

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/278048218>

From Walras' auctioneer to continuous time double auctions: A general dynamic theory of supply and demand

Article in *Journal of Statistical Mechanics Theory and Experiment* · June 2015

DOI: 10.1088/1742-5468/aa4e8e · Source: arXiv

CITATIONS

12

READS

1,875

2 authors:



Jonathan Donier

Sorbonne Université

21 PUBLICATIONS 432 CITATIONS

[SEE PROFILE](#)



Jean-Philippe Bouchaud

680 PUBLICATIONS 39,378 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Fluctuations and amphiphiles [View project](#)



Silica glasses [View project](#)

From Walras' auctioneer to continuous time double auctions

A general dynamic theory of supply and demand

J. Donier^{1,2,*} and J.-P. Bouchaud^{1,3,†}

¹*Capital Fund Management, 23-25 Rue de l'Université, 75007 Paris, France.*

²*Laboratoire de Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie (Paris 6).*

³*CFM-Imperial Institute of Quantitative Finance, Department of Mathematics,
Imperial College, 180 Queen's Gate, London SW7 2RH*

(Dated: June 11, 2015)

In standard Walrasian auctions, the price of a good is defined as the point where the supply and demand curves intersect. Since both curves are generically regular, the response to small perturbations is linearly small. However, a crucial ingredient is absent of the theory, namely transactions themselves. What happens after they occur? To answer the question, we develop a dynamic theory for supply and demand based on agents with heterogeneous beliefs. When the inter-auction time is infinitely long, the Walrasian mechanism is recovered. When transactions are allowed to happen in continuous time, a peculiar property emerges: close to the price, supply and demand vanish quadratically, which we empirically confirm on the Bitcoin. This explains why price impact in financial markets is universally observed to behave as the square root of the excess volume. The consequences are important, as they imply that the very fact of clearing the market makes prices hypersensitive to small fluctuations.

Keywords: Walras' law, Price formation, Supply, Demand, Economic dynamics, Market impact, Liquidity

* jonathan.donier@polytechnique.org

† jean-philippe.bouchaud@cfm.fr

I. INTRODUCTION

One of the most time-worn statement of economic science is that “prices are such that supply matches demand”. In order to explain how this really comes about, one usually invokes a Walras auctioneer, who attempts to measure the supply and demand curves $S(p)$ and $D(p)$, that give the total amount of supply/demand for a given good (or asset), would the price be set to p . The equilibrium price p^* is then such that $D(p^*) = S(p^*)$, which maximizes the amount of good exchanged among agents, given the set of preferences corresponding to the current supply and demand curves Walras (2013). In reality, the full knowledge of $S(p)$ and $D(p)$ is problematic, and Walras envisioned his famous *tâtonnement* process as a mean to observe the supply/demand curves. However, there is a whole aspect of the dynamics of markets that is totally absent in Walras’ framework. While it describes how a pre-existing supply and demand would result in a clearing price, it does not tell us anything about what happens *after* the transaction has taken place. In this sense, the Walrasian price is of very limited scope, since the theory ceases to apply as soon as the price is discovered.

A practical solution to match supply and demand is the so-called “order book” Glosten (1994), Harris et al. (1990), where each agent posts the quantities s/he is willing to buy or sell as a function of the price p . $S(p)$ (resp. $D(p)$) is then the sum of all sell (buy) quantities posted at or above (below) price p . At each time step, the auctioneer can then clear the market by finding the (unique) price such that $D(p^*) = S(p^*)$. This is in fact how most financial markets worked before the advent of electronic matching engines, when market makers played the role of “active” Walrasian auctioneers, in the sense that they would themselves contribute to the order book as to insure orderly trading and stable prices Glosten and Milgrom (1985), Madhavan (2000).

Although close to Walras’ idealization, order book based auctions are still confronted with a fundamental problem: agents do not necessarily reveal their intentions by placing visible orders, for fear of giving away information to the rest of the market – among other reasons Bongiovanni et al. (2006), Handa and Schwartz (1996). It is plausible that only agents with the most urgent need to buy or to sell reveal their intentions. Only close to the transaction price is the order book expected to reveal the true underlying supply and demand curves $S(p)$ and $D(p)$, where they however get intertwined with the orders of market makers/high frequency traders who play strategic “hide and seek” games Bouchaud et al. (2008), Foucault et al. (2013), Handa et al. (2003), Roşu (2006). The visible order book is a sort of *Potemkin village* that reveals only very little about the true underlying supply and demand¹ and whose features strongly depend to the precise design of the market (time priority, pro-rata matching, small or large tick, presence of hidden orders, etc. – see e.g. Kockelkoren (2010)). A direct empirical observations of the dynamics of the full supply and demand curves $S(p)$ and $D(p)$ is therefore difficult (except in particular markets such as Bitcoin, see below and Donier and Bouchaud (2015)). But since the dynamics of prices is essentially governed by that of supply and demand, we need a plausible theoretical framework to model the (unobservable) evolution of the time dependent curves $S(p, t)$ and $D(p, t)$, where t is time, to account for the (observable) evolution of prices. This would allow one to construct a “Walrasian” description of market dynamics, offering a much deeper level of understanding than simply postulating ad-hoc stochastic models for prices, such as the standard (geometric) Brownian motion Bachelier (1900), Black and Scholes (1973).

There are indeed many questions, some of utmost fundamental and practical importance, which cannot be addressed within these stochastic models and require the knowledge of the underlying supply and demand structure and dynamics. One of them is *price impact* Bouchaud (2010), i.e. how much does an additional unconditional buy/sell quantity Q move the price up/down? This is important both for practitioners who want to estimate the costs associated to the impact of their trading strategies Almgren and Chriss (2001), and for regulators who want to understand the stability of markets and the price sensitivity to large “freak” orders (see e.g. Donier and Bouchaud (2015)). It is also

¹ The overwhelming activity market makers/HFT in the visible order book is a distraction from our story here: since their position quickly mean-reverts around zero, they chiefly act as intermediaries (as they should) and only weakly contribute to the supply and demand curves. See the detailed discussion in Donier et al. (2014) and below.

of interest for the general understanding of price discovery and market efficiency: how much noise do “noise traders” Kyle (1985) introduce in markets through their impact on prices? How relevant is market to market accounting? Amihud and Mendelson (1986), Caccioli et al. (2012), etc.

In a Walrasian context, the impact \mathcal{I} of a small buy quantity Q is easily shown to be linear in Q , simply because the slopes of the supply and demand curves around the price p^* (that would prevail for $Q = 0$) are generically non zero. More precisely, writing that $S(p_Q) = D(p_Q) + Q$ and Taylor expanding $S(p)$ and $D(p)$ around p^* to first order in Q , one readily obtains:²

$$S(p^*) + (p_Q - p^*)\partial_p S(p^*) = D(p^*) + (p_Q - p^*)\partial_p D(p^*) + Q \quad \Rightarrow \quad \mathcal{I}(Q) \equiv p_Q - p^* = \lambda Q \quad (1)$$

with $\lambda^{-1} = \partial_p S(p^*) - \partial_p D(p^*) > 0$, since one expects $S(p)$ to be a strictly increasing function of p and $D(p)$ a strictly decreasing function of p (see Fig. 1). Whenever the derivatives of the supply and demand curves do not simultaneously vanish at p^* , the price response to a perturbation must be *linear*. This intuitive result can also be justified using much more elaborate arguments, such as provided by the Kyle model Kyle (1985), where noise traders, market makers and an informed trader interact in the market place. As shown by Kyle, the optimal strategy of market makers is to shift the price linearly in the market imbalance, with the coefficient λ (“Kyle’s lambda”) proportional to the volatility of the asset σ and inversely proportional to the typical volume V traded by the whole market.

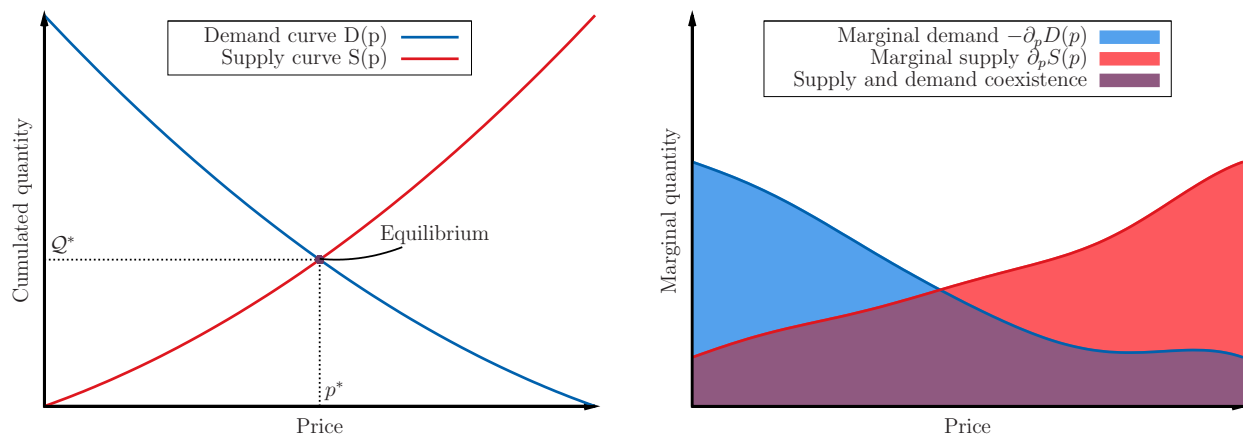


FIG. 1. Left: Illustration of the supply and demand curve (SD), and the resulting price according to Walras’ law. Right: Marginal supply and demand curves (MSD) corresponding to the figure on the left.

Until very recently, the above linear relation between order imbalance and price changes was taken for granted in most academic papers. Remarkably, a series of independent empirical studies of the impact of proprietary orders in financial markets published since the mid-nineties suggests otherwise Almgren et al. (2005), Bacry et al. (2014), Bershova and Rakhlin (2013), Brokmann et al. (2014), Donier and Bonart (2014), Gomes and Waelbroeck (2014), Grinold and Kahn (2000), Mastromatteo et al. (2014a), Moro et al. (2009), Torre and Ferrari (1997), Tóth et al. (2011). All these studies report a *strongly concave* price impact, even in the regime where Q/V is very small (say between 10^{-4} and 0.1). In fact, a simple square-root law

$$\mathcal{I}(Q) = Y\sigma\sqrt{\frac{Q}{V}} \quad (2)$$

where Y is a constant of order unity, accounts surprisingly well for a number of observations, quite independently of the type of markets (stocks, futures, FX, options,...), geographical zones, epochs (pre-2005, before the advent of

² The notation ∂_p means that one takes the derivative with respect to the price p . More generally, ∂_x means the derivative with respect to any variable x .

massive HFT, or post-2005), trading style, etc. The empirical evidence is now so compelling – see in particular the Bitcoin data in Donier and Bonart (2014) – that it is difficult to avoid looking for a consistent theoretical explanation for such a universal non-linear impact law.

There are actually two reasons to believe that usual equilibrium models are not relevant to explain the observed square-root impact. The first one is about orders of magnitude in the Bitcoin case, where the square-root impact is perfectly obeyed for price changes 30 times smaller than transaction fees themselves, and 300 times smaller than the daily volatility – see Donier and Bonart (2014). Imagining that the typical amateur trader on Bitcoin is able to optimize anything with this level of precision seems a total utopia. The second reason is that a square-root impact formally corresponds to $\partial_Q \mathcal{I}(0) = \lambda \rightarrow \infty$, i.e. a situation where $\partial_p S(p^*)$ and $\partial_p D(p^*)$ are both zero. This is a fact that equilibrium models (as in Kyle (1985)) cannot reproduce, as they would always predict a linear impact for small quantities.

A more likely possibility is that this square-root law is not willingly enforced by any market participant but is rather an emergent property. Stepping away from classical ideas, a detailed scenario for the divergence of Kyle’s λ was proposed in Tóth et al. (2011). The main assumption in that paper is that the shape of the supply and demand curves is primarily the result of the interaction between order flow and past transactions themselves. This is in line with the idea that large population of interacting agents, each using heuristic decision rules, can lead to universal emergent behaviour (see e.g. Hommes (2006) for a review of heterogeneous agent models in economics and finance Gualdi et al. (2015) for a recent didactic discussion). This scenario in fact emphasizes the *transient* aspects of market dynamics, which are usually discarded in equilibrium models. The resulting detailed theory, elaborated in Donier et al. (2014), Mastromatteo et al. (2014b), accounts well for the above square-root impact law and for many other empirical observations. In fact, the model presented in Donier et al. (2014), Tóth et al. (2011) and in the present paper is not incompatible with game theoretic approaches as it can be rephrased in the language of the mean field games introduced in Lasry and Lions (2007), see below, Sec. III C.

The aim of the present paper is to revisit and extend the Walrasian theory by proposing a dynamic theory of supply and demand in the light of these recent results. After a brief review of the literature in Sec. II, we propose in Sec. III a dynamical theory for the evolution of the supply and demand curves $S(p, t)$ and $D(p, t)$, including transactions. Our theory only relies on weak, general hypotheses about the behaviour of agents, in particular the assumption of heterogeneous beliefs in a very large population. The Walrasian auctions setting can be seen as a limiting case of our theory, corresponding to an infinitely long time between auctions. We then show (Sec. IV) that as soon as the inter-auction time τ is finite, the impact of small volumes is linear, but with a coefficient λ that diverges as $1/\sqrt{\tau}$ in the $\tau \rightarrow 0$ limit (corresponding to continuous time double auctions) where we recover an exact square-root impact. This reflects the fact that the supply and demand curves both vanish quadratically close to the current price p^* , a property that we validate empirically using Bitcoin data. For small, but finite τ , impact is linear for very small Q s and becomes square-root beyond a crossover value $Q^*(\tau) \propto \tau$ when $\tau \rightarrow 0$. Finally, in Secs. V, VI we discuss some of the conceptual aspects of our framework, as well as some insights concerning market design and market stability, in particular in view of the recent proposals to curb the HFT activity by reintroducing periodic batch auctions Budish et al. (2013), Fricke and Gerig (2014).

II. REVIEW OF THE LITERATURE

The literature on price formation is obviously old and vast. It is divided into two distinct branches: microeconomics and financial economics, with quite different perspectives on the problem. The microeconomic community often attempts to determine how an economy with a pre-defined set of agents and preferences produces equilibrium prices, and study their properties (uniqueness, stability, computability, convergence etc.). This is mostly a static view, whereas financial economists are mostly interested in the dynamics of these prices. However, the assumption that markets are instantaneously arbitrated and efficient imposes (semi-)martingale properties for the price, the dynamics

of which is entirely driven by *news*, subsuming all knowledge of the actual dynamics of supply and demand itself. The quantities of interest are then the volatility of prices, the distribution of returns, etc., as well as the micro-structural properties of the immediate supply and demand visible in the order book, which however corresponds to an infinitesimal fraction of the total supply and demand (see below). The aim of this section is to give a (rough) overview of these two different approaches and position the present work with respect to both of them.

A. Theory of supply and demand in economics

As stated above, the question of how prices emerge from supply and demand has fuelled more than a century of economic research, based on the assertion that prices are such that supply equals demand for every asset in the economy. The immediate questions that arise are whether this equilibrium exists, is unique and is stable (in some sense) Mas-Colell et al. (1995). Whereas these questions are rather subtle in a multi-asset economy Hicks (1946), Samuelson (1983), they become trivial in the case of a single asset economy as soon as the supply and demand curves are strictly monotonous. We will only consider a single asset economy in the present paper, since it fully suffices for our purpose and allows one to better focus on the essential part of our message, as we skip some of the usual problems that arise in multi-dimensional settings (is the equilibrium unique, stable, computable, etc.).

As many have noted, such a static description of prices is not fully satisfying, and a description of the dynamics of prices would be highly valuable. However, several interpretations of what “dynamics” actually means can be found in the literature.

1. Dynamics might refer to the way prices converge towards equilibrium. To address this point and the unrealistic fact that the Walrasian mechanism does not allow agents to trade until the equilibrium is reached, some economists have introduced the concept of *non-tâtonnement* in which agents are allowed to trade before the equilibrium has been reached Fisher (1989). Whether such convergence dynamics, even in the presence of trading, should be identified to the dynamics of market price itself, is far from obvious. In fact, as mentioned in Mas-Colell et al. (1995), such a model should be thought of *not as modelling the evolution of a supply-and-demand driven economy, but rather as a tentative trial-and-error process taking place in fictional time [...] [to find] the equilibrium level of prices*. We are thus speaking here of a *transient* dynamics in an otherwise stable world.
2. A second way of introducing dynamics is to consider a multiple period economy in which the supply and demand may evolve at each period, resulting in a new price (see for example Mankiw (2014), Ch. 14). This collection of static equilibria is a rather weak notion of dynamics, that is closer to a *quasi-static* evolution without transactions, in which the price is always the outcome of an equilibrium supply and an equilibrium demand. Hicks (1946) noted that *we shall find [...] that there is a way of reducing the dynamic problem into terms where it becomes formally identical with that of statics*, showing that something important is somehow missing. As we shall indeed argue below, this figment misses the essential point that when transactions occur, supply and demand curves are both immediately depleted, thus affecting subsequent transactions. Only when the time between market clearing auctions is large enough can the supply and demand curve again be considered in an equilibrium state prior to any further transaction.
3. Following up on the last remark, our view is that a complete dynamic description must account for the evolution of supply and demand in an ever changing world, both in-between auctions/transactions and right when the auction takes place. This appears as a necessity if one wants to understand the formation of real prices, in particular in financial markets where supply and demand permanently interact, and where transactions prevent the supply and demand curves from being in an equilibrium at any point in time. To our knowledge, whereas many papers have worked in the direction of understanding price formation under continuous double auction Biais (1993), Cason and Friedman (1996), Easley and Ledyard (1993), Gjerstad and Dickhaut (1998), such a

general dynamical theory of supply and demand is not available at this stage. This is what the present paper aims at achieving.

As we shall see below, our proposal is at odds with classical approaches, which usually consider “rigid” supply and demand curves, that shift uniformly with respect to each other (see e.g. Mankiw (2014), Ch. 14). In contrast, we propose a partial derivative equation that describes the evolution and deformation of the full price dependent supply and demand curves as the core ingredient of our model.

B. Financial economics: The Kyle model

One strategy to model financial markets that has been popular since Kyle’s seminal paper Kyle (1985) is to consider markets as one- or several-period(s) equilibrium(a) between two or more (representative) agents, each representing a well-identified trading behaviour. In Kyle’s original paper, an informed trader competes with a market maker who provides liquidity for every trade, in the presence of a noise trader who trades at random. This modelling strategy has been followed in many subsequent papers, where a small number of rational agents optimize a given utility function and maintain an ecological equilibrium.

Whereas the results of such models often yield useful qualitative intuition, they miss – in our opinion – two essential features of markets. First, trading occurs in continuous time and reasoning in terms of periods (e.g. one period per trading day) is not appropriate in that respect. Second, the number of (representative) agents is usually pre-defined and small; typically, three strategies in Kyle’s model. This has to be contrasted with real markets where agents are strongly heterogeneous, and the very idea of representative agents dubious. As a result, these equilibrium models do not reproduce some essential market features such as the square-root impact law and therefore probably miss some fundamental aspects of price formation.

At variance with these usual models, the present paper suggests that relevant agent-based modelling should incorporate three essential features: (i) a one-dimensional definition of the price dynamics *via* the order book, (ii) continuous time and (iii) heterogeneous agents. Based on these three ingredients, we define and solve below a particular tractable class of models that appears to capture faithfully some essential feature of price formation in continuous double-auction markets. But we believe that our results would hold for a much wider class of models based on the same ingredients.

C. Financial economics: Models of the limit order book

Instead of the above “macroscopic” considerations on the (unobservable) supply and demand, a whole branch of financial mathematics (concerned with “market microstructure”) has recently emerged. The focus is on the actual evolution of the limit order book and of price formation on financial markets. The limit order book is described as a queueing system, in which buyers (resp. sellers) post quantities on a discrete price grid (the elementary price change is called the “tick”) and wait for being executed. The buyers (resp. sellers) that offer the best price are then executed by aggressive sellers (resp. buyers) according to a *first in, first out* policy. When the whole volume present on the best ask/bid queue is executed, then either some sell/buy volume replenishes the queue and the price is unchanged, or the queue is replenished by opposite side traders and the price moves by one tick. An obvious motivation for such research is to give practical answers to many questions from the financial industry (concerning optimal market making, optimal execution, optimal trading, etc. Cartea and Jaimungal (2014), Cont and Kukanov (2013)) as well as from regulators (tick size, market ecology, market design and stability etc. Laruelle et al. (2011), Robert and Rosenbaum (2011)). Much effort has been devoted to understand and model the mechanics of the limit order books, how it is affected by market design and the ecology of traders (in particular High Frequency Trading), and how it relates to macroscopic variables such as price volatility, etc Foucault et al. (2013), Handa et al. (2003), Hasbrouck

(2006).

The availability of detailed data where all market events are recorded (i.e. trades, quotes, cancellations, etc.) has generated a flurry of empirical papers, describing many aspects of price formation at the microstructural level for a review see Biais et al. (1995), Bouchaud et al. (2008). Correspondingly, a host of stylized models of the order book have appeared, with different starting points and objectives. For example, “zero-intelligence” models Bouchaud et al. (2002), Cont et al. (2010), Farmer et al. (2005), Gareche et al. (2013), Smith et al. (2003) form an important class of models of the order book, where one assumed that agents act mechanically (rather than strategically) leading to simple Poissonian statistics for the order flow. Although obviously too simple to account for what goes on in financial markets, such models reveal some interesting relationships between observables (spreads, volatility, activity, etc.) Cont and De Larrard (2013), Farmer et al. (2005). Much more elaborate models have also been developed, taking into account the heterogeneity, strategies and preferences of market participants Foucault (1999), Lachapelle et al. (2013), Maglaras and Moallemi (2011), Roşu (2014), Roşu (2009), some including the queues behind the best buy/sell prices Huang et al. (2014).

The present paper is clearly partly inspired by the above strand of papers on real limit order books, in particular Bouchaud et al. (2002), Smith et al. (2003). However, we depart from these models on one very fundamental issue. Instead of trying to describe the evolution of the *visible* order book (where only a tiny fraction of the outstanding liquidity is revealed, and whose dynamics is dominated by highly strategic market-makers/HFT), we want to describe the much deeper and much slower “latent” order book, introduced in Tóth et al. (2011), that contains all buy/sell *intentions*, whether displayed or not by market participants. In other words, we model the true underlying supply and demand curves that would materialize if the transaction price was to move closer to the reservation prices. The distinction between the visible limit order book (which, as stated above, gives a very poor indication on liquidity at larger scales) and the true supply and demand curves is absolutely crucial for all that follows. The model described below is a generalisation of the ideas introduced in Mastromatteo et al. (2014a), Tóth et al. (2011) and in Donier et al. (2014). It builds upon the intuition that agents can revise their reservation prices in an heterogeneous manner, introduced long ago in Bak et al. (1997) and recently revisited in completely different contexts in Lasry and Lions (2007), Lehalle et al. (2011) and in Tóth et al. (2011). The motivation of the latter paper was to explain the universal concave (“square-root”) impact of directional trade sequences mentioned in the introduction, that deeply challenges standard equilibrium models.

In summary, the aim of the present paper is to reconcile the insights gained by the financial literature on price formation with a more Walrasian view of supply and demand that provides us with a macroscopic theory of price formation. We believe that this reconciliation has important conceptual consequences from an economic perspective (in particular in emphasizing the *dynamical aspects* of price formation and liquidity), as well as practical implications for market design and regulation (in particular concerning the crucial issue of market stability).

III. A DYNAMIC THEORY OF THE SUPPLY & DEMAND CURVES

A. Definitions

The classical supply and demand curves $S(p, t)$ and $D(p, t)$ (SD) represent respectively the amount of supply and demand that would reveal themselves if the price were to be set to p at time t . In classical Walrasian auctions, the equilibrium price p_t^* is then set to the value that matches both quantities so that $D(p_t^*, t) = S(p_t^*, t)$. This equilibrium is unique provided the curves are strictly monotonous³. The supply and demand curves, as well as the resulting equilibrium price, are represented on Fig. 1 (left).

³ By definition, or simply by common sense, the demand curve is a decreasing function of the price whereas the supply curve is increasing.

In order to define the dynamics of the supply and demand curves, we also introduce the *marginal supply and demand curves* (MSD), on which we will focus in the rest of this paper. They are defined as the derivative of the SD curves

$$\begin{aligned}\rho_S(p, t) &= \partial_p S(p, t) \geq 0; \\ \rho_D(p, t) &= -\partial_p D(p, t) \geq 0,\end{aligned}\tag{3}$$

with the following interpretation: For any price p , $\rho_S(p, t)dp$ (resp. $\rho_D(p, t)dp$) is, at time t , the quantity of supply (resp. demand) that would materialize if the price changed from p to $p + dp$ (resp. $p - dp$). The MSD curves can thus be seen as the density of supply and demand intentions in the vicinity of a given price. Fig. 1 (right) shows MSD curves corresponding to the SD curves: Higher MSD levels correspond to larger slopes for the SD curves.

In the Walrasian story, supply and demand pre-exist and the Walrasian auctioneer gropes (*tâtonne*) to find the price p_t^* that maximizes the amount of possible transactions. The auction then takes place at time t and removes instantly all matched orders. Assuming that all the supply and demand intentions close to the transaction price were revealed before the auction and were matched, the state of the MSD just after the auction is simple to describe, see Figs. 1 & 7:

$$\begin{cases} \rho_S(p, t^+) = \rho_S(p, t^-) & (p > p_t^*) \\ = 0 & (p \leq p_t^*) \\ \rho_D(p, t^+) = \rho_D(p, t^-) & (p < p_t^*) \\ = 0 & (p \geq p_t^*). \end{cases}\tag{4}$$

But what happens next, once the auction has been settled? So far the story does not tell (to the best of our knowledge). The aim of the following is to set up a general framework for the dynamics of the supply and demand curves. This will allow us to describe, among other questions, how the supply and demand curves evolve from the truncated shape given by Eq. (4) up to the next auction at time $t + \tau$ (where τ is the inter-auction time).

B. General hypotheses about the behaviour of agents

The theory that we present here relies on weak and general assumptions on agents behaviours that translate into a simple and universal evolution of the MSD curves, with only very few parameters⁴. The MSD curves aggregate the intentions of all agents, which would materialize in the “real” order book if it was not for fear of being picked off by more informed traders, or of revealing some information to the market. This is why the MSD curves were called the “latent” order book (LOB) in Refs. Donier et al. (2014), Mastromatteo et al. (2014a,b), as initially proposed in Tóth et al. (2011).

We will assume that there is a so-called “fundamental” price process \hat{p}_t which is only partially known to agents, in a sense clarified below (see Sec. V). For simplicity, we will also posit that \hat{p}_t is an additive Brownian motion. In the absence of transactions, the MSD curves evolve according to three distinct mechanisms, that we model as follows:

- New intentions, not present in the supply and demand before time t , can appear. The probability for new buy/sell intentions to appear between t and $t + dt$ and between prices p and $p + dp$ is chosen to be $\omega_{\pm}(p - \hat{p}_t)$, where $\omega_+(x)$ is a decreasing function of x and $\omega_-(x)$ is an increasing function of x .
- Already existing intentions to buy/sell at price p can simply be cancelled and disappear from the supply and demand curves. The probability for an existing buy/sell intention around price p to disappear between t and $t + dt$ is chosen to be $\nu_{\pm}(p - \hat{p}_t)$.

⁴ In fact, as shown in Donier et al. (2014) and below, only two parameters suffice to describe the problem in the vicinity of the price: one is the price volatility, and the other one is related to market activity (traded volume per unit time).

- Already existing intentions to buy/sell at price p can be revised. Between t and $t+dt$, each agent i revises his/her reservation price p^i to $p^i + \beta^i d\xi_t + dW_{i,t}$, where $d\xi_t$ is common to all i , representing some public information. β^i is the sensitivity of agent i to the news, which we imagine to be a random variable from agent to agent, with a mean normalized to 1. Some agents may over-react ($\beta^i > 1$), others under-react ($\beta^i < 1$). The idiosyncratic contribution $dW_{i,t}$ is an independent Wiener noise both across different agents and in time, with distribution of mean zero⁵ and variance $\Sigma_i^2 dt$, that may depend on the agent (some agents might be more “noisy” than others).

We will furthermore assume that the “news” term $d\xi_t$ is a Wiener noise of variance $\sigma^2 dt$, corresponding to a Brownian motion for the fundamental price $\hat{p}_t = \int^t d\xi_{t'}$ with volatility σ . Normalising the mean of the β^i ’s to unity thus corresponds to the assumption that agents are on average unbiased in their interpretation of the news – i.e. their intentions remain centred around the fundamental price \hat{p}_t in the course of time – but see the expanded discussion of this point in Sec. V.

Our central assumptions are *heterogeneity*, together with the hypothesis that idiosyncratic behaviours “average out” in the limit of a very large number of participants, i.e., no single agent accounts for a finite fraction of the total supply or demand. While not strictly necessary, this assumption leads to a deterministic aggregate behaviour and allows one to gloss over some rather involved mathematics.

C. The model in terms of optimizing agents

The above assumptions might appear obscure to those used to think in terms of rational optimizing agents and equilibria. Here we rephrase these assumptions in a language closer to standard economic intuition.

We consider an *open* economic system, in which many heterogeneous, infinitesimal agents operate. Each agent i has a certain utility $\mathcal{U}_i(p, \theta | \hat{p}_t^i, \mathcal{F}_t)$ for buying ($\theta = +1$) or selling ($\theta = -1$) a unit (small) quantity at price p , given his/her estimate of the fundamental price \hat{p}_t^i and all the information about the rest of the world, available at time t , encoded in \mathcal{F}_t . The third option available to agent i is to be inactive ($\theta = 0$), in which case the number of goods s/he owns remains constant. Agents are heterogeneous in the sense that both their utility function and their estimates of the fundamental price are different; one can think of them as random members of some adequate statistical ensembles. For the sake of simplicity, we consider no interest rate and no risk of any kind.

At time t , each agent computes his optimal action p_t^i, θ_t^i as the result of the following optimisation program:

$$(p_t^i, \theta_t^i) = \underset{p, \theta}{\operatorname{argmin}} \mathcal{U}_i(p, \theta | \hat{p}_t^i, \mathcal{F}_t). \quad (5)$$

Because of the random evolution of the outside world summarized by $\hat{p}_t^i, \mathcal{F}_t$, the value of $\theta_t^i \in \{-1, 0, +1\}$ can change between t and $t + dt$. For the sake of simplicity, we assume that the change of the state of the world in time dt is never so large as to induce direct transitions from $\mp 1 \rightarrow \pm 1$ without pausing at 0. Hence, between t and $t + dt$, the following transitions (or absence thereof) are possible:

- $0 \rightarrow 0$: this clearly induces no change in the MSD curves;
- $0 \rightarrow \pm 1$: in this case, agent i previously absent from the market becomes either a buyer or a seller, with reservation price p_t^i given by Eq. (5). The assumption that agents are heterogeneous translates in a model where this event is a Poisson process with some arrival rate $\omega_{\pm}(p)$;
- $\pm 1 \rightarrow 0$: in this case, agent i previously present in the market as a buyer or a seller, decides to become neutral, which is modelled as a Poisson process with some cancellation rate $\nu_{\pm}(p)$;

⁵ One could generalize the calculations below to the case where the mean is non zero (modelling for example the tendency of agents to revise their reservation price in the direction of the traded price). This would affect none of the conclusions below, at least in the limit where the inter-auction time τ becomes very small.

- $\pm 1 \rightarrow \pm 1$: in this case, a buyer/seller remains a buyer/seller, but may change his/her reservation price because the solution of Eq. (5) has changed. Writing $p_t^i = f_i(\hat{p}_t^i, t)$, where f is a regular function if \mathcal{U}_i is regular enough, and applying Itô's lemma, one finds:

$$\begin{aligned} dp_t^i &= \frac{\partial f_i}{\partial t} dt + \frac{\partial f_i}{\partial p} d\hat{p}_t^i + \frac{\sigma_i^2}{2} \frac{\partial^2 f_i}{\partial p^2} dt \\ &= \alpha_t^i d\hat{p}_t^i + \gamma_t^i dt. \end{aligned} \quad (6)$$

The drift term γ_t^i will play little role in the following (see previous footnote), and we neglect it henceforth. In order to recover the specification of the above section, we further decompose the price revision $dp_t^i = \alpha_t^i d\hat{p}_t^i$ into a *common* component $\beta^i d\xi_t$ and an *idiosyncratic* component $dW_{i,t}$ as above.

Therefore, the mechanism proposed in the above section indeed describes the behaviour of an open system of *infinitesimal* and *heterogeneous* market participants. Note that we do not need to distinguish between fundamental investors, noise traders and market makers, as for example in the Kyle model Kyle (1985). This is due to our assumption that the market contains a very large number of participants, in which case the MSD curves are continuous. Discretization effects (in price and in quantity) would open gaps in the MSD curves, and specific market makers would then be needed to ensure continuous, orderly trading. Finally, and quite importantly, the price dynamics in the above setting is arbitrage free (see Donier et al. (2014)). There is therefore no optimal strategic component that is missing from the above utility maximisation program.

D. The “free evolution” equation for the MSD curves

Endowed with the above hypothesis, one can derive stochastic partial differential equations for the evolution of the marginal supply ($\rho_S(p, t) = \partial_p S(p, t)$) and the marginal demand ($\rho_D(p, t) = -\partial_p D(p, t)$) in the absence of transactions Donier et al. (2014). It turns out that, as expected, these equations take a simpler form in the reference frame of the (moving) fundamental price \hat{p}_t . Introducing the shifted price $y = p - \hat{p}_t$, one finds Donier et al. (2014):⁶

$$\begin{aligned} \partial_t \rho_D(y, t) &= \mathcal{D} \partial_{yy}^2 \rho_D(y, t) - \nu_+(y) \rho_D(y, t) + \omega_+(y); \\ \partial_t \rho_S(y, t) &= \underbrace{\mathcal{D} \partial_{yy}^2 \rho_S(y, t)}_{\text{Updates}} - \underbrace{\nu_-(y) \rho_S(y, t)}_{\text{Cancellations}} + \underbrace{\omega_-(y)}_{\text{New orders}}, \end{aligned} \quad (7)$$

where $\mathcal{D} = \frac{1}{2}[\mathbb{E}_i(\Sigma_i^2) + \sigma^2 \text{Var}(\beta^i)]$, i.e. part of the diffusion term comes from the purely idiosyncratic “noisy” updates of agents ($\mathbb{E}_i(\Sigma_i^2)$), and another part comes from the inhomogeneity of their reaction to news ($\sigma^2 \text{Var}(\beta^i)$), which indeed vanishes if all β^i 's are equal to unity.⁷

These equations, that are at the core of the present paper, describe the structural evolution of supply and demand around the fundamental price \hat{p}_t . Notice however that \hat{p}_t has disappeared from the above equations. The dynamics of the MSD curves can be treated independently from the dynamics of the price itself, provided one describes the MSD in the reference frame of the price. There is however a direct relationship between the price volatility σ and the diffusion coefficient \mathcal{D} , as expressed above and noted in Donier et al. (2014).

Interestingly, whereas the price is random and follows a rough path (typically a Brownian motion), the structural part is deterministic and smooth, thanks to the assumption of “infinitesimal” orders (that can be made rigorous by considering an appropriate scaling for system parameters that corresponds to a hydrodynamic limit, see Gao et al. (2014)).

⁶ See also Lasry and Lions (2007), Lehalle et al. (2011) for similar ideas in the context of mean-field games. Note that Eq. (7) is strictly valid when $\rho_S(p, t)$ and $\rho_D(p, t)$ are interpreted as the marginal supply and demand curves averaged over the noise processes. Otherwise some noisy component remains, see e.g. Dean (1996).

⁷ Here we neglect the possibility that buyers and sellers update their price differently, but one could make a distinction between a \mathcal{D}_+ and a \mathcal{D}_- , or even make \mathcal{D} price/time dependent.

The above equations for $\rho_D(y, t)$ and $\rho_S(y, t)$ are linear and can be formally solved in the general case, starting from an arbitrary initial condition such as Eq. (4), using a spectral decomposition of the evolution operator. This general solution is however not very illuminating, and we rather focus here in the special case where $\nu_{\pm}(y) \equiv \nu$ does not depend on y nor on the side of the latent order book. The general solution can then be written in a fairly transparent way, as:

$$\rho_{S,D}(y, t) = \int_{-\infty}^{+\infty} \frac{dy'}{\sqrt{4\pi\mathcal{D}t}} \rho_{S,D}(y', t=0^+) e^{-\frac{(y'-y)^2}{4\mathcal{D}t} - \nu t} + \int_0^t dt' \int_{-\infty}^{+\infty} \frac{dy'}{\sqrt{4\pi\mathcal{D}(t-t')}} \omega_{\pm}(y') e^{-\frac{(y'-y)^2}{4\mathcal{D}(t-t')} - \nu(t-t')}, \quad (8)$$

where $\rho_{S,D}(y, t=0^+)$ is the initial condition, i.e. just after the last auction.

We will now explore the properties of the above solution at time $t = \tau^-$, i.e., just before the next auction, in the two asymptotic limits $\tau \rightarrow \infty$, corresponding to very infrequent auctions, and $\tau \rightarrow 0$, corresponding to continuous time auctions.

IV. DISCRETE AUCTIONS AND PRICE IMPACT

The aim of this section is to show that the shape of the marginal supply and demand curves can be fully characterized in the limit of very infrequent auctions (corresponding to Walras' auctions) and in the opposite limit of nearly continuous time auctions (corresponding to financial markets), and describe the transition between the two limits. The upshot is that while the liquidity around the auction price is in general finite and leads to a linear impact using the standard argument in Eq. (1) above, this liquidity vanishes as $\sqrt{\tau}$ when the inter-auction time $\tau \rightarrow 0$. This signals the breakdown of linear impact and, as shown at the end of the section, its replacement by the square-root law mentioned in the introduction.

A. Walras, or the limit of infrequent auctions

Letting $t = \tau \rightarrow \infty$ in the above Eq. (8), one immediately sees that the first term disappears, meaning that one reaches a *stationary solution* $\rho_{S,D}^{\text{st.}}(y)$ that is independent of the initial condition. The second term can be simplified further to give the following general solution:

$$\rho_{S,D}^{\text{st.}}(y) = \frac{1}{2\sqrt{\nu\mathcal{D}}} \int_{-\infty}^{+\infty} dy' \omega_{\pm}(y') e^{-\sqrt{\frac{\nu}{\mathcal{D}}}|y'-y|}. \quad (9)$$

A particularly simple case is when $\omega_{\pm}(y) = \Omega_{\pm} e^{\mp\mu y}$, meaning that buyers(/sellers) have an exponentially small probability to be interested in a transaction at high/low prices. In this toy-example, one readily finds that a stationary state only exists when $\nu > \mathcal{D}\mu^2$ and reads:

$$\rho_{S,D}^{\text{st.}}(y) = \frac{\Omega_{\pm}}{\nu - \mathcal{D}\mu^2} e^{\mp\mu y}. \quad (10)$$

Other forms for $\omega_{\pm}(y)$ can be investigated as well, for example $\omega_{\pm}(y) = \omega_{\pm}^0 \mathbb{1}_{\{y < > 0\}}$ which yields:

$$\rho_{S,D}^{\text{st.}}(y) = \frac{\omega_{\pm}^0}{2\nu} \left[1 \pm \text{sign}(y)(1 - e^{-\sqrt{\nu/\mathcal{D}}|y|}) \right], \quad (11)$$

that we will use in Figs. 3 and 4. The shape of $\rho_{S,D}^{\text{st.}}(y)$ is generically the one shown in Fig. 1 with an overlapping region where buy/sell orders coexist. The auction price $p_{\tau}^* = \hat{p}_{\tau} + y^*$ is determined by the condition $D(p_{\tau}^*, \tau^-) = S(p_{\tau}^*, \tau^-)$, or else:

$$\int_{y^*}^{\infty} dy \rho_D^{\text{st.}}(y) = \int_{-\infty}^{y^*} dy \rho_S^{\text{st.}}(y) \equiv v^*, \quad (12)$$

where v^* is the volume exchanged during the auction. For the simple exponential case above, this equation can be readily solved as:

$$y^* = \frac{1}{2\mu} \ln \frac{\Omega_+}{\Omega_-}, \quad (13)$$

with a clear interpretation: if the new buy order intentions accumulated since the last auction happen to outsize the new sell intentions during the same period, the auction price will exceed the fundamental price, and vice-versa. This pricing error is expected to be small if the order book is observable during the inter-auction period, since in that case Ω_+ and Ω_- will track each other and remain close. Otherwise, one expects the imbalance to invert in the next period, leading to a kind of “bid-ask bounce” well known in the context of market microstructure. One can also compute the volume exchanged during the auction v^* . One finds:

$$v^* = \frac{\sqrt{\Omega_+ \Omega_-}}{\mu(\nu - \mathcal{D}\mu^2)}. \quad (14)$$

Just after the auction, the MSD curves start again from $\rho_{S,D}^{\text{st.}}(y)$, truncated below (resp. above) y^* , as in Eq. (4).

Let us now turn to price impact in this model. From Eq. (9), it is immediate that for any clearing price y^* , both $\rho_S^{\text{st.}}(y^*)$ and $\rho_D^{\text{st.}}(y^*)$ are strictly positive. This would remain true even if the dependence on y of cancellation rate $\nu_{\pm}(y)$ was reinstalled. The general argument given in the introduction therefore predicts a *linear impact* for an extra buy/sell quantity given by:

$$\mathcal{I}(Q) = \pm \lambda Q; \quad \lambda = \frac{1}{\rho_S^{\text{st.}}(y^*) + \rho_D^{\text{st.}}(y^*)}. \quad (15)$$

For the exponential case, this again takes a simple form:

$$\lambda = \frac{\nu - \mathcal{D}\mu^2}{2\sqrt{\Omega_+ \Omega_-}}, \quad (16)$$

whereas for a general symmetric order flow $\omega_+(y) = \omega_-(-y)$, y^* is obviously equal to zero, leading to:

$$\lambda = \frac{\sqrt{\nu \mathcal{D}}}{\int_{-\infty}^{+\infty} dy' \omega(y') e^{-\sqrt{\frac{\nu}{\mathcal{D}}} |y'|}}. \quad (17)$$

For $\omega_{\pm}(y) = \omega^0 \mathbb{1}_{\{y < > 0\}}$, one obtains the simple and intuitive result:

$$\lambda = \frac{\nu}{\omega_0}, \quad (18)$$

i.e. that the market liquidity, measured by λ^{-1} , grows linearly with the rate of incoming orders and inversely proportionally to the cancellation rate.

The main point of the present section is that when the inter-auction time is large enough, each auction clears an equilibrium supply with an equilibrium demand, with very simple and predictable outcomes. This corresponds to the quasi-static dynamics discussed in item 2., Section II A, and to the standard representation of market dynamics in the Walrasian context, since in this case only the long-term properties of supply and demand matter and the whole transients are discarded. The next section will depart from this limiting case, by introducing a finite inter-auction time such that the transient dynamics of supply and demand becomes a central feature in the theory.

B. High frequency auctions

We will now investigate the alternative limit where the inter-auction time τ tends to zero. Since all the supply (resp. demand) curve left (resp. right) of the auction price is wiped out by the auction process, one expects intuitively

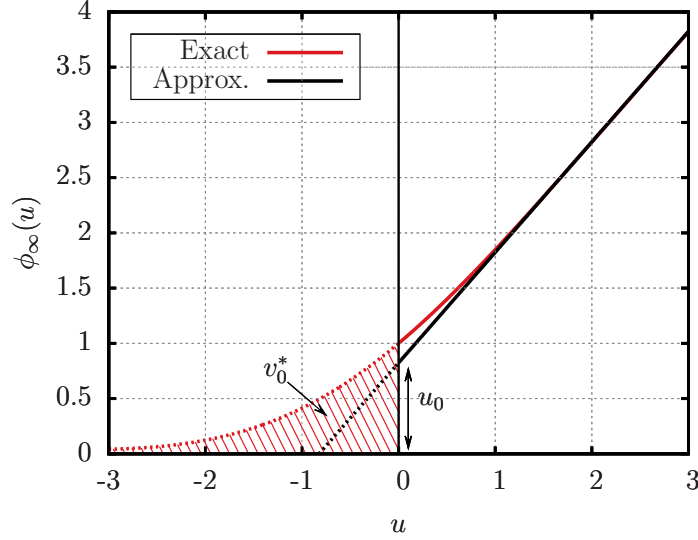


FIG. 2. Graph of the normalized exact solution $\phi_\infty(u)$, and its affine approximation. The whole picture must be rescaled by a factor $\sqrt{\tau}$ to recover the order book when the inter-auction time is τ . The hatched region corresponds to the volume to be executed, and therefore scales with τ .

that after a very small time τ , the density of buy/sell orders in the immediate vicinity of the transaction price will remain small. We will show that this is indeed the case, and specify exactly the shape of the stationary MSD after many auctions have taken place. Consider again Eq. (8) just before the $n+1$ th auction at time $(n+1)\tau^-$, in the case where the flow of new orders is symmetric, i.e. $\omega_+(y) = \omega_-(y)$, such that the transaction price is always at the fundamental price ($y^* = 0$). We will focus on the supply side and postulate that $\rho_S(y, t = n\tau^-)$ can be written, in the vicinity of $y = 0$, as ⁸:

$$\rho_S(y, t = n\tau^-) = \sqrt{\tau} \phi_n \left(\frac{y}{\sqrt{\mathcal{D}\tau}} \right) + O(\tau) \quad (19)$$

when $\tau \rightarrow 0$ (and symmetrically for the demand side). Plugging this ansatz into Eq. (8), making the change of variable $y' \rightarrow \sqrt{\mathcal{D}\tau}w$ and taking the limit $\tau \rightarrow 0$ leads to the following iteration equation, exact up to order $\sqrt{\tau}$:⁹

$$\phi_{n+1}(u) = \int_0^{+\infty} \frac{dw}{\sqrt{4\pi}} \phi_n(w) e^{-(u-w)^2/4} + \sqrt{\tau} \omega(0) + O(\tau). \quad (20)$$

Note that ν has entirely disappeared from the equation (but will appear in the boundary condition, see below), and only the value of ω close to the transaction price is relevant at this order.

After a very large number of auctions, one therefore finds that the stationary shape of the demand curve close to the price and in the limit $\tau \rightarrow 0$ is given by the non-trivial solution of the following fixed point equation:

$$\phi_\infty(u) = \int_0^{+\infty} \frac{dw}{\sqrt{4\pi}} \phi_\infty(w) e^{-(u-w)^2/4}, \quad (21)$$

supplemented by the boundary condition $\phi_\infty(u \gg 1) \approx \mathcal{L}\sqrt{\mathcal{D}u}$, where \mathcal{L} is a constant to be determined below. [Note that the solution of Eq. (21) is determined up to a multiplicative factor that must be fixed by some external condition].

Equation (21) is of the Wiener-Hopf type and its analytical solution can be found in Atkinson (1974), Boersma (1974). We plot numerically this solution in Fig. 2; it is seen to be numerically very close to an affine function for

⁸ This approximation happens to be exact in the particular setting considered in Donier et al. (2014).

⁹ An extra correction of order $\sqrt{\tau}$ would appear if a drift term was added to Eqs. (7).

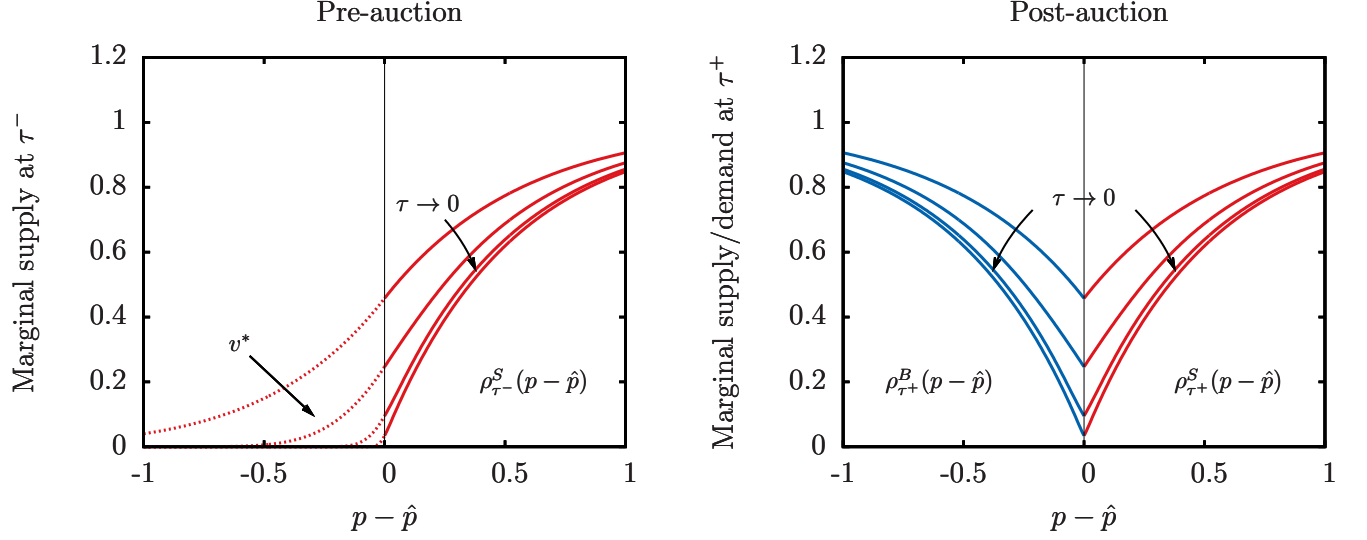


FIG. 3. Left: Shape of the marginal supply curve immediately before the auctions, for different inter-auction times τ , in the case $\omega_{\pm}(y) = \omega_{\pm}^0 \mathbb{1}_{\{y < 0\}}$. Right: Shape of the MSD immediately after the auctions, again for different inter-auction times τ . Note that as $\tau \rightarrow 0$, the MSD acquires a characteristic V-shape.

$u > 0$: $\phi_{\infty}(u) \approx \mathcal{L}\sqrt{\mathcal{D}}(u + u_0)$ with $u_0 \approx 0.824$. In summary, the stationary shape $\rho_{S,\text{st.}}(y)$ of the marginal supply curve in the frequent auction limit $\tau \rightarrow 0$ and close to the transaction price ($y = O(\sqrt{\mathcal{D}\tau})$), has a *universal shape*, independent of the detailed specification of the model (i.e., the functions $\nu_{\pm}(y)$ and $\omega_{\pm}(y)$). This supply curve is given by $\sqrt{\tau}\phi_{\infty}(y/\sqrt{\mathcal{D}\tau})$, which can itself be approximated by a simple affine function that will fully suffice for the purpose of the present paper:

$$\rho_{S,\text{st.}}(y \geq 0) \approx \mathcal{L}(y + y_0); \quad y_0 = u_0\sqrt{\mathcal{D}\tau}; \quad (\tau \rightarrow 0), \quad (22)$$

and similarly for $\rho_{D,\text{st.}}(y)$, see Fig. 3. The detailed interpretation of this result – in terms of market liquidity and price impact – will be given below.

We however still need to find the value of \mathcal{L} . This is done by comparing with the stationary solution $\varphi_{\text{st.}}(y)$ of Eq. (7) that satisfies the boundary solution $\varphi_{\text{st.}}(0) = 0$ (valid for $\tau = 0$). For $\nu_{\pm}(y) = \nu$, $\varphi_{\text{st.}}(y)$ can be computed explicitly and is given by:

$$\varphi_{\text{st.}}(y) = \frac{1}{\mathcal{D}} e^{-\sqrt{\nu/\mathcal{D}}y} \int_0^y dy' e^{2\sqrt{\nu/\mathcal{D}}y'} \int_{y'}^{\infty} dy'' e^{-\sqrt{\nu/\mathcal{D}}y''} \omega(y''). \quad (23)$$

Expanding $\varphi_{\text{st.}}(y)$ for small y (but still much larger than $\sqrt{\mathcal{D}\tau}$) finally leads to:

$$\varphi_{\text{st.}}(y) \approx \mathcal{L}y; \quad \mathcal{L} = \frac{1}{\mathcal{D}} \int_0^{\infty} dy' e^{-\sqrt{\nu/\mathcal{D}}y'} \omega(y'), \quad (24)$$

where \mathcal{L} can be seen as a measure of the market liquidity (see Donier et al. (2014) and below). Again in the simple case $\omega_{\pm}(y) = \omega^0 \mathbb{1}_{\{y < 0\}}$, one finds:

$$\mathcal{L} = \frac{\omega_0}{\sqrt{\nu\mathcal{D}}}. \quad (25)$$

Therefore, liquidity increases with the order arrival rate and decreases with their cancellation rate, as above, but also decreases with the diffusion constant \mathcal{D} that can be loosely identified with market volatility (see the discussion above).

Coming back to Eq. (22), one notes that Kyle's λ behaves as $\lambda^{-1} \equiv 2\rho_{S,\text{st.}}(y = 0) \propto \sqrt{\tau}$, which is the pivotal result of the present paper. It means that the marginal supply and demand at the transaction price becomes very

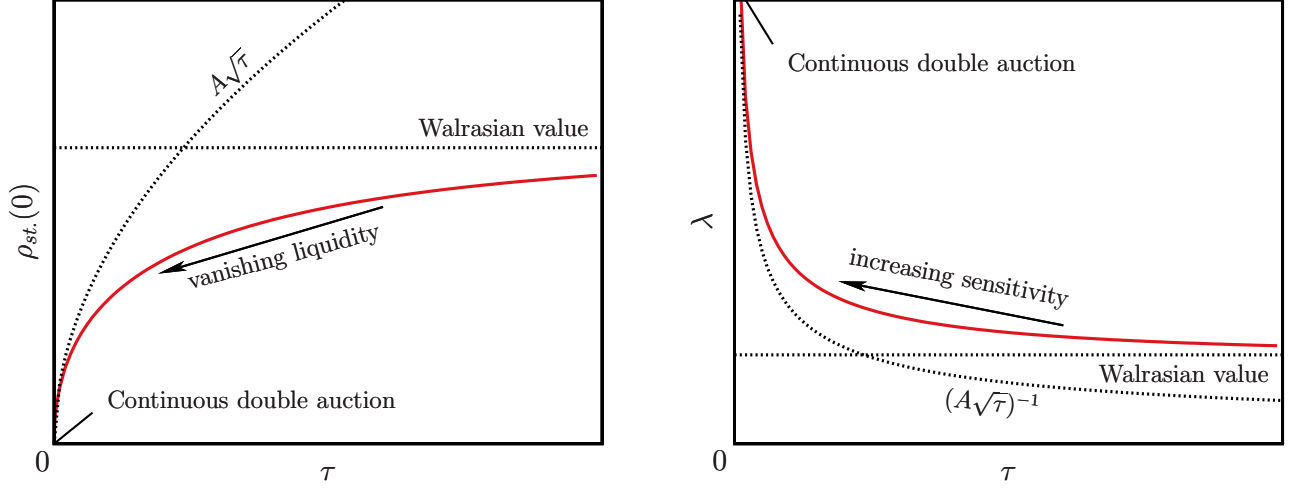


FIG. 4. Left: As the inter-auction time τ decreases, the liquidity close to the price, $\rho_{st.}(y = 0)$ decreases from its finite Walrasian value to zero in the case of continuous double auctions. Right: Conversely, the relative impact of (small) additional volumes λ diverges for continuous double auctions ($\lambda \rightarrow \infty$ when $\tau \rightarrow 0$), eventually leading to the square-root impact law.

small around the transaction price as the auction frequency increases. Intuitively, this is due to the fact that close to the transaction price, liquidity has no time to rebuild between two auctions. From the point of view of impact, the divergence of Kyle's λ as $1/\sqrt{\tau}$ means that the auction price becomes more and more susceptible to any imbalance between supply and demand. We show in Fig. 4 (left) $\rho_{st.}(y = 0)$ in the special case where $\omega_{\pm}(y) = \omega^0 \mathbb{1}_{\{y < > 0\}}$, illustrating how liquidity vanishes as $\tau \rightarrow 0$ as well as (right) the corresponding impact parameter $\lambda(\tau)$ that diverges in this limit. One can thus see how, by increasing the auction frequency, one smoothly departs from the Walrasian equilibrium scenario to reach the limit of continuous double auction corresponding to modern financial markets.

The last item we need is the shape of the supply curve *below* the transaction price just before the next auction, that gives the amount of supply/demand on the “wrong” side of the price, i.e. precisely the volume exchanged at the auction. Using the simple affine approximation Eq. (22), one finds:

$$\rho_{S,st.}(y < 0) \approx \mathcal{L} \int_0^{+\infty} \frac{dy'}{\sqrt{4\pi\mathcal{D}\tau}} (y' + y_0) e^{-\frac{(y' - y)^2}{4\mathcal{D}\tau}}, \quad (26)$$

or, again setting $y = -u\sqrt{\mathcal{D}\tau}$ and $y' = w\sqrt{\mathcal{D}\tau}$,

$$\rho_{S,st.}(y < 0) \approx \mathcal{L}\sqrt{\mathcal{D}\tau} \int_0^{+\infty} \frac{dw}{\sqrt{4\pi}} (w + u_0) e^{-\frac{(w+u)^2}{4}} = \mathcal{L}\sqrt{\mathcal{D}\tau} \left[\frac{e^{-u^2/4}}{\sqrt{\pi}} + \frac{1}{2}(u_0 - u)(1 - \text{Erf}(u/2)) \right]. \quad (27)$$

From this expression, the total volume v^* exchanged during each auction is found to be:

$$v^* = \int_{-\infty}^0 dy \rho_{S,st.}(y) = \mathcal{L}\mathcal{D}\tau \left[\frac{1}{2} + \frac{u_0}{\sqrt{\pi}} \right] \approx 0.965\mathcal{L}\mathcal{D}\tau, \quad (28)$$

whereas the exact result (that can be obtained directly from the diffusion equation in the $\tau \rightarrow 0$ limit) is $v^* = \mathcal{L}\mathcal{D}\tau$. The error induced by our simple affine approximation is thus only a few percents. Interestingly, one sees that the total transacted volume V in a finite time interval T , given by $V = v^*T/\tau$, remains finite when $\tau \rightarrow 0$, and equal to $V = \mathcal{L}DT$. This observation should be put in perspective with the recent evolution of financial markets, where the time between transactions τ has become very small, while the volume of each transaction has simultaneously decreased, in such a way that the daily volume has remained roughly constant.

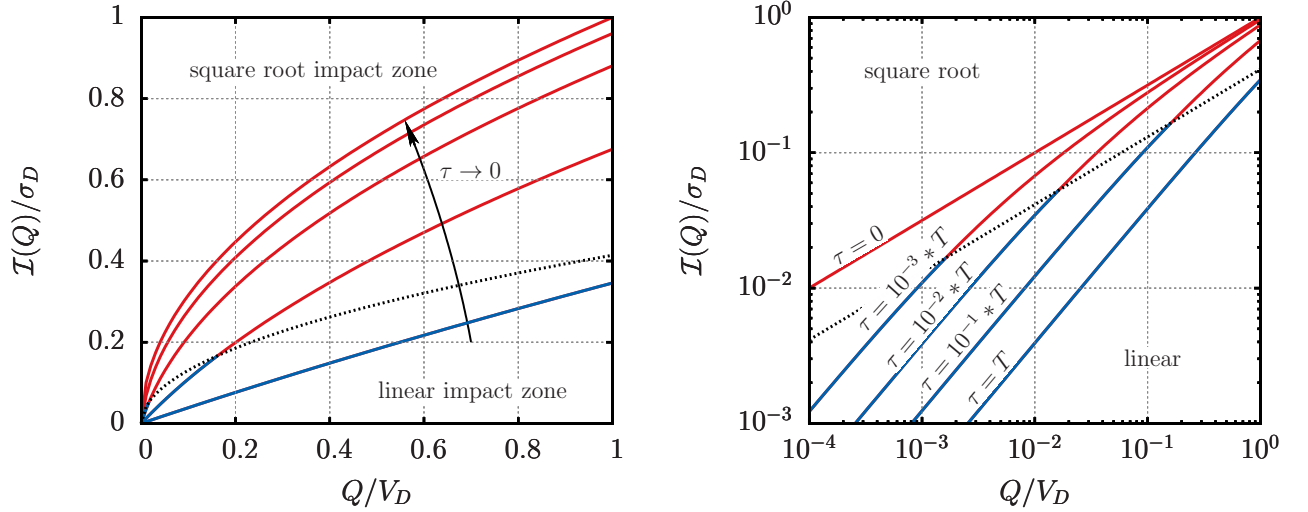


FIG. 5. The impact of traded volumes Q for a given inter-auction time τ is linear for $Q \ll v^* = \mathcal{L}\mathcal{D}\tau$ and then square-root for $Q \gg v^* = \mathcal{L}\mathcal{D}\tau$. The linear impact zone shrinks to zero when $\tau \rightarrow 0$, when one recovers a pure square-root impact, i.e. a diverging Kyle's λ .

C. The vanishing liquidity limit: From linear to square-root impact

From the shape of the MSD close to transaction price given by Eq. (22), it is immediate to compute the supply and demand curves just before an auction when the inter-auction time τ tends to 0. Denoting again as y the difference between the price level p and the fundamental price \hat{p}_t , one finds:

$$\begin{aligned} S(p \geq \hat{p}_t) &= \mathcal{L}(y_0 y + \frac{1}{2} y^2) + v^* \\ D(p \leq \hat{p}_t) &= \mathcal{L}(-y_0 y + \frac{1}{2} y^2) + v^* \end{aligned} \quad (29)$$

where, as found in the previous section, $y_0 \equiv u_0 \sqrt{\mathcal{D}\tau} \approx 0.824 \sqrt{\mathcal{D}\tau}$. From Eq. (27) above, it is readily seen that the supply (resp. demand) curve below (resp. above) \hat{p}_t can be written as $v^* F(y/\sqrt{\mathcal{D}\tau})$, where $F(u)$ is a certain function that goes from $F(0) = 1$ to $F(\infty) = 0$.¹⁰

The above equation Eq. (29) immediately allows us to compute the impact $\mathcal{I}(Q) \equiv y^*$ of an extra buy quantity Q , as the solution of $\mathcal{L}(y_0 y^* + \frac{1}{2} y^{*2}) + v^* = Q + v^* F(y^*/\sqrt{\mathcal{D}\tau})$. It is clear that the solution can be written as $y^* = \sqrt{\mathcal{D}\tau} Y(Q/\mathcal{L}\mathcal{D}\tau)$, where $Y(q)$ obeys $u_0 Y + \frac{1}{2} Y^2 + (1 - F(Y)) = q$. The limiting behaviours of Y in the limits $q \ll 1$ and $q \gg 1$ are easy to compute, and read:

$$Y(q) \approx_{q \ll 1} 0.555q; \quad Y(q) \approx_{q \gg 1} \sqrt{2q}. \quad (31)$$

One therefore deduces that the impact $\mathcal{I}(Q)$ is linear in a region where the volume Q is much smaller than $v^* \sim \mathcal{L}\mathcal{D}\tau$, i.e. when the extra volume is small compared to the typical volume exchanged during auctions, as expected. In the other limit, however, one recovers the *square-root impact* observed empirically (as found in Donier et al. (2014)¹¹) :

$$\mathcal{I}(Q \gg v^*) \approx \sqrt{\frac{2Q}{\mathcal{L}}}, \quad (32)$$

¹⁰ This function reads, explicitly:

$$\left[\frac{1}{2} + \frac{u_0}{\sqrt{\pi}} \right] F(u) = \frac{1}{2} (1 - \text{Erf}(u/2)) \left(\frac{u^2}{2} - u_0 u + 1 \right) - \frac{e^{-u^2/4}}{\sqrt{\pi}} \left(\frac{u}{2} - u_0 \right). \quad (30)$$

¹¹ In that paper, the study of market impact in the $\tau \rightarrow 0$ limit has been investigated much more in depth, in particular in the case of a progressive execution over some finite time window.

The impact in the universal small Q region¹², with a linear regime for $Q < v^*$ and a crossover to a square root regime when Q becomes greater than v^* , is shown in Fig. 5. Clearly, for $\tau = 0^+$, the auction volume $v^* = \mathcal{L}\mathcal{D}\tau$ also tends to zero, so that the region where impact is linear in volume shrinks to zero. In other words, when the interaction time becomes infinitely small, the impact of small trades is *never* linear. This comes from the fact that the MSD curves tend to zero exactly at the trading price when $\tau \rightarrow 0$.

D. Empirical confirmation: the shape of supply and demand on the Bitcoin

Remarkably, we can check directly the above prediction on the shape of the MSD curves using Bitcoin data, where traders are much less strategic than in more mature financial markets and display their orders in the visible order book even quite far from the current price. The graph of the average shape of the Bitcoin limit order book (LOB) and of its integral, that are assumed to be good proxies for the MSD/SD curves at least not too far from the price, is shown on Fig. 6.

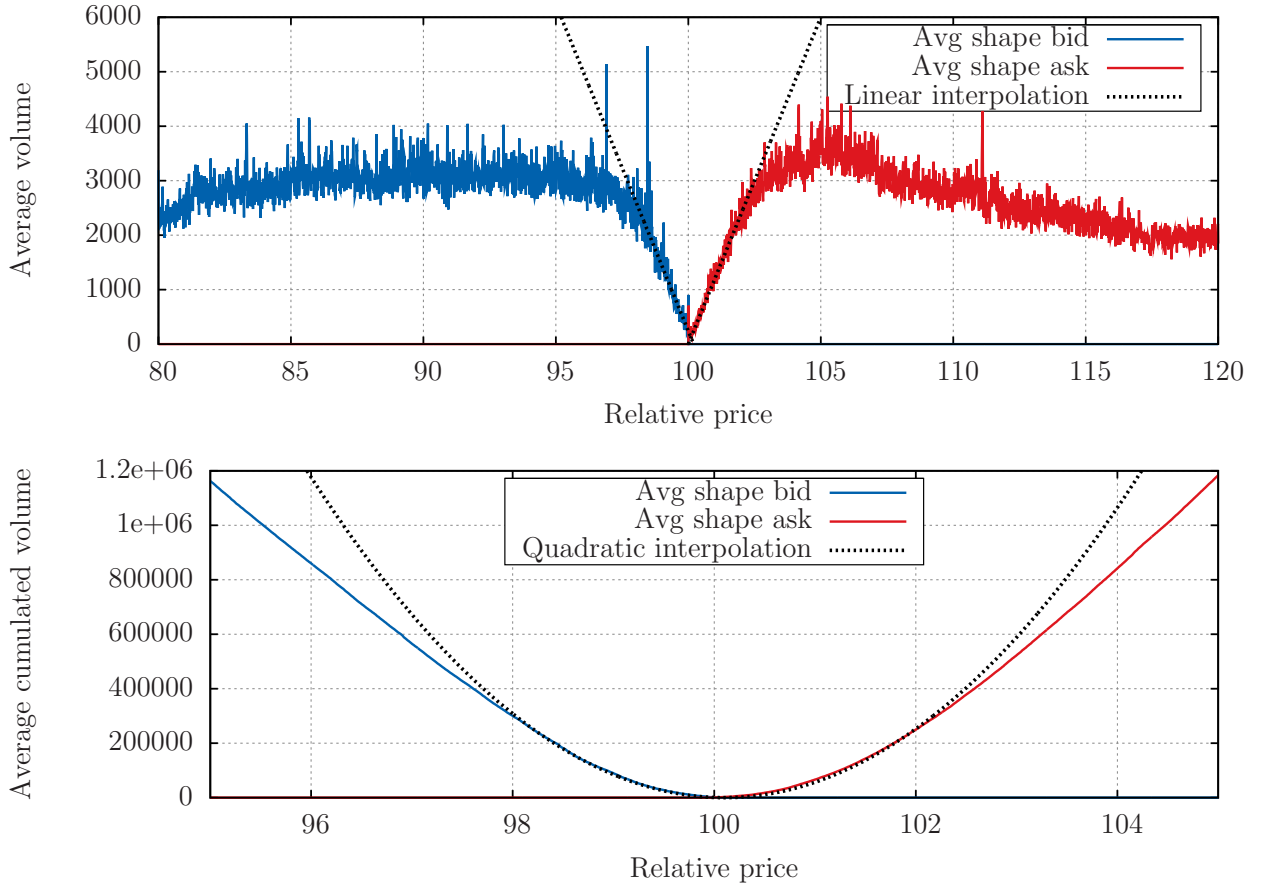


FIG. 6. Top: The average shape of the limit order book on the Bitcoin market, which we believe to be a representative sample of the MSD curve, in particular close to the price. The data comes from successive snapshots of the full order book of the Bitcoin market, every 15 minutes from May 2013 to September 2013, centred around the current mid-price. Bottom: Integrated shape of the visible order book, as a proxy for the supply and demand curves on the Bitcoin. The LOB/MSD curves grow linearly with respect to the distance to the price, resulting in a quadratic shape for supply and demand.

¹² Corresponding to the region where ϕ_∞ is a good approximation of the order book, see Eq. 19. For very large Q 's, the linear approximation describing the MSD curves breaks down, and one enters a presumably non-universal regime that is beyond the scope of the present discussion (cf. e.g. the shape of the MSD on the Bitcoin, Fig. 6).

Quite strikingly, the MSD curves are indeed *linear* in the vicinity of the price that corresponds to about 5% range, in perfect agreement with our dynamical theory of supply and demand in the limit of frequent auctions (note in particular that $\partial_y S(p^*) = \partial_y D(p^*) \approx 0$). Correspondingly, we do expect that impact of meta-orders should be well accounted by a square-root law in this region, which is indeed also found empirically (see Donier and Bonart (2014) for the special case of Bitcoin case, and Almgren et al. (2005), Bacry et al. (2014), Bershova and Rakhlin (2013), Brokman et al. (2014), Gomes and Waelbroeck (2014), Grinold and Kahn (2000), Mastromatteo et al. (2014a), Moro et al. (2009), Torre and Ferrari (1997), Tóth et al. (2011) for more mainstream markets). Further away from the price, the non-universal region clearly appears, where the shape of the MSD (here, approximately saturating to some constant value) depends on the detailed characteristics of the order flow, modelled in this paper by the functions $\nu(y)$ and $\omega(y)$.

E. Summary

The above section presented our story in mathematical terms. The punchline is however quite simple, and well summarized by the graphs plotted in Fig. 7, where we show (a) the standard Walrasian supply and demand curves just before the auction, from which the equilibrium price p^* can be deduced; (b) the supply and demand curves just after an auction, when the inter-auction time τ is large enough – in which case the marginal supply and demand are both finite at p^* ; and (c) the supply and demand curves in the continuous time limit $\tau \rightarrow 0$, for which the marginal supply and demand curves vanish linearly around the current price, as found in the Bitcoin market.

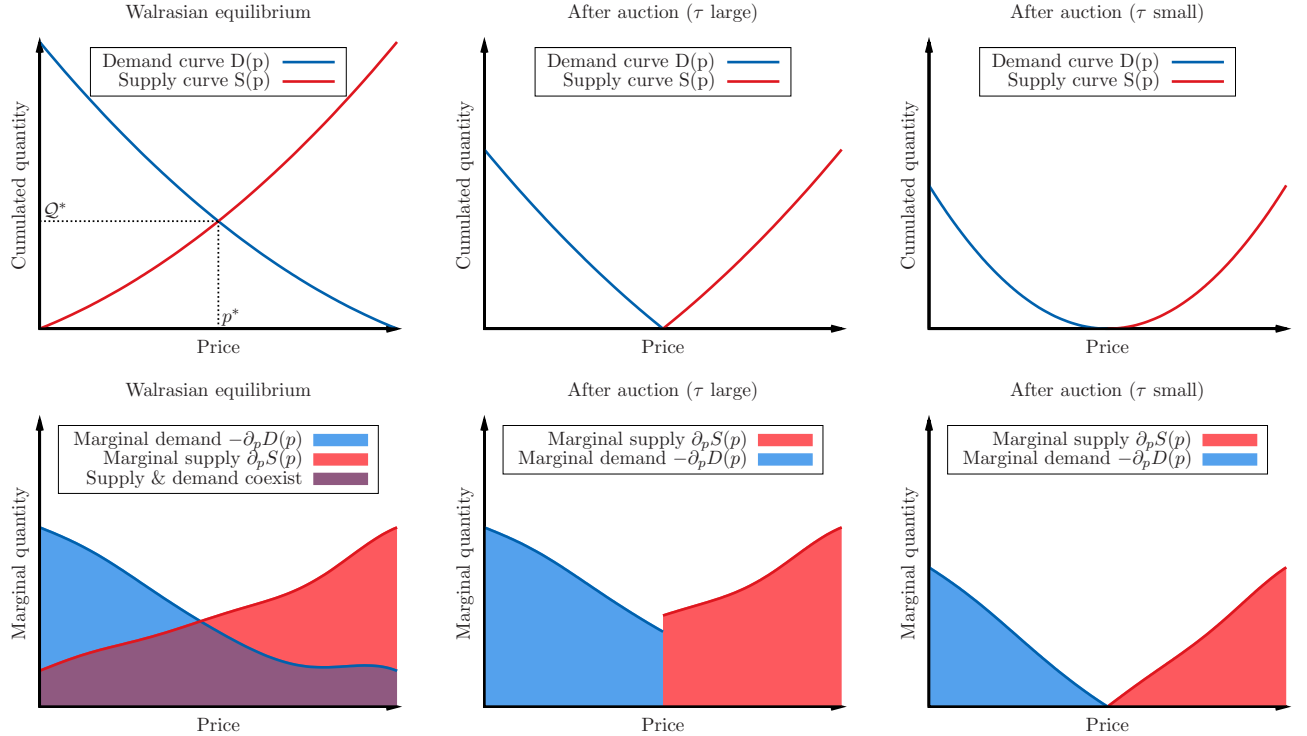


FIG. 7. Top: Supply and demand curves in (left) Walrasian auctions, (center) immediately after infrequent auctions and (right) immediately after frequent auctions. Bottom: Corresponding MSD curves. When transactions occur, supply and demand cannot cross (center and right). When the market is cleared frequently, supply and demand are depleted close to the price and exhibit a characteristic V-shape (right).

V. DISCUSSION

Up to now, our presentation has been fairly technical, with the aim of establishing our main results. Still, many points of general interest have glossed over for the sake of readability. We feel that these deserve a more detailed discussion that we provide now.

A. Price discovery vs. price formation

The interpretation of price moves in financial markets has generated endless theoretical debates, culminating in the split 2013 Nobel prize between the hero of efficient market theory (Fama (1970)), and the beacon of behavioural economics (Shiller (1980)). For the former school of thought, the *fundamental price* of an asset pre-exists and only waits to be “discovered” by aggregating the unbiased opinions of rational market participants. Up to small self-correcting errors, markets do clear at the right price. This assumption is a pre-requisite for many economic models, e.g. the classical Kyle model Kyle (1985) within which some agents are *informed*: they are supposed to know the fundamental price for the next period and invest accordingly so as to make a maximum profit from of this information.

This perfect Platonian view of markets is however hard to swallow for many followers of Shiller. After all, financial markets are driven by humans who have a very imperfect knowledge of the fundamental price, prone to many behavioural biases. The market is a device that merely aggregates all participants’ intentions, regardless on whether they are justified or not, and spits out the “market price”. In this view, markets allow *price formation* rather than *price discovery*.

Our modelling strategy provides a very intuitive framework to think about the difference between the two viewpoints. In Sec. III B, we postulated that on any time interval $(t, t + dt)$ an information item is released, leading to a change in the (unobservable) fundamental price $d\xi_t$, of variance $\sigma^2 dt$. This piece of news is however interpreted by agent i as predicting a price change $\beta_t^i d\xi_t$ plus idiosyncratic noise. As noted above, this means that the news is over-interpreted by some ($\beta^i > 1$) and under-interpreted by others ($\beta^i < 1$). If for any news $d\xi_t$, the β_t^i ’s average *exactly* to 1, the market on aggregate perfectly digests the news and the (permanent) increment in perceived price is precisely $d\xi_t$. If however this is not the case, $\mathbb{E}_i[\beta_t^i] d\xi_t \neq d\xi_t$, and this leads to two possibly different definitions of a reference price: the *fundamental price* p_t^F (unknown to agents) and the *market price* \hat{p}_t (encoding agents perceptions):¹³

$$\begin{aligned} p_t^F &\equiv p_0 + \int_0^t d\xi_s \\ \hat{p}_t &\equiv p_0 + \int_0^t \mathbb{E}_i[\beta_s^i] d\xi_s, \end{aligned} \tag{33}$$

where p_0 is an arbitrary reference price, assuming that the market price is equal to the fundamental price initially. Only if the average market reaction $\beta_t \equiv \mathbb{E}_i[\beta_t^i]$ is unbiased at any time can one speak of efficient markets and *price discovery*. Otherwise, *price formation* prevails, and the market price errs away from the fundamental price, as envisaged by Black (1986). More precisely, one can compute the pricing error as:

$$\text{Var}(\hat{p}_t - p_t^F) = \mathbb{E} \left[\sigma^2 \int_0^t (\beta_s - 1)^2 ds \right] = \text{Var}(\beta_t) \text{Var}(p_t^F), \tag{34}$$

as well as the variance of the market price as:

$$\text{Var}(\hat{p}_t) = \mathbb{E} \left[\sigma^2 \int_0^t \beta_s^2 ds \right] = [1 + \text{Var}(\beta_t)] \text{Var}(p_t^F) \tag{35}$$

¹³ Here and below, we assume that agents perceptions are symmetrically distributed around \hat{p}_t , that we identify with the auction clearing price p^* . In other words, we neglect here any “bid-ask” bounce that may affect the short-time price dynamics.

showing that even if there is no bias on average (i.e. $\mathbb{E}_s[\beta_s] \equiv 1$), $\text{Var}(\hat{p}_t) \geq \text{Var}(p_t^F)$ with a strict inequality as soon as the market reaction is not perfectly unbiased *at all times* – a highly plausible situation.¹⁴ This embeds Shiller’s famous excess volatility puzzle Shiller (1980) in an interesting formal framework.

We are here at the core of a crucial question in financial economics: what is “information”? The discussion above naturally leads to two definitions of information:

- (a) information on the *fundamental price* (often called fundamental information), corresponds to some (perhaps noisy) knowledge about the value of p_t^F , while
- (b) information on the *market price*, corresponds to some (perhaps noisy) knowledge about the future value of the market price \hat{p}_t .

In the latter case, information is only about *correctly anticipating the behaviour of others*, exactly as Keynes envisioned Keynes (2006). The notion of *information* should then rather be replaced by the notion of *correlation* – if all the market participants’ β were negative, the correct information for an arbitrageur would correspond to also interpreting the news with a negative β , even if it did not make sense. In this context, the difference between a “noise” trader and an “informed” trader in Kyle’s model is merely the level of correlation with the crowd of other market participants: informed traders are positively correlated with it, whereas noise traders are simply uncorrelated with the crowd¹⁵.

Finally, let us give an alternative interpretation of the market price equation Eq. (33), in terms of the individual estimate of the fundamental price \hat{p}_t^i and the fraction of market shares F^i of agent $i \in \{1, \dots, N\}$, such that $\sum_i F^i = 1$. The individual estimate of the fundamental price evolves as $d\hat{p}_t^i = d\xi_t^i$. Assuming that expectations are symmetrically distributed around \hat{p}_t , the market clearing price \hat{p}_t is given by the weighted average of individual prices (whether they are truly “informed” or not):

$$\hat{p}_t = \sum_i F^i \hat{p}_t^i, \quad (36)$$

which of course coincides with Eq. (33). However, one now clearly sees that the *permanent* impact of an agent with market share F^i on the price is $F^i d\xi^i$ when his view on the price changes by $d\xi^i$. When the market share of an individual investor is small, the permanent impact of his/her isolated orders on market prices is itself small (in particular much smaller than the transient square-root impact that depends on his/her fraction of the daily volume), except if his/her trade is correlated with the rest of the market, in which case it is often said to have an *alpha* – i.e. a predictive signal on the price (see previous footnote).

B. Market stability and marked-to-market valuation

The bottom line of the model and analysis presented in the above sections is the possible effect of market design itself on *price stability*. Indeed, the highly singular square root impact consistently measured on financial markets implies that small perturbation (e.g. noise trading) may result in abnormally high returns, questioning the robustness of the price formation and the very stability of markets. Indeed, if one freak order, whose volume remains small in comparison to market daily volume, can move the price by a large fraction of the daily volatility, then it is clear that one should not put too much faith in the reliability and resiliency of market prices. As we argued above, this market fragility is the result of continuous market clearing, that leads to the following property:

¹⁴ Note however that on long time scales, some weak mean reversion towards the fundamental price should take place, as discussed in e.g. Black (1986), Bouchaud and Cont (1998), De Bondt and Thaler (1990).

¹⁵ See the detailed discussion in Donier et al. (2014), and Donier and Bonart (2014), Gomes and Waelbroeck (2014) for empirical studies on how “informed” and “noise” trades impact the price differently on the long run. While both have similar instantaneous impacts, on the long run the price reverts to its initial value for noise trades whereas a permanent component remains for “informed” trades. Interestingly, data suggests that the permanent impact is of the same order of magnitude as the transient impact.

The price is the point at which a *vanishing* supply meets a *vanishing* demand.

An implication of particular importance is the relevance of mark-to-market accounting rules for the valuation of large portfolios (so as to leave aside the problem of liquidation costs that usually enters in the discussion, one can think of the assets of an insurance firm that are kept until expiry). From the previous discussion, it is meaningless – or even dangerous – to mark too closely portfolios to the market prices, as some noise traders, fat-fingers or even ill-intentioned manipulators can trigger large re-balancing with not-so-large volumes, resulting in inappropriate profits and in unstable prices.

It would be more significant to assess market prices on the basis of the local supply and demand (around the market price) so as to smooth out fluctuations. A practical way to do so is to monitor the market *liquidity* \mathcal{L} , that is well proxied by the ratio $\sigma_D/\sqrt{V_D}$ (where σ_D is the daily volatility and V_D is the daily traded volume), as proposed and tested on the Bitcoin market in Donier and Bouchaud (2015). Several other measures of liquidity might also be relevant, see Amihud and Mendelson (1986), Corradi et al. (2015), Foucault et al. (2013). In the case when the asset is not assumed to be kept until expiry but might incur liquidation costs, the monitoring of liquidity would be a good indicator of the *ex-post* value of the portfolio, once liquidated. This is the idea of impact-discounted mark-to-market value proposed and discussed in Caccioli et al. (2012).

C. Would batch auctions be beneficial?

In order to curb potentially nefarious and socially wasteful HFT activities, a possibility that is currently hotly debated Budish et al. (2013) is to change the continuous trading system to frequent batch auctions, that would occur in discrete time with a time interval between auctions of order 100ms-1s. While we are not necessarily convinced that high frequency trading is such an evil¹⁶ and the investment in speed technologies is such a waste on the long run, we believe that the issue of market stability should be of primary concern. Our theory provides a natural framework for studying the effect of such market design changes on the supply and demand curves, and on the stability of the resulting price.

From the above analysis, we concluded that the singularity at the immediate vicinity of the price is regularized when auctions occur in discrete time τ , leading to a reduced price impact (linear instead of square-root). However, one should not rejoice too fast, since we also saw that this regularisation only concerns very small volumes, less than the average volume traded in the market during time τ . In order to have a substantial effect and reduce the impact of trade size Q of – say – 1–10% of the average daily volume, the inter-auction time should be, unsurprisingly, of the order of five minutes to one hour and not milliseconds or even seconds. Trying to improve market stability through frequent batch auctions does not seem very useful in the light of our theory, except if one accepts to clear the market every hour or so, which would potentially imply other problems, such as and new source of liquidity risk and the corresponding development of secondary markets where transactions would take place in-between auctions.

VI. CONCLUSION

In this paper, we have developed a fully dynamic theory of liquidity, based on weak and general assumptions on investors' behaviours: in a nutshell, heterogeneous reactions to incoming news of a large number of “infinitesimal” investors. Addressing the inability of the Walrasian theory to take transactions into consideration, we allow for auctions to clear the market periodically, and show how the market clearing mechanism itself affects the structural properties

¹⁶ The overall profit of HFT firms was estimated to be around 5B\$/year at its peak in 2010. This corresponds to an estimated cost of 1 basis point (10^{-4}) per transaction in the US equity markets, compatible with a fraction of the average bid-ask spread in the same period Hendershott et al. (2011). This is probably at least 10 times smaller than the profits of “old school” market makers: the average bid-ask spread on US markets was fluctuating around 60 basis points (!) from 1900 to 1980, before declining sharply Jones (2002).

of supply and demand. In the case when the time between auctions is very large, we recover classical Walrasian auctions, in which market prices and liquidity are determined by the long-term (im-)balance between the incoming supply and the incoming demand. When auctions are allowed to happen at high frequency, the liquidity around the price mechanically vanishes, which leads to an anomalous, square-root impact of small orders that increases with market *volatility*. This accounts for the universally observed square root impact of small orders on modern financial markets and on the Bitcoin. In order to obtain a direct confirmation of the theory, we measured the shape of the Bitcoin order book (that appears to be a faithful reflection of low frequency supply and demand, at least close enough to the price), which indeed displays a striking “V-shape” for the marginal supply and demand (see Fig. 6).

Our results highlight an apparent paradox: the more frequent the transactions are allowed to occur (thereby increasing, in theory, market efficiency), the more *fragile* the resulting price is! In continuous double auctions markets, the price can be seen as the point at which a *vanishing* supply meets a *vanishing* demand, challenging the Platonian view of financial markets that prices are well-defined and stable. Our framework allows us to draw two further conclusions of general interest. First, the local estimates of supply and demand needs to be taken into account when one watches market prices, with important implications on portfolio valuation and stability monitoring Donier and Bouchaud (2015), Kyle and Obizhaeva (2012). Second, as soon as the reaction to incoming news is not unbiased at all times, the volatility of the market price exceeds the fundamental volatility, embedding Shiller’s famous excess volatility puzzle Shiller (1980) in an interesting formal framework.

Although we only considered market design and price stability within a particular angle, we believe that our framework can be extended to address many other practical problems, such as the market maker’s problem, cross-impact between markets, optimal execution issues or the effect of taxation and other changes in market design. We also restricted to stationary market conditions, but our formalism can be readily adapted to include fluctuations in liquidity and/or market volatility (that would translate in some stochastic evolution of \mathcal{D}). Finally, the surprisingly good agreement between theory and empirical data suggests to extend our set of hypotheses to agent-based models of markets (cf Sec. II B), with the aim of producing realistic emergent properties from microscopic agent behaviour. More generally, our results vindicate an approach to dynamics in economic sciences as resulting from complex interactions between many heterogeneous agents, that may – or may not – be rational, in the general vein of, e.g. Gualdi et al. (2015), Hommes (2006) and refs. therein. The resulting partial differential equation that governs the evolution of agents’ preferences offers a much deeper level of understanding than simply postulating ad-hoc stochastic models for prices. We believe that this modelling strategy can be extended to many other situations of economic relevance.

ACKNOWLEDGEMENTS

We warmly thank M. Abeille, J. Bonart, R. Cont, J. De Lataillade, D. Delli Gatti, M. Gould, T. Hendershott, J. Kockelkoren, C. A. Lehalle, Y. Lempérière, I. Mastromatteo, M. Potters, I. Rosu and B. Tóth for many crucial discussions and collaborations on these issues – and in particular G. Zerah for his help on the Wiener-Hopf method. We also thank A. Tilloy for useful remarks on the manuscript.

-
- R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, 18(7):5762, 2005.
- Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.
- Yakov Amihud and Haim Mendelson. Asset pricing and the bid-ask spread. *Journal of financial Economics*, 17(2):223–249, 1986.
- C Atkinson. A wiener-hopf integral equation arising in some inference and queueing problems. *Biometrika*, 61(2):277–283, 1974.
- Louis Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.

- Emmanuel Bacry, Adrian Iuga, Matthieu Lasnier, and Charles-Albert Lehalle. Market impacts and the life cycle of investors orders. *Available at SSRN 2532152*, 2014.
- Per Bak, Maya Paczuski, and Martin Shubik. Price variations in a stock market with many agents. *Physica A: Statistical Mechanics and its Applications*, 246(3):430–453, 1997.
- N. Bershova and D. Rakhlin. The non-linear market impact of large trades: Evidence from buy-side order flow. *Quantitative Finance*, 13:1759–1778, 2013.
- Bruno Biais. Price formation and equilibrium liquidity in fragmented and centralized markets. *The Journal of Finance*, 48(1):157–185, 1993.
- Bruno Biais, Pierre Hillion, and Chester Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *the Journal of Finance*, 50(5):1655–1689, 1995.
- Fischer Black. Noise. *The journal of finance*, 41(3):529–543, 1986.
- Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973.
- Joop Boersma. Note on a wiener-hopf integral equation arising in some inference and queueing problems. 1974.
- Steve Bongiovanni, Milan Borkovec, and Robert D Sinclair. Let’s play hide-and-seek: The location and size of undisclosed limit order volume. *The Journal of Trading*, 1(3):34–46, 2006.
- J-P Bouchaud and Rama Cont. A langevin approach to stock market fluctuations and crashes. *The European Physical Journal B-Condensed Matter and Complex Systems*, 6(4):543–550, 1998.
- Jean-Philippe Bouchaud. Price impact. *Encyclopedia of quantitative finance*, 2010.
- Jean-Philippe Bouchaud, Marc Mézard, Marc Potters, et al. Statistical properties of stock order books: empirical results and models. *Quantitative finance*, 2(4):251–256, 2002.
- Jean-Philippe Bouchaud, J Doyne Farmer, and Fabrizio Lillo. How markets slowly digest changes in supply and demand. *Fabrizio, How Markets Slowly Digest Changes in Supply and Demand (September 11, 2008)*, 2008.
- X. Brokmann, E. Sérié, J. Kockelkoren, and J.-P. Bouchaud. Slow decay of impact in equity markets. arXiv:1407.3390, submitted to Market Microstructure and Liquidity, 2014.
- Eric B Budish, Peter Cramton, and John J Shim. The high-frequency trading arms race: Frequent batch auctions as a market design response. *Fama-Miller Working Paper*, pages 14–03, 2013.
- F Caccioli, JP Bouchaud, and D Farmer. Impact-adjusted valuation and the criticality of leverage. Technical report, technical report, 2012.
- Álvaro Cartea and Sebastian Jaimungal. Optimal execution with limit and market orders. *Forthcoming: Quantitative Finance*, 2014.
- Timothy N Cason and Daniel Friedman. Price formation in double auction markets. *Journal of Economic Dynamics and Control*, 20(8):1307–1337, 1996.
- Rama Cont and Adrien De Larrard. Price dynamics in a markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013.
- Rama Cont and Arseniy Kukanov. Optimal order placement in limit order markets. *Available at SSRN 2155218*, 2013.
- Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations research*, 58(3):549–563, 2010.
- Francesco Corradi, Andrea Zaccaria, and Luciano Pietronero. Liquidity crises on different time scales. *arXiv preprint arXiv:1504.02956*, 2015.
- Werner FM De Bondt and Richard H Thaler. Do security analysts overreact? *The American Economic Review*, pages 52–57, 1990.
- David S Dean. Langevin equation for the density of a system of interacting langevin processes. *Journal of Physics A: Mathematical and General*, 29(24):L613, 1996.
- J. Donier and J. Bonart. A million metaorder analysis of market impact on bitcoin. arXiv:1412.4503, submitted to Market Microstructure and Liquidity, 2014.
- Jonathan Donier and Jean-Philippe Bouchaud. Why do markets crash? bitcoin data offers unprecedented insights. *SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2583743*, 2015.
- Jonathan Donier, Julius Friedrich Bonart, Iacopo Mastromatteo, and Jean-Philippe Bouchaud. A fully consistent, minimal model for non-linear market impact. *Quantitative finance, to appear*, 2014.

- David Easley and John Ledyard. Theories of price formation and exchange in double oral auctions. *The double auction market: Institutions, theories, and evidence*, pages 63–97, 1993.
- Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.
- J Doyne Farmer, Paolo Patelli, and Ilija I Zovko. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6):2254–2259, 2005.
- Franklin M Fisher. *Disequilibrium foundations of equilibrium economics*. Number 6. Cambridge University Press, 1989.
- Thierry Foucault. Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial markets*, 2(2):99–134, 1999.
- Thierry Foucault, Marco Pagano, and Ailsa Röell. *Market liquidity: theory, evidence, and policy*. Oxford University Press, 2013.
- Daniel Fricke and Austin Gerig. Liquidity risk, speculative trade, and the optimal latency of financial markets. 2014.
- Xuefeng Gao, JG Dai, Ton Dieker, and Shijie Deng. Hydrodynamic limit of order book dynamics. *Available at SSRN 2530306*, 2014.
- A Gareche, G Disdier, J Kockelkoren, and J-P Bouchaud. Fokker-planck description for the queue dynamics of large tick stocks. *Physical Review E*, 88(3):032809, 2013.
- Steven Gjerstad and John Dickhaut. Price formation in double auctions. *Games and economic behavior*, 22(1):1–29, 1998.
- Lawrence R Glosten. Is the electronic open limit order book inevitable? *The Journal of Finance*, 49(4):1127–1161, 1994.
- Lawrence R Glosten and Paul R Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100, 1985.
- C. Gomes and H. Waelbroeck. Is market impact a measure of the information value of trades? Market response to liquidity vs informed trades. *Quantitative Finance*, pages 1–21, 2014. doi:10.1080/14697688.2014.963140.
- Richard C Grinold and Ronald N Kahn. Active portfolio management. 2000.
- Stanislao Gualdi, Marco Tarzia, Francesco Zamponi, and Jean-Philippe Bouchaud. Tipping points in macroeconomic agent-based models. *Journal of Economic Dynamics and Control*, 50:29–61, 2015.
- Puneet Handa and Robert A Schwartz. Limit order trading. *The Journal of Finance*, 51(5):1835–1861, 1996.
- Puneet Handa, Robert Schwartz, and Ashish Tiwari. Quote setting and price formation in an order driven market. *Journal of financial markets*, 6(4):461–489, 2003.
- Lawrence Harris et al. Liquidity, trading rules and electronic trading systems. Technical report, 1990.
- Joel Hasbrouck. *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*. Oxford University Press, 2006.
- Terrence Hendershott, Charles M Jones, and Albert J Menkveld. Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1–33, 2011.
- John R Hicks. *Value and capital*, volume 2. Clarendon press Oxford, 1946.
- Cars H Hommes. Heterogeneous agent models in economics and finance. *Handbook of computational economics*, 2:1109–1186, 2006.
- Weibing Huang, Charles-Albert Lehalle, and Mathieu Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, (just-accepted):00–00, 2014.
- Charles M Jones. A century of stock market liquidity and trading costs. *Available at SSRN 313681*, 2002.
- John Maynard Keynes. *General theory of employment, interest and money*. Atlantic Publishers & Dist, 2006.
- Julianus Kockelkoren. *Order types*. Encyclopedia of Quantitative Finance, 2010.
- Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.
- Albert S Kyle and Anna A Obizhaeva. Large bets and stock market crashes. In *AFA 2013 San Diego Meetings Paper*, 2012.
- Aimé Lachapelle, Jean-Michel Lasry, Charles-Albert Lehalle, and Pierre-Louis Lions. Efficiency of the price formation process in presence of high frequency participants: a mean field game analysis. *arXiv preprint arXiv:1305.6323*, 2013.
- Sophie Laruelle, Charles-Albert Lehalle, and Gilles Pages. Optimal split of orders across liquidity pools: a stochastic algorithm approach. *SIAM Journal on Financial Mathematics*, 2(1):1042–1076, 2011.
- Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.
- Charles-Albert Lehalle, Olivier Guéant, and Julien Razafinimanana. High-frequency simulations of an order book: a two-scale approach. In *Econophysics of Order-driven Markets*, pages 73–92. Springer, 2011.

- Ananth Madhavan. Market microstructure: A survey. *Journal of financial markets*, 3(3):205–258, 2000.
- Costis Maglaras and Ciamac Moallemi. A multiclass model of limit order book dynamics and its application to optimal trade execution. Technical report, Working paper, Columbia Business School, New York, 2011.
- N.Gregory Mankiw. *Principles of macroeconomics*. Cengage Learning, 2014.
- Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.
- I. Mastromatteo, B. Tóth, and J.-P. Bouchaud. Agent-based models for latent liquidity and concave price impact. *Phys. Rev. E*, 89:042805, 2014a.
- I. Mastromatteo, B. Tóth, and J.-P. Bouchaud. Anomalous impact in reaction-diffusion models, 2014b.
- E. Moro, J. Vicente, L. G. Moyano, A. Gerig, J. D. Farmer, G. Vaglica, F. Lillo, and R. N. Mantegna. Market impact and trading profile of hidden orders in stock markets. *Physical Review E*, 80(6):066102, 2009.
- Christian Y Robert and Mathieu Rosenbaum. A new approach for the dynamics of ultra-high-frequency data: The model with uncertainty zones. *Journal of Financial Econometrics*, 9(2):344–366, 2011.
- Ioanid Roşu. Multi-stage game theory in continuous time. Technical report, Technical report, Working Paper, University of Chicago, 2006.
- Ioanid Roşu. Liquidity and information in order driven markets. *Available at SSRN 1286193*, 2014.
- Ioanid Roşu. A dynamic model of the limit order book. *Review of Financial Studies*, page hhp011, 2009.
- Paul A Samuelson. *Foundations of Economic Analysis, enlarged edition*. Harvard University Press Harvard, 1983.
- Robert J Shiller. Do stock prices move too much to be justified by subsequent changes in dividends?, 1980.
- Eric Smith, J Doyne Farmer, László Gillemot, and Supriya Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative finance*, 3(6):481–514, 2003.
- N. Torre and M. Ferrari. Market impact model handbook, BARRA inc., Berkeley. <http://www.barra.com/newsletter/nl166/miminl166.asp>, but see also Grinold and Kahn (2000), 1997.
- B. Tóth, Y. Lempérière, C. Deremble, J. De Lataillade, J. Kockelkoren, and J.-P. Bouchaud. Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2):021006, 2011.
- Leon Walras. *Elements of pure economics*. Routledge, 2013.