

Market Microstructure knowledge needed to control an intra-day trading process

Charles-Albert Lehalle*

Abstract

A lot of academic and theoretical works have been dedicated to optimal liquidation of large orders these last twenty years. The optimal split of an order through time (“optimal trade scheduling”) and space (“smart order routing”) is of high interest for practitioners because of the increasing complexity of the market micro structure since recent evolutions of regulations and liquidity worldwide. This article is translating in quantitative terms these regulatory issues and more broadly the current market design.

It confronts the recent advances in optimal trading, order-book simulation and optimal liquidity seeking to the reality of trading in an emerging global network of liquidity.

1 Market micro-structure modelling and payoff understanding are key elements of quantitative trading

As it is widely known, optimal (or quantitative) trading is about finding the proper balance between providing liquidity to minimise the impact of the trades, and consuming liquidity to minimise the market risk exposure, while taking profit of potential instantaneous trading signals, supposed to be triggered by liquidity inefficiencies.

The mathematical framework required to solve this kind of optimisation needs a model of the consequences of the different ways to interact with liquidity (like a market impact model [Almgren et al., 2005] [Wyart et al., 2008] [Gatheral, 2010]), a proxy for the “market risk” (the most natural of them being the high frequency volatility [Aït-Sahalia and Jacod, 2007, Zhang et al., 2005, Robert and Rosenbaum, 2011]) and a model to quantify the likelihood of the liquidity state of the market [Bacry et al., 2009, Cont et al., 2010]. A utility function allows then to consolidate these different effects with respect to the goal of the trader: minimising the impact of large trades under price, duration and volume constraints (typical for brokerage trading [Almgren and Chriss, 2000]), providing as liquidity as possible under inventory constraints (typical for market-makers [Avellaneda and Stoikov, 2008] or [Lehalle-Gueant-Frenandez]), or following a belief on the trajectory of the market (typical of arbitrageurs [Lehalle, 2009]).

*Global Head of Quantitative Research (clehalle@cheuvreux.com), Crédit Agricole Cheuvreux - 9 Quai Paul Doumer, Paris-La Defense, France

Once these key elements are defined, rigorous mathematical optimisation methods can be used to derive an optimal behaviour ([Bouchard et al., 2011, Predoiu et al., 2011]). Since the optimality of the result deeply relies on the modelled phenomena, the understanding of the market micro-structure is a pre-requisite to ensure the applicability of a given theoretical framework.

The *market micro-structure* is the ecosystem where buying interests meet selling ones, giving birth to trades. Seen from outside of the micro-structure, the prices of the traded shares are often uniformly sampled to build time series that are modelled via martingales [Shiryaev, 1999] or studied using econometrics. Seen from the inside of electronic markets, buy and sell open interests (i.e. passive limit orders) form *limit order books*, where an impatient trader can find two different prices: the highest of the resting buy orders if he needs to sell, and the lowest of the selling ones if he needs to buy (cf Figure 1). The price to sell and the price to buy are thus not equal. Moreover, the price will monotonically increase (for impatient buy orders) or decrease (for impatient sell orders) with the quantity to trade following a concave function [Smith et al., 2003]: the more you trade, the worst price you will obtain.

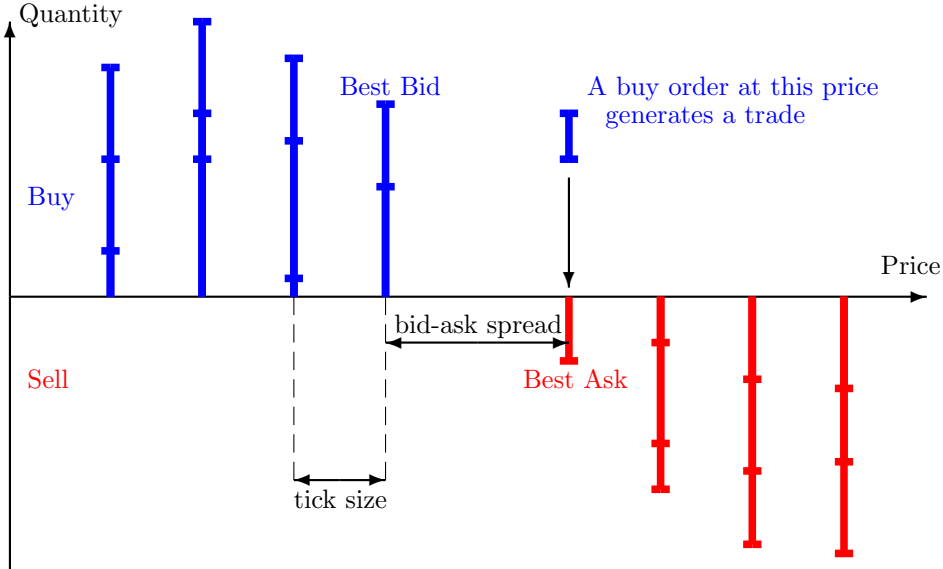


Figure 1: Stylized order-book

The market micro-structure is deeply conditioned by the *market design*, that is the set of explicit rules governing the *price formation process* (PFP): the type of auctions (fixing or continuous ones), the tick size (i.e. the minimum allowed difference between two consecutive prices), the interactions between trading platforms (like “trade-through rules”, pegged orders, interactions between visible and hidden orders, etc), are typical elements of the market design.

The market micro-structure of an asset class is a mix of the market design, the trading behaviours of trading agents, the regulatory environment, and the availability of correlated instruments (like Equity Traded Funds, Futures or any kind of derivative products). Formally, the micro-structure of a market can be seen as several sequences of auction mechanisms taking place in parallel, each of them having its specificity. For instance the German market place is mainly

composed (in 2011) of the Deutsche Börse regulated market, the Xetra mid-point, the Chi-X visible order-book, Chi-delta (the Chi-X hidden mid-point), Turquoise Lit and Dark Pools, BATS pools. The regulated market implements a sequence of fixing auctions and continuous auctions (one open fixing, one continuous session, one mid-auction and one closing auction), others implement only continuous auctions, and Turquoise mid-point implement optional random fixing auctions.

To optimise his behaviour, a trader has to choose an abstract description of the micro-structure of the markets he will interact with: this will be his model of market micro-structure. It can be a statistical “macroscopic” one like in the widely used Almgren-Chriss framework [Almgren and Chriss, 2000], in which the time is sliced in 5 or 10 minutes long time intervals during which the interactions with the market are aggregated in two statistical phenomena: the market impact as a function of the “participation rate” of the trader and the volatility as a proxy of the market risk. It can also be a microscopic description of the order-book behaviour like in the Alfonsi-Schied proposal [Alfonsi et al., 2010] in which the shape of the order-book and its resiliency to liquidity consuming orders is modelled.

This article will thus expose some relationships between the market design and the market micro-structure using examples of the European and the USA since they have seen regulatory changes (in 2007 for Europe with the MiFI Directive and in 2005 for the USA with the regulation NMS) as much as behavioral changes with the financial crisis of 2008. A detailed description of some important elements of the market micro-structure will be conducted: dark pools, impact of fragmentation on the price formation process, tick size, auctions, etc. Key events like the 2010 May 6 flash crash in the US market and some European market outages will be commented too.

To obtain an optimal trading trajectory, a trader needs to define its payoff. Here also choices have to be made from a mean-variance criteria [Almgren and Chriss, 2000] to stochastic impulse control [Bouchard et al., 2011] going through stochastic algorithms [Pagès et al., 2012]. This article comments on a statistical viewpoint of the Almgren-Chriss framework, showing how practitioners can use it to take into account a large variety of effects. It ends with comments on an order-flow oriented view of optimal execution, dedicated to smaller time scales problems, like “*Smart Order Routing*” (SOR).

2 From market design to market micro-structure: practical examples

The recent history of the French equity market is archetypal in the sense that it went from a highly concentrated design with only one electronic platform hosting in Paris [Muniesa, 2003] to a fragmented pan-European one with four visible trading pools and more than twelve “dark ones” located in London in less than four years.

Seen by economists and from the outside of the micro-structure, the equity market is a place where *listed firms* raise capital offering shares to buy. Once

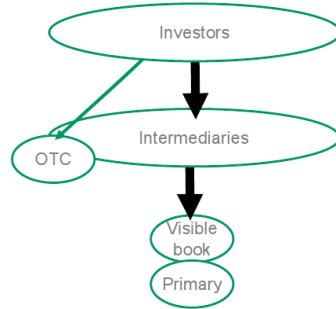


Figure 2: Stylized pre fragmentation market micro-structures

shares are available to buy and sell in the market place, the mechanism of balance between offer and demand (in terms of intentions to buy and intentions to sell) forms a *fair price*.

At the micro-structure scale, the market place is more sophisticated. Market participants are no more just listed firms and investors making rational investment decisions; micro-structure focuses on the process allowing an investor to buy or sell to another one, putting emphasis on the *Price Formation Process* also named *Price Discovery*. Moreover, recent regulations promote the use of electronic markets, being comfortable with the recording and traceability levels provided by such platforms, leading to *fragmented markets*. It is worthwhile to make the difference between two states of the micro-structure: pre and post fragmented one (Figure 2 and 4):

- *Pre fragmented micro-structure*: before Reg NMS in the US and MiFID in Europe, the micro-structure could be stylized in three distinct layers:
 - investors, taking buy or sell decisions;
 - intermediaries, giving unconflicted advices (through financial analysts or strategists) and providing access to trading pools they are members of; low frequency market makers (or maker-dealers) can be considered to be part of this layer;
 - market operators: hosting the trading platforms, NYSE Euronext, NASDAQ, BATS, Chi-X, belong to this layer. They are providing matching engines to other market participants, hosting the *Price Formation Process*.

These three layers are simply connected: intermediaries concentrate a fraction of the buying and selling flows in a (small) *Over the Counter* (OTC) market, the remaining open interests are confronted in the order-books of the market operators. Facilitators (i.e. low frequency market makers or specialists), localised in the same layer than the intermediaries provide liquidity, thus minimizing the *Market Impact* of the orders of not coordinated enough investors (i.e. when a large buyer comes in the market two hours after a large seller, any liquidity provider being able to sell to the first one and buy to the later will prevent a price oscillation; on the

one hand he will be “rewarded” for this service through the bid-ask spread he will demand to the two investors; on the other hand he will take the risk of a large change of the *fair price* in between the two transactions [Gabaix et al., 2006], see Figure 3).



Figure 3: Stylized kinematics of market impact caused by bad synchronisation (A1-A2-A3 sequence) and preservation of the market depth thanks to a market maker agreeing to support market risk (B1-B2-B3 sequence).

- *Post fragmented markets*: regulations evolved in a direction implementing more competition across each layer of the previous diagram (especially across market operators) and allowing more transparency:
 - in the US, Reg NMS decided to keep the competition inside the layer of market operators: it demands to an Exchange or an Electronic Communication Network (ECN) to route an order to the platform that offers the best match (it is called the *trade-through rule*). For instance, if a trader send a buy order at \$10.00 to BATS where the best ask price is \$9.75 and if the best ask for this stock is \$9.50 on NYSE, BATS has to re-route the order to NYSE. This regulation needs two important elements: (1) a way to push to all market operators the best bid and ask of any available market with accuracy (it raises concerns linked to the latency of market data); (2) that buying at \$9.50 on NYSE is always better for a trader than buying at \$9.75 on BATS, meaning that the other trading costs (especially clearing and settlement costs) are the same. The data conveying all the best bid and asks is named the *consolidated pre-trade tape* and its best bid and offer is named the *National Best-Bid and Offer* (NBBO).

- in Europe, mainly because of the diversity of the clearing and settlement channels, MiFID allowed to extend the competition to the intermediaries: they are in charge of defining their *Execution Policies* describing how and why they will route and split orders across market operators. The European Commission thus relies on competition between execution policies to select the best way to split orders taking into account all trading costs. As a consequence, Europe does not have any officially consolidated pre-trade tape.

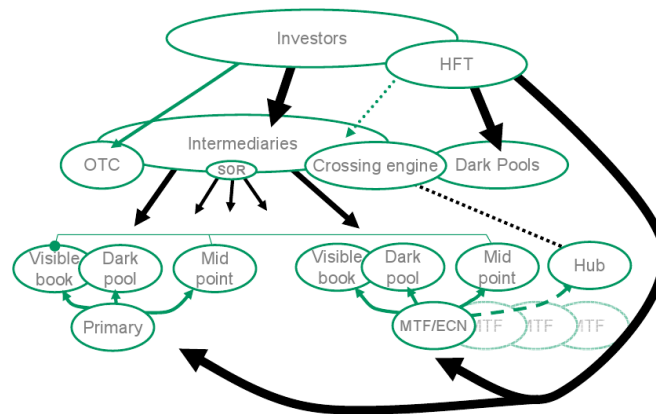


Figure 4: Stylized post fragmentation market micro-structure.

Despite these differences, European and US electronic markets have a lot in common: their micro-structures evolved similarly to a state where latency is crucial and *High Frequency Market-Makers* (also called *High Frequency Traders*) became the main liquidity providers of the market.

Diagram of Figure 4 gives a stylized view of this fragmented micro-structure:

- A specific class of investors: the *High Frequency Traders* (HFT) became an essential part of the market; investing more than other market participants in technology, thus reducing their latency to markets, they succeeded in implementing market-making-like behaviours at high frequency: providing liquidity at the bid and ask prices when the market has few chances to move (thanks to statistical models) and being able to cancel very fast resting orders to minimize the market risk exposure of their inventory, they are said to be part of 70% of the transactions on US Equity markets, 40% in Europe and 30% in Japan in 2010. Their interactions with the market has been intensively studied by Menkveld in [Menkveld, 2010].

- Because they are the main customers of market operators, HFTs obtained new features easing their activity: low latency access to matching engines (better quality of service and *co-hosting*; i.e. the ability to locate their computers physically close to the ones of the matching engines), and even *flash orders* (the capability to know before other market participants that an order is being inserted in the order-book).
- Market participants being not proprietary high-frequency traders asked specific features of the order books too, mainly to hide their interests to high frequency traders: *Dark Pools*, implementing anonymous auctions (i.e. partially observable), are part of this offer.
- On the one hand, the number of market operators as firms does not increase that much when a market goes from non-fragmented to fragmented, because of high technological costs linked to a fragmented micro-structure. On the other hand, each operator offers more products (order-books) to clients when fragmentation increases. The BATS and Chi-X Europe merged and the London Stock Exchange - Milan Stock Market - Turquoise trading formed an unique group too. Looking at the European order-books offered by NYSE-Euronext only in 2011, we have:
 - * several visible (i.e. *Lit*) *order books*: one for Paris-Amsterdam-Brussels stocks, another (Nyse-Arca Europe) for other European names;
 - * *Mid-points*: an order book with only one queue *pegged* at the mid-price of a reference market (SmartPool);
 - * *Dark pools*: an anonymous order book (i.e. market participants can send orders like in a Lit book, but no-one can read the state of the book);
 - * *Fixing auctions*, opening and closing the continuous auctions on visible books.

The result is an interconnected network of liquidity in which each market participant is no more located in one layer only: HFTs are at the same time investors and very close to market operators too, intermediaries are offering *Smart Order Routers* to split optimally orders across all available trading pools taking into account the specific liquidity needs of each investor, thus being close to operators too, market operators are close to technology providers.

The regulatory task is thus more sophisticated in a fragmented market rather than in a concentrated one:

- the *Flash Crash* of May the 6, 2010 on US markets raised concerns about the stability of such a micro-structure (see Figure 5);
- the cost of surveillance of trading flows across a complex network is higher than in a concentrated one.

Moreover, elements of the market design play a lot of different roles: the *tick size* for instance, is not only the minimum between two consecutive different prices,

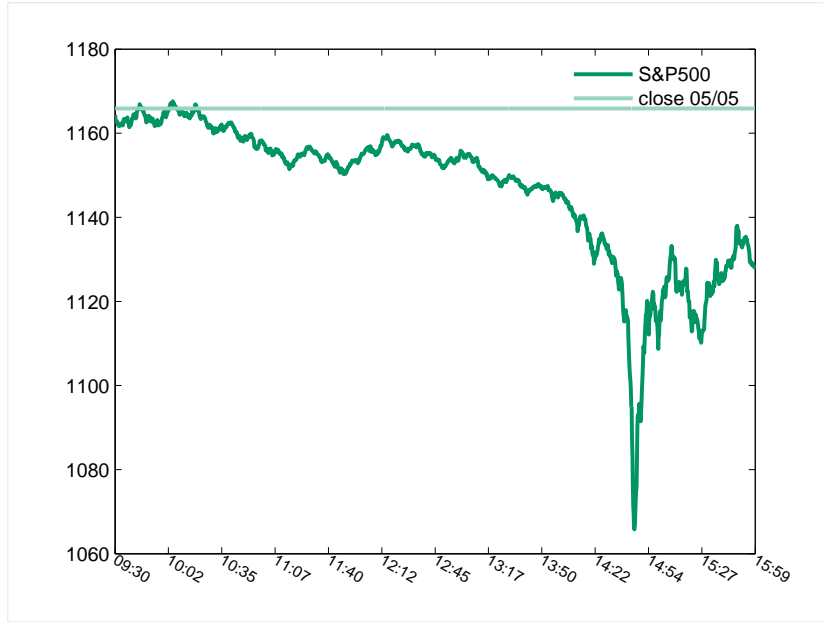


Figure 5: The “Flash Crash”: May, the 6, 2010, US market rapid down-and-up move by almost 10% was only due to market micro-structure effects.

or a constraint on the bid-ask spread; it is a key in the competition between market operators too. In June 2009, European market operators tried to gain market shares in reducing the tick size on their order-books. Each time one of them offered a lower tick than others, it gained around 2% of market shares (see Figure 6). After few weeks of competition on the tick, they limited this kind of infinitesimal decimalisation of the tick thanks to a gentleman agreement obtained under the umbrella of the FESE (Federation of European Security Exchanges): such a decrease has been harassing in CPU and memory demand for their matching engines.

A stylized view of the “Flash Crash”. The flash crash has been precisely described in [Kirilenko et al., 2010]. The sequence of events that conducted to a negative jump of price and huge increase of traded volumes in few minutes, followed by a return to normal in less than 20 minutes can be stylized that way:

1. A final investor decided to sell a large amount v^* of shares of the E-Mini future contracts, he asked to a broker to take care of this sell by electronic means on his behalf.
2. The broker decided to use a *PVOL* (i.e. Percentage of Volume) algo, with the instruction to follow almost uniformly 9% of the market volume without regard of price or time. This participation rate is not uncommon (it is usual to see PVOL algos with the instruction to follow 20% of the market volume).
3. the trading algorithm can be seen as a trade scheduler splitting the order in slices of 1 minutes, expecting to see a traded volume V_t during the t th

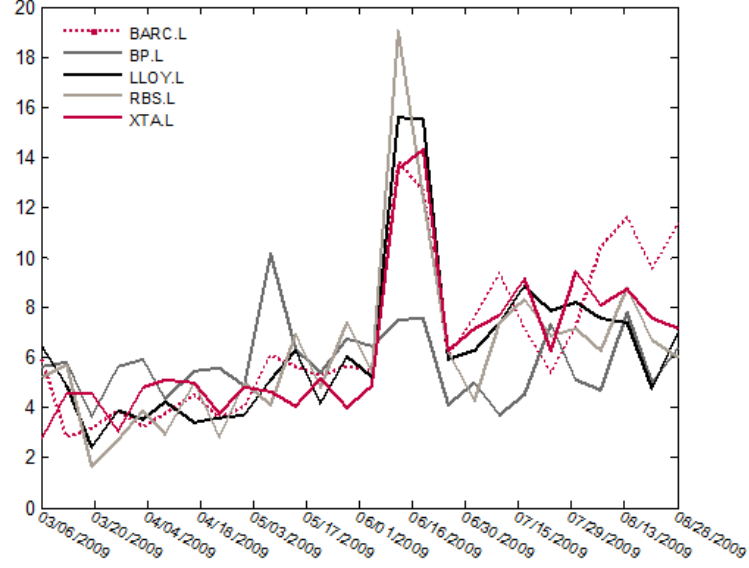


Figure 6: The “Tick war” in June 2009, in Europe. The increase of market share of Turquoise (an European Multilateral Trading Facility; MTF) on 5 Stocks listed on the London Stock Exchange following a decrease of the tick size. When other MTFs lowered the tick size, the market share get back to the previous level.

slice (meaning that $\mathbb{E}(V_t) \simeq \bar{V}/500$, where \bar{V} is the expected daily traded volume).

4. for its first step, the algo begun to sell on the future market around $v_0 = \mathbb{E}(V_0) \times 9/(100 - 9) \simeq \bar{V}/500 \times 0.09$ shares,
5. the main buyers of these shares have been intra-day market makers; say that they bought $(1 - q)$ of them,
6. because the volatility was quite high the 6th of May 2010, the market makers did not feel comfortable with such an imbalanced inventory, they decided to hedge it on the cash market, selling $(1 - q) \times v_t$ shares of a properly weighted basket of Equities,
7. unfortunately the buyers of most of these shares (say $(1 - q)$ of them again), were intra-day market makers themselves, who decided at their turn to hedge their risk on the future market,
8. it immediately increased the traded volume on the future market by $(1 - q)^2 v_0$ shares;
9. assuming that intra-day market makers could play this *hot potato game* (as it has been named in the SEC-CFTC report), N times in 1 minutes; the volume traded on the future market is now $\sum_{n \leq N} (1 - q)^{2n} v_0$ larger than expected by the brokerage algo;

10. back to step 4 at $t + 1$, the PVOL algo is now late by $\sum_{n \leq N} (1 - q)^{2n} v_0 \times 8/(100 - 8)$, and has to sell $\bar{V}/500 \times 8/(100 - 8)$ again; i.e. selling

$$v_{t+1} \simeq \left(N \times v_t + \frac{\bar{V}}{500} \right) \times 0.08$$

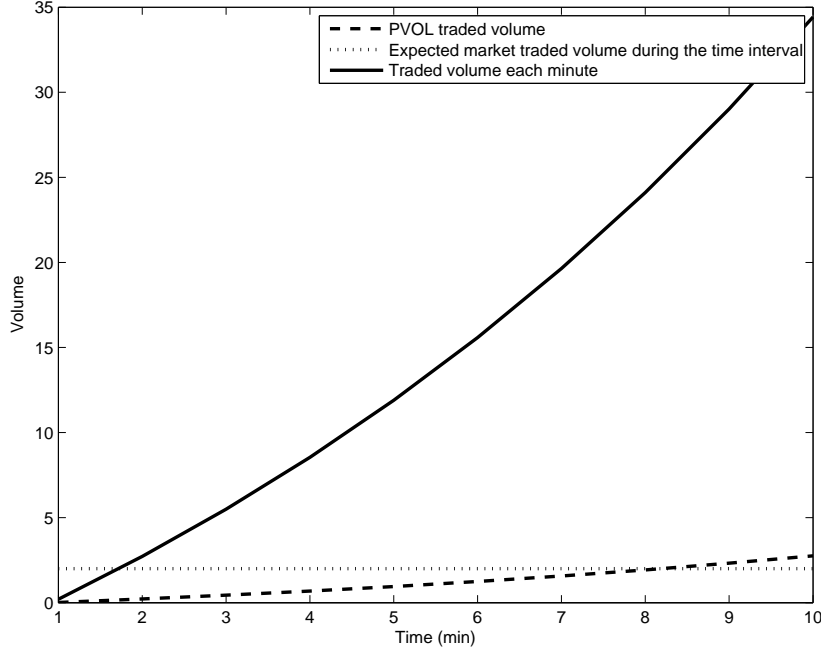


Figure 7: Traded volume of the future market according to the simple stylized model with $\bar{V} = 100$, $T = 10$ and $N = 2$.

Figure 7 shows how explosive can be the *hot potato game* between intra-day market makers, even with not a very high frequency trading rate (here $N = 1.1$). Most of this trading flow has been a selling flow, pushing most of US prices to very low levels. For instance Procter and Gamble quoted from \$60 to a low of \$39.37 in approximately 3.5 minutes.

In reality other effects contributed to the flash crash:

- only some trading pools implemented circuit breakers that should freeze the matching engines in case of sudden liquidity event,
- most market participants only looked at the *consolidated tape* for market data, preventing them to notice that on some pools the trading was frozen,
- in the US, most retail flow is internalized by market makers, at one point in the day these intermediaries decided to hedge their positions on the market at their turn, pushing more the prices.

This glitch in the electronic structure of markets is not an isolated case, even if it has been the largest one. The conjunction of a failure in each layer of

the market (an issuer of large institutional trades, a broker, HF market-makers, market operators) with an highly uncertain market context is for sure a crucial element of this crash. It has moreover shown that most orders do reach the order-books through electronic means only.

European markets did not suffered from such *flash crashes*, but they have not seen many months of 2011 without an outage of a matching-engine.

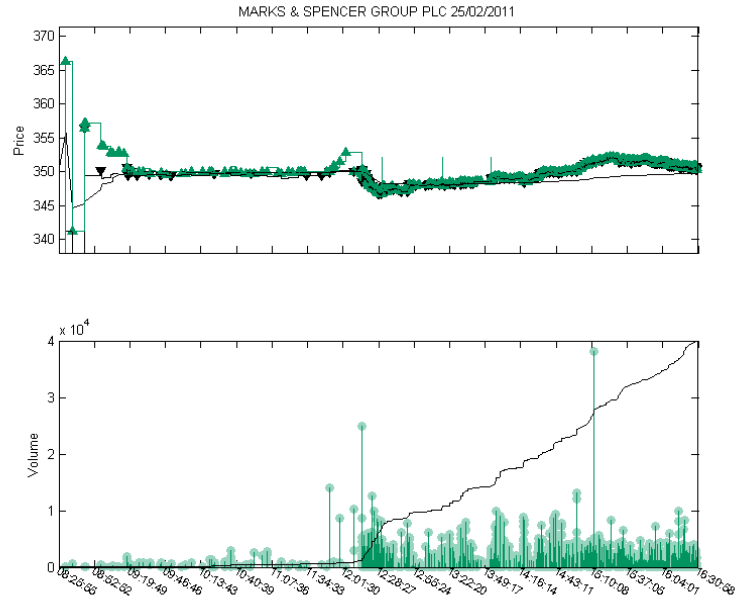


Figure 8: Examples of outages in European equity markets the 25th of February, 2011. The price (top) and the volumes (bottom) when the primary market did only open after 12:15 (London time). The price did not move a lot.

European outages. Outages are “simply” bugs in matching engines. In such cases, the matching engines of one or more trading facilities can be frozen, or just stop to publish market data, becoming true *Dark Pools*. From a scientific viewpoint, and because in Europe there is no *consolidated pre-trade tape* (i.e. each member of the trading facilities need to be connected explicitly to each of them, and has a proprietary methodology to build its *consolidated view* of the current European best bid and offer), they then provide examples of behaviour of market participants when they do not all share the same level of information on the state of the offer and demand.

For instance :

- when no information is available on primary markets but trading remains open: two price formation processes can take place in parallel, one for market participants having access to other pools, and the other for participants who just looked at the primary market,

- (Figure 8) when the primary market does not start trading at the very beginning of the day: the price does not really move on alternative markets; no “real” price formation process takes place during such European outages.

The flash crash in US and the European outages emphasize the *role of information in the price formation process*. When market participants are confident that they have access to a reliable source of information (during the flash or during some European outages), they continue to *mimic* a price formation process which output can be far from efficient. On the contrary, if they do not believe in the information they have, they just freeze their price discovering behaviour and trade at the *last confident price*, waiting for reliable updates.

3 Forward and backward components of the price formation process

The literature on market micro-structure can be split in two generic subsets:

- papers with a *Price Discovery* viewpoint, in which the market participants are injecting into the order book their views on a fair price. In these papers (see for instance [Biais et al., 2005, Ho and Stoll, 1981, Cohen et al., 1981]) the *fair price* is assumed to exist for fundamental reasons (at least in the mind of investors) and the order books are implementing a Brownian-bridge-like trajectory targeting this evolving fair price. This is a *backward* view of the price dynamics: the investors are updating assumptions on the future value of tradeable instruments, and send orders in the electronic order books according to the distance between the current state of the offer and demand and this value, driving the quoted price to the aggregation of their anticipations.

Figure 9 shows a price discovery pattern: the price of the stock changes for fundamental reasons, and the order book dynamics react accordingly generating more volume, more volatility, and a price jump.

- Other papers rely on a *Price Formation Process* viewpoint. For their authors (most of them being econophysicists, see for instance [Smith et al., 2003, Bouchaud et al., 2002] or [Chakraborti et al., 2011] for a review of agent based models of order books) the order books are *building the price* in a forward way. The market participants take decisions with respect to the current orders in the books making assumptions of the future value of their inventory; it is a *forward* process.

Following [Lehalle et al., 2010], it is possible to attempt a crude modelling of these two dynamics simultaneously. In a framework with an infinity of agents (using a Mean Field Game approach, see [Lasry and Lions, 2007] for more details), the order book at the bid (respectively at the ask), is a density $m_B(t, p)$ (resp. $m_A(t, p)$) of agents agreeing at time t to buy (resp. sell) at price p . In such a continuous framework, there is no bid-ask spread and the *trading price* $p^*(t)$ is such that there is no offer at a price lower than $p^*(t)$ (and no demand

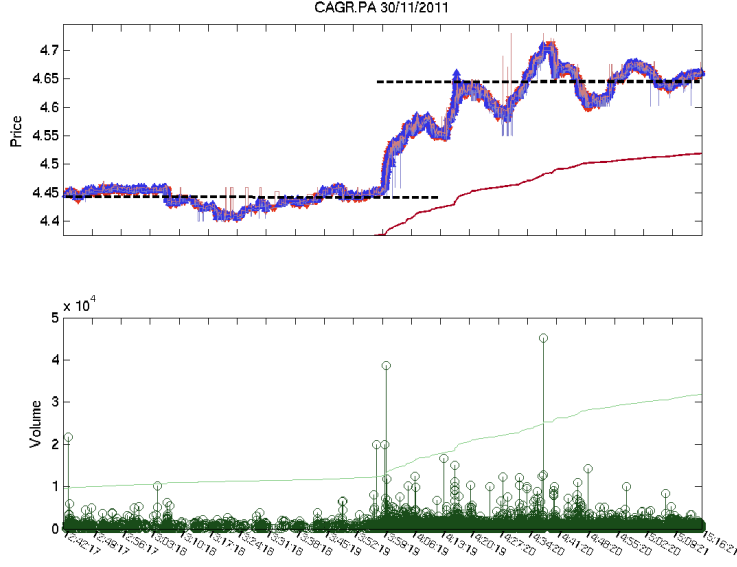


Figure 9: A typical *Price Discovery* exercise: the 30th of November, 2011 on the Cr dit Agricole share price (French market). The two stable states of the price are materialized using two dark dotted lines, one before and the other after the announcement by major European central banks a coordinated action to provide liquidity.

at a price greater then $p^*(t)$. Under diffusive assumptions, the two sides of the order book are submitted to this simple partial differential equation:

$$\begin{aligned} \partial_t m_B(t, p) - \frac{\varepsilon^2}{2} \partial_{pp}^2 m_B(t, p) &= \lambda(t) \delta_{p=p^*(t)} \\ \partial_t m_A(t, p) - \frac{\varepsilon^2}{2} \partial_{pp}^2 m_A(t, p) &= \lambda(t) \delta_{p=p^*(t)} \end{aligned}$$

moreover, the trading flow at $p^*(t)$ is clearly defined as:

$$\lambda(t) = -\frac{\varepsilon^2}{2} \partial_p m_B(t, p^*(t)) = \frac{\varepsilon^2}{2} \partial_p m_A(t, p^*(t))$$

It is then possible to define a regular order book m joining the bid side to the ask one by:

$$m(t, p) = \begin{cases} m_B(t, p) & , \quad \text{if } p \leq p^*(t) \\ -m_A(t, p) & , \quad \text{if } p > p^*(t) \end{cases}$$

which satisfies a unique parabolic equation:

$$\partial_t m(t, p) - \frac{\varepsilon^2}{2} \partial_{pp}^2 m(t, p) = -\frac{\varepsilon^2}{2} \partial_p m(t, p^*(t)) (\delta_{p=p^*(t)-a} - \delta_{p=p^*(t)+a}) \quad (1)$$

with a limit conditions $m(0, \cdot)$ given on the domain $[p_{min}, p_{max}]$ and, for instance, Neumann conditions at p_{min} and p_{max} .

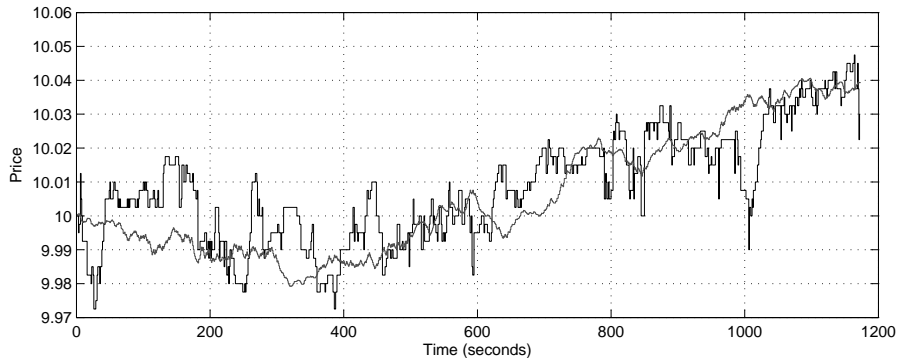


Figure 10: Simulation of the dynamics of an order book modelling with a forward-backward approach: the “fair price” is the continuous grey line and the realized price is the stepwise dark one.

Such a *forward process* describes the order-book dynamics without any impact of investors’ fundamental views (it is a *price formation process* model).

Authors of the same paper then introduce a more complex source to re-inject the orders in the books containing market participants forward views on the price. For instance, a trend follower with a time horizon of h buying at price $p^*(t)$ at time t targets to unwind its position at an higher (i.e. “trend targeted”) price and thus insert an order in the book accordingly (around $p^*(t) + (p^*(t) - p^*(t - h))$) see the paper for more details). Figure 10 shows an example of such a dynamic.

This is a way to introduce investors-driven views in the model, which are essentially *backward*: a trend follower accepts to be part of a transaction because he believes that the price will continue to move in the same direction at his investment time scale. This future price of the share is at the root of his decision. This is an injection of a *price discovery* component in the model.

4 From statistically optimal trade scheduling to microscopic optimisation of order-flows

Modelling the price formation dynamics is of interest for the regulators and the policy makers. It enables them to understand the potential effects of a regulatory or rule change on the efficiency of the whole market (see for instance [Foucault and Menkveld, 2008] for an analysis of the introduction of competition among trading venues on the efficiency of the markets). It thus helps in understanding potential links between market design and systemic risk.

In terms of risk management inside a firm hosting trading activities, it is more important to understand the trading cost of a position, which can be understood as its *liquidation risk*.

From the viewpoint of one trader confronted to the whole market, three key phenomena have to be controlled:

- the *market impact* (see [Kyle, 1985, Lillo et al., 2003, Engle et al., 2012, Almgren et al., 2005, Wyart et al., 2008]) is the market move generated

by selling or buying a large amount of shares (all else being equal); it comes from the forward component of the price formation process, and can be temporary if other market participants (being part of the backward component of the price discovery dynamics) provide enough liquidity to the market to bring back the price at its previous level;

- *adverse selection*, capturing the fact that providing too much (passive) liquidity via limit orders, the trader can maintain the price at an artificial level; not a lot of literature is available on this effect, which has been nevertheless identified by practitioners [Altunata et al., 2010];
- and the uncertainty on the fair value of the stock that can move the price during the trading process; it is often referred as the *intra-day market risk*.

4.1 Replacing market impact by statistical costs

A now widely used framework to control the overall costs of the liquidation of a portfolio has been proposed by Almgren and Chriss in the late nineties [Almgren and Chriss, 2000]. Applied to trade a single stock, this framework:

- cut the trading period into an arbitrary number of intervals N of a chosen duration δt ,
- models the *fair price* moves thanks to a Gaussian random walk:

$$S_{n+1} = S_n + \sigma_{n+1} \sqrt{\delta t} \xi_{n+1} \quad (2)$$

- models the *temporary market impact* η_n inside each time bin using a power law of the trading rate (i.e. the ratio of the traded shares v_n by the trader over the market traded volume during the same period V_n):

$$\eta(v_n) = a \psi_n + \kappa \sigma_n \sqrt{\delta t} \left(\frac{v_n}{V_n} \right)^\gamma \quad (3)$$

where a , κ and γ are parameters, and ψ is the half bid-ask spread;

- the *permanent market impact* is taken linear in the participation rate;
- uses a mean-variance criterion and minimise it to obtain the optimal sequence of shares to buy (or sell) through time.

It is first important to notice that there is an implicit relationship between the time interval δt and the temporary market impact function: without changing η and simply by choosing a different slicing of time, the cost of trading is changed. It is in fact not possible to choose (a, κ, γ) and δt independently; they have to be chosen accordingly to the decay of the market impact on the stock, provided that most of the impact is kept in a time bin of size δt . Not all the decay functions are compatible with this view (see [Gatheral and Schied, 2012] for details about available market impact models and their interactions with trading). Up to now the terms in $\sqrt{\delta t}$ has been ignored. It will also be considered that the parameters (a, κ, γ) exist at this time scale.

It is not mandatory to see this framework as if it was based on structural model assumptions (i.e. that the market impact really have this shape, or that the price moves are really Brownian), but as a statistical one. With such a viewpoint, any practitioner can use the database of its past executed orders and perform an econometric study of its “trading costs” on any δt interval of time (see [Engle et al., 2012] for an analysis of this kind on the whole duration of the order). If a given time scale succeeds into capturing the parameters of a model of trading cost with enough accuracy, then this model can be used to optimise trading. Formally, the result of such a statistical approach will be the same as the structural one, as it will be shown later in this paper, it is besides possible to go one step further, taking into account the statistical properties of the variables (and parameters) of interest.

Going back to the simple case of the liquidation of one stock without any permanent market impact, the value (which is a random variable) of a buy of v^* shares in N bins of size v_1, v_2, \dots, v_N is (see [Bouchard et al., 2011] for a more sophisticated model and more generic utility functions; since a stylized model is used here to obtain easier illustrations of phenomena of interest) is:

$$\begin{aligned}
W(v_1, v_2, \dots, v_N) &= \sum_{n=1}^N v_n (S_n + \eta_n(v_n)) \\
&= S_0 v^* + \underbrace{\sum_{n=1}^N \sigma_n \xi_n x_n}_{\text{market move}} + \underbrace{\sum_{n=1}^N a \psi_n(x_n - x_{n+1}) + \kappa \frac{\sigma_n}{V_n^\gamma} (x_n - x_{n+1})^{\gamma+1}}_{\text{market impact}} \quad (4)
\end{aligned}$$

using the *remaining quantity to buy*: $x_n = \sum_{k \geq n} v_k$ instead of the instantaneous volumes v_n . To obtain as much closed form formula as possible, γ will be taken equal to 1 (i.e. linear market impact).

To add a practitioner-oriented flavor to our upcoming optimisation problems, just introduce a set of independent random variables $(A_n)_{1 \leq n \leq N}$ to model the *arbitrage opportunities* during time slices. It will reflect the anticipation that the trader will be able to buy shares at price $S_n - A_n$ during slice n rather than at price S_n .

Such an effect can be used to inject a statistical arbitrage approach into optimal trading or to take into account the anticipation of the opportunity to cross orders at mid price in Dark Pools or Broker Crossing Networks (meaning that the expected trading costs should be smaller during given time slices). Now the cost to buy v^* shares is:

$$W(\mathbf{v}) = S_0 v^* + \sum_{n=1}^N \sigma_n \xi_n x_n + \sum_{n=1}^N (a \psi_n - A_n) v_n + \kappa \frac{\sigma_n}{V_n} v_n^2 \quad (5)$$

Conditioned expectation optimisation. The expectation of this cost $\mathbb{E}(W|(V_n, \sigma_n, \psi_n)_{1 \leq n \leq N})$ given the market state writes:

$$C_0 = S_0 v^* + \sum_{n=1}^N (a \psi_n - \mathbb{E}A_n) v_n + \kappa \frac{\sigma_n}{V_n} v_n^2 \quad (6)$$

a simple optimisation under constraint (to ensure $\sum_{n=1}^N v_n = v^*$) gives:

$$v_n = w_n \left(v^* + \frac{1}{\kappa} \left(\left(\mathbb{E}A_n - \sum_{\ell=1}^N w_\ell \mathbb{E}A_\ell \right) - a \left(\psi_n - \sum_{\ell=1}^N w_\ell \psi_\ell \right) \right) \right) \quad (7)$$

where w_n are weights proportional to the inverse of the market impact factor:

$$w_n = \frac{V_n}{\sigma_n} \left(\sum_{\ell=1}^N \frac{V_\ell}{\sigma_\ell} \right)^{-1}$$

Simple effects can be deduced from this first stylized result:

1. without any arbitrage opportunity and without any bid-ask cost (i.e. $\mathbb{E}A_n = 0$ for any n and $a = 0$), the optimal trading rate is proportional to the inverse of the market impact coefficient: $v_n = w_n \cdot v^*$. Moreover, when the market impact has no intra-day seasonality, $w_n = 1/N$ implying that the optimal trading rate is linear.
2. following formula (7) it can be seen that: the largest the expected arbitrage gain (or the lower the spread cost) on a slice compared to the market-impact-weighted expected arbitrage gain (or spread cost) over the trading full interval, the largest quantity to trade during this slice. More quantitatively:

$$\frac{\partial v_n}{\partial \mathbb{E}A_n} = \frac{w_n}{2\kappa} (1 - w_n) > 0, \quad \frac{\partial v_n}{\partial \psi_n} = -\frac{a}{2\kappa} (1 - w_n) w_n < 0$$

This result gives the *adequate weight* to apply to the expected arbitrage gain to translate it in an adequate trading rate to take profit on arbitrage opportunities on average. Just note that usually the expected arbitrage gains increase with market volatility, the w_n -weighting is consequently of interest to balance this effect optimally.

Conditioned mean-variance optimisation. Going back to a mean-variance optimisation of the cost to buy progressively v^* shares, the criteria to minimise (using a risk aversion parameter λ) writes:

$$\begin{aligned} C_\lambda &= \mathbb{E}(W|(V_n, \sigma_n, \psi_n)_{1 \leq n \leq N}) + \lambda \mathbb{V}(W|(V_n, \sigma_n, \psi_n)_{1 \leq n \leq N}) \\ &= S_0 v^* + \sum_{n=1}^N (a \psi_n - \mathbb{E}A_n) (x_n - x_{n+1}) + \left(\kappa \frac{\sigma_n}{V_n} + \lambda \mathbb{V}A_n \right) (x_n - x_{n+1})^2 \end{aligned} \quad (8)$$

To minimise C_λ being only constrained by terminal conditions on x (i.e. $x_0 = v^*$ and $v_{N+1} = 0$) it is enough to cancel its derivatives with respect to any x_n , leading to a recurrence relation:

$$\begin{aligned} \left(\frac{\sigma_n}{V_n} + \frac{\lambda}{\kappa} \mathbb{V}A_n \right) x_{n+1} &= \frac{1}{2\kappa} (a(\psi_{n-1} - \psi_n) - (\mathbb{E}A_{n-1} - \mathbb{E}A_n)) \\ &+ \left(\frac{\lambda}{\kappa} \sigma_n^2 + \left(\frac{\sigma_n}{V_n} + \frac{\lambda}{\kappa} \mathbb{V}A_n + \frac{\sigma_{n-1}}{V_{n-1}} + \frac{\lambda}{\kappa} \mathbb{V}A_{n-1} \right) \right) x_n \\ &- \left(\frac{\sigma_{n-1}}{V_{n-1}} + \frac{\lambda}{\kappa} \mathbb{V}A_{n-1} \right) x_{n-1} \end{aligned} \quad (9)$$

It shows that the variance of the arbitrage has a similar effect than the market impact (through a risk-aversion rescaling), and that the risk aversion parameter acts as a multiplicative factor to the market impact, meaning that within an arbitrage-free and spread-costs-free framework (i.e. $a = 0$ and $\mathbb{E}A_n = 0$ for all n), the market impact model by any constant b has no effect on the final result as far as λ is replaced by $b\lambda$.

Figure 11 compares optimal trajectories coming from different criteria and parameter values.

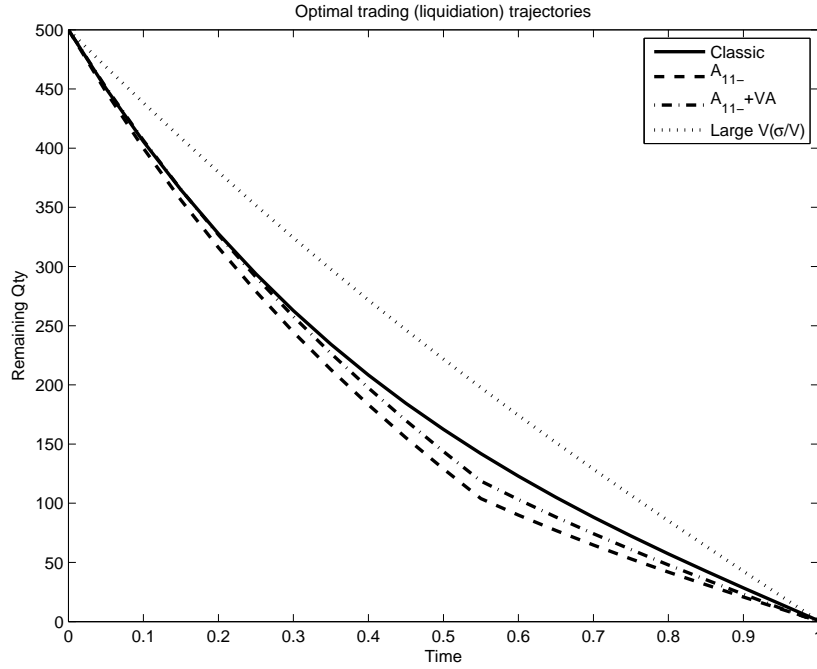


Figure 11: Examples of optimal trading trajectories for mean-variance criteria: the classical result (Almgren-Chriss) in solid line, the dotted line is for high variance of the variable of interest (σ/V), the semi-dotted ones for an arbitrage opportunity (A_{11+} means: after the 11th period and $A_{11+} + VA$ means: adding expected variance to the arbitrage opportunity).

A statistical viewpoint. The two previous examples show how easy it is to include effects in this sliced mean-variance framework, the implicit assumptions are:

- on one time-slice, it is possible to capture the market impact (or *trading costs*) using model (3),
- the trader has a view on the traded volumes and market volatility at the same time-scale.

Practically, the two assumptions come from statistical modelling:

- the market impact parameters a, κ and γ are estimated on a large database of trades using a maximum likelihood or MSE methods ; the reality is consequently that the market model has the following shape:

$$\eta(v_n) = a \psi_n + \kappa \sigma_n \sqrt{\delta t} \left(\frac{v_n}{V_n} \right)^\gamma + \varepsilon \quad (10)$$

where ε is an i.i.d. noise.

- moreover, the market volatility and traded volumes are estimated using historical data and market context assumptions (to take into account at least the scheduled news, like the impact of the expiry of derivative products on the volume of the cash market, see Figure 12 for typical estimates).

Taking these statistical modelling steps into account in the classical mean-variance criteria of equation (8) changes it into its unconditioned version:

$$\begin{aligned} \tilde{C}_\lambda &= \mathbb{E}(W) + \lambda \mathbb{V}(W) \\ &= S_0 v^* + \sum_{n=1}^N (a \mathbb{E} \psi_n - \mathbb{E} A_n) (x_n - x_{n+1}) \\ &\quad + \left(\kappa \mathbb{E} \left(\frac{\sigma_n}{V_n} \right) + \lambda (a^2 \mathbb{V} \psi_n + \mathbb{V} A_n + \mathbb{V} \varepsilon) \right) (x_n - x_{n+1})^2 \\ &\quad + \lambda \sigma_n^2 x_n^2 + \lambda \kappa^2 \mathbb{V} \left(\frac{\sigma_n}{V_n} \right) (x_n - x_{n+1})^4 \end{aligned} \quad (11)$$

The consequence of using this criteria rather than the conditioned one are clear:

- the simple plug-in of empirical averages of volumes and volatility in criteria (8) instead of the needed expectation of the overall trading costs leads to use $(\mathbb{E} \sigma_n)/(\mathbb{E} V_n)$ instead of $\mathbb{E}(\sigma_n/V_n)$, Figure 12 shows typical differences between the two quantities.
- if the uncertainty on the market impact is huge (i.e. the $\mathbb{V} \varepsilon$ term dominates all others), then the optimal trading strategy is to trade linearly, which is also the solution of a pure expectation-driven minimisation with no specific market behaviour linked with time.

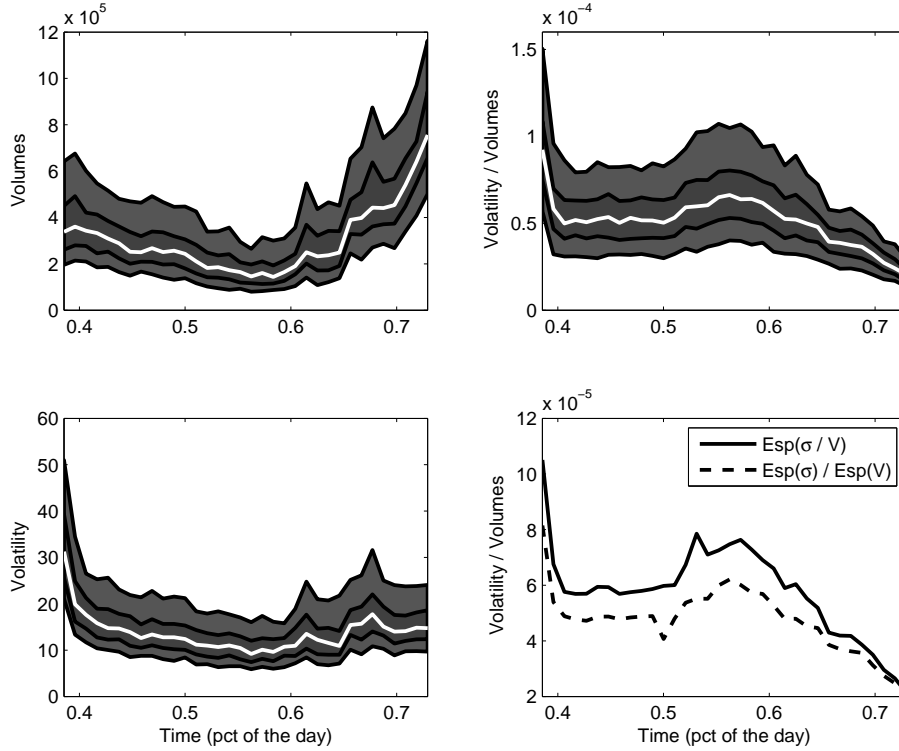


Figure 12: Typical intra-day traded volume (top-left) and realized volatility (bottom-left) profiles (i.e. intra-day seasonalities on traded volumes and market volatility) with their quantiles of level 25% and 75%. X-axis is time. The top-right chart figures the quantiles of the ratio of interest σ/V . The bottom-right ones shows the difference between the expectation of the ratio (solid line) and the ratio of the expectations (dotted line).

Within this new statistical trading framework, the inaccuracy of the models and the variability of the market context are taken into account: the obtained optimal trajectories will no more have to follow sophisticated behaviours if the models are not realistic enough.

Moreover, it is not difficult to solve the optimisation program associated to this new criteria; the new recurrence equation is a polynomial of degree 3. Figure 11 gives illustrations of obtained results.

A lot of other effects can be introduced in the framework, like auto-correlations on the volume-volatility couple. This statistical framework does not embed recent and valuable proposals like the decay of market impact [Gatheral and Schied, 2012] or a set of optimal stopping times to do not stick to an uniform and a priori sampled time [Bouchard et al., 2011]. It is nevertheless simple enough so that most practitioners can use it to include their views of the market conditions and the efficiency of their interactions with the market at a given time scale; it can be compared to the Markowitz approach for quantitative portfolio allocation [Markowitz, 1952].

4.2 An order-flow oriented view of optimal execution

If the price dynamics are often modelled using diffusive processes in quantitative finance, just looking at prices of transactions in a limit order book convinces that a more discrete and event-driven class of model has to be used; at a time scale of several minutes or more, the diffusive assumptions used in equation (2) to model the price are not that false, but even at this scale, the “*bid-ask bounce*” has to be taken into account to be able to estimate with enough accuracy the intra-day volatility. The effect on volatility estimates of the rounding of a diffusion has been first studied in [Jacod, 1996], then other effects have been taken into account, like an additive *micro-structure noise* [Zhang et al., 2005], sampling [Aït-Sahalia and Jacod, 2007] or *liquidity thresholding* (or uncertainty zones) [Robert and Rosenbaum, 2011]. Thanks to these modelling efforts, it is now possible to use high frequency data to estimate the volatility of an underlying diffusive process generating the prices without being polluted by the signature plot effect (i.e. an explosion of the usual empirical estimates of volatility when high frequency data are used).

Similarly advances have been made to obtain accurate estimates of correlations between two underlying prices avoiding the drawback of the *Epps effect* (i.e. a collapse of usual estimates of correlations at small scales [Hayashi and Yoshida, 2005]).

To optimise the interactions of trading strategies with the order-books, it is mandatory to zoom as much as possible and to model most known effects taking place at this time scale (see [Wyart et al., 2008, Bouchaud et al., 2002]). Point processes have been used with success that for, especially because they can embed short memory modelling [Large, 2007, Hewlett, 2006]. Hawkes-like processes have most of these interesting properties and exhibit a diffusive behaviour when the time scale is zoomed out [Bacry et al., 2009]. To model the prices of the transactions at the bid N_t^b and at the ask N_t^a , two coupled Hawkes processes can be used. Their intensities Λ_t^b and Λ_t^a are stochastic and governed by (μ^b and μ^a are constants):

$$\Lambda_t^{a/b} = \mu^{a/b} + c \int_{\tau < t} e^{-k(t-\tau)} dN_t^{b/a}$$

In such a model: the more transactions at the bid (resp. ask), the more probable to have one at the opposite price in a short future.

The next qualitative step has been to link the prices with the traded volumes. It has recently been shown that under some assumptions that are almost true for very liquid stocks in a calm market context (a constant bid-ask spread and no dramatic change in the dynamics of liquidity providing orders), there is a correspondence between a two dimensional point process of the quantities available at the first limits and the price of the corresponding stock [Cont et al., 2010, Cont and De Larrard, 2011].

To understand the mechanism underlined by such an approach, just notice that the set of stopping times defined by the instants when the quantity at the first ask crosses zero (i.e. $\mathcal{T}^a = \{\tau : Q_\tau^a = 0\}$) exactly maps the increases of prices (if the bid-ask spread is constant). Similarly the set $\mathcal{T}^b = \{\tau : Q_\tau^b = 0\}$ maps the decreases of the price.

Despite these valuable proposals to model the dynamics of the order book at small time scales, no direct use of such effects in an optimal trading framework is currently available. The most sophisticated approaches for optimal

trading including order-book dynamics are based on continuous and martingale assumptions [Alfonsi and Schied, 2010] or on Poisson like point processes [Guéant et al., 2011].

On another hand, focusing on the optimality of very short term trading strategies (like *Smart Order Routing*) allows to build optimal tactics with assumptions in accordance with the recent upper views on order book dynamics. Smart Order Routers (SOR) are software devices dedicated to split an order across all available trading venues to obtain the desired quantity as fast as possible implementing a so-called *liquidity capturing scheme*. With the raise of fragmentation of electronic equity markets (see Figure 13), it is impossible to access to more than 60% of European liquidity without a SOR.

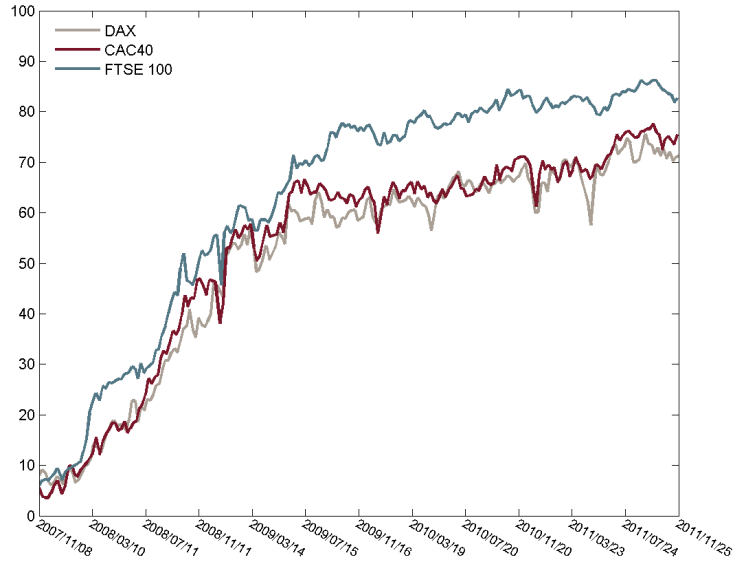


Figure 13: Fragmentation of European markets: the market share of the primary market decreases continuously since the entry in force of the MiFI Directive; the use of a SOR is mandatory to access to the liquidity of the whole market. This graph monitors the *normalized entropy* of the fragmentation: if the market shares (summing to one) over K exchanges are q_1, \dots, q_K , the indicator is $-\sum_k q_k \ln q_k$ renormalized so that its maximum is 100% (i.e. divided by $\ln(K)$).

Optimal policies for SOR have been proposed in [Pagès et al., 2012] and [Ganchev et al., 2010]. The later one used censored statistics to estimate liquidity available and build an optimization framework at the top of it; the former builds a stochastic algorithm and proves that it asymptotically converges to a state minimising a given criteria.

To give the sketch of the methodology associated to the stochastic algorithm viewpoint, just consider the following optimization problem.

Optimal liquidity seeking: the expected fast end criteria. To define the criteria to optimize, first assume that K visible order-books are available (for instance BATS, Chi-X, Euronext, Turquoise for an European stock). At time t , a buy order of size V_t has to be split over the order books according to a key (r_1, \dots, r_K) , given that on the k th order book:

- the resting quantity “cheaper” than a given price S is I_t^k (i.e. the quantity at the bid side posted at a price higher or equals to S , or at a lower price at the ask);
- the incoming flow of sell orders consuming the resting quantity at prices cheaper than S follows a Poisson process N_t^k of intensity λ^k , i.e.

$$\mathbb{E}(N_{t+\delta t}^k - N_t^k) = \delta t \cdot \lambda^k$$

- the waiting time on the k th trading destination to consume a volume v added on the k th trading destination at price S in t is noted $\Delta T_t^k(v)$; it is implicitly defined by:

$$\Delta T_t^k(v) = \arg \min_{\tau} \{N_{t+\tau}^k - N_t^k \geq I_t^k + v\}$$

Assuming that there is no specific toxicity in available trading platforms, a trader would like to split an incoming order at time t of size V_t according to an allocation key (r^1, \dots, r^K) to minimise a waiting time criteria:

$$\mathcal{C}(r^1, \dots, r^K) = \mathbb{E} \max_k \{\Delta T_t^k(r^k V_t)\} \quad (12)$$

It means that the trader targets to optimise the following process:

1. an order of size $V_{\tau(u)}$ is to be split at time $\tau(u)$, the set of order arrival times is $\mathcal{S} = \{\tau(1), \dots, \tau(n), \dots\}$;
2. it is split over the K available trading venues thanks to an “allocation key” $\mathbf{r} = (r^1, \dots, r^K)$: $r^k V_{\tau(u)}$ is sent to the k th order book (all the quantity is spread, i.e. $\sum_{k \leq K} r_k = 1$);
3. the trader waits the time needed to consume all the sent orders.

The criteria $\mathcal{C}(\mathbf{r})$ defined in (12) reflects that the faster the allocation key allows to obtain liquidity, the better: the obtained key is well suited for a *liquidity seeking algorithm*.

First note $k_t^*(\mathbf{r})$ the last trading destination to consume the order sent at t :

$$k_t^*(\mathbf{r}) = \arg \max_k \{\Delta T_t^k(r^k V_t)\}$$

a gradient approach to minimize $\mathcal{C}(\mathbf{r})$ will require to compute $\partial \Delta T_t^u(r^u V_t) / \partial r^k$ for any pair (k, u) . To respect the constraint, just replace an arbitrary r^ℓ by $1 - \sum_{u \neq \ell} r^u$, consequently:

$$\frac{\partial \Delta T_t^u(r^u V_t)}{\partial r^k} = \frac{\partial \Delta T_t^k(r^k)}{\partial r^k} \mathbb{1}_{k_t^*(\mathbf{r})=k} + \frac{\partial \Delta T_t^\ell(r^\ell)}{\partial r^k} \mathbb{1}_{k_t^*(\mathbf{r})=\ell}$$

with the notation $\Delta N_t^k = N_t^k - N_{t-}^k$, we can write that any allocation key \mathbf{r} such that, for any pair (ℓ, k) :

$$\mathbb{E}(V_t \cdot D_t^k(r^k) \cdot \mathbb{1}_{k_t^*(\mathbf{r})=k}) = \mathbb{E}(V_t \cdot D_t^\ell(r^\ell) \cdot \mathbb{1}_{k_t^*(\mathbf{r})=\ell}) \quad (13)$$

where

$$D_t^k(r^k) = \frac{1}{\Delta N_{t+\Delta T_t^k(r^k)}^k} \mathbb{1}_{(\Delta N_{t+\Delta T_t^k(r^k)}^k > 0)}$$

is a potential minimum for the criteria $\mathcal{C}(\mathbf{r})$ (the verification theorem will not be provided here). Equation (13) also writes:

$$\mathbb{E}(V_t \cdot D_t^k(r^k) \cdot \mathbb{1}_{k_t^*(\mathbf{r})=k}) = \frac{1}{K} \sum_{\ell=1}^K \mathbb{E}(V_t \cdot D_t^\ell(r^\ell) \cdot \mathbb{1}_{k_t^*(\mathbf{r})=\ell})$$

It can be shown (see [Lelong, 2011] for generic results of this kind) that the asymptotic solutions of the following stochastic algorithm on the allocation weights through time (provided strong enough ergodicity assumptions on the $(V, (N^k)_{1 \leq k \leq K}, (I^k)_{1 \leq k \leq K})$ multidimensional process):

$$\begin{aligned} \forall k, r^k(n+1) &= r^k(n) - \gamma_{k+1} \left(V_{\tau(n)} \cdot D_{\tau(n)}^k(r^k(n)) \cdot \mathbb{1}_{k_{\tau(n)}^*(\mathbf{r}(n))=k} - \right. \\ &\quad \left. \frac{1}{K} \sum_{\ell=1}^K V_{\tau(n)} \cdot D_{\tau(n)}^\ell(r^\ell(n)) \cdot \mathbb{1}_{k_{\tau(n)}^*(\mathbf{r}(n))=\ell} \right) \end{aligned} \quad (14)$$

minimizes the expected fast end criteria $\mathcal{C}(\mathbf{r})$.

Qualitatively, it can be read in this update rule that if a trading venue k demands more time to execute the fraction of the volume that it receives (taking into account the conjunction of I and N) than the average waiting time on all venues, it is needed to decrease the fraction r^k of the orders to send to k has to be decreased for future uses.

5 Perspectives and future works

The needs of intra-day trading practitioners are currently focused on optimal execution and trading risk control. For sure some improvements of what is actually available can be proposed by academics, mainly:

- provide optimal trading trajectories taking into account *multiple trading destinations* and different type of orders: liquidity providing (i.e. limit) ones and liquidity consuming (i.e. market) ones;
- the *analysis of trading performances* is also an important topic; models are needed to understand what part of the performance and risk are due to the planned scheduling, the interactions with order-books, the market impact and the market moves;
- *stress testing*: before putting a trading algorithm on real markets, it is needed to understand its exposure to different market conditions, from volatility or momentum to bid-ask spread or trading frequency. The study

of “Greeks” of the payoff of a trading algorithm is not straightforward since it is inside a closed loop of liquidity: its “psi” should be its derivative with respect to the bid-ask spread, its “phi” with respect to the trading frequency, and its “lambda” with respect to the liquidity available in the order-book.

For the special case of portfolio liquidation studied in this paper (using the payoff \tilde{C}_λ defined by equality (11)), these trading Greeks would be:

$$\Psi = \left(\frac{\partial \tilde{C}_\lambda}{\partial \psi_\ell} \right)_{1 \leq \ell \leq N}, \quad \Phi = \frac{\partial \tilde{C}_\lambda}{\partial N}, \quad \Lambda = \frac{\partial \tilde{C}_\lambda}{\partial \kappa}$$

The progresses in these three directions will provide a better understanding of the price formation process and the whole cycle of asset allocation and hedging, taking into account execution costs, closed-loops with the markets, and portfolio trajectories at any scales.

Acknowledgments. Most of the data and graphics used here come from the work of Cr dit Agricole Cheuvreux Quantitative Research group.

References

- [Aït-Sahalia and Jacod, 2007] Aït-Sahalia, Y. and Jacod, J. (2007). Volatility Estimators for Discretely Sampled Lévy Processes. *Annals of Statistics*, 35:355–392.
- [Alfonsi et al., 2010] Alfonsi, A., Fruth, A., and Schied, A. (2010). Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157.
- [Alfonsi and Schied, 2010] Alfonsi, A. and Schied, A. (2010). Optimal Execution and Absence of Price Manipulations in Limit Order Book Models. *SIAM J. Finan. Math.*, 1:490–522.
- [Almgren et al., 2005] Almgren, R., Thum, C., Hauptmann, E., and Li, H. (2005). Direct Estimation of Equity Market Impact. *Risk*, 18:57–62.
- [Almgren and Chriss, 2000] Almgren, R. F. and Chriss, N. (2000). Optimal execution of portfolio transactions. *Journal of Risk*, 3(2):5–39.
- [Altunata et al., 2010] Altunata, S., Rakhlin, D., and Waelbroeck, H. (2010). Adverse Selection vs. Opportunistic Savings in Dark Aggregators. *Journal of Trading*, 5:16–28.
- [Avellaneda and Stoikov, 2008] Avellaneda, M. and Stoikov, S. (2008). High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224.
- [Bacry et al., 2009] Bacry, E., Delattre, S., Hoffman, M., and Muzy, J. F. (2009). Deux modèles de bruit de microstructure et leur inférence statistique. presentation.
- [Biais et al., 2005] Biais, B., Glosten, L., and Spatt, C. (2005). Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications. *Journal of Financial Markets*, 2(8):217–264.
- [Bouchard et al., 2011] Bouchard, B., Dang, N.-M., and Lehalle, C.-A. (2011). Optimal control of trading algorithms: a general impulse control approach. *SIAM J. Financial Mathematics*, 2:404–438.
- [Bouchaud et al., 2002] Bouchaud, J. P., Mezard, M., and Potters, M. (2002). Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4).
- [Chakraborti et al., 2011] Chakraborti, A., Toke, I. M., Patriarca, M., and Abergel, F. (2011). Econophysics review: II. Agent-based models. *Quantitative Finance*, 11(7).
- [Cohen et al., 1981] Cohen, K. J., Maier, S. F., Schwartz, R. A., and Whitcomb, D. K. (1981). Transaction Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread. *The Journal of Political Economy*, 89(2):287–305.
- [Cont and De Larrard, 2011] Cont, R. and De Larrard, A. (2011). Price Dynamics in a Markovian Limit Order Book Market. *Social Science Research Network Working Paper Series*.

- [Cont et al., 2010] Cont, R., Kukanov, A., and Stoikov, S. (2010). The Price Impact of Order Book Events. *Social Science Research Network Working Paper Series*.
- [Engle et al., 2012] Engle, R. F., Ferstenberg, R., and Russell, J. R. (2012). Measuring and Modeling Execution Cost and Risk. *The Journal of Portfolio Management*, 38(2):14–28.
- [Foucault and Menkveld, 2008] Foucault, T. and Menkveld, A. J. (2008). Competition for Order Flow and Smart Order Routing Systems. *The Journal of Finance*, 63(1):119–158.
- [Gabaix et al., 2006] Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2006). Institutional Investors and Stock Market Volatility. *Quarterly Journal of Economics*, 121(2):461–504.
- [Ganchev et al., 2010] Ganchev, K., Nevmyvaka, Y., Kearns, M., and Vaughan, J. W. (2010). Censored exploration and the dark pool problem. *Commun. ACM*, 53(5):99–107.
- [Gatheral, 2010] Gatheral, J. (2010). No-Dynamic-Arbitrage and Market Impact. *Quantitative Finance*, 10(7).
- [Gatheral and Schied, 2012] Gatheral, J. and Schied, A. (2012). Dynamical models of market impact and algorithms for order execution. In Fouque, J.-P. and Langsam, J., editors, *Handbook on Systemic Risk (Forthcoming)*. Cambridge University Press.
- [Guéant et al., 2011] Guéant, O., Lehalle, C. A., and Fernandez-Tapia, J. (2011). Optimal Execution with Limit Orders. *Working paper*.
- [Hayashi and Yoshida, 2005] Hayashi, T. and Yoshida, N. (2005). On Covariance Estimation of Non-synchronously Observed Diffusion Processes. *Bernoulli*, 11(2):359–379.
- [Hewlett, 2006] Hewlett, P. (2006). Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes*. Ecole Polytechnique.
- [Ho and Stoll, 1981] Ho, T. and Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1):47–73.
- [Jacod, 1996] Jacod, J. (1996). *Hommage a P. A. Meyer et J. Neveu*, volume 236, chapter La variation quadratique moyenne du brownien en présence d’erreurs d’arrondi. Asterisque.
- [Kirilenko et al., 2010] Kirilenko, A. A., Kyle, A. P., Samadi, M., and Tuzun, T. (2010). The Flash Crash: The Impact of High Frequency Trading on an Electronic Market. *Social Science Research Network Working Paper Series*.
- [Kyle, 1985] Kyle, A. P. (1985). Continuous Auctions and Insider Trading. *Econometrica*, 53(6):1315–1335.

- [Large, 2007] Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25.
- [Lasry and Lions, 2007] Lasry, J.-M. and Lions, P.-L. (2007). Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260.
- [Lehalle, 2009] Lehalle, C.-A. (2009). Rigorous Strategic Trading: Balanced Portfolio and Mean-Reversion. *The Journal of Trading*, 4(3):40–46.
- [Lehalle et al., 2010] Lehalle, C.-A., Guéant, O., and Razafinimanana, J. (2010). High Frequency Simulations of an Order Book: a Two-Scales Approach. In Abergel, F., Chakrabarti, B. K., Chakraborti, A., and Mitra, M., editors, *Econophysics of Order-Driven Markets*, New Economic Windows. Springer.
- [Lelong, 2011] Lelong, J. (2011). Asymptotic normality of randomly truncated stochastic algorithms. *ESAIM: Probability and Statistics (forthcoming)*.
- [Lillo et al., 2003] Lillo, F., Farmer, J. D., and Mantegna, R. (2003). Econophysics - Master Curve for Price - Impact Function. *Nature*, (421).
- [Markowitz, 1952] Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1):77–91.
- [Menkveld, 2010] Menkveld, A. J. (2010). High Frequency Trading and The New-Market Makers. *Social Science Research Network Working Paper Series*.
- [Muniesa, 2003] Muniesa, F. (2003). *Des marchés comme algorithmes: sociologie de la cotation électronique à la Bourse de Paris*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris.
- [Pagès et al., 2012] Pagès, G., Laruelle, S., and Lehalle, C.-A. (2012). Optimal split of orders across liquidity pools: a stochastic algorithm approach. *SIAM Journal on Financial Mathematics (Forthcoming)*.
- [Predoiu et al., 2011] Predoiu, S., Shaikhet, G., and Shreve, S. (2011). Optimal Execution of a General One-Sided Limit-Order Book. *SIAM Journal on Financial Mathematics*, 2:183–212.
- [Robert and Rosenbaum, 2011] Robert, C. Y. and Rosenbaum, M. (2011). A New Approach for the Dynamics of Ultra-High-Frequency Data: The Model with Uncertainty Zones. *Journal of Financial Econometrics*, 9(2):344–366.
- [Shiryaev, 1999] Shiryaev, A. N. (1999). *Essentials of Stochastic Finance: Facts, Models, Theory*. World Scientific Publishing Company, 1st edition.
- [Smith et al., 2003] Smith, E., Farmer, D. J., Gillemot, L., and Krishnamurthy, S. (2003). Statistical Theory of the Continuous Double Auction. *Quantitative Finance*, 3(6):481–514.
- [Wyart et al., 2008] Wyart, M., Bouchaud, J.-P., Kockelkoren, J., Potters, M., and Vettorazzo, M. (2008). Relation between Bid-Ask Spread, Impact and Volatility in Double Auction Markets. Technical Report 1.

[Zhang et al., 2005] Zhang, L., Mykland, P. A., and Sahalia, Y. A. (2005). A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data. *Journal of the American Statistical Association*, 100(472).