Efficiency of the Price Formation Process in Presence of High Frequency Participants: a Mean Field Game analysis

Aimé Lachapelle[1]   Jean-Michel Lasry[2]   Charles-Albert Lehalle[3]   Pierre-Louis Lions[4]

Abstract

This paper deals with a stochastic order-driven market model with waiting costs, for order books with heterogenous traders. Offer and demand of liquidity drives price formation and traders anticipate future evolutions of the order book. The natural framework we use is mean field game theory, a class of stochastic differential games with a continuum of anonymous players. Several sources of heterogeneity are considered including the mean size of orders. Thus we are able to consider the coexistence of Institutional Investors and High Frequency Traders (HFT). We provide both analytical solutions and numerical experiments. Implications on classical quantities are explored: order book size, prices, and effective bid/ask spread. According to the model, in markets with Institutional Investors only we show the existence of inefficient liquidity imbalances in equilibrium, with two symmetrical situations corresponding to what we call *liquidity calls for liquidity*. During these situations the transaction price significantly moves away from the fair price. However this macro phenomenon disappears in markets with both Institutional Investors and HFT, although a more precise study shows that the benefits of the new situation go to HFT only, leaving Institutional Investors even with higher trading costs.

Keywords: order book modeling, mean field games, order-driven market, waiting cost, liquidity equilibrium, high frequency trading

1 Introduction

With the recent changes in regulation on financial markets (MiFID, 2007, in Europe and Reg NMS, 2005, USA) the competition across trading venues favored the appearance of new trading rules, in a global attempt to capture most of the decreasing liquidity available in the post-2008 financial crisis world. Trading venues thus proposed innovative ways to trade in electronic order books (that have the favor of regulators and policy makers because of their native traceability):

- tiny tick sizes (i.e. the minimum price change between two consecutive quotes [Hall et al., 2005]) to attract automated market orders using SORs (Smart Order Routers [Pagès et al., 2012], [Foucault and Menkveld, 2008]),

- low latency networks and matching engines, to allow high frequency players to decrease their exposure to market risk, in attempts to give them incentive to provide more liquidity [Madhavan, 2011],

- maker/taker fee schedules to pay Liquidity Provider orders inserted in order books, in order to attract liquidity,

- creation of Dark Pools of various kinds (see [Ganchev et al., 2010]), to promote anonymous liquidity seeking so that large investors can continue to exchange blocks in an electronic manner,

- size-priority and pro-rata matching rules [Mendelson and Amihud, 1991] to complement the usual time-priority models,

are among these changes in market microstructure.

The analysis of the efficiency of the emerging ecology of partially connected trading pools is questioned, especially since the flash crash [Madhavan, 2011], [Kirilenko et al., 2010] during which the US equity market

---
[1]MFG Labs
[2]CEREMADE and Crédit Agricole Corporate and Investment Bank
[3]Capital Fund Management (corresponding author, charles-albert.lehalle@cfm.fr)
[4]CEREMADE and Collège de France

has lost around 10% of its value in 10 minutes, regaining it in 20 minutes. The resiliency of the liquidity provided by HFT (High Frequency Traders [Menkveld, 2010]) raised concerns. So did the spread of liquidity on such an heterogeneous network of pools.

Addressing these points is difficult because the market microstructure is not only a set of trading rules that could be studied statically, it changes with market participants behaviors, each of them trying to optimize her own utility function and anticipating others' moves.

This article provides an order-driven market modeling, where the volume of arriving flows is the risk source and where the key driver is the demand/offer of liquidity.

Yet only a limited number of papers have explored such models, the most notables being [Roşu, 2009] (modeling the order book queue dynamics) and [Gareche et al., 2013] (empirically studying order book data to extract the main components of the dynamics). This paper can be seen as a very good complement to those two very interesting ones: with an accurate economic modeling on the one hand, and empirical results on the other hand. The MFG approach links them together since we provide for instance Partial Derivative Equation formulations (compatible with the Fokker-Planck equation described in the second paper) arising from a structural modeling (compatible with the modeling of the first one). The problem is very complex, and the topic deserves for more studies and publications. Indeed, such liquidity models involve a very large number of traders who arrive and leave the system at different times and strategically interact. Such components lead necessarily to complex situations.

In our model we consider smart traders (we call them *players* as soon as we use the game theory environment) that arbitrate between limit and market orders. That is they have to choose between the immediate transaction price and expected later transaction prices. The dynamic model is in continuous time, in infinite horizon. Since patience is at the heart of our model, the present approach belongs to the family of waiting cost order-driven market models. Closely related papers are the work of [Foucault et al., 2005], and the more recent paper by [Roşu, 2009]. The former is the seminal waiting cost based model (as opposed to asymmetry information models) in discrete time. The latter is a continuous time approach where traders have the possibility to cancel their orders for free. This late assumption greatly simplifies the problem and allows the author to describe the equilibrium in an elegant manner.

Our paper is most closely related to [Roşu, 2009] since it is also a continuous time model with Poisson processes used to model newcomers' arrivals. On the other hand, our model present some important dissimilarities. First the patience structure of traders is more endogenous since no cancellation of orders are permitted. Choices made by the players are thus irreversible and traders' anticipations of future events become a core issue. Their is a deep impact on the equilibrium equations: the problem becomes nonlinear. Secondly, the goal of our paper is to study the case with heterogenous traders, in particular to model the interactions of Institutional Investors and High Frequency Traders. In [Roşu, 2009] several types of traders are considered, but the strategical arbitrage between market and impact orders is allowed only for one of the types. This is not the case in our work, where all types make choices.

Game theory is necessary as soon as markets are incomplete. When markets are complete, strategy is unnecessary and the only task agents have to perform is to optimize in regards to the price. Order-driven markets are by essence incomplete since the source of risk is the random arrival of traders, and it is impossible to hedge this risk because choices made by traders are irreversible (note that the more realistic modeling where the modeler considers costly cancellation of orders must lead to a similar incompleteness).

Consequently we are convinced that game theory offers a proper framework.

Mean Field Games (MFG monotone systems, as detailed in the next section) are the suitable class of games that naturally allow to take into account the specific components of the order-driven market we consider, that is: a continuum of anonymous players, irreversibility of the actions, recursivity (anticipation of future prices).

The resulting dynamics is thus a mix of backward-driven behaviors (based on actualized anticipations of future values of trades) and forward-driven ones (resulting from the immediate actions taken by agents).

The MFG framework has been built to capture this two way dynamics, therefore this paper uses it to render the dynamics of a stylized order book, allowing to obtain results on different market configurations.

The paper is organized as follows. In section 2 we provide a quick introduction to Mean Field Game theory. In section 3 we introduce the modeling approach. We start with a one-sided order book as a base camp towards the two-sided order book exposed later. Section 4 and 5 are dedicated to the introduction and theoretical study of the recursive equations characterizing the equilibrium. Finally we conclude the paper in section 6, where we apply the model to several markets: markets with Institutional Investors only *versus* markets with both Institutional Investors and High Frequency Traders.

2   Mean Field Games: a quick introduction

Mean Field Games (MFG for short) are a class of stochastic differential games with a continuum of agents. They have been introduced by [Lasry and Lions, 2007]. Similar ideas have been introduced from an engineering viewpoint by [Huang et al., 2007] and [Adlakha et al., 2013]. From then on, MFG have known

numerous developments and applications to various fields, mainly in economics [Lucas and Moll, 2013, Guéant et al., 2010], statistics [Pequito et al., 2011a, Pequito et al., 2011b], and human crowd behaviors [Lachapelle and Wolfram, 2011]. The mathematics and numerics of MFG have been widely developed. Most of the mathematical tools for MFG have been the purpose of a 5 years course at Collège de France [Lions, 2012], and recent developments are described in [Cardaliaguet et al., 2012] from an analysis viewpoint and [Carmona et al., 2012] with a probabilistic approach.

In the continuum, agents are atomized, which means that their influence on the global state is reduced to nil. In economics, this aspect has to be linked with the notion of *price taker* agents as opposed to the case of a *price maker* monopolist for instance. This is precisely the whole continuum that makes the equilibrium. The nil influence may have other sources than the presence of infinitely many players in the game. Indeed, a game with stochastic continuous entries and exits of players leads to the same property. This will be in particular the case of the model we propose in the present paper, where we consider Poisson entries and exits.

The complete information consists of a measure on the space of states S. Being a measure, it is often denoted by m in the literature, but to be consistent with the notations of the model, we rather call it x. Then $x(s)$ quantifies the density of agents having state s.
In a MFG, players individually optimize (by choosing actions) their expected pay-off, considering the evolution of the global dynamic of the collectivity as an observable parameter (and they *anticipate* its evolution). Simultaneously, the statistical evolution of the collective dynamic follows from the individual optimal behaviors. The equilibrium takes place as soon as the anticipated evolution coincides with the statistical evolution.

A core characteristic of Mean Field Games is that they are anonymous games. This notion is well known and means that the game is invariant for any permutation of the players. In other words, the players are not labelled. This assumption is very natural in complex systems involving numerous players.
Mean Field Games are approximations of anonymous games with finitely many players. But things are getting much simpler in MFG. The strategical powerlessness of individuals (i.e. the atomized characteristic of players) dramatically shrinks the traditional complexity (materialized by numerous coupling of the equilibrium equations) of N-player games, which is well-known as being Achilles' heel of classical stochastic differential games. Players interact with others only *via* the global state of the collectivity.

In N-player stochastic differential games, each player optimizes her value function u_i, depending upon every individual states of agents (including herself). The equilibrium is then characterized by a complex system of coupled differential equations.
In a Mean Field Game, the N value functions become a single value function $\mathcal{U}$ depending upon the the state s of a generic player and the density x of the continuum.
The MFG equilibrium is then characterized by a *master equation* verified by $\mathcal{U}$. The master equation is in general very tricky and mathematically challenging. Their is a natural classification of cases in term of risk structure.

- Individual risk: in this case, the stochasticity of each player's dynamic is independent of each other. This particular case was firstly introduced. A major simplification is that the value function does not depend on the density x, but only on the state s. Consequently the master equation reduces to a system of two coupled partial differential equations having a forward-backward structure. The dynamic of the collectivity is deterministic.

- Shared risk: here the only risk that agents face is common to all of them. When agent's space S is finite (that is $x := (x_1, ..., x_M)$), then the value function can be discretized

$$\mathcal{U} := (u_j), \ u_j(x_1, \ldots, x_M), \ j = 1, \ldots, M.$$

This class of cases have been deeply investigated and is referred to as the case of *monotone systems* (see [Lions, 2012]). The monotone system takes the following form:

$$(1) \qquad 0 = -ru_j - \sum_{k=1}^{N} \alpha_k(u, x) \frac{\partial u_j}{\partial x_k} + \beta_j(u, x), \ \text{ for } j = 1, ..., M,$$

where

$$u \to \alpha_j(u, x) \text{ is monotone for all } j$$
$$u \to \beta_j(u, x) \text{ is monotone for all } j.$$

We will see later that our order book model falls into this class of MFG.
Note that there is also a time dependent version of (1) with a time derivative term added.

- Mix models: some classes of cases that mix both shared and individual risks are needed for economic modeling (e.g. for solving the Krussel-Smith problem).

3 Model

The stylized order book used here is a two-sided one. We start with a simple single-queue model as a base camp towards the two-sided one that is exposed later.

3.1 A simple single-queue model with anticipations

The purpose of introducing first a single queue model is didactic and does not aim at directly providing insights on order book modeling. However we believe this single queue is the occasion to introduce some key concepts, such as endogenous strategic entries of agents that anticipate the future. Consequently, sellers entering the system are also called players since we locate the modeling approach in the game theoretic framework (agents perform actions optimizing their respective pay-off).
In particular, when new sellers arrive, they look at the queue size and decide whether to enter the queue or not (action), after considering their expected pay-off (value function assessment).

The simplified model is the occasion to introduce anticipation behaviors in a very stylized one-sided order book, where patient sellers arrive at exogenous Poisson rate and where the arrival rate of impatient buyers increases as soon as the queue size increases. We will finally use it to provide insights on the modeling of distinct execution protocols, namely process sharing and First In First Out protocol.

The model. The arrival rate of players is continuous and stochastic. In this simplified model, it is exogenous. As usual, they arrive following a Poisson process with intensity λ.
Impatient buyers arrive at rate $\mu(x) \geq 0$, a given increasing function of x.
The unit size of an order in the queue is q. The queuing discipline is a process sharing one (with no priority), i.e. individual service in a queue of size x is worth q/x.
The pay-off gained by a player per unit of order is a nonnegative decreasing function of the queue size: $P(x)$. Typical cases are $P(x) := p > 0$ and $P(x) = 1/x$. On the other hand, there is a cost c of waiting in the queue.
Now, as usual in game theory, there is a value function u for any player. The value function depends upon the queue size x. It is the expected Profit & Loss (P&L) of a player entering the queue. Note that we assume that agents are risk neutral and that their reservation utility is set to 0, which means that an agent decides to enter the queue as soon as the value function is positive: $u(x) > 0$.

The value function dynamic comes from an infinitesimal expression of events impacting it:

- a newcomer enters the queue as soon as $u(x) > 0$ (remember u is the "*expected value to pay if you enter the queue*").

- the queue is consumed by a Poisson process of intensity $\mu(x)$. We consider an order already waiting in the queue, it will be partially executed (according to a prorata rule): its owner will sell q/x shares for a price $P(x)$. The new expected value for a participant waiting in the queue in this case is thus $q/x \cdot P(x) + (1 - q/x) \cdot u(x - q)$ (i.e. the first part of the expression comes from the sell of q/x shares and the second one from the expected value of the queue that is now of size $x - q$).

- in all other cases, the expected value does not change.

- the waiting cost is proportional to q (the size of the orders); it decreases the expected value of u by $c\,q\,dt$.

Equilibrium : the value function equation. In this paragraph we introduce the equation verified by the value function at the equilibrium. Below we detail the equilibrium equation for each probability event.

$$
\begin{aligned}
(2) \qquad u(x) =\ & (1 - \lambda\mathbf{1}_{u(x)>0}dt - \mu(x)dt) \cdot u(x) && \leftarrow \text{ nothing happens} \\
& + \lambda\mathbf{1}_{\{u(x)>0\}}dt \cdot u(x+q) && \leftarrow \text{ new queue entrance} \\
& + \mu(x)dt \cdot \left(\frac{q}{x}P(x) + (1 - \frac{q}{x})u(x-q)\right) && \leftarrow \text{ service} \\
& - cqdt && \leftarrow \text{ waiting cost}
\end{aligned}
$$

We can perform a Taylor expansion for small q in the discrete equation above. In this way we derive the following differential equation:

$$
0 = \frac{\mu(x)}{x}(P(x) - u) - c + (\lambda\mathbf{1}_{\{u>0\}} - \mu(x))u' + q\left(\frac{1}{2}(\lambda\mathbf{1}_{\{u>0\}} - \mu(x))u'' + \frac{\mu(x)}{x}u'\right),
$$

where the second order term is the last one (blue term).

First order analysis. Before approximating numerically the solution to (2), we propose to get some insights on the shape of the solution by doing a first order analysis. More precisely, the solution to the queuing system described above is characterized by the sign of the value function u. Consequently we are interested in finding potential sign switching points of u.

The core modeling ingredient is the value of the Poisson arrival rate λ relative to $\mu(x)$.

For the first order analysis we look at the first order equation:

$$(3) \qquad 0 = \frac{\mu(x)}{x}\Big(P(x) - u(x)\Big) - c + \Big(\lambda \mathbf{1}_{\{u(x)>0\}} - \mu(x)\Big)u'(x).$$

Remark 3.1. *Let us remark that equation (3) corresponds to a trivial shared risk Mean Field Game monotone system with $N = 1$, as described in the previous section. Note that in the framework of this model, the mean field aspect does not come from the continuum of agents (for every instant, the number of players is finite), but rather to the stochastic continuous structure of entries and exits of players.*

Now we look at the case where the stylized limit order book presented here has an *infinite resiliency*, meaning once the orderbook is partially consumed by a marketable order, the remaining liquidity rearranges itself to fill the gap. Moreover, we will consider the non degenerated case where sellers arrive at rate λ, larger than the exogenous consuming rate $\mu(x)$, for all x. Then the term multiplying the derivative in (3) changes sign as u does. It is a simple matter to see that this happens for all points x^* such that:

$$(4) \qquad x^* = \mu(x^*)P(x^*)/c.$$

In this specific case these points fully characterize the solution.

Equation (4) has a (unique) solution under usual fixed point theorem assumptions, such as continuity and contraction mapping assumptions. Remark that if P and μ are constant parameters then there is a single point verifying (4). Also, we notice that these points are sensitive to the granularity ratio μ/c.

An example with anticipation behavior. Assume the arrival rate of buyers has the specificity to take two values:

- a low value μ_1 below a certain queue size threshold S,
- a higher value μ_2 ($\mu_2 > \mu_1$), above the threshold S.

As a function depending upon the queue size variable x, it reads:

$$\mu(x) = \mu_1 \mathbf{1}_{x<S} + \mu_2 \mathbf{1}_{x\geq S}, \ 0 \leq \mu_1 < \mu_2.$$

Here there are at least two points where u changes sign:

$$(5) \qquad x_1^* = \mu_1 P(x_1^*)/c \text{ and } x_2^* = \mu_2 P(x_2^*)/c.$$

Figure 1 shows the plot of the solution (numerical approximation of the solution to equation (2)) for a certain set of parameters (for P constant). We can see that the first switching point is close to the first order approximation x_1^*, while the last sign switch significantly deviates from the first order approximation x_2^*. It means that higher order terms have a non-negligible effect.

But most importantly, we observe that there is another sign switch strictly below the threshold S. The existence of such a switching point means that players anticipate the improved service before the threshold is reached. Indeed, their value function becomes positive meaning that players enter the queue strictly before the improved service starts. This is why we talk about an *anticipation switching point*. Consequently, we can conclude that at the equilibrium, the strategical players adopt an anticipation behavior.

First In First Out model. Finally we want to show that our approach allows to model distinct execution processes, and how the resulting equilibrium equations are impacted.

To do so, we consider the First In First Out (FIFO) protocol. This is the only change we make in the model. To consider such a priority protocol, we have to introduce a new variable z denoting the position of a trader in the queue of size x. Consequently the problem becomes bi-dimensional.

The equation becomes:

$$(6) \qquad \begin{aligned} u(z,x) = \ & (1 - \lambda \mathbf{1}_{u(x,x)>0}dt - \mu(x)dt) \cdot u(z,x) && \leftarrow \text{ nothing happens} \\ & + \lambda \mathbf{1}_{\{u(x,x)>0\}}dt \cdot u(z, x+q) && \leftarrow \text{ new queue entrance} \\ & + \mu(x)dt \cdot u(z-q, x-q) && \leftarrow \text{ execution of the first order} \\ & - cqdt && \leftarrow \text{ waiting cost,} \end{aligned}$$

in the domain $q < z < x$, and the boundary condition for $z = q$ is:

$$(7) \qquad \begin{aligned} u(q,x) = \ & (1 - \lambda \mathbf{1}_{u(x,x)>0}dt - \mu(x)dt) \cdot u(q,x) && \leftarrow \text{ nothing happens} \\ & + \lambda \mathbf{1}_{\{u(x,x)>0\}}dt \cdot u(q, x+q) && \leftarrow \text{ new queue entrance} \\ & + \mu(x)dt \cdot P(x) && \leftarrow \text{ execution of the first order} \\ & - cqdt && \leftarrow \text{ waiting cost.} \end{aligned}$$
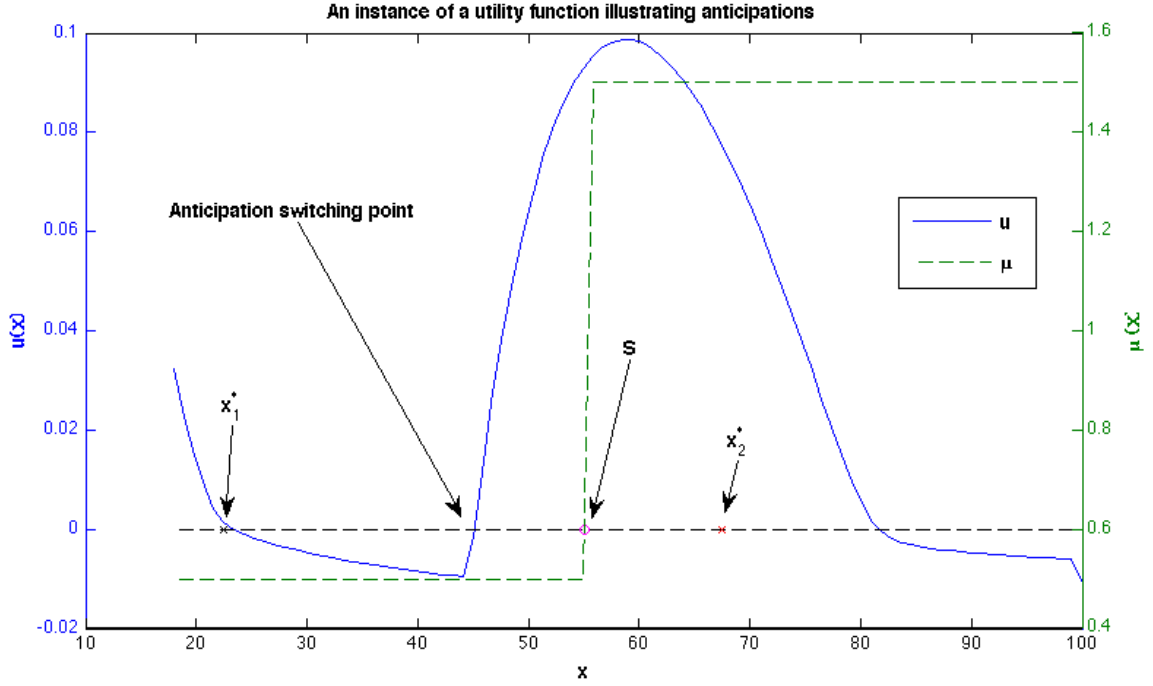
System (6-7) can be easily solved numerically.

Figure 1: Here we notice that there is a point strictly before S where u switches from negative to positive. It means that players anticipate improved service μ_2 and therefore are newly interested in entering the queue.

3.2 The order book model

The matching mechanisms of order books. One of the roles of financial markets is to *form prices* according to the balance between offer and demand. In modern markets, this mechanism takes place inside *electronic order books*. They implement the following dynamic:

1. buyers and sellers can send electronic messages to a *"matching engine"*. These messages, called *orders*, contain a side ("buy" or "sell"), a limit price and a quantity.

2. The matching engine contains a list of all pending orders it received in its memory. When it receives a new buy (respectively sell) order, it looks if pending sell (resp. buy) orders at a lower (resp. higher) price are available.

   - If it is the case, it generates transactions between the owner of the incoming order and the owners of the compatible opposite orders, and removes the corresponding quantities in its list of pending orders;
   - if the incoming order has a remaining quantity, it is inserted in the list of pending orders.

The list of pending orders waiting in the matching engine is called its *"order book"*.

During the matching process, it is possible that the quantity of an incoming order does not match exactly the quantity made available at a compatible price (i.e. lower prices for a buy order and higher prices for a sell order) by opposite orders in the order book. To handle such cases, matching engines need to implement a priority mechanism. The most used (see [Mendelson and Amihud, 1991] for more details) are:

   - *time priority*: the "oldest" pending orders in the order book are matched first;
   - *size priority*: the largest pending orders are matched first in case of competition between resting orders at the same price;
   - *pro rata*: pending orders are matched for a fraction of their quantity linear in their relative size.

Each trading platform discloses its matching mechanism in detail to market participants in a *rulebook* (like [Euronext, 2006]).

Matching dynamics and trading styles. Market participants thus have to cope with rules of the matching engine they trade into while fulfilling their day-to-day goals. Recent regulatory discussions raised questions on the potential negative interactions between the following classes of market participants in the same order book:

   - *Institutional investors*, that buy and sell large quantities of shares to manage their portfolios on the long term. They take the decision to buy or sell independently from the immediate state of the order

book. They are in essence *impatient* since they interact with other participants in the order book with the final goal to really buy or sell given quantities before a given deadline. They will not change their mind during the trading process given the state of the liquidity in the order book.

- *High Frequency Traders* are far more opportunistic. Even if they do not have all the same behavior (see [Brogaard et al., 2012] for more details), they have in common the fact that: (1) they send very small orders to trading platforms, (2) they do it very often (i.e. at *high frequency*), (3) they have no other reason to trade than the immediate state of the order book.

Concerns raised focused on the integrity of the price dynamics when so different participants are mixed in order books. The "Flash Crash" [Kirilenko et al., 2010] has shown that liquidity glitches could cause large variations of prices formed in electronic order books with no fundamental reasons. Academics studying the price formation process in order books usually name *"temporary market impact"* the way prices temporally deviate from their stable value due to high consumption of liquidity (i.e. of pending orders) in an order book (see [Almgren et al., 2005], [Gatheral, 2010]).

Recent regulatory changes unexpectedly favored HFTs activity [Burgot et al., 2013] (they are said to now be part of 70% of transactions in the US, 40% in Europe and 30% in some Asian markets, like Japan).

Dedicating a model to study liquidity games in order books. The way market participants interact in order books is sophisticated, due to the fact that they continuously try to anticipate actions of other participants to take an adequate decision. Their classical dilemma is the following: on the one hand they want to trade as slow as possible to avoid to be detected nor consume liquidity too fast thus moving the price an unfavorable way (i.e. adverse selection costs). On the other hand they cannot afford to trade too slow to avoid to be exposed to adverse market moves (i.e. opportunity costs).

A large literature proposes mathematical frameworks for market participant to optimize their trading kinematics: first mean-variances approaches [Almgren and Chriss, 2000], then stochastic control ones [Bouchard et al., 2011], and more recently stochastic algorithms have been designed to capture optimally liquidity at the smallest time scale [Pagès et al., 2012]. In all these approaches, each market participant tries to optimize her behavior assuming that the aggregation of other players is "martingale" in the sense that it is submitted to price moves and to some order books characteristics (like the volatility, the market depth, the intensity of orders reaching the matching engine, etc.) emerging from the activity of other participants without influencing it (in most cases a *market impact function* is introduced, exogenously from the activity of other participants).

The MFG approach presented here takes into account the way strategies of market participants change the dynamics of the order book. It opens the door to more endogenous models. The previous section is a simple illustration of this approach: the mean field is the state of the one-sided order book, and since each player implements an optimal strategy (in the sense that she values the time to wait in the queue and compares it to an immediate price to pay), it is possible to understand the dynamics of the value function u shared by all market participants.

In this section we will go one step further: the consuming rate $\mu(x)$ of the one-sided order book (say it is the queue of sellers) of section 3.1 will be linked to the size of the queue x, but in an endogenous way: that is via optimal strategies followed by participants in the other queue (the one of buyers). The flow consuming the selling queue is the one of buyers deciding on their side to pay immediately instead of waiting in the queue (of buyers). It will enable the emergence of coupled dynamics taking into account the states of the two queues.

To render a market impact effect, we will model the way impatient buy or sell orders consume the queue of sellers or buyers. For the ease of presentation, in this paper we will consider that at our time scale the "fair price" (that can be understood as a *latent price* like in [Robert and Rosenbaum, 2011] or [Bacry et al., 2012], or as a *fundamental price* like in [Ho and Stoll, 1983]) does not change significantly. But the reader can note that extending this model pegging a diffusive behavior on this fair price will do no more than adding an Ito term to the considered dynamics. Mean Field-inspired models at a largest time scale, targeting the understanding of the latent price dynamics have been already proposed, but not at the level of the order books (for instance in [Lehalle et al., 2010], the dynamics of a latent order book is submitted to an MFG like mechanism, but the realizations of the order book is modeled via a forward only scheme).

Hence, we introduce here a market-impact like relation between the size of the order book queues and the transaction prices around the *fair price* P: consuming a quantity q of the queue of pending selling orders of size Q^a will temporally move the price from P to $P + \delta \cdot q/(Q^a - q)$. Qualitatively, it implies an almost linear market impact with elasticity δ (i.e. δ can be compared to Kyle's lambda [Kyle, 1985]). Moreover, the modelled orderbook will have an *infinite resiliency*: once liquidity is consumed in a queue, the remaining quantity will reshape itself to fill the created gap.

The details of the MFG model are exposed in the following sections. In short, it contains these following ingredients:

- market participants are able to act strategically, anticipating others' moves;
- the dynamics of the two queues (patient buyers and patient sellers) are coupled thanks to the fact that the flow consuming each of them is provided by agents of the other side choosing to be impatient (either because they do not use a smart routing strategy, or because the outcome of their smart strategy is to send a market order);

7

- market impact is introduced dynamically (related to the size of the queues), modifying the premium to be paid by impatient traders, thus influencing their choices.

Moreover, our order book model needs a priority rule, for simplicity reasons we will use a pro-rata rule (since it keeps the dimensionality of the model tractable). As it will be seen later, it allows to render enough complexity to obtain meaningful results.

Rendering different trading styles in an order book model. To understand the features of our MFG model, we will first study its dynamics when only one class of market participants interacts in one order book. Since we are in a MFG framework, it will render a continuum of agents, at this stage they share the same macroscopic parameters:

- the same messaging intensity λ,
- the same size of orders they send q,
- the same waiting cost c.

Beside, we enrich the model with one more feature: the use of SOR (Smart Order Router). A Smart Order Router (see [Foucault and Menkveld, 2008] for an efficiency study or [Burgot et al., 2013] for a generic presentation) is a device containing a software dedicated to "smartly route" orders. In our model, only SOR users will be able to act strategically instead of being blindly impatient.

It can be considered that agents not using a Smart Order Router (SOR) have an infinite waiting cost. Since institutional investors take decisions independently of the current state of the orderbook, it is reallistic to consider that a fraction of them will not take time to implement sophisticated microscopic strategies on some of their orders.

The proportion of market participants using a SOR will be parametrized thanks to a specific flow of intensity λ^-.

	Instit. Investors	HFT
Order size	large	small
Speed	normal	fast
SOR	often used	always used

Table 1: Qualitative modeling of Institutional Investors and HFT.

In a second stage we will mix two different classes of agents (Table 1):

1. *Institutional investors*, trading large quantities not using systematically a SOR;

2. *HFT (High Frequency Traders)*, faster than the former participants, using smaller orders, more patient (in the sense that they bare a lower cost per share waiting in a queue), and all of them using a SOR.

Transaction price. The *market price* will be centered on a constant P. The *market depth* is δ, meaning that no transaction will take place at a price lower than $P - \delta$ or higher than $P + \delta$. The (time varying) size of the bid queue (waiting buy orders) is Q_t^b and the size of the ask one (waiting sell orders) is Q_t^a.

When a market (buying) order hits the ask queue, the transaction price is p^{buy} and when the bid queue is lifted by a market (selling) order, the transaction price is p^{sell}. The price takes into account instantaneous queue size adjustments depending upon the order size q.

$$(8) \qquad p_q^{\text{buy}}(Q_t^a) := P + \frac{\delta q}{Q_t^a - q}, \quad p_q^{\text{sell}}(Q_t^b) := P - \frac{\delta q}{Q_t^b - q}$$

Qualitatively, it means that the market impact is linear. Boundary conditions, to be introduced later, impose $Q_t^a, Q_t^b > q$, so that there is no definition problem of the transaction prices

Value functions. The value function for a trader submitting a buy order in the bid queue is $v(Q_t^a, Q_t^b)$ and the one of a sell order in the ask queue is $u(Q_t^a, Q_t^b)$. In the model agents have risk-neutral preferences, thus the utility functions coincide with price expectations.

Orders arrival rates. We distinguish between SOR and non-SOR orders. The proportion of these two types of orders is exogenous, and set as an input of the model.

Buy and sell SOR orders arrive according to two Poisson processes with intensity λ_{buy} and λ_{sell}. Several cases can be considered:

1. Homogeneous Poisson processes:

$$(9) \qquad \lambda_{\text{buy}} = \lambda_{\text{sell}} = \lambda,$$

2. Heterogeneous (in space) Poisson processes

$$\lambda_{\text{buy}} = \lambda f(Q_t^b), \ \lambda_{\text{sell}} = \lambda f(Q_t^a),$$

where $f(x)$ is a decreasing function. Typical instances are $f(x) = 1/x$, $f(x) = \mathbf{1}_{x \leq \bar{Q}}$ likewise.

However, we will focus in this paper on the homogenous case.
Let us remark that the previous rates could be endogenized and set as the result of an optimization problem involving the utility functions, consequently depending upon the queue sizes $Q_t^\bullet$.

Non-SOR orders (i.e. belonging to very impatient investors or traders) are always liquidity remover, with arriving rate $2\lambda^-$ (equally distributed between buyers and sellers).

Market participants decision processes. When a buy (resp. sell) order arrives, its owner has to make a *routing decision*:

- if $v(Q_t^a, Q_t^b + q) < p^{\text{buy}}(Q_t^a)$ (resp. $u(Q_t^a + q, Q_t^b) > p^{\text{sell}}(Q_t^b)$) it is more valuable to route the order to the bid (resp. ask) queue (i.e. sending a limit order). In such a case the order will be a *Liquidity Provider (LP)*. We define symmetrically *Liquidity Consumer (LC)* orders. This decision is formalized in the model by setting the variable $R_{\text{buy}}^{\oplus}(v, Q_t^a, Q_t^b + q)$ to 1 when $v(Q_t^a, Q_t^b + q) < p^{\text{buy}}(Q_t^a)$, and to zero otherwise:

(10)
$$R_{\text{buy}}^{\oplus}(v, Q_t^a, Q_t^b + q) := \mathbf{1}_{v(Q_t^a, Q_t^b + q) < p^{\text{buy}}(Q_t^a)}, \ \text{LP buy order}$$
$$R_{\text{sell}}^{\oplus}(u, Q_t^a + q, Q_t^b) := \mathbf{1}_{u(Q_t^a + q, Q_t^b) > p^{\text{sell}}(Q_t^b)}, \ \text{LP sell order}.$$

- otherwise the order goes Liquidity Consumerly to the ask (resp. bid) queue to obtain a trade. It will be a liquidity remover in this case:

$$R_{\text{buy}}^{\ominus}(Q_t^a, Q_t^b) := 1 - R_{\text{buy}}^{\oplus}(Q_t^a, Q_t^b), \ \text{LC buy order}$$
$$R_{\text{sell}}^{\ominus}(Q_t^a, Q_t^b) := 1 - R_{\text{sell}}^{\oplus}(Q_t^a, Q_t^b), \ \text{LC sell order}.$$

The price of such a transaction is p^{buy} (resp. p^{sell}) as defined by equality (8). Note that we omit the dependence on u, v when it is unnecessary for the understanding of the equations.

We impose the following boundary conditions:

(11)
$$\text{Min liquidity condition} : R_{\text{buy}}^{\oplus}(r, Q_t^b) = 1, \ R_{\text{sell}}^{\oplus}(Q_t^a, r) = 1, \ \forall r \leq q,$$
$$\text{Technical condition} : R_{\text{buy}}^{\oplus}(Q_t^a, r) = 1, \ R_{\text{sell}}^{\oplus}(r, Q_t^b) = 1, \ \forall r < q.$$

In particular, conditions (11) ensure that $(Q_0^a, Q_0^b) \geq (q, q) \Rightarrow (Q_t^a, Q_t^b) \geq (q, q), \ \forall t > 0$.

Matching process. The matching process is close to a pro-rata one [Field and Large, 2008]: in case of a liquidity removing buy order of size Q to be matched, all market participants having a quantity q resting in the ask queue will obtain a transaction for a fraction $Q \cdot q/Q_t^a$ of its order is traded at price $p^{\text{buy}}(Q_t^a)$, the remaining quantity staying in the orderbook.
At a first glance one may think that this matching process will induce intricate terms in the equations, but in fact it will not since we only consider utilities by units of good transactions.

- The orderbook shape is assumed to be linear (in the price), meaning that if a newcomer decide to provide liquidity to the market, her order will be split proportionally to the liquidity already present in the book: the orderbook will remain linear in price with a higher slope.

- Hence when a Liquidity Consumer order occurs, it will partially fill all Liquidity Provider orders according to a proportional rule;

4 The PFP (Price Formation Process) dynamics

4.1 Introducing the equations

We characterize an equilibrium via recursive equations of the expected value of future payoffs (value functions).

$$(12) \quad u(Q_t^a, Q_t^b) =$$

$$(1 - \lambda_{\text{buy}}dt - \lambda_{\text{sell}}dt - 2\lambda^- dt)\, u(Q_t^a, Q_t^b) \qquad\qquad \leftarrow \text{nothing}$$

$$+ (\lambda_{\text{sell}}R_{\text{sell}}^\ominus(u, Q_t^a + q, Q_t^b) + \lambda^-)dt\, u(Q_t^a, Q_t^b - q) \qquad \leftarrow \text{sell order, LC}$$

$$+ \ \lambda_{\text{sell}}R_{\text{sell}}^\oplus(u, Q_t^a + q, Q_t^b)dt\, u(Q_t^a + q, Q_t^b) \qquad\qquad \leftarrow \text{sell order, LP}$$

$$+ (\lambda_{\text{buy}}R_{\text{buy}}^\ominus(v, Q_t^a, Q_t^b + q) + \lambda^-)dt \cdot \big[\qquad\qquad \leftarrow \text{buy order, LC}$$

$$\underbrace{\frac{q}{Q_t^a}p^{\text{buy}}(Q_t^a)}_{\text{trade part (ask)}} + \underbrace{(1 - \frac{q}{Q_t^a})\, u(Q_t^a - q, Q_t^b)}_{\text{removing (ask)}} \big]$$

$$+ \ \lambda_{\text{buy}}\, R_{\text{buy}}^\oplus(v, Q_t^a, Q_t^b + q)dt\, u(Q_t^a, Q_t^b + q) \qquad\qquad \leftarrow \text{buy order, LP}$$

$$- \ c_a q\, dt. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \leftarrow \text{cost to maintain inventory}$$

Symmetrically, we have :

$$(13) \quad v(Q_t^a, Q_t^b) =$$

$$(1 - \lambda_{\text{buy}}dt - \lambda_{\text{sell}}dt - 2\lambda^- dt)\, v(Q_t^a, Q_t^b) \qquad\qquad \leftarrow \text{nothing}$$

$$+ (\lambda_{\text{buy}}R_{\text{buy}}^\ominus(v, Q_t^a, Q_t^b + q) + \lambda^-)dt\, v(Q_t^a - q, Q_t^b) \qquad \leftarrow \text{buy order, LC}$$

$$+ \ \lambda_{\text{buy}}R_{\text{buy}}^\oplus(v, Q_t^a, Q_t^b + q)dt\, v(Q_t^a, Q_t^b + q) \qquad\qquad \leftarrow \text{buy order, LP}$$

$$+ (\lambda_{\text{sell}}R_{\text{sell}}^\ominus(u, Q_t^a + q, Q_t^b) + \lambda^-)dt \cdot \big[\qquad\qquad \leftarrow \text{sell order, LC}$$

$$\underbrace{\frac{q}{Q_t^b}p^{\text{sell}}(Q_t^b)}_{\text{trade part (bid)}} + \underbrace{(1 - \frac{q}{Q_t^b})\, v(Q_t^a, Q_t^b - q)}_{\text{removing (bid)}} \big]$$

$$+ \ \lambda_{\text{sell}}\, R_{\text{sell}}^\oplus(u, Q_t^a + q, Q_t^b)dt\, v(Q_t^a + q, Q_t^b) \qquad\qquad \leftarrow \text{sell order, LP}$$

$$- \ c_b q\, dt. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \leftarrow \text{cost to maintain inventory}$$

Remind that R_{buy} and R_{sell} are functionals of Q^a and Q^b and also implicitly depends on u and v. Of course the previous principles hold for $Q_t^a, Q_t^b > q$, which is always the case thanks to conditions (11). In the equations above, c_a and c_b are positive constants modeling the cost to maintain inventory per unit, that is the cost of never being processed once waiting in the queue.

Remark : *The model, as described above, involves two standard queues in queueing theory, both have Poisson arrivals and generalized distributions of service time. The latter are endogenous and result from agent expectations. Classical results applies provided the mean of the endogenous distributions are known. From the queueing theory viewpoint, the novel point lies in the coupling of the queues (the service rate of a given queue depends upon the size of the other queue).*

4.2 Symmetric case

In the case where $\lambda_{\text{sell}} = \lambda_{\text{buy}} = \lambda$, and $c_a = c_b = c$, we have the following results.
For the sake of simplicity we will often use new notations for the queue size variables: x and y stand for Q_a and Q_b.

Lemma 4.1.
$$\forall(x, y), \ R_{\text{sell}}^\oplus(u, x, y) = R_{\text{buy}}^\oplus(2P - v, y, x)$$

This simple symmetry result is useful to get a necessary condition for the solution.

Proposition 4.2. *If system (12)-(13) has a unique solution (u, v), then*

$$\forall(x, y), \ u(x, y) + P = P - v(y, x).$$

That is, u and v are antisymmetric up to the constant P.

Proof. Take Equation (12) then perform the change of variable $w(y, x) = 2P - u(x, y)$, then apply the previous Lemma, switch the roles of x and y and multiply by -1. Then you get equation (13), hence the conclusion. $\qquad\square$

4.3 Continuous approximation

In this paragraph we formally derive differential equations corresponding to the PFP dynamic discrete equations (12-13) as presented in the previous section. Hopefully, this will lead us to get easily some qualitative insights on the solutions u and v.

To do so, we write the Taylor expansion of order 1 at the point (x, y) in system (12-13). After a quick computation, we get the following system of Partial Differential Equations (PDEs). Note that for the sake of simplicity we shorten the notations as follows: *sell* becomes s, *buy* becomes b, Q^a becomes x and Q^b becomes y.

$$\text{(Ask)} \quad 0 = [(\lambda_b R_b^\ominus + \lambda^-)\frac{1}{x}(p^b(x) - u) - c_a]$$
$$+ [\lambda_s R_s^\oplus - \lambda_b R_b^\ominus - \lambda^-] \cdot \partial_x u + [\lambda_b R_b^\oplus - \lambda_s R_s^\ominus - \lambda^-] \cdot \partial_y u,$$

$$\text{(Bid)} \quad 0 = [(\lambda_s R_s^\ominus + \lambda^-)\frac{1}{y}(p^s(y) - v) + c_b]$$
$$+ [\lambda_s R_s^\oplus - \lambda_b R_b^\ominus - \lambda^-] \cdot \partial_x v + [\lambda_b R_b^\oplus - \lambda_s R_s^\ominus - \lambda^-] \cdot \partial_y v.$$

Recall that u, v, R_b, R_s are estimated at (x, y) and R_b depends upon v, resp. R_s depends upon u. Consequently, R_b and R_s are the coupling terms in the PDE system (Ask)-(Bid).

The system has to be understood locally in the four regions

$$R^{++} = \{(x, y), R_s^\oplus(x, y) = R_b^\oplus(x, y) = 1\}, R^{--} = \{(x, y), R_s^\ominus(x, y) = R_b^\ominus(x, y) = 1\},$$
$$R^{+-} = \{(x, y), R_s^\oplus(x, y) = R_b^\ominus(x, y) = 1\}, R^{-+} = \{(x, y), R_s^\ominus(x, y) = R_b^\oplus(x, y) = 1\}.$$

Now we can write the general form of the first order system of coupled PDEs.

$$(14) \qquad 0 = \gamma_a(u, v, x, y) + \alpha(u, v, x, y)\partial_x u + \beta(u, v, x, y)\partial_y u$$
$$(15) \qquad 0 = \gamma_b(u, v, x, y) + \alpha(u, v, x, y)\partial_x v + \beta(u, v, x, y)\partial_y v,$$

where $\gamma_a, \gamma_b, \alpha, \beta$ have some good symmetry properties to be described later on.

The MFG framework. The model is of course a Mean Field Game. As mentioned in section 2, there are continuous entries and exits of players (modeled with Poisson processes). Therefore the basis assumptions are fulfilled: continuum of atomized and anonymous players.

Comparing equations (14)-(15) and (1), it is easy to notice that the equilibrium equations have the same form as the monotone system characterizing some MFG equilibria.

Second order terms. We kept only the first order terms in the equations. The second order terms to be added to the equations are:

$$\text{In (Ask)} \quad \frac{q^2}{2}\left[\frac{2}{x}(\lambda_b R_b^\ominus + \lambda^-)\partial_x u + \lambda^-\Delta u + (\lambda_s R_s^\oplus + \lambda_b R_b^\ominus)\partial_{xx} u + (\lambda_s R_s^\ominus + \lambda_b R_b^\oplus)\partial_{yy} u\right],$$

$$\text{In (Bid)} \quad \frac{q^2}{2}\left[\frac{2}{y}(\lambda_s R_s^\ominus + \lambda^-)\partial_y v + \lambda^-\Delta v + (\lambda_s R_s^\oplus + \lambda_b R_b^\ominus)\partial_{xx} v + (\lambda_s R_s^\ominus + \lambda_b R_b^\oplus)\partial_{yy} v\right].$$

5 Equilibrium analysis

5.1 Change of variables

From now on we focus on the symmetric case where $\lambda_s = \lambda_b = \lambda$ and $c_a = c_b = c$. First it is convenient to notice that in this important case, we have the following property:

$$\alpha = \beta = [\lambda(R_s^\oplus(u, x, y) - R_b^\ominus(v, x, y)) - \lambda^-].$$

We will see later that this property allows to solve the problem thanks to the characteristics method.

There is a very welcome change of variables that we will use throughout this section. We define

$$(16) \qquad \tilde{u} = (u - P)/q \quad \text{and} \quad \tilde{v} = (v - P)/q.$$

Then the (Ask)-(Bid) system reads

$$(17) \qquad 0 = [(\lambda \tilde{R}_b^\ominus + \lambda^-)\frac{1}{x}(\frac{\delta}{x - q} - \tilde{u}) - \frac{c}{q}] + [\lambda \tilde{R}_s^\oplus - \lambda \tilde{R}_b^\ominus - \lambda^-] \cdot (\partial_x \tilde{u} + \partial_y \tilde{u}),$$
$$0 = [(\lambda \tilde{R}_s^\ominus + \lambda^-)\frac{1}{y}(\frac{-\delta}{y - q} - \tilde{v}) + \frac{c}{q}] + [\lambda \tilde{R}_s^\oplus - \lambda \tilde{R}_b^\ominus - \lambda^-] \cdot (\partial_x \tilde{v} + \partial_y \tilde{v}).$$

Proposition 5.1. *Assume that system (17) admits a unique solution $(\tilde{u}, \tilde{v})$, then it is antisymmetric, that is:*

$$\forall(x, y), \ \tilde{v}(x, y) = -\tilde{u}(y, x).$$

The general form of the system (17) is as follows:

$$(18) \qquad 0 = \gamma(\tilde{u}, \tilde{v}, x, y) + \alpha(\tilde{u}, \tilde{v}, x, y)(\partial_x \tilde{u} + \partial_y \tilde{u})$$

$$(19) \qquad 0 = -\gamma(\tilde{v}, \tilde{u}, y, x) + \alpha(\tilde{u}, \tilde{v}, x, y)(\partial_x \tilde{v} + \partial_y \tilde{v}).$$

5.2 First Order Analysis

Here we explore formally some aspects of the first order approximation to the solution.

The key point of the analysis is that in the two equations of system (17), the derivative terms are the same, so that we conclude that the characteristics satisfy

$$\dot{x} = \dot{y} = \alpha \Rightarrow x = y + k.$$

Note that the reasoning of this paragraph holds on the region below the diagonal, but can be trivially extended to the whole domain by symmetry arguments.

We heuristically suppose that for a given k, and along the characteristic line $y = x - k$, there is a first point $M_0 = (x_0, y_0)$ where the sellers become Liquidity Consumer, that is M_0 is a point at the boundary of the regions R^{++} and R^{-+}.

Then there is a second point $M_1 = (x_1, y_1)$, with $x_1 \geq x_0$ and $y_1 \geq y_0$ where the buyers become Liquidity Consumer, that is M_1 is a point at the boundary of the regions R^{-+} and R^{--}.

First recall that:

$$R^{++} \text{ is defined by } R_s^{\oplus} = 1 \text{ and } R_b^{\oplus} = 1,$$
$$R^{-+} \text{ is defined by } R_s^{\oplus} = 0 \text{ and } R_b^{\oplus} = 1,$$
$$R^{--} \text{ is defined by } R_s^{\oplus} = 0 \text{ and } R_b^{\oplus} = 0.$$

We can write the differential equations on the three regions mentioned above:

$$(A_{R^{++}}) \quad 0 = \left[\frac{\lambda^-}{x}\left(\frac{\delta}{x-q} - \tilde{u}\right) - \frac{c}{q}\right] + [\lambda - \lambda^-] \cdot (\partial_x \tilde{u} + \partial_y \tilde{u}),$$

$$(B_{R^{++}}) \quad 0 = \left[\frac{\lambda^-}{y}\left(\frac{-\delta}{y-q} - \tilde{v}\right) + \frac{c}{q}\right] + [\lambda - \lambda^-] \cdot (\partial_x \tilde{v} + \partial_y \tilde{v}),$$

$$(A_{R^{-+}}) \quad 0 = \left[\frac{\lambda^-}{x}\left(\frac{\delta}{x-q} - \tilde{u}\right) - \frac{c}{q}\right] + [-\lambda^-] \cdot (\partial_x \tilde{u} + \partial_y \tilde{u}),$$

$$(B_{R^{-+}}) \quad 0 = \left[\frac{\lambda + \lambda^-}{y}\left(\frac{-\delta}{y-q} - \tilde{v}\right) + \frac{c}{q}\right] + [-\lambda^-] \cdot (\partial_x \tilde{v} + \partial_y \tilde{v}),$$

$$(A_{R^{--}}) \quad 0 = \left[\frac{\lambda + \lambda^-}{x}\left(\frac{\delta}{x-q} - \tilde{u}\right) - \frac{c}{q}\right] + [-\lambda - \lambda^-] \cdot (\partial_x \tilde{u} + \partial_y \tilde{u}),$$

$$(B_{R^{--}}) \quad 0 = \left[\frac{\lambda + \lambda^-}{y}\left(\frac{-\delta}{y-q} - \tilde{v}\right) + \frac{c}{q}\right] + [-\lambda - \lambda^-] \cdot (\partial_x \tilde{v} + \partial_y \tilde{v}).$$

The equations are relatively simple in each region. The tricky point is, as always, to stick together the solutions of each region. First we compute the boundaries of the regions.

Search for M_0: first order boundary between R^{++} and R^{-+}. Looking at equations $(A_{R^{++}})$ and $(A_{R^{-+}})$ we notice that at the boundary there is a jump causing a change of sign of the coefficient multiplying the derivatives (under the basic assumption $\lambda \geq \lambda^-$). Therefore, at this point we must have

$$(20) \qquad \frac{\lambda^-}{x_0}\left(\frac{\delta}{x_0 - q} - \tilde{u}\right) = \frac{c}{q}.$$

On the other hand, as the seller's routing decision $R_s^{\oplus}$ jumps from 1 to 0, we must have

$$(21) \qquad \tilde{u} = \frac{-\delta}{y_0 - q}.$$

Combining (20) and (21) we get the equality:

$$\eta x_0 = \frac{1}{x_0 - q} + \frac{1}{y_0 - q},$$

where

$$(22) \qquad \eta := \frac{c}{\delta q \lambda^-}.$$

It follows that the diagonal point of the boundary M_0 is the point

$$(23) \qquad (x_0^*, x_0^*) = (q + \sqrt{q^2 + 8/\eta})/2$$

and that the boundary is defined by the parametric equation:

$$(24) \qquad (x_0, y_0) = \left(x_0, l(x_0) := q + \left(\eta x_0 - \frac{1}{x_0 - q}\right)^{-1}\right), \ \forall x_0 \geq x_0^*.$$

Search for M_1: first order boundary between R^{-+} and R^{--}. Unfortunately, looking at equations (B_{R-+}) and (B_{R--}) we conclude that we cannot adopt the same reasoning since the sign of the coefficients multiplying the derivative terms does not change.

We use another strategy. We solve $\tilde{v}$ analytically all along the characteristic line $y_1 = x_1 - k$, and then intersect the solution $\tilde{v}$ with $\frac{\delta}{x_1-q}$.

Along the characteristic $x = y + k$, we introduce the function

$$f(y) = \tilde{v}(y + k, y).$$

Looking at equation (17), we get the generic form of the ordinary differential equation (ODE for short) satisfied by f:

$$(25) \qquad f' + \frac{a}{y}f + \left(\frac{b}{y(y-q)} + d\right) = 0,$$

where

$$a = 1 + \lambda/\lambda^-, \ b = \delta(1 + \lambda/\lambda^-), \ d = -\delta\eta, \text{ on } R^{-+}.$$

We use the variation of constant method to solve equation (25).

The homogeneous solution is $f(y) = y^{-a} \times constant$. Now let the constant varies as a function $g(x)$. We have $f' = g'y^{-a} - agy^{-a-1}$. Substituting in (25) we obtain:

$$g'(y) = -b\frac{y^{a-1}}{y-q} - dy^a.$$

This function is easy to integrate numerically. However in order to stay working with analytical formulas, we make the approximation $y - q \approx y$ for small q (recall that all this analytical part focus on the small q first order approximation). Now we are in the position to integrate the derivative g'.

We get

$$f(y) = g(y)y^{-a} = (constant \ y^{-a} - \frac{b}{a-1}y^{-1} - \frac{d}{a+1}y).$$

Now we have to compute the *constant*. Recall that we are working on the line $(y + k, y)$ and on the region R^{-+} so that we are solving the ODE with an initial condition on M_0, which is known to be $(x_0, l(x_0))$. Consequently we have to look at f as a family (f_k) of functions indexed by $k \in \mathbb{R}^+$. On the characteristic line starting at $x_0 - l(x_0)$, the function is given by

$$(26) \qquad f_{x_0-l(x_0)}(y) = (\mathcal{C}(x_0) \ y^{-a} - \frac{b}{a-1}y^{-1} - \frac{d}{a+1}y), \ \forall y \geq l(x_0).$$

The core argument to compute the constant parameter $\mathcal{C}(x_0)$ for the solution on the characteristic $(y+k, y)$, with $k = x_0 - l(x_0)$, is to remark that:

$$f_k(y) = \tilde{v}(y + k, y) = -\tilde{u}(y, y + k) = -f_k(y + k).$$

Then, the initial condition equality

$$f_{x_0-l(x_0)}(l(x_0)) = -f_{x_0-l(x_0)}(x_0),$$

automatically gives the expression of $\mathcal{C}$:

$$(27) \qquad \mathcal{C}(x_0) = \delta\frac{(1 + \lambda^-/\lambda)[x_0^{-1} + l(x_0)^{-1}] - \frac{\eta}{1+\lambda/\lambda^-}[x_0 + l(x_0)]}{x_0^{-(1+\lambda/\lambda^-)} + l(x_0)^{-(1+\lambda/\lambda^-)}},$$

where the last equality holds since the equation of $\tilde{u}$ on R^{+-} matches the equation of $\tilde{v}$ on R^{+-}. Consequently, the analytical solution is given by (26)-(27).

Finally we are in the position to compute the parametric curve of the boundary between the two regions R^{-+} and R^{--}.

To do so we look for the point $(x_1, y_1) = (y_1 + k, y_1)$ such that $\tilde{v}(x_1, y_1) = \frac{\delta}{x_1-q}$.

More precisely, M_1 is defined by: $(y_1 + x_0 - l(x_0), y_1)$, $\forall x_0 \geq x_0^*$, where

$$y_1 \text{ verifies } f_{x_0-l(x_0)}(y_1) = \frac{\delta}{y_1 + x_0 - l(x_0) - q}.$$

Figure 2 exhibits an instance of the first order curves. We observe that near the diagonal, there is a region where several solutions could happen. The first order analysis thus shows the global form of the shape of the solution (since it is based on the curves M_0, M_1), and that considering higher order terms is necessary to understand what happens in the region near the diagonal.
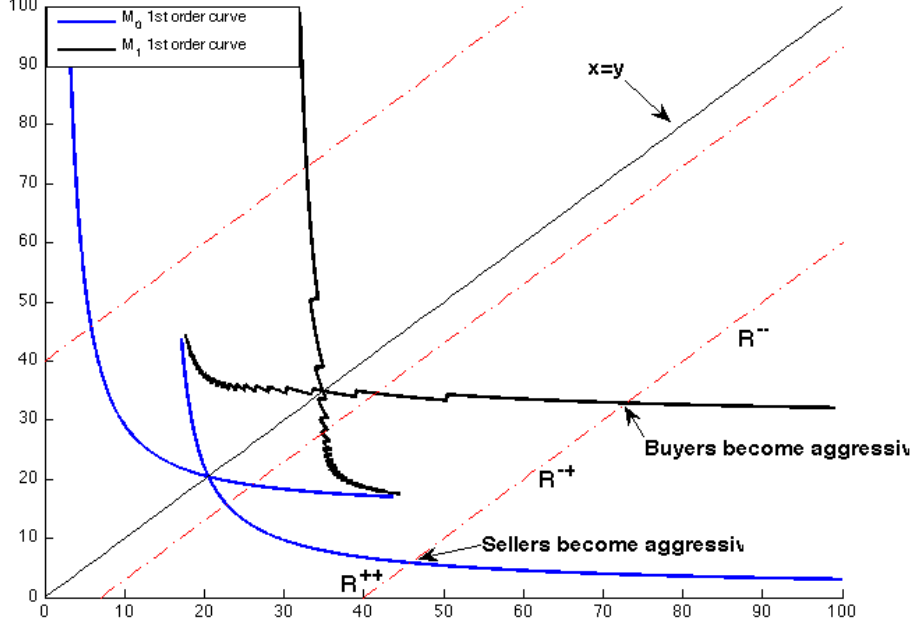
Figure 2: First order decision curves

5.3 Second Order Equations

According to section 4.3, the general form of the second order equations is:

$$
\begin{aligned}
0 =\ & \gamma(\tilde{u},\tilde{v},x,y) + \alpha(\tilde{u},\tilde{v},x,y)(\partial_x\tilde{u}+\partial_y\tilde{u}) \\
& + q\Big(\rho(\tilde{v},x,y))\partial_x\tilde{u} + \xi_1(\tilde{u},\tilde{v},x,y)\partial_{xx}\tilde{u} + \xi_2(\tilde{u},\tilde{v},x,y)\partial_{yy}\tilde{u}\Big), \\
0 =\ & -\gamma(\tilde{v},\tilde{u},y,x) + \alpha(\tilde{u},\tilde{v},x,y)(\partial_x\tilde{v}+\partial_y\tilde{v}) + \\
& + q\Big(\rho(\tilde{u},y,x))\partial_y\tilde{v} + \xi_1(\tilde{u},\tilde{v},x,y)\partial_{xx}\tilde{v} + \xi_2(\tilde{u},\tilde{v},x,y)\partial_{yy}\tilde{v}\Big),
\end{aligned}
$$

(28)

where:
$\rho = \frac{1}{x}(\lambda R_b^{\ominus} + \lambda^-)$, $\xi_1 = (\lambda(R_s^{\oplus}+R_b^{\ominus}) + \lambda^-)/2$, and $\xi_2 = (\lambda(R_s^{\ominus}+R_b^{\oplus}) + \lambda^-)/2$.

Define $\Lambda = \lambda + \lambda^-$. Let us now give the local equations on the same four regions.

$$
\begin{aligned}
(A_{R++})\quad & 0 = \Big[\frac{\lambda^-}{x}(p^b(x)-u)-c\Big] + [\lambda-\lambda^-](\partial_x u+\partial_y u) + q\Big(\frac{\lambda^-}{x}\partial_x u + \frac{\Lambda}{2}\Delta u\Big), \\
(B_{R++})\quad & 0 = \Big[\frac{\lambda^-}{y}(p^s(y)-v)+c\Big] + [\lambda-\lambda^-](\partial_x v+\partial_y v) + q\Big(\frac{\lambda^-}{y}\partial_y v + \frac{\Lambda}{2}\Delta v\Big), \\
(A_{R-+})\quad & 0 = \Big[\frac{\lambda^-}{x}(p^b(x)-u)-c\Big] + [-\lambda^-](\partial_x u+\partial_y u) + q\Big(\frac{\lambda^-}{x}\partial_x u + \frac{\lambda^-}{2}\Delta u + \lambda\partial_{yy}u\Big), \\
(B_{R-+})\quad & 0 = \Big[\frac{\Lambda}{y}(p^s(y)-v)+c\Big] + [-\lambda^-](\partial_x v+\partial_y v) + q\Big(\frac{\Lambda}{y}\partial_y u + \frac{\lambda^-}{2}\Delta u + \lambda\partial_{yy}v\Big), \\
(A_{R+-})\quad & 0 = \Big[\frac{\Lambda}{x}(p^b(x)-u)-c\Big] + [-\lambda^-](\partial_x u+\partial_y u) + q\Big(\frac{\Lambda}{x}\partial_x u + \frac{\lambda^-}{2}\Delta u + \lambda\partial_{xx}u\Big), \\
(B_{R+-})\quad & 0 = \Big[\frac{\lambda^-}{y}(p^s(y)-v)+c\Big] + [-\lambda^-](\partial_x v+\partial_y v) + q\Big(\frac{\lambda^-}{y}\partial_y u + \frac{\lambda^-}{2}\Delta u + \lambda\partial_{xx}v\Big), \\
(A_{R--})\quad & 0 = \Big[\frac{\Lambda}{x}(p^b(x)-u)-c\Big] + [-\Lambda](\partial_x u+\partial_y u) + q\Big(\frac{\Lambda}{x}\partial_x u + \frac{\Lambda}{2}\Delta u\Big), \\
(B_{R--})\quad & 0 = \Big[\frac{\Lambda}{y}(p^s(y)-v)+c\Big] + [-\Lambda](\partial_x v+\partial_y v) + q\Big(\frac{\Lambda}{y}\partial_y v + \frac{\Lambda}{2}\Delta v\Big).
\end{aligned}
$$

Remark that both a diffusion term **and a drift term** appear.

In the next part, we provide several example of markets based on the model.

14

6   Applications

This section is dedicated to applications of the MFG model to study the outcome of a combination of different trading behaviors in the same order book.

The purpose here is not to study how the price discovery operates on the long term, but how microstructure effects can deviate transaction prices from the fair price. This model will thus explain how the state of the liquidity can change the dynamics of the price and potentially form an equilibrium price. This equilibrium can potentially deviate from the latent (or fair) price. The main drivers of these modifications will be the behavior of trading agents, and specifically the average size of their orders, their speed, their waiting cost, and how often they use optimized strategies (see Table 1 for a qualitative description of the main parameters of their strategies).

In this section, we will investigate theoretically and using simulations the reasons why the price deviates or not from the "*fair price*" (exogenously fixed). The variables of interest are:

- the asymptotic state of the liquidity offer (i.e. the size of the bid queue and ask queue): are they large or small? are they balanced?

- The average transaction price: how far away it is from the fair price?

- The average value of the bid-ask spread; in which conditions is it high or low?

Having in mind that each time an agent buys or sells she suffers from market impact, the strategy of each market participant affects her price. We will thus be able to compute an average price for each class of market participants, answering the question: do the institutional investors pay more than high frequency traders?

It will also allow us to compute an *effective bid-ask spread* being twice the difference between the mid price and the transaction price; it will not be the same for each market participant.

Definition 6.1 (Effective bid-ask spread). *The effective bid ask spread of an agent A is the expected transaction price of its liquidity removing buying orders minus the one of its liquidity removing selling orders:*

$$(29) \qquad \psi^e(A) := \delta \cdot \mathbb{E}\left(\frac{q}{Q_t^a} \,\middle|\, R_{\text{buy}}^{\ominus}(A) \right) + \delta \cdot \mathbb{E}\left(\frac{q}{Q_t^b} \,\middle|\, R_{\text{sell}}^{\ominus}(A) \right).$$

The effective spread is higher for an impatient agent if the spread is larger when he consume liquidity than when he provides liquidity.

Another important tool is the *invariant measure*, being the probability of having the system in a specific region of the state space. Our state space is mostly captured by the sizes of the two queues (the bid queue and the ask queue).

Thanks to the results obtained in the previous sections, we will be able not only to observe discrepancies between agents' behaviour and their outcome, but also to explain and understand them in details.

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
q_{ii}	1	0.25	1	1	1	1
λ_{ii}	1	1	1	0.5	0.5	0.6
λ_{ii}^-	0.2	0.2	0.2	0.5	0.5	0.4
$c_{ii} \cdot q_{ii}$	$2.5 \cdot 10^{-3}$	$2.5 \cdot 10^{-3}$	10^{-2}	$2.5 \cdot 10^{-3}$	$2.5 \cdot 10^{-3}$	$2.5 \cdot 10^{-3}$
q_{HFT}	-	-	-	-	0.25	0.25
λ_{HFT}	-	-	-	-	4	3.6
λ_{HFT}^-	-	-	-	-	0	0.4
$c_{\text{HFT}} \cdot q_{\text{HFT}}$	-	-	-	-	10^{-2}	10^{-2}

Table 2: Parameters defining the studied models.

A first subsection is dedicated to applications with models including one class of agents only, to understand and explain in details the mechanisms that our MFG model can render. In a second subsection we will use a multi-agent model, allowing to understand the result of putting together more than one class of market participants. Here we mix Institutional Investors and High Frequency Traders. Section **??** present a theoretical expansion of Section 3.2 needed to handle more than one agent class. Table 2 summarizes the different models and their parameters.

6.1   Markets with Institutional Investors only

6.1.1   Modeling Institutional Investors

Since we just want to model one class of market participants, their specification is not very important. It will become crucial when we mix different agents: the relative speed, the relative sizes of orders, etc., will play a crucial role in the multi-agent simulations.

With one type of investors only we mainly focus on using realistic values and exploring the sensitivities of the emerging dynamics to the values of the parameters.

Note first that some parameters define the framework of the simulation and not the market participant themselves:

- the *fair price P* has no impact on the dynamics, it is taken as a constant,

- the *market depth* δ, playing a role in the expression of the market impact of one trade (at the first order it is homogenous to Kyle's lambda). Looking carefully at the market impact expression (8), it can be read that δ is homogenous to the inverse of a quantity: dividing it by two and multiplying quantities by two will not change the dynamics but increase the waiting costs (that are proportional to the order size q).

Other parameters are directly associated with the agent:

- the size of her orders q,

- the intensity λ of the Poisson process governing the arrival rate of smart routed orders;

- the intensity λ^- of the Poisson process governing the arrival of not smart routed orders (i.e. blindly sending market orders);

- the cost of waiting per share c: waiting dt seconds is worth $cq\,dt$.

Some simple statistics on equity markets can give reasonable figures for these parameters (see [Burgot et al., 2013] for more details about evolution of trading behaviors):

- the intensity $\Lambda = \lambda + \lambda^-$ can be roughly estimated by the average number of trades per time unit ;

- the size q has no unit (it will have a role when compared to the size of HFT orders); for the sake of simplicity we will take it equals to one. For information the table gives the *average trade size* and the *average size at first limit*.

6.1.2 Simulations and results

First we consider the case of a single group of traders all with the same order size q. The elementary algorithm we use to compute the equilibrium is as follow:

1. Initialize u^0 and v^0 (e.g. to the constant function equal to P)

2. Step k:
   compute u^k the solution to equation (12) using the inputs u^{k-1} and v^{k-1},
   compute v^k the solution to equation (13) using the inputs u^k and v^{k-1}.

Equilibrium as an invariant measure. Equilibrium visualization is made of the level sets of the stationary measure of queue sizes. The previous measure is computed from the transition probability depending upon u and v.
More precisely,the transition process at a certain state (Q_a, Q_b):

$$
\begin{array}{lcccl}
(Q_a, Q_b) & \rightarrow & (Q_a, Q_b) & \text{with probability} & 1 - 2\lambda dt - 2\lambda^- dt \\
(Q_a, Q_b) & \rightarrow & (Q_a + q, Q_b) & \text{with probability} & \lambda R_s^\oplus dt \\
(Q_a, Q_b) & \rightarrow & (Q_a - q, Q_b) & \text{with probability} & \lambda R_b^\ominus dt + \lambda^- dt \\
(Q_a, Q_b) & \rightarrow & (Q_a, Q_b + q) & \text{with probability} & \lambda R_b^\oplus dt \\
(Q_a, Q_b) & \rightarrow & (Q_a, Q_b - q) & \text{with probability} & \lambda R_s^\ominus dt + \lambda^- dt.
\end{array}
$$

The resulting process has of course the Markov property.

**Test 1: institutional traders with many SOR arrivals (i.e. very few impatient traders).** The first numerical test corresponds to the following set of parameters: $q = 1, \delta = 2, c = 2,5 \times 10^{-3}, \lambda = 1$, and $\lambda^- = 0.2$.
   We show the results in Figure 3. We observe that the decision regions R^{++}, R^{-+}, R^{+-}, and R^{--} have the expected form. We also remark that the second order term selects a particular solution amongst all order one solutions. This has to be linked to the notion of viscosity solutions.
   Below the diagonal, that is for values of Q_b smaller than Q_a, the region where both sellers and buyers are Liquidity Provider corresponds to small Q_a and Q_b, then the sellers turn to be Liquidity Consumer while the buyers remain liquidity adders, and finally they also turn to be Liquidity Consumer.

The invariant measure is almost concentrated on the points where both sellers and buyers turn to be Liquidity Consumer, i.e. it is concentrated on the boundary curve M_0 describing the frontier between R^{++} and R^{-+} & R^{+-}. From now on, we refer to this curve as the *PC switching curve* for Provider to Consumer switching curve. We symmetrically define the *CP switching curve* as the frontier between R^{+-} & R^{-+} and R^{--}.
   However, the most remarkable point is that the invariant measure shows two bumps, located in the cavities of the PC switching curve. In the new reference frame after a $\pi/4$ axis rotation, the cavities corresponds to the global minimum points of the PC switching curve.
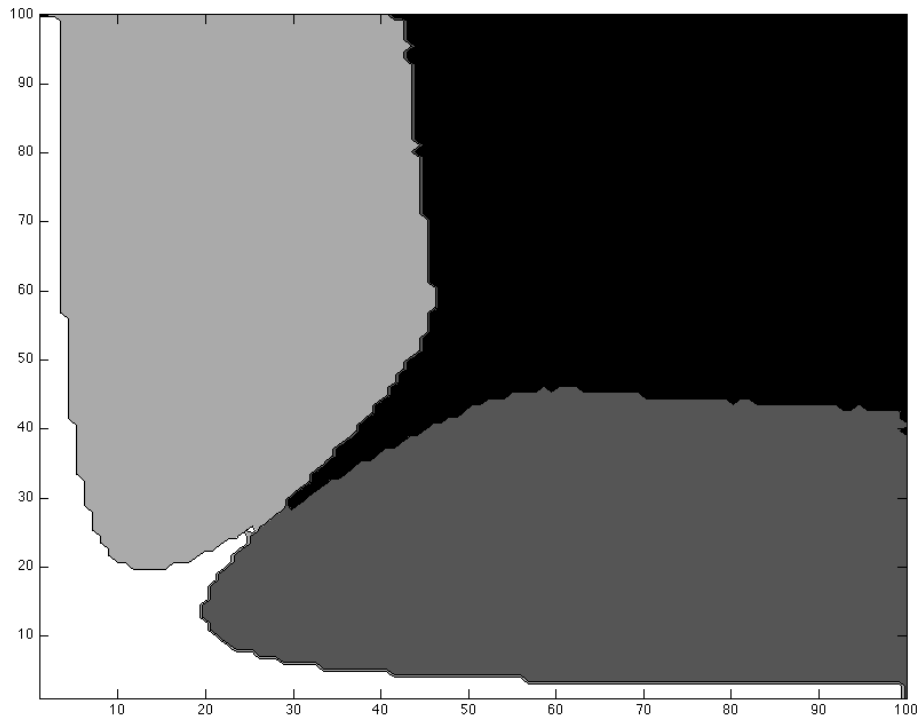
(a) Mapping of the decision regions
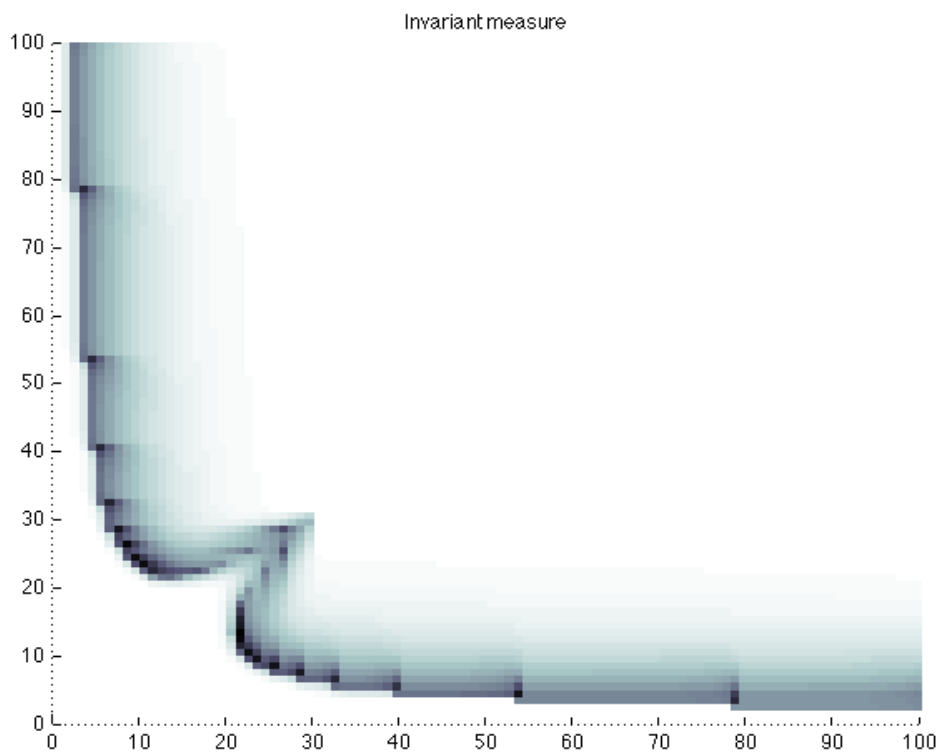
(b) Corresponding invariant measure

Figure 3: The numerical solution for a single homogeneous specie of traders. The routing decision regions have the expected form. The invariant measure exhibits two symmetric bumps.
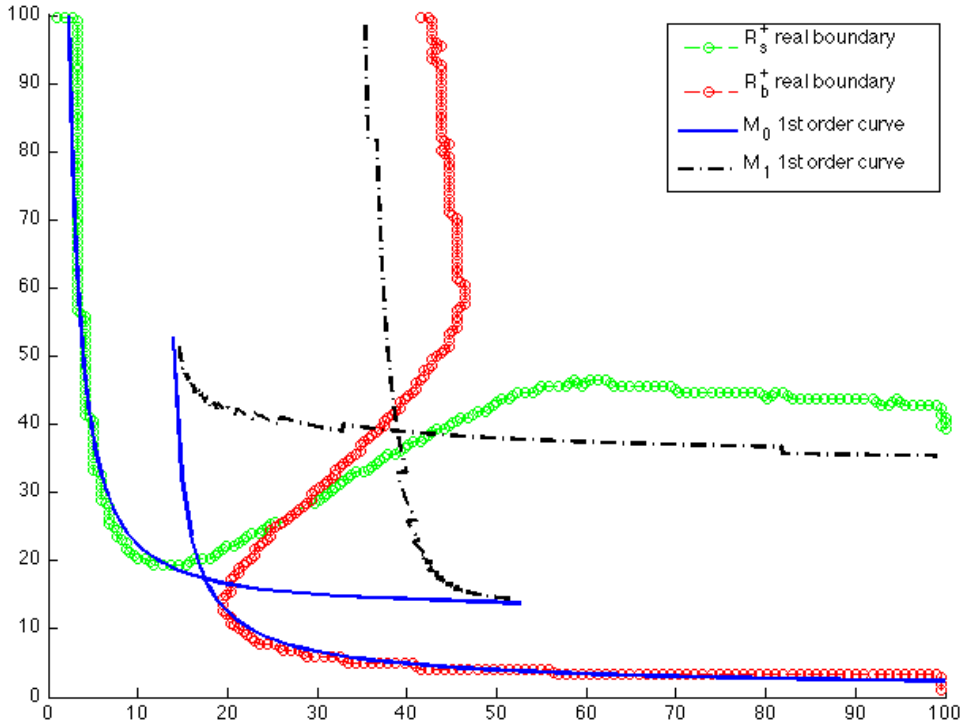
Figure 4: The *real* switching curves tries to conciliate the curves analytically computed at order 1.

Here the economic intuition is that there are two symmetric liquidity pools, one on the buy side, where only buy orders are completed, and conversely for the sell side.

Figure 4 shows the form of the routing decision boundaries *versus* the first order analytical curves derived earlier. It is noteworthy to observe that the *real* switching curves tries to conciliate the curves analytically computed at order 1. However, it approximates better the PC switching curve than the CP switching curve.

Several visualizations of system trajectories are possible. We display the time evolution (for 600 instants) of various quantities in Figure 5. In this example there are mainly two distinct regimes: from instant t_{ini} to $t_{switch} = 330$, the activity is mainly concentrated on the ask queue. During the second period, most of the activity holds on the bid side.

The coupled trajectory of queue sizes in the space $Q_a \times Q_b$ is another possible visualization. In the plot, dots are colored from yellow to red, according to the number of time the queue system passes through the corresponding size configuration. Here also we see the hange of regime at t_{switch} where the process goes through the diagonal, jumping from the ask activity zone to the bid activity one. One more time both liquidity configurations are visible. Above the diagonal $Q_a = Q_b$, red dots are more likely to be horizontally distributed (meaning that most of transactions hold on the ask side), and symmetrically below the diagonal.

Both visualizations confirm the phenomenon that one could expect after looking at the invariant measure plot.

Test 2: order size impact. Now we only change the value of q, and take it smaller than in test 1: $q = 0, 25$. Figure 6 shows that the real CP switching curve is closer to the 1st order curve, which is natural since taking q smaller means that the second order term impacts are shrunk.

Test 3: risk aversion impact. Figure 7 shows the impact of a bigger c on the solution. We compare the results obtained for the set of parameters of Test 1 with the results obtained for the same parameters exepct the value of the new risk aversion $c_h = 4 \times c = 10^{-2}$. We observe mainly two effects. First queue sizes are shrunk (from about 30 to 10). Second the invariant measure maximum is now on the diagonal. Therefore, the two antisymmetric liquidity pools progressively disappear.

Test 4: the case with half non-SOR orders. Here we provide a stationary equilibrium instance in a case where half of the order arrivals are non-SOR. Figure 8 shows that at the equilibrium there are still two symmetric regions of concentration at the neighborhood of the PC switching curves. Naturally we observe more density on low queue sizes.
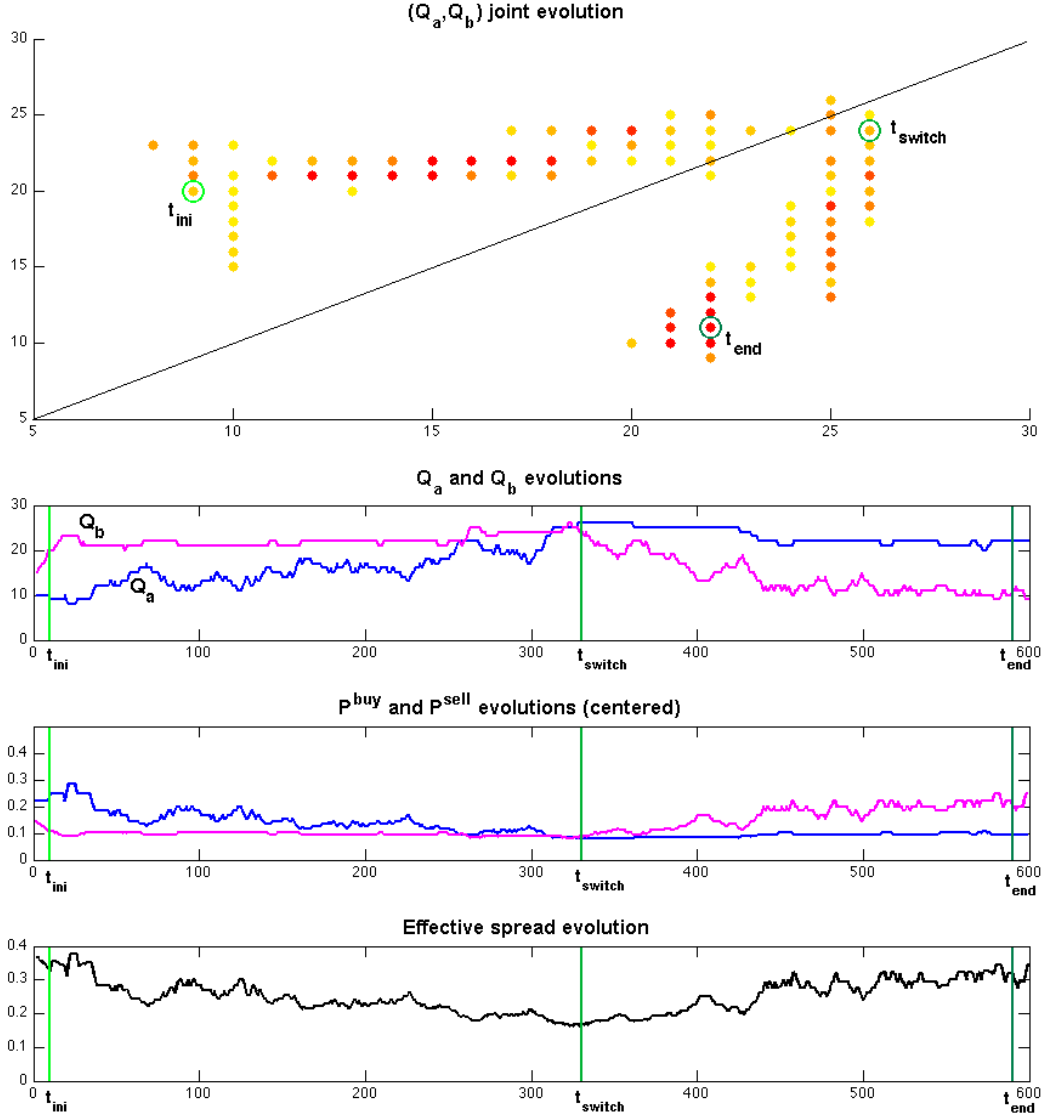
Figure 5: Here we show a particular simulation of the order book. We plot the evolution of various quantities for 600 instants. A sample trajectory of coupled queue sizes is plotted. Dots are colored from yellow (1 visit) to red (about 15 visits). Note that 3 milestones are introduced. We observe a change of regime at instant t_{switch} where the market activity switches from the ask queue to the bid queue.

6.1.3 Possible liquidity imbalance with one class of participant only

First remind that the *type of a participant* is described by the way she interacts with order books. Hence a pension fund taking long term positions, a low frequency statistical arbitrageur, and the hedging desk of an investment bank will have the same *type*. The important elements being they all:

- take a decision before starting to interact with the order book,
- do not use a smart order router systematically,
- trade with relatively large orders, even once their *meta orders* have been split thanks to an optimal trading scheme (like in [Bouchard et al., 2011],[Almgren and Chriss, 2000],[Alfonsi et al., 2009]).

The outcome of the application to one class of investors is that the market can suffer for long and stable liquidity imbalances. We have seen that is such typical cases the bid and ask queues are in an asymmetric configuration:

- one of the queues (the ask one, for instance) is significantly shorter than the other,
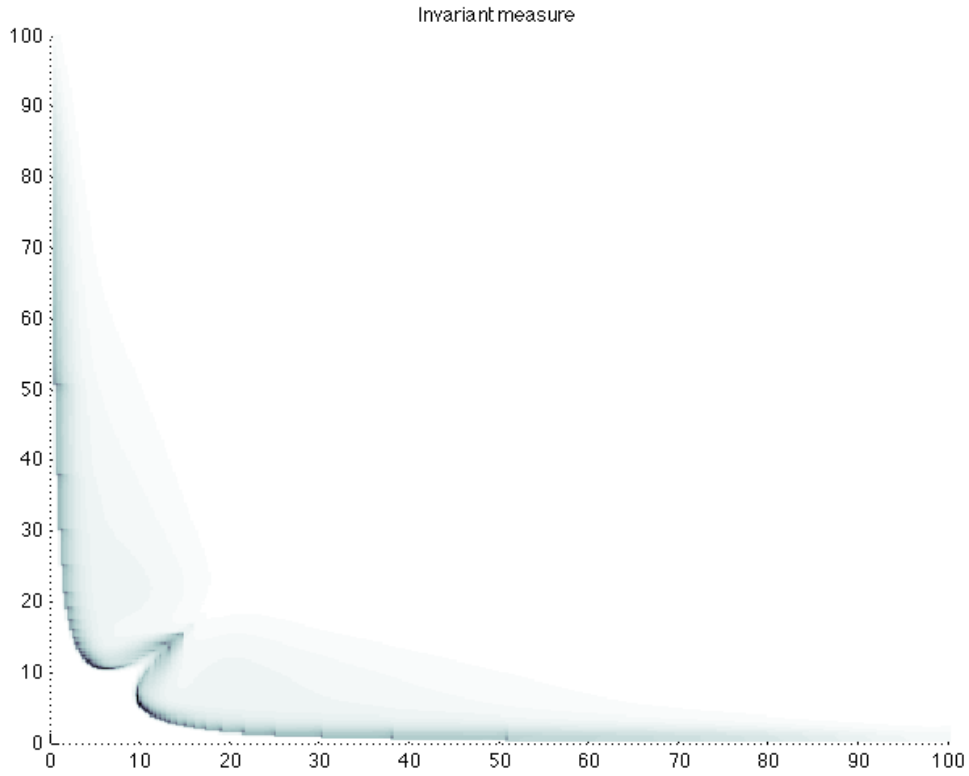
Invariant measure

(a) The invariant measure has the same form, it is concentrated on smaller values.
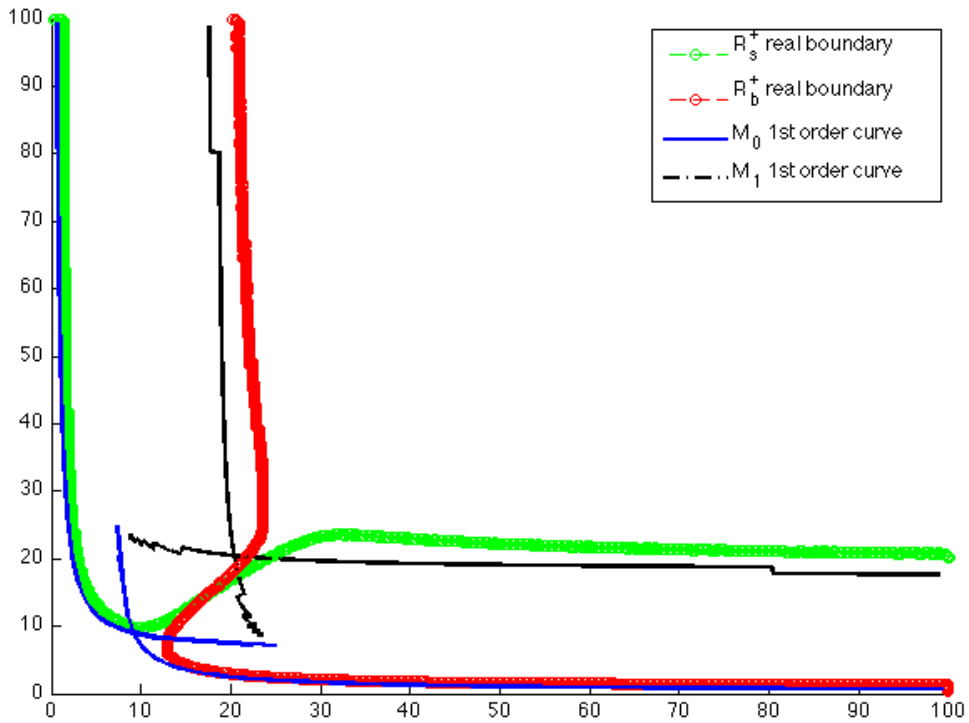
(b) The real PC switching curves is closer to the 1st order switching curve.

Figure 6: Comparing with the previous case, we deduce the second order terms impacts.
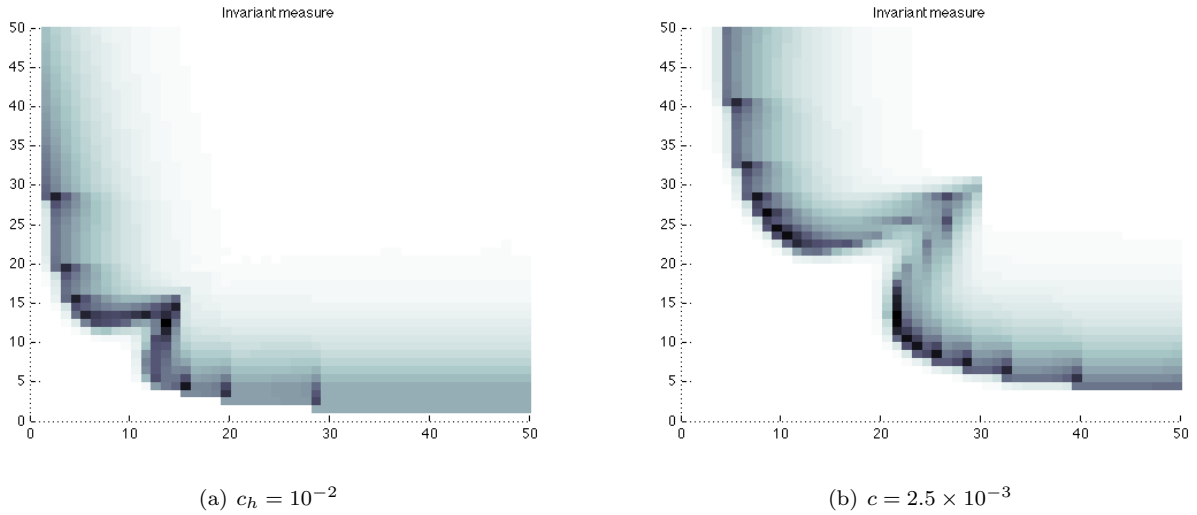
(a) $c_h = 10^{-2}$

(b) $c = 2.5 \times 10^{-3}$

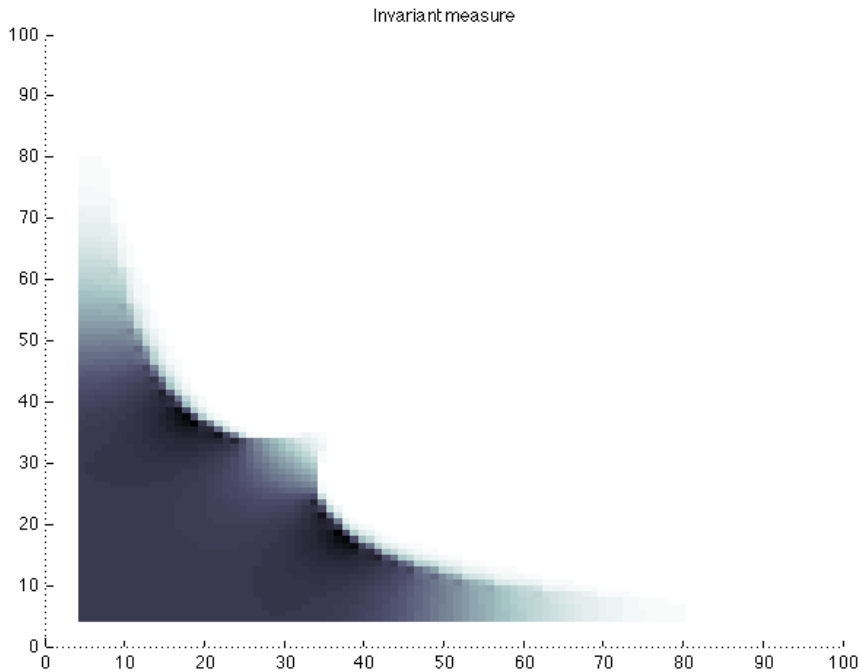Figure 7: The two symmetric bumps vanish and a single bump appears on the diagonal

Figure 8: The case with half SOR and half non-SOR arrivals.

- the flow of buyers considers that the price to pay to wait is too high and accept to pay the market impact on a small queue,
- the flow of seller notices that they can obtain a fast trade being passive (i.e. going into the bid queue), since 100% of the buyers are now impatient.

This leads to a stable state of the order book: the invariant measure sees two symmetrical concentrations of such configurations, dominating states on the diagonal (see Figure 3).

In such a situation, we can say that *liquidity calls for liquidity*: the conjunction of a high rate of consuming orders at the smaller queue and of an high arrival rate of liquidity on the same queue feeds an equilibrium.

During such a configuration the transaction price is significantly different from the *fair price*. Since the model is stochastic its state will nevertheless evolve to explore other configuration (see Figure 5 for a trajectory instance). Nevertheless the form of the invariant measure indicates that the fraction of time during which the model is in such inefficient configurations dominates.

21

When the behaviors of participants are so similar that they create liquidity imbalances, it is often proposed to add a population of market makers (see [Ho and Stoll, 1983]), hoping that it will break these "liquidity circles" and bring back the invariant measure on the diagonal.

6.2 Introducing High Frequency Traders

6.2.1 Equations and quantities for two groups

The model is such that considering several types of traders is not a big deal. This is good news since our aim is to get insights on the role of High Frequency Traders in the scope of our model.
We therefore split the agents into two subsets:

- Institutional Investors, with a smaller intensity λ_1, but with bigger sizes q_1 and risk aversion c_1.

- HFTs, with a higher intensity λ_2 and smaller sizes q_2 and risk aversion c_2,

The two groups also differentiate by having specific λ_1^-, λ_2^- and c_1, c_2. This leads to twice the value functions we had.
We also have to consider group's 2 routing decisions $R_{buy}^{\oplus}(v_2, x, y + q_2) := \mathbf{1}_{v_2(x, y+q) < p_{q_2}^{buy}(x)}$ (symmetrically $R_{sell}^{\oplus}$).
Now we write the equation for one of sellers' value functions u_1, u_2. As before, buyers' value function equations can be easily derived by simple symmetry arguments.

$$k \cdot u_1(x, y) =$$
$$(\lambda_1 R_{\text{sell}}^{\ominus}(u_1, x + q_1, y) + \lambda_1^-) u_1(x, y - q_1) + \lambda_1 R_{\text{sell}}^{\oplus}(u_1, x + q_1, y) u_1(x + q_1, y)$$
$$+ (\lambda_1 R_{\text{buy}}^{\ominus}(v_1, x, y + q_1) + \lambda_1^-)[\frac{q_1}{x} p_1^{\text{buy}}(x) + (1 - \frac{q_1}{x}) u_1(x - q_1, y)]$$
(30)
$$+ (\lambda_1 R_{\text{buy}}^{\oplus}(v_1, x, y + q_1) u_1(x, y + q_1)) + (\lambda_2 R_{buy}^{\oplus}(v_2, x, y + q_2) u_1(x, y + q_2)$$
$$+ (\lambda_2 R_{sell}^{\ominus}(u_2, x + q_2, y) + \lambda_2^-) u_1(x, y - q_2) + \lambda_2 R_{sell}^{\oplus}(u_2, x + q_2, y) u_1(x + q_2, y)$$
$$+ (\lambda_2 R_{buy}^{\ominus}(v_2, x, y + q_2) + \lambda_2^-)[\frac{q_2}{x} p_2^{buy}(x) + (1 - \frac{q_2}{x}) u_1(x - q_2, y)]$$
$$- c_1 q_1),$$

$$k \cdot u_2(x, y) =$$
$$(\lambda_1 R_{\text{sell}}^{\ominus}(u_1, x + q_1, y) + \lambda_1^-) u_2(x, y - q_1) + \lambda_1 R_{\text{sell}}^{\oplus}(u_1, x + q_1, y) u_2(x + q_1, y)$$
$$+ (\lambda_1 R_{\text{buy}}^{\ominus}(v_1, x, y + q_1) + \lambda_1^-)[\frac{q_1}{x} p_1^{\text{buy}}(x) + (1 - \frac{q_1}{x}) u_2(x - q_1, y)]$$
(31)
$$+ (\lambda_1 R_{\text{buy}}^{\oplus}(v_1, x, y + q_1) u_2(x, y + q_1)) + (\lambda_2 R_{buy}^{\oplus}(v_2, x, y + q_2) u_2(x, y + q_2))$$
$$+ (\lambda_2 R_{sell}^{\ominus}(u_2, x + q_2, y) + \lambda_2^-) u_2(x, y - q_2) + \lambda_2 R_{sell}^{\oplus}(u_2, x + q_2, y) u_2(x + q_2, y)$$
$$+ (\lambda_2 R_{buy}^{\ominus}(v_2, x, y + q_2) + \lambda_2^-)[\frac{q_2}{x} p_2^{buy}(x) + (1 - \frac{q_2}{x}) u_2(x - q_2, y)]$$
$$- c_2 q_2,$$

where $k = 2(\lambda_1 + \lambda_2 + \lambda_1^- + \lambda_2^-)$.
At this stage, it is important to remark that the only difference in equations (30) and (31) is the term $c_1 q_1$. As a consequence, u_1 and u_2 coincide as soon as $c_1 q_1 = c_2 q_2$, which is the reference case we study in the present work.
Note that in this case, a first order resolution can be performed using the methodology of the previous section.

In this section we will compare equilibria in terms of average transaction prices and spread. Let us detail the way we define average prices. Before writing down the average price equations, we need to introduce some notations. For the sake of simplicity we only work with sellers. Buyers notations and equations can be easily derived by symmetry.

Empirical stationary measure: $\hat{m}(Q_a, Q_b)$

Type's i stationary proportion: $\hat{\gamma}_i(Q_a, Q_b)$

As previously, there are several regions defined by the LP or LC behavior of traders. Figure 9 shows an instance with 9 regions. Two cases may happen, depending upon the relative size of $\lambda_i^-, \lambda_i, i = 1, 2$. We do not want to enter the (technical) details nor review all possible cases, but we would like to mention that a necessary condition for the existence of several regions is that $\sum \lambda_i > \sum \lambda_i^-$, that is there is globally more SOR than non-SOR traders in the system. In the first case, Institutional Investors switch first from LP to LC. In the second case, HFT switch first (which is the case in Figure 9). Note that the proportion of traders of type i is constant in each region defined in Figure 9. Consequently we can define the marginal distribution of traders of type i as:

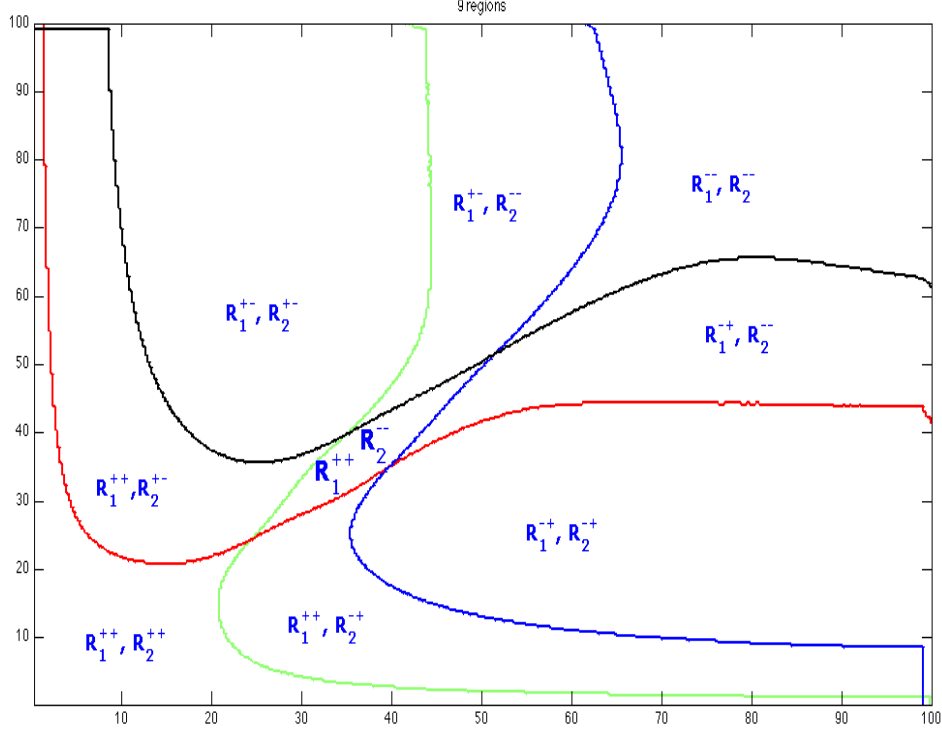$$\hat{m}_i(Q_a, Q_b) := \hat{\gamma}_i(Q_a, Q_b)\hat{m}(Q_a, Q_b).$$

Figure 9: There are nine regions in terms of trader type (HFT vs Institutional Investor) and trader action (LP vs LC)

Now, in each region, there is a certain traded quantity ξ. In table 3 we provide the corresponding values of ξ and γ_i.

	γ_i	ξ
$\mathcal{R}_1 := (R_1^{++}, R_2^{++})$	$\frac{\lambda_i}{\lambda_1 + \lambda_2}$	$(\lambda_1^- q_1/Q_a, \lambda_2^- q_2/Q_a)$
$\mathcal{R}_2 := (R_1^{++}, R_2^{--})$	$\mathbf{1}_{i=1}$	$(\lambda_1^- q_1/Q_a, \Lambda_2 q_2/Q_a)$
$\mathcal{R}_3 := (R_1^{--}, R_2^{--})$	0	-
$\mathcal{R}_4 := (R_1^{++}, R_2^{-+})$	$\mathbf{1}_{i=1}$	$(\lambda_1^- q_1/Q_a, \lambda_2^- q_2/Q_a)$
$\mathcal{R}_5 := (R_1^{-+}, R_2^{+-})$	0	-
$\mathcal{R}_6 := (R_1^{-+}, R_2^{--})$	0	-
$\mathcal{R}_7 := (R_1^{++}, R_2^{+-})$	$\frac{\lambda_i}{\lambda_1 + \lambda_2}$	$(\lambda_1^- q_1/Q_a, \Lambda_2 q_2/Q_a)$
$\mathcal{R}_8 := (R_1^{+-} R_2^{+-})$	$\frac{\lambda_i}{\lambda_1 + \lambda_2}$	$(\Lambda_1 q_1/Q_a, \Lambda_2 q_2/Q_a)$
$\mathcal{R}_9 := (R_1^{+-}, R_2^{--})$	$\mathbf{1}_{i=1}$	$(\Lambda_1 q_1/Q_a, \Lambda_2 q_2/Q_a)$

Table 3: Values of various quantities in each of the 9 regions

The general formula of the average prices are:

• Type's i LC proportion:

$$M_{s,i}^- := \int_{(Q_a, Q_b) \in R_i^{-+} \cup R_i^{--}} \Lambda_i q_i d\hat{m}(Q_a, Q_b) + \int_{(Q_a, Q_b) \in R_i^{++} \cup R_i^{+-}} \lambda_i^- q_i d\hat{m}(Q_a, Q_b)$$

• Type's i LP proportion:

$$M_{s,i}^+ := \sum_{i=1}^9 \int_{(Q_a, Q_b) \in \mathcal{R}_i} < \xi(Q_a, Q_b), (1, 1) > d\hat{m}_i(Q_a, Q_b)$$

• Type's i price for Liquidity Consumer traders:

$$\bar{p}_{s,i}^- := \left(\int_{(Q_a, Q_b) \in R_i^{-+} \cup R_i^{--}} \Lambda_i q_i p_{q_i}^{\text{sell}}(Q_b) d\hat{m}(Q_a, Q_b) + \int_{(Q_a, Q_b) \in R_i^{++} \cup R_i^{+-}} \lambda_i^- q_i p_{q_i}^{\text{sell}}(Q_b) d\hat{m}(Q_a, Q_b) \right) / M_{s,i}^-$$

23

• Type's i price for Liquidity Provider traders:

$$\bar{p}_{s,i}^{+} := \Big(\sum_{i=1}^{9} \int_{(Q_a,Q_b)\in\mathcal{R}_i} < \xi(Q_a,Q_b), (p_{q_1}^{\text{buy}}(Q_a), p_{q_2}^{\text{buy}}(Q_a)) > d\hat{m}_i(Q_a,Q_b) \Big) / M_{s,i}^{+}$$

Finally the average price for sellers can be simply deduced:

$$\bar{p}_{s,i} = \frac{\bar{p}_{s,i}^{-} M_{s,i}^{-} + \bar{p}_{s,i}^{+} M_{s,i}^{+}}{M_{s,i}^{-} + M_{s,i}^{+}}.$$

6.2.2 Numerical tests.

Test 5. Our aim is to model a market opened to HFT, and to observe the effects of the arrival of HFTs. To do so, we consider the following case:

• HFT order sizes are four times smaller than Institutional Investors orders

• All HFT arbitrate between being Liquidity Providers or Liquidity Consumers while half of Institutional Investors are Liquidity Consumers anyway

The selected parameters corresponding to such a market are:

• General parameters : $\delta = 2, P = 100$

• Institutional Investors: $\Lambda_1 = \lambda_1 + \lambda_1^{-} = \frac{1}{2} + \frac{1}{2}$, $q_1 = 1$, $c_1 = 0.25\%$

• HFT: $\Lambda_2 = \lambda_2 + \lambda_2^{-} = 4 + 0$, $q_2 = 0.25, c_2 = 1\%$

Note that we have chosen to set $c_1 q_1 = c_2 q_2$.
Another modeling assumption is that orders are equally split between both types, that is $\Lambda_1 = \Lambda_2$.

As in section 5.2, we can distinguish several cases depending upon the action type of the traders. This lead to 9 distinct regions in the $Q_a \times Q_b$ space, which are depicted in Figure 9. Note that the upper-index denotes the action (+ stands for Liquidity Provider, − for Liquidity Consumer) and the lower-index denotes the type of trader.

We can compare the two following situations: the market stationary equilibrium with a single specie of traders (Institutional Investors), and the market after the arrival of HFTs.
Figure 10 shows the corresponding stationary measure of states (size of ask and bid queues).
We notice that in the case with Institutional Investors only, there are stable liquidity imbalances with two symmetric configurations (one favorable to buyers, and the other one to sellers). On the other hand, in the case with both Institutional Investors & HFT, we observe a liquidity stabilization and a concentration on a single balance equilibrium.

Recall that in the present case with $c_1 q_1 = c_2 q_2$, the value functions of Inst. Inv. and HFT coincide. Consequently the existence of nine regions and the stabilizing effect described above are only explained by market impact heterogeneity. Which is a noteworthy numerical result.

In Table 4 we display the numerical values of the average transaction prices (only for sellers since the prices for buyers are symmetrical, the *fair price* being 20).
We remark that HFT trade at better prices than Institutional Investors and that Institutional Investors average selling price decreases in the market with HFT. Consequently, in this case HFT traders capture the difference, and even more.

	Test 4	Test 5		
	II only	II in the mix	HFT in the mix	mix
Liquidity Consumers	99.849	99.842	99.938	99.89
Liquidity Providers	100.238	100.103	100.189	100.146
Average	99.876	99.852	99.981	99.916
Spread ψ	0.248	0.296	0.038	0.167
Spread ψ (bps)	25	30	4	17

Table 4: Expected transaction prices and spread in the model

The last row shows the impact on the spread. In the framework of our model we define the expected bid-ask spread as the difference $\psi = \mathbb{E}(p^{\text{buy}}) - \mathbb{E}(p^{\text{sell}})$. We conclude that the spread increase for Institutional Investors is 20%, while the global spread decrease of the market is worth 33%. Consequently, the spread reduction clearly profits to HFTs.
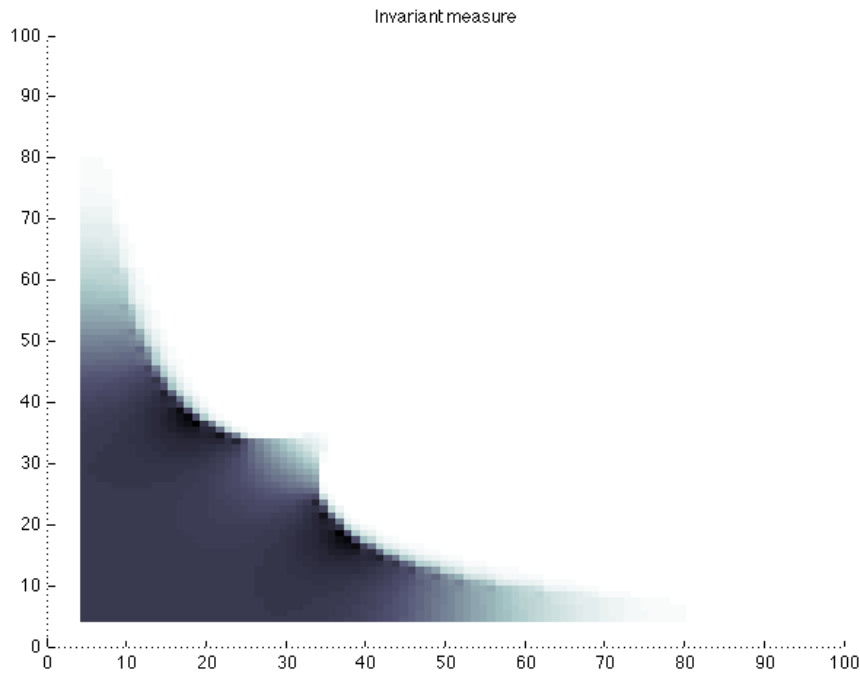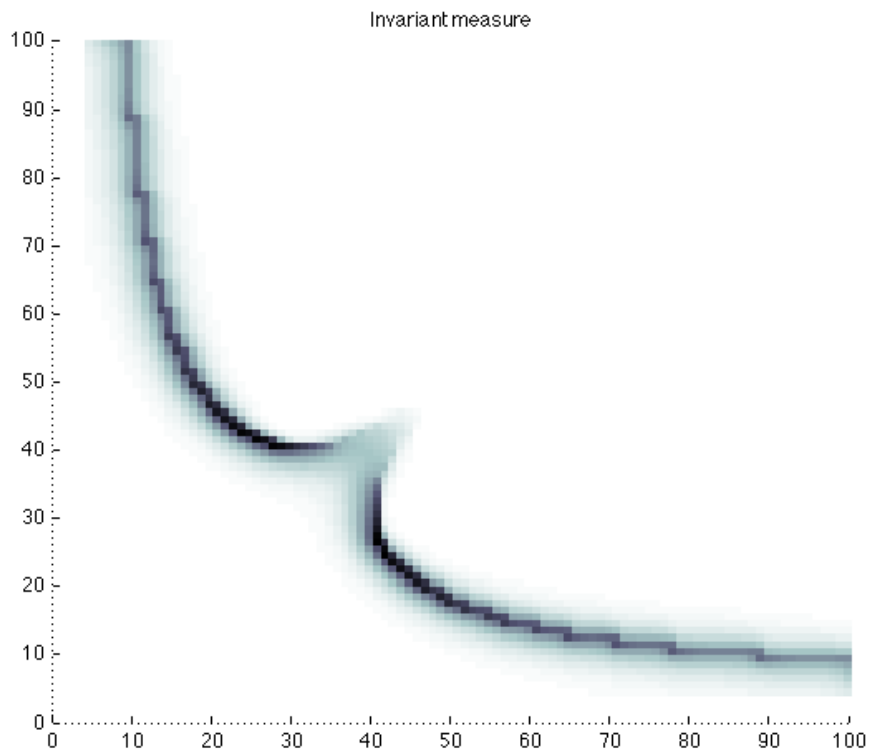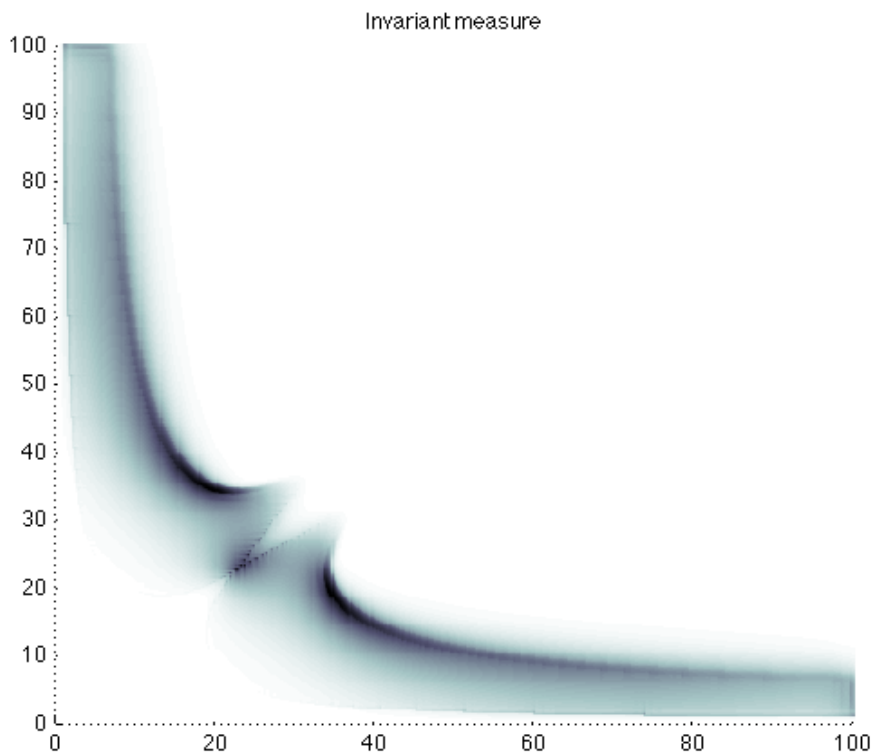
(a) Institutional Investors only
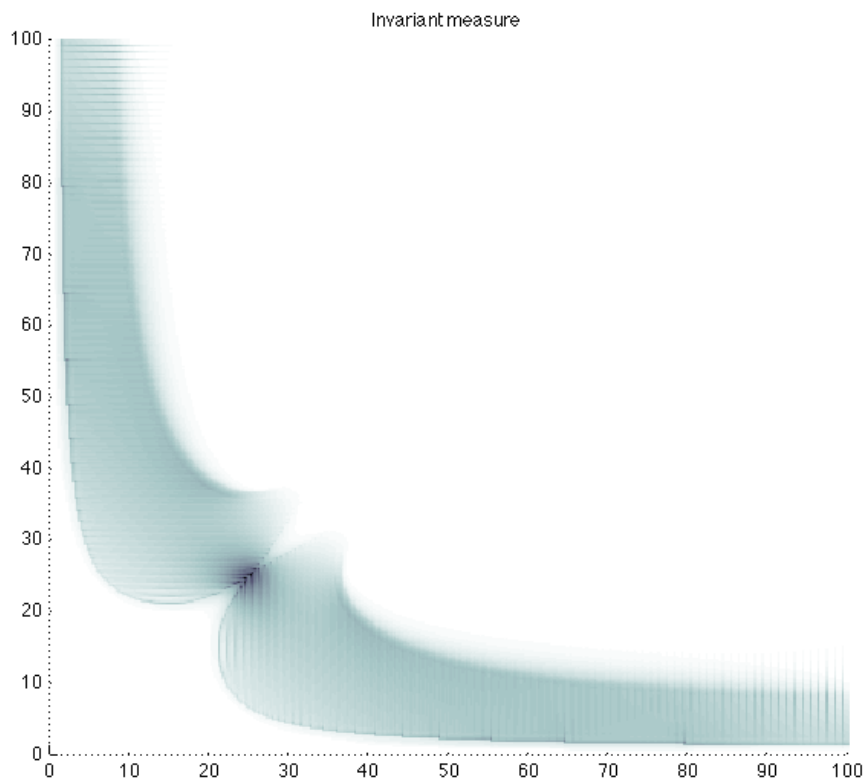
(b) Institutional Investors and HFT

Figure 10: Comparison of invariant measures, test 5

Test 6. For the sake of completeness we end this section with another example, for other proportions of SOR and non-SOR traders. Here we take the same parameters as in Test 1 except that $\Lambda_1 = \lambda_1 + \lambda_1^- = 90\% \times 4 + 10\% \times 4$ and $\Lambda_2 = \lambda_2 + \lambda_2^- = 60\% + 40\%$. Thus we look at a situation where there are 10

(a) Institutional Investors only

(b) Institutional Investors and HFT

Figure 11: Comparison of invariant measures, test 6

points more SOR in the Institutional Investor population and 10 points less SOR amongst HFTs. In Figure

11 we compare the situation between a market with Institutional Investors only and with a mix of both Institutional Investors and HFTs.

Table 5 shows the corresponding quantities.

	Test 4	Test 6		
	II only	II in the mix	HFT in the mix	mix
Liquidity Consumers	99.898	99.854	99.994	99.924
Liquidity Providers	100.168	100.094	100.157	100.125
Average	99.91	99.864	99.974	99.919
Spread ψ	0.180	0.272	0.052	0.162
Spread ψ (bps)	18	27	5	16

Table 5: Expected transaction prices and spread in the model

7 Conclusion

This paper demonstrates how MFG (Mean Filed Games, [Lasry and Lions, 2007]) can be used to model orderbook dynamics. At the junction of structural approaches (see [Roşu, 2009] and [Foucault et al., 2005]) and flow driven ones (see [Gareche et al., 2013]), the mean field games render the strategic behaviour of traders, leading to partial derivative equations that can be numerically solved, and partly reduced to simpler dynamics.

The application presented here used an stylized orderbook model in which:

- each side of the orderbook (buy or sell) is captured by one variable: its size (i.e. the number of orders waiting in each queue);
- the "fair" or "latent" price is stable, since we focus on microstructure effects on the traded price;
- the market impact of a trade is close to linear;
- the orderbook has an infinite resiliency, in the sense that its shape does not change through time: when liquidity is consumed in one queue, its shape readjusts immediately to a linear one (even if its size changes).

The trading strategies of the investors are described by:

- their arrival rate λ, following an homogenous Poisson process;
- the average size of their orders q;
- their waiting cost c: the larger they are, the more impatient the investor;
- a fraction of the orders of investors have an infinite impatience: we call them "non SOR (Smart Order Router) users" since they are not patient enough to follow a liquidity-driven microscopic trading strategy.

Section 4 studies the dynamics of such a model and Section 6.1 shows that such a stylized modelling give birth to realistic dynamics: with one class only of investors, stable states of liquidity imbalance can appear. This can be read as a justification for the introduction of the role of market makers.

In a third step of our reasoning, we introduce HFT (High Frequency Traders) with the hope they will assume this market making role. Consistently with [Menkveld, 2010] and [Brogaard et al., 2012], they are modelled as: fast, using smaller orders than institutional investors, and taking decisions according to the immediate state of the orderbook (in our vocabulary, they are "Smart Order Router" users). It is important to underline that they have not a different impatience (i.e. waiting cost per share) than other investors.

Section 6.2.1 extends the approach developed in Section 4 to a model with two types of investors (to be applied to institutional ones and HFT). We then study numerically the properties of markets with institutional investors and HFTs, looking for an answer to regulators and policy makers questions about the effect of mixing two so different classes of market participants. First note that our results in terms of invariant measure distribution are consistent with data explorations conducted in [Gareche et al., 2013].

Qualitatively, our conclusions are that *the introduction of HFT improves the usual measures of the efficiency of the price formation process*: the stable states of offer and demand are more balanced and the effective bid-ask spread is smaller than without HFTs. But *the observed improvement is at the exclusive advantage of the HFTs*: the effective bid-ask spread paid by institutional investors is largest than before the introduction of HFT. Of course these conclusions are conditioned to the accuracy of our stylized model; nevertheless they can explain the disjunction between the claims of institutional investors (that, for them, the price formation process is more difficult to deal with in presence of HFTs) and the objective improvement of measurements of the state of liquidity since HFT activity increased.

Hence this paper is not only a contribution to the modelling of orderbook dynamics, showing how a MFG-approach can conciliate structural and flow-driven approaches. It provides a qualitative analysis of the role of High Frequency Trading in electronic markets. It also underlines the lack of liquidity measurements adapted to the current market microstructure.

References

[Adlakha et al., 2013] Adlakha, S., Johari, R., and Weintraub, G. Y. (2013). Equilibria of Dynamic Games with Many Players: Existence, Approximation, and Market Structure. *Journal of Economic Theory*, to appear.

[Alfonsi et al., 2009] Alfonsi, A., Fruth, A., and Schied, A. (2009). Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157.

[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.

[Bacry et al., 2012] Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J. F. (2012). Modeling microstructure noise with mutually exciting point processes. *Quantitative Finance (forthcoming)*.

[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.

[Brogaard et al., 2012] Brogaard, J., Baron, M., and Kirilenko, A. (2012). The Trading Profits of High Frequency Traders. In *Market Microstructure: Confronting Many Viewpoints*.

[Burgot et al., 2013] Burgot, R., Laruelle, S., Lasnier, M., Lehalle, C.-A., and Pelin, S. (2013). *Market Microstructure in Practice*. World Scientific publishing.

[Cardaliaguet et al., 2012] Cardaliaguet, P., Lasry, J.-M., Lions, P.-L., and Porretta, A. (2012). Long time average of mean field games. *Networks and heterogeneous media*, 7(2):279–301.

[Carmona et al., 2012] Carmona, R., Delarue, F., and Lachapelle, A. (2012). Control of McKean-Vlasov Dynamics versus Mean Field Games. *Mathematics and Financial Economics*.

[Euronext, 2006] Euronext (2006). EuroNext rules - book I. Technical report, EuroNext.

[Field and Large, 2008] Field, J. and Large, J. (2008). Pro-rata matching and one-tick futures markets. CFS working paper 2008,40, Frankfurt, Main.

[Foucault et al., 2005] Foucault, T., Kadan, O., and Kandel, E. (2005). Limit Order Book as a Market for Liquidity. *Review of Financial Studies*, 18(4).

[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.

[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.

[Gareche et al., 2013] Gareche, A., Disdier, G., Kockelkoren, J., and Bouchaud, J. P. (2013). A Fokker-Planck description for the queue dynamics of large tick stocks.

[Gatheral, 2010] Gatheral, J. (2010). No-Dynamic-Arbitrage and Market Impact. *Quantitative Finance*, 10(7).

[Guéant et al., 2010] Guéant, O., Lasry, J.-M., and Lions, P.-L. (2010). *Paris-Princeton Lectures on Mathematical Finance 2010*, chapter Mean Field Games and Applications. Springer-Verlag.

[Hall et al., 2005] Hall, A., Kofman, P., and Mcculloch, J. (2005). True Spreads Censored by Tick Size. Technical report.

[Ho and Stoll, 1983] Ho, T. and Stoll, H. R. (1983). The dynamics of dealer markets under competition. *Journal of Finance*, 38(4):1053–1074.

[Huang et al., 2007] Huang, M. Y., Caines, P. E., and Malhame, R. P. (2007). Large Population Cost-Coupled LQG Problems with Non-uniform Individuals. *IEEE Trans. on Automatic Control*, 52(9):1560–1571.

[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.

[Lachapelle and Wolfram, 2011] Lachapelle, A. and Wolfram, M.-T. (2011). On a mean field game approach modeling congestion and aversion in pedestrian crowds. *Transportation Research Part B*, 15:10.

[Lasry and Lions, 2007] Lasry, J.-M. and Lions, P.-L. (2007). Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260.

[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.

[Lions, 2012] Lions, P.-L. (2007 - 2012). Jeux à champ moyen et applications. Technical report, Collège de France.

[Lucas and Moll, 2013] Lucas, R. E. and Moll, B. (2013). Knowledge Growth and the Allocation of Time. *Journal of Political Economy*.

[Madhavan, 2011] Madhavan, A. (2011). Exchange-Traded Funds, Market Structure and the Flash Crash. *Social Science Research Network Working Paper Series*.

[Mendelson and Amihud, 1991] Mendelson, H. and Amihud, Y. (1991). *How (Not) to Integrate the European Capital Markets*. Cambridge University Press.

[Menkveld, 2010] Menkveld, A. J. (2010). High Frequency Trading and The New-Market Makers. *Social Science Research Network Working Paper Series*.

[Pagès et al., 2012] Pagès, G., Laruelle, S., and Lehalle, C.-A. (2012). Optimal split of orders across liquidity pools: a stochatic algorithm approach. *SIAM Journal on Financial Mathematics (Forthcoming)*.

[Pequito et al., 2011a] Pequito, S., Aguiar, A. P., Sinopoli, B., and Gomes, D. A. (2011a). Nonlinear estimation using Mean Field Games. *: Network Games, Control and Optimization (NetGCooP), 2011 5th International Conference on*.

[Pequito et al., 2011b] Pequito, S., Aguiar, A. P., Sinopoli, B., and Gomes, D. A. (2011b). Unsupervised Learning of Finite Mixture using Mean Field Games. *9th Allerton Conference on Communication, Control and Computing*.

[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.

[Roşu, 2009] Roşu, I. (2009). A Dynamic Model of the Limit Order Book. *Review of Financial Studies*, 22(11):4601–4641.