $R^N \times \{0; 1\}$. Select and fix any non-negative continuous function $Q(\mu, a, b)$ of $\mu \in \mathcal{M}$, $0 \leq a \leq 1, 0 \leq b \leq 1$ which is a *separately increasing* function of a and b.

We then define the *predictive power* $\pi(\phi) = \pi(\mu, \phi)$ of each randomized binary predictors $\hat{Y}_\phi$ of Y by

$$\pi(\phi) = Q(\mu, P^1, P^0) = Q(\mu, \mu^1 p^1, \mu^0 p^0) \tag{5}$$

where $\mu^1 = \mathbb{P}(Y = 1)$ and $mu^0 = \mathbb{P}(Y = 0)$. Note that in our benchmark study below, due to our adaptive binarization of trading performance, the probabilities μ^1 and μ^0 will be constant in time and will have known pre-assigned fixed values such as 3 and 97%.

The predictive power $\pi(\mu, \phi)$ is then clearly continuous in (μ, ϕ) for the weak convergence topologies of $\mathcal{M}$ and Φ.

We shall see in the next paragraph that this definition of predictive power is compatible, and these definitions actually extend the predictive power quantification by relative entropy.

The functional $Q(\mu, P^1, P^0)$ will be called a *predictor quality* function. Here are basic examples of functions Q often used in the accuracy analysis of predictors.

- $Q_{min} = \min\{P^1, P^0\}$,
- $Q_{weight} = u\,P^1 + (1 - u)P^0$, for some $0 < u(\mu) < 1$.

Compared to the common use of a '*ROC curve*' (see for instance the very good tutorial in Flach (2004)), Q characterizes the predicting power of a predictor once it is fully specified.

In this impact analysis the predictor is only restricted to be an arbitrary binary valued function of a vector of market variables. In the context of this paper, the quality of a given predictor is predefined by a fixed but quite general functional $Q(p^+, p^-)$ of the two probabilities of correct prediction. By contrast, the ROC curve characterizes a whole range of predictors, for instance binary predictors (see Section 5.3) derived by all possible thresholds θ. ROC curves quality evaluators such as the classical *Area under ROC curve* (AUROC) criterion would then attribute a single quality level to the collection of all the $(p^+(\theta), p^-(\theta))$, characterizing the quality of the whole package of predictors indexed by θ. It gives clues to choose an adequate value for θ, with respect to the AUROC criterion. In this paper we produce a way to obtain the best possible predictor, and then use it to understand the root of bad performances. Qualitatively, it means that if the best possible predictor using a specific market variable (say the volatility score) has a good predictive power of bad performance, we can deduce that the volatility regime is at the root of the bad performance.

5.4. *Predictive power based on mutual information and/or relative entropy*

An information theoretic characterization for the predictive power of a predictor $Z = \hat{Y}$ of Y is the amount of information which Z reveals on the yet unknown variable Y. This has classically been quantified by *relative entropy* criteria such as the mutual information ratio $\mathrm{MIR}(Z, Y)$ (see for instance Brillinger (2004), Azencott *et al.* (2007), Mougeot and Azencott (2011), Billingsley (1965), Khinchin (1957), Shannon (1948)). Recall that the entropy $H(U)$ of a random variable U taking only a finite number of values u_i is given by

$$H(U) = -\sum_i \mathbb{P}(U = u_i) \log \mathbb{P}(U = u_i),$$

The *mutual information ratio* $\mathrm{MIR}(Z, Y)$ between Y and its predictor $Z = \hat{Y}$ is defined by

$$\mathrm{MIR}(Z, Y) = \frac{H(Z) + H(Y) - H(Z, Y)}{H(Y)},$$

where $H(Z)$, $H(Y)$, $H(Z, Y)$ are the respective entropies of the three random variables Z, Y and (Z, Y). Here Z and Y are binary valued and (Z, Y) takes only four values.

The ratio $\mathrm{MIR}(Z, Y)$ which is directly related to the relative entropy of Z with respect to Y lies between 0 and 1, and reaches the value 1 if and only if Y is a deterministic function of Z. Good predictors Z of Y should thus achieve high values of $MIR(Z, Y)$. Indeed we have the following result .

PROPOSITION 5.1 *Fix any random vector of 'explanatory variables' $X \in \mathbb{R}^N$ and any binary random variable Y jointly defined on a probability space Ω, and call μ the joint probability distribution of (X, Y). Assume that Y is not deterministic, so that $H(Y) > 0$. Then for all randomized binary predictor $Z = \hat{Y}_\phi$ of Y defined by arbitrary Borel decision functions $0 \leq \phi \leq 1$, the mutual information ratio $\mathrm{MIR}(Z, Y)$ is a separately increasing functional of the two conditional probabilities of correct decisions $p^1 = \mathbb{P}(\hat{Y} = 1|Y = 1)$ and $p^0 = \mathbb{P}(\hat{Y} = 0|Y = 0)$, provided $p^1 \geq 1/2$ and $p^0 \geq 1/2$.*

Proof The joint distribution of (Z, Y) on $\{0; 1\}^2$ is easily seen to be determined by the three parameters p^1, p^0, and $\mu^1 = \mathbb{P}(Y = 1)$. Since $H(Y)$ is fixed, $\mathrm{MIR}(Z, Y)$ is an increasing linear function of $H(Z) - H(Z, Y)$. An elementary computation easily proves the identity

$$H(Z) - H(Z, Y) = -\mu^1 Ent(p^1) - \mu^0 Ent(p^0)$$

where $Ent(p) = -p \log(p) - (1 - p) \log(1 - p)$ and $\mu^0 = 1 - \mu^1$. Since the entropy $Ent(p)$ decreases with p for $p \geq 1/2$, this concludes the proof. □

5.5. *Generic optimal randomized predictors*

We now characterize the randomized predictors achieving optimal predictive power.

PROPOSITION 5.2 *Fix a random vector $\mathbb{X} \in R^N$ of explanatory factors and a target binary variable Y. Let $0 \leq v(x) \leq 1$ be any Borel function of $x \in R^N$ such that $v(\mathbb{X}) = \mathbb{P}(Y = 1 \,|\, X)$ almost surely.*

For any Borel decision function $\phi \in \Phi$, define the predictive power of the randomized predictor $\hat{Y}_\phi$ by $\pi(\phi) = Q(\mu, P^1(\phi), P^0(\phi))$, where Q is a fixed continuous and increasing function of the probabilities of correct decisions P^1, P^0 . Then there exists $\psi \in \Phi$ such that the predictor $\hat{Y}_\phi$ has maximum predictive power

$$\pi(\psi) = \max_{\phi \in \Phi} \pi(\phi)$$

Any such optimal Borel function $0 \leq \psi(x) \leq 1$ must almost surely verify, for some suitably selected constant $0 \leq c \leq 1$.

$$\psi(X) = 1 \quad \text{for } v(X) > c\,; \qquad \psi(X) = 0 \quad \text{for } v(X) < c \tag{6}$$

Proof Predicting the actual value of the yet 'unknown' binary random variable Y given the random vector $\mathbb{X}$ of explanatory variables is clearly equivalent to deciding between the two formal 'hypotheses':

$$H^0 : \{Y = 0\} \text{ versus } H^1 : \{Y_t = 1\}$$

on the basis of the observed $\mathbb{X}$, which is a standard *testing problem* (see Lehmann and Romano (2005)). Any borelian 'rejection region' $B \subset \mathbb{R}^k$ defines the binary valued predictor $\mathbf{1}_B(YX$ of Y which rejects H^0 whenever $\mathbb{X} \in B$. More generally any Borel decision function $0 \leq \phi(x) \leq 1$ defines the binary predictor $\hat{Y}_\phi$ which, given the observed $\mathbb{X}$ rejects H^0 with probability $\phi(X)$. This test has 'confidence level' $\alpha = 1 - P^0$, and 'detection power' P^1 where P^1, P^0 are the probabilities of correct decisions for the predictor $\hat{Y}_\phi$.

Classical testing of H^0 versus H^1 involves the *likelihood function* defined with probability 1 by

$$L(\mathbb{X}) = \frac{\mathbb{P}(Y = 1 \,|\, \mathbb{X})}{\mathbb{P}(Y = 0 \,|\, \mathbb{X})} = \frac{v(X)}{1 - v(X)}$$

By Neymann-Pearson theorem (see Lehmann and Romano (2005)), for each "confidence level" $0 \leq (1 - \alpha) \leq 1$ there exists a randomized binary predictor $\hat{Y}_\phi$ which maximizes $P^1(\phi)$ among all predictors verifying $P^0(\phi) \geq 1 - \alpha$. Morover one can find $c \geq 0$ such that the Borel function $0 \leq \psi(x) \leq 1$ verifies almost surely

$$\psi(X) = 1 \quad \text{for } L(X) > c\,; \qquad \psi(X) = 0 \quad \text{for } L(X) < c \tag{7}$$

Since both $P^0 = P^0(\phi)$ and $P^1 = P^1(\phi)$ are weakly continuous functions of $\phi \in \Phi$, the predictive power $\pi(\phi) = Q(\mu, P^1(\phi), P^0(\phi))$ of $\hat{Y}_\phi$ is also weakly continuous in ϕ and thus must reach its maximum on the weakly compact set Φ for some Borel function $\phi^* \in \Phi$.

Since $Q(\mu, P^1, P^0)$ is an increasing function of P^1 and P^0, we see that the optimal predictor $\hat{Y}_{\phi^*}$ must necessarily maximize $P^1(\phi)$ among all predictors verifying $P^0(\phi) \geq P^0(\phi^*)$. Select the confidence level $1 - \alpha = P^0(\phi^*)$ and apply the Neyman-Pearson theorem just recalled above to conclude that there exists a threshold $c \geq 0$ and an associated ψ of the form (7) such that $P^0(\psi) \geq P^0(\phi^*)$ and $P^1(\psi) \geq P^1(\phi^*)$. This implies the following inequality between predictive powers

$$\begin{aligned} \pi(\psi) &= Q(\mu, P^1(\psi), P^0(\psi)) \\ &\geq Q(\mu, P^1(\phi^*), P^0(\phi^*)) = \pi(\phi^*) \end{aligned}$$

and hence $\pi(\psi) = \pi(\phi^*)$ since $\hat{Y}_{\phi^*}$ has maximal predictive power. This clearly achieves the proof. □

Definition 5.3 In the preceding situation we will quantify the capacity of the random vector $\mathbb{X}$ to 'explain' the target binary variable Y by an 'influence coefficient' $\mathcal{I}(X, Y)$ defined as the predictive power $\pi(\psi)$ of an optimal randomized binary predictor $\hat{Y}_\psi$ of Y. More precisely the influence coefficient of $\mathbb{X}$ on Y is given by

$$\mathcal{I}(X, Y) = \pi(\psi) = \max_{\phi \in \Phi} \pi(\phi)$$

Clearly, once the quality function Q is selected and fixed, the influence $\mathcal{I}(X, Y)$ depends only on the joint probability distribution μ of (X, Y).

The notion of influence coefficient is immediately extended to arbitrary subsets of explanatory factors. To any subset of indices $G \subset \{1, 2, \ldots, N\}$, we associate the random vector $X^G = \{X^j \mid j \in G\}$ of explanatory factors, with $\#(G) \leq N$, and we define as above the influence coefficient by

$$J(G) = \mathcal{I}(X^G, Y) \leq \mathcal{I}(X, Y) \tag{8}$$

6. Quantifying on line the influence of groups of explanatory factors

6.1. Benchmark study context

In our intraday data study below, at each fixed time slice t, we observe simultaneously on $K_t < 700$ trading lines the current values of our random vector $\mathbb{X} \in R^N$ of $N = 28$ explanatory factors (see Section 3), and the corresponding current values of the binarized trading performances Y. Ideally, for each subgroup $X^G = \{X^j \mid j \in G\}$ of explanatory factors, where $G \subset \{1, 2, \ldots, N\}$, we want to estimate the current *influence coefficient* $J(G)$ of X^G on Y by the formula (8), using only the current sample of K_t jointly observed values of (X^G, Y). Statistical reliability of the $J(G)$ estimates will lead us below to consider only groups G of small cardinal.

Our goal was to determine, at each time slice t, one or possibly several groups G of explanatory factors having small cardinal and high influence $J(G)$ on current trading performances, and to specifically focus on detecting small groups of current 'major causes' for the trading performance *degradations* just observed at time t.

To this end we selected a class of asymmetric predictive power functionals parameterized by one parameter $70\% \leq r \leq 100\%$, called here the 'floor predictive power'. For each predictor $\hat{Y}$ of Y with current conditional probabilities of correct predictions p^1, p^0, the predictive power of $\hat{Y}$ was computed by

$$\begin{aligned} Q_r(p^1, p^0) &= p^1, \quad \text{if } \min(p^1, p^0) \geq r\%, \\ Q_r(p^1, p^0) &= 0, \quad \text{if } \min(p^1, p^0) < r\%. \end{aligned} \tag{9}$$

Note that Q_r emphasizes strongly the probability p^1 of correctly predicting bad trading performance. The associated influence coefficients $J(G)$ then quantify the current impact of the explanatory factors X^G on 'performance degradation'. In our benchmark study of intraday datasets, systematic tests led us to fix $r = 85\%$.

6.2. Influence computation: accuracy analysis

As in the preceding subsection, the time slice t is fixed and we keep the same notations. We now analyze how to implement a numerical computation of the current influence coefficients $J(G)$ for small groups X^G of explanatory factors. Let m be the cardinal of G and denote $X^G = Z = [Z^1, \ldots, Z^m]$. To compute $J(G)$ we need to compute an optimal decision function $0 \leq \psi(Z) \leq 1$, maximizing the predicting power of the predictor $\hat{Y}$ defined by $\mathbb{P}\{\hat{Y} = 1|Z\} = \psi(Z)$. By proposition 5.2 , for each value z of Z currently observed at time t, this requires first to estimate by empirical frequencies $\hat{v}(z)$ the probabilities

$$v(z) = \mathbb{P}[\,(Z = z) \cap (Y = 1)\,]$$

and then to find an optimal threshold $0 < c < 1$ for the $\hat{v}(z)$ values. At time t the estimates $\hat{v}(z)$ are derived only from the moderately sized sample of $K_t \equiv 700$ of currently observed joint values for the pair (Z, Y). By construction of the binarized trading performance (see Section 2) the empirical frequency $\{Y = 1\}$ is kept constant equal to $q = 3\%$. So the number of currently observed values z of Z for which $\hat{v}(z)$ is non-zero will always be inferior to $K_t \times 3/100 \simeq 20$. Empirical thresholding of the $\hat{v}(z)$ at time t can then obviously be restricted to exploring at most 20 values of c.

We seek then an optimal decision function $0 \leq \psi(Z) \leq 1$, which according to formula (6), should be associated to some threshold $1 > c > 0$, with $\psi(Z) = 1$ when $v(Z) > c$ and $\psi(Z) = 0$ when $v(Z) < c$. To achieve statistical robustness and fast online computation, we restrict $\psi(Z)$ to only take the values 0 or 1, and we thus impose $\psi(Z) = \mathbf{1}_{v(Z) \geq c}$. At time t, the predictive power $g(c) = \pi(\psi) = Q(P^1, P^0)$ of the estimator $\hat{Y} = \psi(Z)$ depends only on its current probabilities of correct prediction

$$\begin{aligned} P^1 &= \mathbb{P}\{(v(Z) > c) \cap (Y = 1)\} \\ P^0 &= \mathbb{P}\{(v(Z) < c) \cap (Y = 0)\} \end{aligned}$$

At time t, these probabilities are approximated by current empirical frequencies $\hat{P}^1, \hat{P}^0$ after replacing $v(Z)$ by the current estimates $\hat{v}(Z)$. This yields an estimated predictive power $\hat{g}(c)$ and the optimal threshold c^* is selected from at most 20 threshold values by a trivial maximization of $\hat{g}(c)$. The current estimate of the influence coefficient of the group of factors G is then $\hat{J}(G) = \hat{g}(c^*)$.

The practical accuracy of the estimates $\hat{J}(G)$ is strongly determined by the accuracy of the estimators $\hat{v}(z)$ of the probability $v(z)$. Given the moderate size of K_t , we deliberately *discretize* all our explanatory factors so that the vector $Z = X^G$ is restricted to take only a finite number s of distinct values. Under the favourable assumption of approximate independence of the K_t trading orders observed at time t the errors of estimations $\varepsilon(z) = \hat{v}(z) - v(z)$ have standard deviations $v(z)(1 - v(z))/K_t$ and hence the relative errors $\varepsilon(z)/v(z)$ on $v(z)$ are of the order of $1/\sqrt{v(z)K_t}$, which is inferior to $1/\sqrt{wK_t}$ where $w = \min_z v(z)$.

The optimal threshold c^* is computed above by empirically thresholding a set of at most 20 estimated $\hat{v}(z)$, since $3\% \times K_t$ is of the order of at most 20 in our data set. The relative error of estimation on c^* will then be roughly of the order of $1/\sqrt{wK_t}$. Thus to obtain a relative error inferior to, say, 11% for the estimation of the optimal c^*, one needs to at least impose the constraint $\frac{1}{wK_t} < 1.21/100$, which yields $1/w < \frac{1.21K_t}{100}$. Since $\mathbb{P}(Z = z) \geq w$ for each z in the currently observed range of $Z = X^G$ which after discretization contains s values, we have $1 \geq sw$ and hence $s < 1/w < \frac{1.21K_t}{100}$. In our benchmark intraday data, the number K_t of orders active at time t is $K_t \equiv 700$, and hence the cardinal of observable values for the discretized random vector $Z = X^G$ should be less than eight.

Maximal discretization of each explanatory factor is reached when each factor is binarized. But even in this case, since the vector $Z = X^G$ is of dimension $m = \text{cardinal(G)}$, the number s of distinct values for X^G is $s = 2^m$ and hence we must still impose the constraint cardinal(G) ≤ 3.

These cardinality constraints show that:

- for cardinal(G) equal to three, all three factors in X^G must be binarized;
- for cardinal(G) equal to two, one of the two factors in X^G must be binarized and the other one must be discretized with at most three values; and
- for cardinal(G) equal to one, the single factor involved must be discretized with at most eight values.

In our benchmark study of intraday data the number K_t of trading orders concretely ranged anywhere between 400 and 700. The preceding analysis thus showed that at each time t and for any group G of explanatory factors, the computation of the influence coefficient $J(G)$ could only be statistically robust if cardinal(G) was equal to one or two and if each explanatory factor was binarized. In practice, at each time t, a key computing task is to select an optimal binarization of each explanatory factor, as indicated below.

7. Influence computation: optimized discretization of explanatory factors

7.1. Influence computation for a single explanatory factor

Consider any real valued single explanatory factor Z having a continuous conditional density function $w(z)$ given $Y = 1$. To predict Y given Z the best binarization of Z should select a subset B of S maximizing the predictive power of the predictor $\mathbf{1}_B(Z)$. An easy extension of the proposition 5.2 proved above,

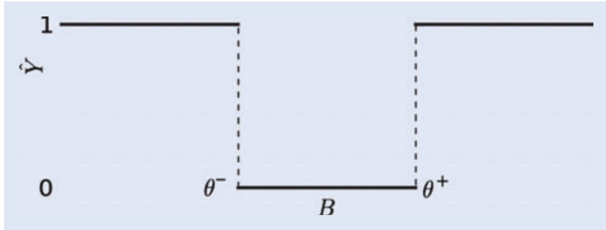

Figure 6. Typical predictor defined by two intervals $(-\infty, \theta^-)$ and $(\theta^+, +\infty)$.

shows that any optimal B should be the set of all $z \in \mathbb{R}$ such that $w(z) > c$ for some $c > 0$. Thus an optimal B must be a closed level set L of the unknown conditional density function $w(z)$. The family of all closed sets in $\mathbb{R}$ is well known to have infinite Vapnik-Cervonenkis dimension (see Vapnik and Chervonenkis (1971), Vapnik (2010)). So, in view of Vapnik's theorems on automated learning (see Vapnik (2006), Cristianini and Shawe-Taylor (2000)), empirical optimal selection of B among all closed sets will have weak generalization capacity, increasing extremely slowly with the number K_t of data. Other classes of predictors commonly used in statistical learning (like artificial neural networks, logistic regressions, random forests, etc.—see REF for more) thus cannot be more optimal than the Borel functions defined by propositions of Section 5.5, they have nevertheless a lower Vapnik-Cervonenkis dimension. Use such functions will not be enough to guarantee decent generalization capabilities, as the short accuracy analysis of Section 6.2 underlined. Focusing on extreme or rare events (3% of the worst performing trading algorithms of a pool of few hundreds) demand to use predictors with very few parameters. This has naturally led us to select sub-optimal but much more robust classes of predictors, with radically reduced Vapnik-Cervonenkis dimension.

In the cases where $w(z)$ can be considered as roughly unimodal or monotonous, the level sets of w are unions of at most two disjoint intervals. We thus deliberately restrict our class of binary predictors of Y to two-sided ones:

Definition 7.1 (Two-sided binary predictor) *Two-sided binary predictors* are of the form $h_\theta = \mathbf{1}_B(Z)$ where B is the union of the two disjoint intervals $(-\infty, \theta^-)$ and $(\theta^+, +\infty)$, indexed by the vector $\theta = (\theta^-, \theta^+) \in \mathbb{R}^2$, with $\theta^- < \theta^+$.

Note that h_θ predicts bad trading performances if and only if the explanatory factor Z takes sufficiently large or sufficiently small values. Hence these estimators of trading performance degradation have *an immediate interpretability* for natural users of online trading performance monitoring.

At time t, given the current K_t joint observations of the explanatory factor Z and of the binarized trading performance Y, an immediate counting provides for each θ the empirical estimates $\hat{P}^1$ and $\hat{P}^0$ of the probabilities of correct prediction P^1, P^0 for the estimator h_θ, given by

$$P^1 = \mathbb{P}\{(Z < \theta^-) \cap (Y = 1)\} + \mathbb{P}\{(Z > \theta^+) \cap (Y = 1)\}$$

$$P^0 = \mathbb{P}\{(\theta^- < Z < \theta^+) \cap (Y = 0)\}$$

The predictive power $\pi(\theta)$ of h_θ is then readily estimated by the explicit formula

$$\hat{\pi}(\theta) = Q_r(\hat{P}^1, \hat{P}^0).$$

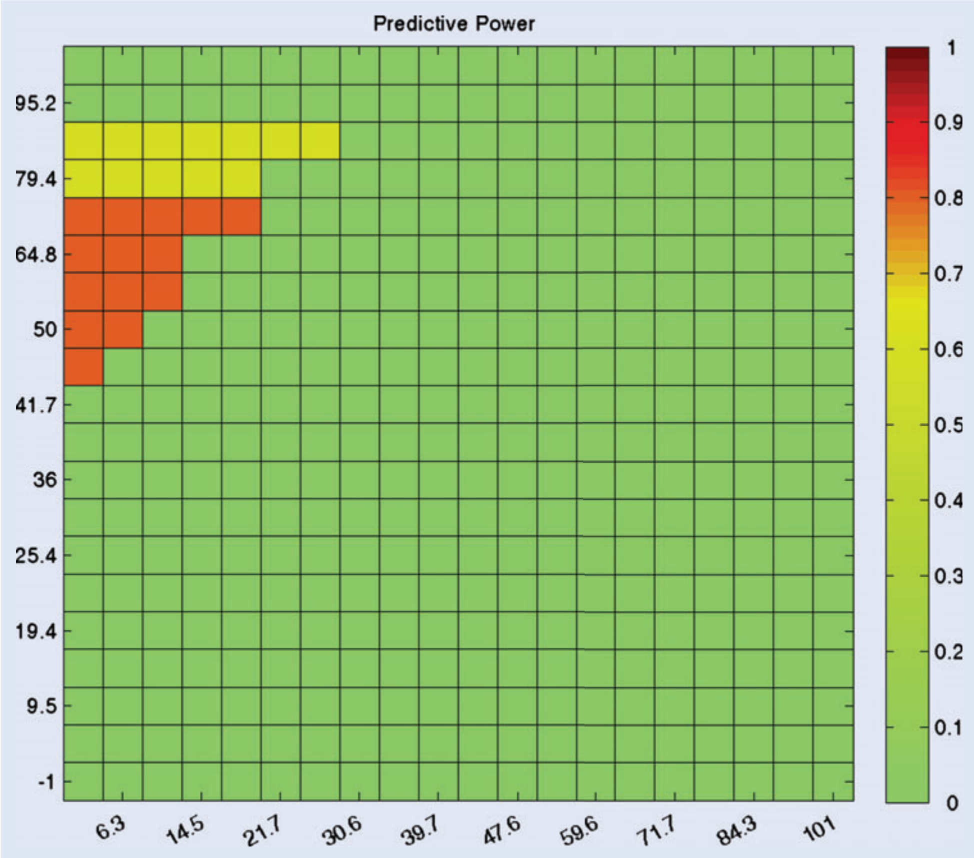

Figure 7. Predictive power as a functional of the two thresholds θ^+ (y-axis) and θ^- (x-axis) of the market variable *Volume Score* at time slice $t = 72$. It can be seen that θ^- lower than 10 and θ^+ around 65 generate an efficient predictor of bad trading performance during this time slice.

The influence coefficient $\mathcal{I}(Z, Y)$ of Z on Y at time t is then estimated by maximizing $\hat{\pi}(\theta)$ over all θ in $\mathbb{R}^2$. At time t, the set S of currently observed values of Z has cardinal inferior or equal to K_t. The previous formulas show that to maximize $\hat{\pi}(\theta)$, we may in fact restrict both θ^- and θ^+ to belong to S, so one needs only explore at most $K_t^2/2$ values of θ.

Clearly this computation tends to underestimate the influence $\mathcal{I}(Z, Y)$. Nevertheless in our benchmark studies we have systematically applied this approach for the following reasons.

- the set of binary predictors h_θ has the merit of having finite Vapnik-Cervonenkis dimension equal to two, so that our empirical estimate of the maximum of $\pi(\theta)$ will be statistically robust even for moderate realistic values of $K_t \equiv 700$;
- the immediate interpretability of the predictors h_θ enables user friendly online graphic displays of the explanatory factors currently having high influence on performance degradation; and
- at each time t, at most K_t^3 basic operations suffice to implement the brute force maximization of $\pi(\theta)$ which generates our current evaluation of $\mathcal{I}(Z, Y)$.

Note that the two optimal thresholds (θ^-, θ^+) will of course strongly depend on the time slice t.

When the explanatory factor Z is one of the 21 *smoothed anomaly detectors* $[[A_t]] \geq 0$ introduced above in Section 3, the preceding implementation can be simplified. Recall that $[[A_t]]$ records the maximal gravity of very recent anomalies of fixed type affecting the dynamics of a fixed market variable, and that for 'most' time slices, $[[A_t]]$ takes the values 0. Thus it is natural to expect that only higher values of $[[A_t]]$ to be potential explanations for currently degraded trading performances. So for practical applications to $Z = [[A_t]]$ of the preceding approach, we may actually impose the constraint $\theta^- = 0$, with essentially no loss of predictive power.

7.2. A few examples for single explanatory factors

The empirical strategy just presented to estimate the influence $J(G)$ when cardinal (G) equals one has been numerically validated on our benchmark set of intraday data. We now outline a few examples. Recall that our benchmark study used the predictive power functional $\pi = Q_r(P^1, P^0)$ given above by formula (5), which is specifically sensitive to predictors capacity to detect *degradations* of trading performances. On our intraday data sets, we have methodically tested the values r = 70, 75, 80, 85, 90 and 95% for the 'floor predictive power' r ; the value $r = 85\%$ turned out to be the best choice for these data sets, and was adopted for all results presented below.

Figure 7 illustrates for the fixed time slice $t = 45$, the predictive power of the predictors h_θ based on the single market variable Z = 'Momentum in Bid-Ask Spread'.

The trading performance evaluator PE is the 'slippage in bid-ask spread'. The PE-thresholds $\ell_t = \ell_{45}$ determining low trading performance is fixed at the 3% -quantile of all performance evaluations observed at time $t = 45$.

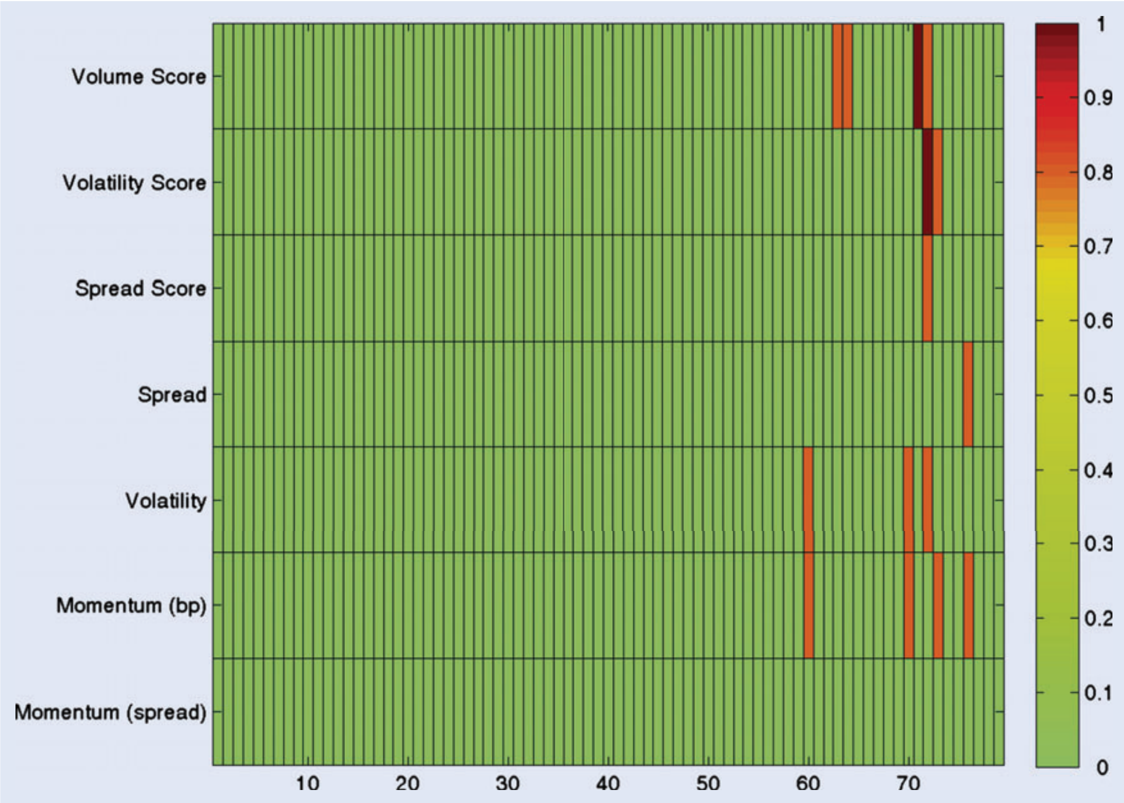

Figure 8. Predictive powers of some explanatory variables (horizontal scale is time in slices of 5 min).

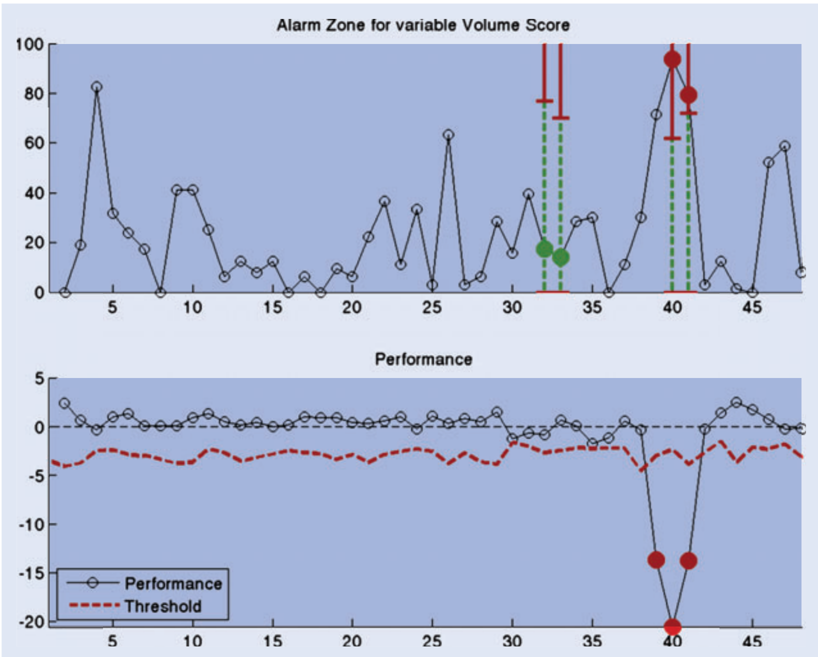

Figure 9. Auto adaptive alarm zones on the *Volume Score* explanatory variable for the order $T(139)$ displayed in figure 3; Top: four alarm zones are active, two realizations of the Volume Score exceed the auto adaptive thresholds and thus emerge as a highly likely explanation for the bad performance exhibited by this order (see Bottom graph).

The x and y axes in the graph (figure 7) indicate the threshold values (θ^-, θ^+) for the market variable Z = 'Momentum in Bid-Ask Spread'. The z axis displays the predictive power $\pi(\theta)$ of h_θ. The red marker indicates at time $t = 45$, the estimated influence coefficient $\mathcal{I}(Z, Y)$ on Y for this specific market variable Z, which turns out to be equal to 100%. The threshold vector $\theta = \theta_{45}$ which achieves maximum predictive power at time 45 is equal to (66.76, 3.87). At each time t, our seven basic market variables can then be ranked on the basis of their approximate influence values computed as above, which provides a ranking of their respective capacity to explain current bad trading performances.

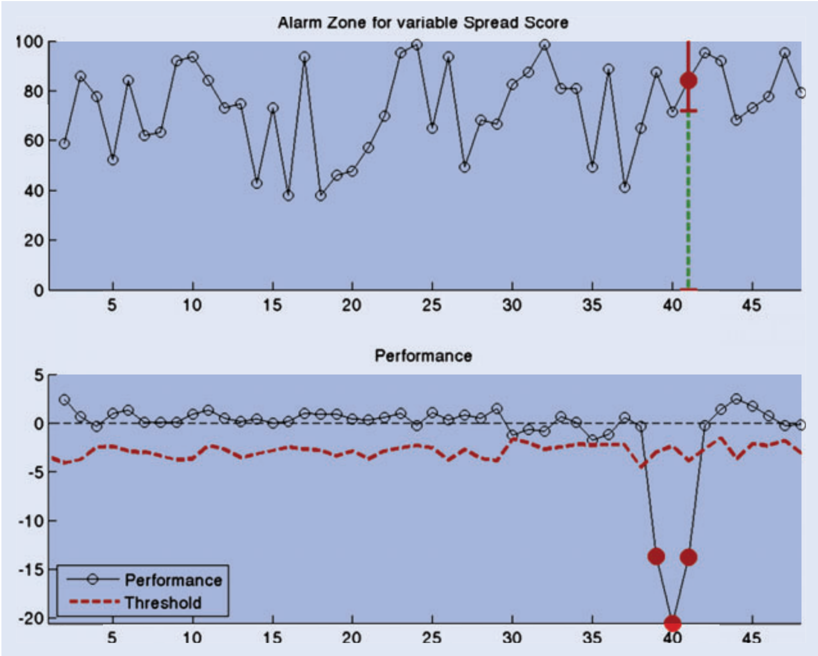

Figure 10. Auto adaptive alarm zones on the *Bid-Ask spread Score* explanatory variable for the order of figure 3; Top: one alarm zone is active, since the Bid-Ask Score exceeds the auto adaptive thresholds, and thus emerges as a quite likely explanation of bad trading performance (as validated on the bottom figure).

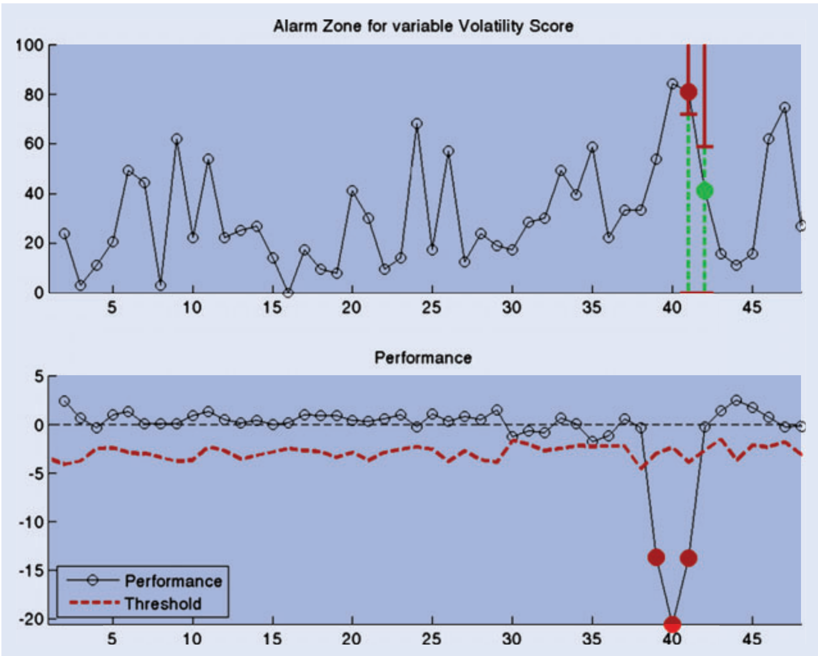

Figure 11. Auto adaptive alarm zones for the *Volatility Score* explanatory variable for the order of figure 3; Top: two alarm zones are active, one realization of the Volatility Score exceeds the auto adaptive thresholds giving and thus provides a highly likely explanation of the current bad trading performance (as displayed on the bottom graph).

8. Influence computation for pairs of explanatory factors

Again at fixed time t, we now sketch our online 'optimized fusion' of predictors to estimate the influence coefficient $J(G)$ when G is a group of two explanatory factors $Z = [Z^1, Z^2]$. Our statistical robustness analysis above indicates the necessity to consider only classes of trading performance predictors having radically low Vapnik-Cervonenkis dimension. So our predictive power maximization among predictors based on Z is deliberately restricted to the following class of predictors.

Let $\mathfrak{P}_2$ be the set of all 16 functions mapping $\{0; 1\}^2$ into $\{0; 1\}$. The class $\mathcal{H}$ will be the set of all predictors of the form

$$h(Z) = \mathfrak{m}(f(Z_1), g(Z_2))$$

where $\mathfrak{m} \in \mathfrak{P}_2$, and the indicator functions $f(z)$ and $g(z)$ are both two-sided binary predictors in the sense of definition 7.1. The class of binary predictors $\mathcal{H}$ has Vapnik-Cervonenkis dimension equal to four. Hence the estimation of maximal predictive power within $\mathcal{H}$ by empirical estimation of probabilities P^1 and P^0 on the basis of the current K_t joint observations of

Y, Z_1, Z_2 will be statistically robust. This provides at time t a stable estimator of $J(G)$, which as above tends to undervalue the true $J(G)$.

In concrete implementation of this approach at fixed time t, we first select only pairs of predictors $f(Z_1)$, $g(Z_2)$ which already have reasonably high probabilities of correctly predicting Y. To maximize the predictive power of $h(Z) = \mathfrak{m}(f(Z_1), g(Z_2))$ we need to select the best binary polynomial $\mathfrak{m}$ among the 16 elements of $\mathfrak{P}_2$. We then impose $\mathfrak{m}(0,0) = 0$ and $\mathfrak{m}(1,1) = 1$, so that whenever the predictions of $f(Z_1)$, and $g(Z_2)$ agree, we also have $h(Z) = f(Z_1) = g(Z_2)$. This 'accelerated fusion' is fairly classical in multi-experts fusion (see Alkoot and Kittler (1999)) and obviously provides an acceleration multiplier of four in the online computation of $J(G)$.

For groups G of $k = 3$ or $k = 4$ explanatory factors, one could estimate $J(G)$ by similar sub-optimal but implementable strategies. However the corresponding predictor classes have Vapnik-Cervonenkis dimensions six and eight, and their statistical robustness is hence much weaker in the concrete context of our intraday datas set, since at each time t the key Vapnik ratios $K_t/6$ and $K_t/8$ were resp. inferior to 120 and 70, values which are much too small and strongly suggested to avoid the estimation of $J(G)$ for cardinal $(G) \geq 3$.

9. Numerical results

We now present the numerical results obtained by applying the above methodology to our benchmark dataset of intra-day trading records.

9.1. *Dataset (portfolio) description*

Recall that our intra-day benchmark data involve a total of 79 time slices of 5 min each (i.e. this portfolio has been traded from 8:55 to 15:30 London time), and that we are monitoring a portfolio of 1037 trading orders, with a maximum of 700 trading orders active simultaneously at each time slice.

At each fixed time slice t, we compute the current influence coefficient for each one of our 28 explanatory factors, namely the seven market descriptors M^j themselves and the 3×7 smoothed anomaly detectors monitoring the dynamic of these market descriptors.

These 28 explanatory factors generate $378 = 28 \times 27/2$ pairs of factors. By accelerated fusion as above, we compute, at each time slice t, the influence of each one of these 378 pairs of explanatory factors.

Among these $406 = (378 + 28)$ groups of explanatory factors, at each time t, we retain only those having both conditional probabilities of correct predictions (p^1, p^0) larger than $r\%$. Here $r\% > 70\%$ is the user selected 'floor predictive power'. Note that each single factor or pair of factors retained at time t can predict current degraded performance degradations with a false alarm rate $FAR = 1 - p^0$ inferior to $(100 - r)\%$.

Among the retained groups of explanatory factors, we compute the maximal influence max J_t achievable at time t. We also determine the set $\mathcal{D}_t$ of dominating groups of explanatory factors, defined as the groups of one or two factors having an influence equal to $\mathfrak{J}_t$ and achieving the minimal false alarm rate $(100 - r)\%$.

9.2. *Predictive power of market descriptors*

Figure 8 gives a heat map of the predictive power of a selected subset of market descriptors on the whole portfolio. The display shows that:

- no market descriptor is used before slice 60 (i.e. 14:00), meaning that there are no significant predictive links between bad trading performances and specific values of the descriptors.
- Then the *Volume Score* has the capability to explain bad trading performance from 14:10 to 14:20 and from 14:50 to 15:00. It means that during these two time intervals, bad performances occurred simultaneously with quite unusual levels of traded volumes.
- The *Volatility Score* emerges as a complementary explanatory factor between 14:55 and 15:05; orders with bad trading performances focused on stocks having unexpectedly high volatility levels during these 10 min.
- The *Bid-Ask Spread Score* conforts this automated diagnosis: a rare event did indeed degrade trading performances around 15:00. Keeping in mind that scores are computed according to historical values during last weeks, it means that for this portfolio, the worst performances occurred on stocks for which volumes, volatility, and bid-ask spread had abnormal values.

It is interesting to note that, before scoring, market performances do not explain that well bad performances.

9.3. *How alarm zones explain bad trading performance*

Since the two-sided binary predictors are built to explain the degraded performances of the worst trading orders, it is easy to identify the most impacted orders one given market descriptor. Here we consider the trading order $\mathbf{T}(139)$ whose lifecycle is shown on figure 3, in order to visualize the impact of its *Volume Score*, *Volatility Score*, and *Bid-Ask Score* on the order performance. First note that this order has been active from 10:55 to 15:30 London time, with a start time two hours after the launch of the portfolio. It means that 5 min slices on this order are numbered from 0 to 49; they have to be shifted by 30 to be synchronized with the time scale of the other portfolio orders.

9.3.1. Alarm zones on the volume score. Figure 9 shows alarm zones for the *Volume Score* of order $\mathbf{T}(139)$: the top subplot draws the value of the volume score through time and the associated alarm zones have been added on top of it, when triggered.

For the whole portfolio (figure 8), alarm zones are triggered on the *Volume Score* from 14:10 to 14:20 and from 14:50 to 15:00 (i.e. slices 32 to 32 and 40 to 41 for this specific order). They are drawn like 'gates' from the low threshold (θ_t^-) to the high one (θ_t^+); if the value of the Volume Score is outside bounds for the given order: we thus state that '*the Volume Score contributed to the degradation of trading performance for order* $\mathbf{T}(139)$'.

It is important to note that if if our thresholds had not been adaptive as proposed by our '*influence analysis*' methodology, they would have generated false explanations at the start of the order (around slice four, i.e. 11:15 London time).

Specifically for this order, the Volume Score does not enter the first alarm zone (around 14.15); it is in line with the performance of order $\mathbf{T}(139)$ that is normal (bottom subplot of figure 9). The boundaries of the next alarm zone (around 15:00) are crossed by the order Volume Score; indeed this order performs quite poorly at that time.

9.3.2. Alarm zones on the bid-ask spread score. Only one alarm zone has been activated on the portfolio (around 15:00) and the advantage of the auto adaptive approach proposed in this paper is straightforward: a unique threshold for all the duration of the portfolio would clearly not have been able to separate the 41st slice of order $\mathbf{T}(139)$ from the others (figures 10 and 11).

9.3.3. Alarm zones on the volatility score. Once again it is clear that the alarm zones succeeded in isolating slices to efficiently explain the bad trading performance around 15:00.

9.3.4. To summarize. When applied to our real portfolio of 1037 orders traded during 6 h and 35 min, the *automated influence analysis methodology* presented and studied from a theoretical viewpoint in this paper efficiently selects quite pertinent explanatory factors for degraded trading performance:

- our alarm zones use thresholds that are automatically adapted online to successive time slices, as computed via the *predicting power* of *two-sided binary predictors* (see Definition 7.1) based on *market descriptors*.
- Our approach generates generates auto adaptive thresholds taking into account currently observed synchronicity between the user selected performance criterion (chosen according to the trading goal, see Section 2.1) with market descriptors *market descriptors*.
- At each time slice, the computed adaptive thresholds on market descriptors apply to the whole portfolio, bad trading performance of orders for which market descriptors take values outside *alarm zones* are said to be explained or *influenced* by the given descriptors. We commented real examples to illustrate these automated selection of explanatory factors.
- The added value of augmenting the state space of market descriptors using *scores* has been illustrated on several examples within our benchmark data set o trading orders.

10. Conclusions

The approach presented in this paper quantifies in real-time, the negative *influence* (on trading performances) of currently detected abnormal behaviour of market factors. The paper presents a theoretical framework to be used for TCA and uses it for online TCA. Dynamic influence-based ranking of market factors provides in real-time the most likely causality links between current bad trading performance and currently detected abnormal behaviour of market variables. To accurately capture the effect of anomalies detected on the dynamics of market variables, binary performance predictors based on anomalies have been extended to cover lagged recent occurrences of anomalies. In particular, this captures and quantifies the predictive power of crenels, jumps over multiple time steps, etc. Our algorithms provide real-time evaluations for the current influence of specific market variables on currently observed trading performance degradations, with a small time-delay dependent on market liquidity. Moreover, *influence analysis* can generate efficient real-time answers to online queries by multiple traders as well as post-trade analysis. Our methodology enables fast in-depth dynamic evaluation of trade scheduling algorithms, and should help to quantify the comparative analysis of various trading algorithms. This will enhance new automated approaches to optimize the parameters of trading algorithms by intensive testing on historical data. Note that we do not address here the issue of picking potential explanatory market factors, possibly from a very large pool of available factors. We consider here seven basic market variables as given arbitrarily, or as pre-identified based on expert knowledge. However, new variables could be automatically included as potential explanatory factors by scanning a large set of market variables and ranking them after computation of their respective influence on performance degradation. For instance, it is advantageous to include factors which are uncorrelated and have historically shown to provide reliable explanation for performance degradation.

Acknowledgements

This research has been supported by the Crédit Agricole Cheuvreux *Trading and Microstructure* Research Initiative, in partnership with the Louis Bachelier Institute and the Collège de France.

References

Alkoot, F.M. and Kittler, J., Experimental evaluation of expert fusion strategies. *Pattern Recogn. Lett.*, 1999, **20**(11), 1361–1369.

Almgren, R., Thum, C., Hauptmann, E. and Li, H., Direct estimation of equity market impact. *Risk*, 2005, **18**, 57–62.

Almgren, R.F. and Chriss, N., Optimal execution of portfolio transactions. *J. Risk*, 2000, **3**(2), 5–39.

Aviv, D., Detection of abrupt changes in statistical models. PhD Thesis, Naval Postgraduate School, Monterey, CA, 1991. Available online at: https://archive.org/details/detectionofabrup00aviv.

Azencott, R., Kreiss, J.P., Mougeot, M., Pastor, C., Pfeiffer, M., Siebert, S. and Zangmeister, T., Analysis methods for accident causation studies, Synthesis of the European Research Project TRACE (No 027763), 2007.

Basseville, M., Detecting changes in signals and systems—A survey. *Automatica*, 1988, **24**(3), 309–326.

Basseville M. and Nikiforov I.V., *Detection of Abrupt Changes: Theory and Application*, Prentice Hall Information and System Sciences Series, 1993 (Prentice Hall: Englewood Cliffs, NJ).

Bertsimas, D. and Lo, A.W., Optimal control of execution costs. *J. Financ. Markets*, 1998, **1**(1), 1–50.

Billingsley P., *Ergodic Theory and Information*, 1965 (Wiley: New York).

Borovkov A.A., *Probability Theory*, 1998 (Gordon and Breach: Amsterdam).

Bouchard, B., Dang, N.-M. and Lehalle, C.-A., Optimal control of trading algorithms: A general impulse control approach. *SIAM J. Financ. Math.*, 2011, **2**(1), 404–438.

Brillinger, D.R., Some data analyses using mutual information. *Brazilian J. Prob. Stat.*, 2004, **18**(6), 163–183.

Cristianini, N. and Shawe-Taylor, J., *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, 1 ed., 2000 (Cambridge University Press: Cambridge).

Flach, P., The many faces of ROC analysis in machine learning. ICML Conference, Banff, AB, 4–8 July, 2004.

Garman, M.B. and Klass, M.J., On the estimation of security price volatility from historical data. *J. Bus.*, 1980, **53**(1), 67–78.

Guéant O., Lehalle C.A. and Fernandez-Tapia J., Dealing with the inventory risk: A solution to the market making problem. *Math. Financ. Econ.*, 2013, **7**(4), 477–507.

Gustafsson F., *Adaptive Filtering and Change Detection*, 1 ed., 2000 (Wiley: Chichester).

Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, 2nd ed., 2009 (Springer: New York), corr. 7th printing 2013 edition.

Khinchin, A.I.A., *Mathematical Foundations of Information Theory*, 1957 (Dover Publications: New York).

Laruelle, S., Lehalle, C.-A. and Pagès, G., Optimal posting price of limit orders: Learning by trading. *Math. Financ. Econ.*, 2013, **7**(3), 359–403.

Lehalle, C.-A., Rigorous strategic trading: Balanced portfolio and mean-reversion. *J. Trading*, 2009, **4**(3), 40–46.

Lehalle, C.-A., Market microstructure knowledge needed to control an intra-day trading process. In *Handbook on Systemic Risk*, edited by J.-P. Fouque and J. Langsam, 2013 (Cambridge University Press: Cambridge).

Lehalle, C.A. and Azencott, R., On the fly health monitoring of mechanical hazards from under sampled signals in formula one. FISITA World Automotive Congress, Barcelona, May 2007.

Lehalle, C.A., Laruelle, S., Burgot, R., Pelin, S., Lasnier, M., *Market Microstructure in Practice*, 2013 (World Scientific publishing).

Lehmann, E.L. and Romano, J.P., *Testing Statistical Hypotheses*. Springer Texts in Statistics, 3rd ed., 2005 (Springer: New York).

Mallat, S., *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed., 2008 (Academic Press: New York).

Moro, E., Vicente, J., Moyano, L.G., Gerig, A., Farmer, J.D., Vaglica, G., Lillo, F., Mantegna, R.N., Market impact and trading profile of hidden orders in stock markets. *Phys. Rev. E* 2009, **80**, 066102+.

Mougeot, M. and Azencott, R., Traffic safety: Non-linear causation for injury severity. *Saf. Secur. Eng. IV*, 2011, **117**, 241–252.

Pagès, G., Laruelle, S. and Lehalle, C.-A., Optimal split of orders across liquidity pools: A stochatic algorithm approach. *SIAM J. Financ. Math.*, 2011, **2**, 1042–1076.

Rosenthal, D.W., Performance metrics for algorithmic traders. Social Science Research Network Working Paper Series, 2009.

Shannon, C.E., A mathematical theory of communication. *Bell Syst. Tech. J.* 1948, **27**(3), 379–423.

Sitter, R.R. and Wu, C., A note on Woodruff confidence intervals for quantiles. *Stat. Prob. Lett.*, 2001, **52**(4), 353–358.

Vapnik, V., *Estimation of Dependences Based on Empirical Data*, Information Science and Statistics, 2006 (Springer: New York).

Vapnik, V., *The Nature of Statistical Learning Theory*, Information Science and Statistics, 2nd ed., 2010 (Springer: New York), softcover reprint of hardcover, 2000 edition.

Vapnik, V.N. and Chervonenkis, A.Y., On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Probab. Appl.*, 1971, **16**(2), 264–280.

Woodruff, R.S., Confidence intervals for medians and other position measures. *J. Am. Stat. Assoc.*, 1952, **47**(260), 635–646.