

# Relation between Bid-Ask Spread, Impact and Volatility in Double Auction Markets

Matthieu Wyart<sup>\*,†</sup>, Jean-Philippe Bouchaud<sup>\*</sup>, Julien Kockelkoren<sup>\*</sup>,  
Marc Potters<sup>\*</sup>, Michele Vettorazzo<sup>\*</sup>

March 13, 2006

<sup>\*</sup> Science & Finance, Capital Fund Management, 6 Bvd Haussmann, 75009 Paris, France

<sup>†</sup> now at: Physics Department, Harvard University, Cambridge, 02 138 MA, USA

## Abstract

We argue that on electronic markets, limit and market orders should have equal effective costs on average. This symmetry implies a linear relation between the bid-ask spread and the average impact of market orders. Our empirical observations on different markets are consistent with this hypothesis. We then use this relation to justify a simple, and hitherto unnoticed, proportionality relation between the spread and the volatility *per trade*. We provide convincing empirical evidence for this relation. This suggests that the main determinant of the bid-ask spread is adverse selection, if one considers that the volatility per trade is a measure of the amount of ‘information’ included in prices at each transaction. Symmetry between market and limit orders stems from the self-organization of liquidity in electronic markets. Our results appear to hold approximately on liquid specialist markets as well, although the spread is significantly larger.

## 1 Introduction

One of the most important attribute of financial markets is to provide liquidity to investors [1], who may convert cash into stocks and vice-versa nearly instantaneously whenever they choose to do so. Of course, some markets are more liquid than others and the liquidity of a given market varies in time and can in fact dramatically dry up in crisis situations. How should markets be organized, at the micro-structural level, to optimize liquidity, and to favor steady and orderly

trading by avoiding these liquidity crises? In the past, the burden of providing liquidity was given to market makers, or specialists. In order to ensure steady trading, the specialists alternatively sell to buyers and buy to sellers, and re-tribute themselves through the so-called bid-ask spread – i.e. the price at which they sell to the crowd is always slightly larger than the price at which they buy. The determinants of the value of the spread in specialists markets have been the subject of many studies in the economics literature [2, 3, 4, 5, 6, 7, 8, 9]. However, most financial markets have nowadays become electronic (with the notable exception of the New-York Stock Exchange, NYSE). In these markets, liquidity is *self-organized*, in the sense that all agents can choose to be liquidity consumers or liquidity providers. More precisely, any agent can provide liquidity by posting *limit orders*: these are propositions to sell (or buy) a certain volume of shares or lots at a minimum (maximum) price fixed by the emitter. Limit orders are stored in the order-book. At a given instant in time, the best offer on the sell side (the ‘ask’) is higher than the best price on the buy side (the ‘bid’) and no transaction takes place. For a transaction to occur, an agent must consume liquidity by launching a *market order* to buy or to sell a certain number of shares; the transaction occurs at the best available price, provided the volume in the order book at that price is enough to absorb the incoming market order. Otherwise, the price ‘walks up’ (or down) the ladder of offers in the order book, until the order is fully satisfied. The liquidity of the market is partially characterized by the bid-ask spread  $S$ , which sets the cost of an instantaneous round-trip of one share (a buy instantaneously followed by a sell).<sup>1</sup> A liquid market is such that this cost is small. A question of both theoretical and practical crucial importance is to know what fixes the magnitude of the spread in the self-organized set-up of electronic markets.

In large fraction of the economics literature [2, 3, 4, 5], liquidity providers are described as market makers who earn profit from the spread. In a competitive framework the spread should be determined by requiring that such a strategy has zero gain on average. The resulting spread is non zero because this strategy has costs. Three types of cost are discussed in the literature: (i) order processing costs; (ii) adverse selection costs: sometimes, liquidity takers have superior information on the future price of the stock, in which case the market maker loses money;<sup>2</sup> (iii) inventory risk: market makers may temporarily accumulate large long or short positions which are risky. If agents are risk-sensitive this adds extra-costs. Theoretical models that account for these costs typically introduce a rather large amount of free parameters (such as risk-aversion to risk, fraction of informed trades, fraction of patient/ impatient traders, etc.) which cannot be measured directly. In order to extract the different determinants of the spread

---

<sup>1</sup>Other determinants of liquidity discussed in the literature are the depth of the order book and market resiliency [10, 11].

<sup>2</sup>This is also discussed as the free option trading problem in the literature, see e.g. [12] and refs. therein.

from empirical data, some drastic assumptions must be made. For example, assuming the order flow to be short-ranged correlated, Huang and Stoll [8] find that 90% of the spread is associated to order processing costs, and not to adverse selection (which is often found to have, within this framework, a negative contribution to the spread!). This is surprising, both because these processing costs are objectively at least ten times smaller than the spread, in particular on electronic markets, where the spread is found to be comparable to the spread in markets with specialists. One reason for the discrepancy is the assumption that the order imbalance has only short-ranged correlation, and therefore that impact is permanent, in striking disagreement with empirical data, where the order flow is found to be a long-memory process [13, 14], and price impact transient, but decaying very slowly [13, 15].

Some important ingredient thus seems to be missing. Still, both adverse selection and inventory risk suggest a positive correlation between the spread and the volatility of the traded asset. This makes perfect intuitive sense, and the aim of the present paper is to clarify in detail the origin of this relation. Positive correlation between spread and volatility is indeed found empirically (see [16, 4, 17, 18, 19, 20, 21]), but is not particularly spectacular and stands as one among other correlations, e.g., with traded volume, flow of limit orders, market capitalization, etc. Here, we want to argue theoretically, and demonstrate empirically on different markets, that there is in fact a very strong correlation between the spread and the volatility *per trade*, rather than with the volatility *per unit time*. Such a relation was first noted on the case of France-Telecom [13], and independently on the stocks of the FTSE-100 [22], but no theoretical argument was given in favor of this relation.

From a theoretical point of view, several statistical models of limit and market order flows have been analyzed to understand the distribution of the bid-ask spread, and relate its average value to flow and cancellation rates [3, 23, 24, 25, 27, 26, 28, 29]. Some models include strategic considerations in order placement and look for a trade-off between the cost of delayed execution and that of immediacy, but suppose that the price dynamics is bounded in a finite interval [24], therefore neglecting the long term volatility of the price (see also [28, 29]). As such, these finite band models have nothing to say about the spread-volatility relationship. Another line of models discards all strategic components (“Zero intelligence models”) and assume Poisson rates for limit orders, market orders and cancellation [25, 27, 26]. One can then compute both the average bid-ask spread and the long-term volatility as a function of these Poisson rates, and compare these predictions with empirical data [30]. The problem with such models is that although the *order flow* is completely random, the persistence of the order book leads to strong non-diffusive short term predictability of the *price*, which would be very easily picked up by high frequency execution engines by adequately conditioning the order flow (limit versus market orders) to optimize

execution costs.<sup>3</sup> In other words, as shown in [13, 14, 32], there are in fact very strong high frequency correlations in the order flow coming from the ‘hide and seek’ game played by buyers and sellers around the order book. For small tick stocks, the total available volume in the order book at any instant in time is in fact extremely small, on the order of  $10^{-5} - 10^{-4}$  of the market capitalization, because liquidity providers want to avoid giving a free trading option to informed traders. Conversely, this state of affairs forces liquidity takers to cut their total order in small chunks [33]. This necessarily creates long term correlation in the order flow, which in turn entices more latent liquidity to reveal itself – all traders seeking to optimize their execution costs (see e.g. [34]). This precisely means that on electronic markets the cost of limit and market orders should be very similar.

In the following, we show that this important average *symmetry* between market and limit orders allows us to relate the average price impact of a market order and the bid-ask spread, a relation that we check empirically. We then argue that the impact function must be related to the volatility per trade, a property that we again check on data. This allows us to establish a proportionality relation between the spread and volatility *per trade*, which holds both across different stocks and for a given stock across time, both on electronic markets *and* on the NYSE. Our symmetry argument should be very general and robust, and shows that in a competitive electronic market the bid-ask spread can in fact only come from “adverse selection”, provided one extends this notion to account for the fact that a trade can be uninformed but still impact the price. What is relevant here is that *any* unexpected component of the market order flow, whether it is truly informed or just random, impacts the price and creates a cost for limit orders, which must be compensated by the spread (see [13, 15]).

## 2 Limit orders *vs* market orders and market impact

### 2.1 A market making strategy

Our aim is to compare the relative costs of limit and market orders. To do so we compute the gain of a simple market making strategy which consists in participating to all trades through limit orders. If this strategy has a positive gain, then limit orders are clearly more favorable than market orders. Otherwise market orders are preferable. More precisely, we consider a market maker with a time horizon  $T$  (where  $T$  is measured in trade time) who provides an infinitesimal fraction  $\phi$  of the total available liquidity. The market maker posts limit orders both at the bid and at the ask. We assume that he participates to a fraction  $\phi$

---

<sup>3</sup>More elaborated ‘weak intelligence’ models have been studied recently, see [31].

of all traded volume. His gains will come from the spread  $S$ , which is in general time dependent, except on large tick markets where the spread is nearly always equal to one tick. Since this market making strategy generates typical inventory imbalances of order  $\sqrt{T}$  which is risky, we impose that the market maker has a perfectly balanced inventory at the end of the period  $T$ . As we show now, this has a cost.

The optimal strategy to use to ensure that the market maker's position is flat at time  $T$  is an interesting problem in itself, but its solution depends on the statistics and correlations in the order flow. We will thus assume a sub-optimal, but universal strategy, which is to use a market order to sell back (or buy back) at time  $T$  the whole position he has accumulated over the trading period  $[0, T - 1]$ . This suboptimal strategy will lead to a lower bound on the gains, and therefore an upper bound on the spread which will turn out to be in very good agreement with empirical data.<sup>4</sup> We will call  $v_i$  the volume of the  $i$ th market order, and  $\epsilon_i$  the sign of that market order ( $\epsilon = +1$  for a buy and  $\epsilon = -1$  for a sell). The volume  $V$  and its sign  $\varepsilon$  accumulated by the market maker is defined as:

$$\varepsilon V \equiv -\phi \sum_{i=0}^{T-1} \epsilon_i v_i; \quad \varepsilon = \pm 1 \quad (1)$$

We assume that  $\phi \ll 1$  is sufficiently small to neglect the impact of the market maker when he sells (buys) back at time  $T$  at the current mid-quote  $P_T$  (plus or minus half the current spread  $S_T$ ). Even in that case, unwinding the position is costly on average because the change of price  $P_T - P_0$  is *anti-correlated* with  $\varepsilon V$ : when the crowd buys, the price is impacted and goes up while the market making strategy accumulates a short position which is costly to buy back at time  $T$ , and vice-versa.

How can one measure the impact of a market order? Suppose a market order of size  $v_i$  and of sign  $\epsilon_i$  is launched at a trade-time  $i$ ; the subsequent mid-point price change at time  $i + \ell$  is a random variable  $P_{i+\ell} - P_i$  which is slightly biased in the direction of the initial trade. (We define  $P_i$  as the mid-quote immediately *before* the  $i$ th trade.) We shall define, following [13], the average impact  $\mathcal{R}(\ell, v)$  as:

$$\mathcal{R}(\ell, v) = \langle \epsilon_i \cdot (P_{i+\ell} - P_i) \rangle|_{v_i=v}, \quad (2)$$

where  $\langle \dots \rangle$  means that we take an empirical average of the bracketed quantity. The function  $\mathcal{R}(\ell, v)$  was studied in detail in [13]. To a good level of approximation,  $\mathcal{R}(\ell, v)$  can be written in a factorized form:  $\mathcal{R}(\ell, v) \equiv R(\ell)f(v)$ , where  $f(v)$  is a strongly concave function, and  $R(\ell)$  an increasing function of  $\ell$  which varies by a factor of  $\sim 2$  when  $\ell$  increases from 1 to several thousands (corresponding to a few days of trading). The shape of  $R(\ell)$ , averaged over a collection of different

---

<sup>4</sup>Determining the optimal market-making strategy when the order flow is long-ranged correlated would be an interesting result; we however expect that the final expression of the gain would only change by prefactors of order unity.

stocks, is shown in Fig. 1 (see caption for more details). The gain  $\mathcal{G}$  of the market maker can be obtained by noting that if  $\epsilon_i$  is the sign of the market order, the transaction price is  $P_i + \epsilon_i S_i/2$ , where  $S_i$  is spread prevailing just before the market order is emitted.<sup>5</sup> Therefore the total gain of this strategy is given by:

$$\begin{aligned} \mathcal{G} &= \phi \left[ \sum_{i=0}^{T-1} \epsilon_i v_i (P_i + \epsilon_i \frac{S_i}{2}) - (\sum_{i=0}^{T-1} \epsilon_i v_i) (P_T - \epsilon \frac{S_T}{2}) \right] \\ &\equiv \phi \left[ \frac{1}{2} \sum_{i=0}^{T-1} v_i S_i - \sum_{i=0}^{T-1} \epsilon_i v_i (P_T - P_i) \right] - \frac{V S_T}{2}. \end{aligned} \quad (3)$$

This expression shows clearly that the gain of the market maker is determined by the competition between the spread (first term) and the impact of market orders (second term). We will neglect the last term for  $T$  large, since  $V$  grows sub-linearly with  $T$ ,<sup>6</sup> while the first two terms grow linearly with  $T$ . Using the above definition of the average impact function, one obtains for the average gain of the market making strategy, per unit of traded volume and per unit time, the following upper bound:

$$g = \frac{\langle \mathcal{G} \rangle}{T \phi} \leq \frac{1}{2} \langle v S \rangle_v - \tilde{R}(T) \quad (4)$$

where we introduced the notation  $\tilde{R}(\ell) \equiv \ell^{-1} \sum_{k=1}^{\ell} \langle v \mathcal{R}(k, v) \rangle_v$ ;  $\tilde{R}(\ell)$  is plotted in Fig. 1 and has a shape similar to  $R(\ell)$  itself.

Eq.(4) is our central result; it shows that the market maker gain can be computed entirely from empirical data, without having to make any assumption on the fraction of informed trades. In fact, trades need not be informed at all; provided their average impact is non zero, these trades inflict losses upon market makers.

From the factorization property of  $\mathcal{R}(\ell, v)$ , one deduces that  $\tilde{R}(\ell)$  can be expressed as:

$$\tilde{R}(\ell) \equiv \lambda(\ell) \langle v \mathcal{R}(1, v) \rangle_v, \quad (5)$$

where  $\lambda$  is a  $\ell$  dependent number between 1 and  $\approx 2$  (see Fig. 1). Therefore, the gain of the market making strategy with horizon  $T$  can be put in a form that we will use below:

$$g \leq \frac{1}{2} \langle v S \rangle_v - \lambda(T) \langle v \mathcal{R}(1, v) \rangle_v. \quad (6)$$

Both the half spread term  $V S_T/2$  and any market impact, neglected in the above formula, reduce the gain of the market maker and therefore reinforce the inequality.

---

<sup>5</sup>The following computation neglects the fact that one single large market order may trigger transactions at several different prices, up the order book ladder. Nevertheless this situation is empirically quite rare, and corresponds to only a few percents of all cases [35].

<sup>6</sup>If the  $\epsilon_i$  were short term correlated, one would have  $V \sim \sqrt{T}$ . But because the  $\epsilon_i$  have correlations decaying with lag  $\ell$  as  $\ell^{-\gamma}$  with  $0 < \gamma < 1$  [13, 14], one rather finds  $V \sim T^{1-\gamma/2} \ll T$ .

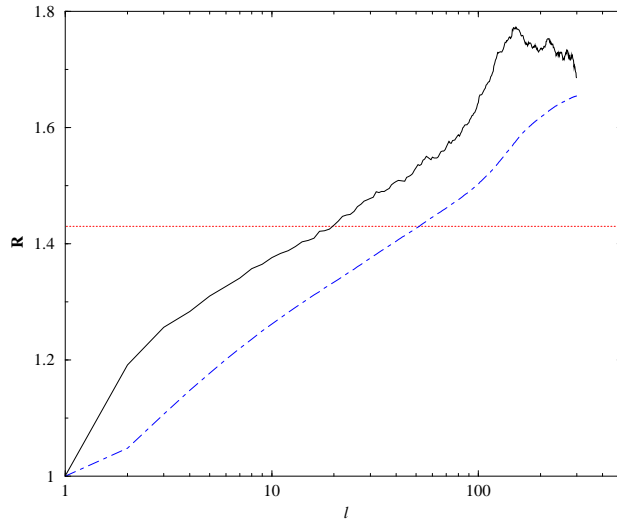


Figure 1: Average over 68 PSE stocks of the impact function  $R(\ell)$  as a function of  $\ell$  (plain line). The average is performed by rescaling the individual  $R(\ell)$  such that  $R(\ell = 1) \equiv 1$ , and by rescaling  $\ell$  by the average daily number of trades. Dash-dotted line: average integrated response  $\tilde{R}(\ell)$ , again normalized to unity for  $\ell = 1$ . The value  $\lambda = 1.43$ , extracted from Fig. 3 below, is also shown as a horizontal line.

## 2.2 Limit/market order symmetry

The gain  $g$  of the above simple market making strategy also corresponds, by definition, to the average cost of a market order. In electronic markets, any agents can choose to execute through limit or market orders, or any mixture of those. Thus, both types of orders must have very similar cost. Any imbalance is easy to detect because of the very large number of orders per day that allows quantitative trading houses to obtain reliable estimates of these costs. If market orders were too costly, limit orders would be preferred, leading to a reduction of the spread and therefore a reduction of the cost of market orders.<sup>7</sup> If limit orders were too costly, more market orders would be used, leading to a widening of the average spread. Therefore the spread should stabilize such that one average limit orders and market orders have an equal cost. This *symmetry* assumption appears reasonable in markets without privileged operators who can manipulate the order flow, which means that it should even hold for liquid stocks of the NYSE where the large flow of limit orders not coming from the specialists insures some sort of equilibrium. This equal cost assumption is equivalent to  $g \equiv 0$ , or, using the above result:

$$\langle Sv \rangle_v \geq 2\lambda(T) \langle v\mathcal{R}(1, v) \rangle_v \quad (7)$$

<sup>7</sup>This argument is valid for small tick contracts. On contracts with large ticks, the spread is artificially bounded from below. See the discussion in section 2.5

Before comparing this prediction with empirical results, let us consider a simple market where the spread would be constant and all market orders would have the same size  $v$  with independent signs  $\epsilon$ . In this case, the impact function is time independent [13], and Eq.(7) reduces to:

$$\mathcal{R} = \frac{1}{2}S \quad (8)$$

In economical terms, this equality has a very simple meaning: it indicates that on average, the new mid-price after the transaction  $P_{i+1} = P_i + \epsilon_i \mathcal{R}$  is equal to the last transaction price  $P_i + \epsilon_i S/2$ .

### 2.3 Comparison with empirical data: small tick electronic markets

We first consider small tick electronic markets, such as the Paris Stock Exchange (PSE) or Index Futures. The case of large tick stocks is different since in this case the spread is (nearly) always one tick, with huge volumes at both the bid and the ask. The case of such markets will be considered below.

We studied extensively the set of the 68 most liquid stocks of the PSE during the year 2002. The summary statistics describing these stocks is given in the Appendix. From the Trades and Quotes data, one has access the the bid-ask just before each trade, from which one can obtain the sign and the volume of each trade (depending on whether the trade happened at the ask or at the bid) and the mid-point just before the trade. From this information, one computes the quantities of interest, such as the instantaneous impact function  $\mathcal{R}$  or the spread  $S$ . Note that we have removed ‘block trades’, which appear as transactions with volumes larger than what is available at the best price that are not followed by a change of quotes. This represent typically a 5 – 10% fraction of the total number of trades.

We test Eq.(7) in two different ways – for a given stock across time, and across all different stocks. Since both the spread and impact vary with time, one can measure ‘instantaneous’ quantities by averaging for a given stock  $\langle Sv \rangle_v / \langle v \rangle_v$  and  $\langle \mathcal{R}(1, v)v \rangle_v / \langle v \rangle_v$  over a number of successive trades. In the example of Fig. 2, each point corresponds to an average over 10000 non overlapping trades, corresponding to 2 days of trading in the case of France Telecom in 2002. Doing so we obtain quantities that vary by a factor 5 that allows us to test the linear dependence predicted by Eq.(7). Our result shown in Fig.2 is in very good agreement with the theoretical bound; note that the bound is violated only on rather rare occasions, even on averages over rather short time scales. A linear fit gives a slope equal to 2.14, corresponding to  $\lambda = 1.07$ .<sup>8</sup> As expected for highly liquid

---

<sup>8</sup>Note that in the rest of the paper we will always refer to linear fits as meaning linear regressions with imposed zero intercept, except stated otherwise.



stocks, the bound  $\lambda \geq 1$  is nearly saturated, meaning that providing liquidity is not well rewarded in this case.

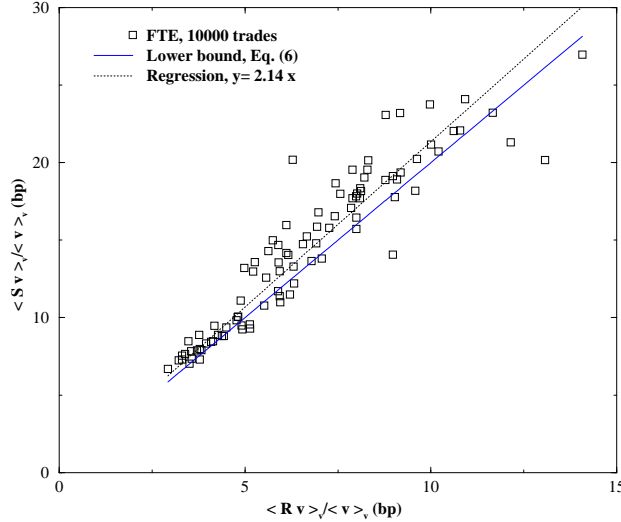


Figure 2: Test of Eq. (7) for France Telecom in 2002. Each point corresponds to a pair  $(\langle Sv \rangle / \langle v \rangle, \langle Rv \rangle / \langle v \rangle)$ , computed by averaging over 10000 non overlapping trades ( $\sim$  two trading days). Both quantities are expressed in basis points. We also show our bound, Eq. (7), with  $\lambda = 1$ , and a linear fit that gives an effective value of  $\lambda \approx 1.07$ .

We also test Eq.(7) cross sectionally in Fig. 3, using the above 68 different stocks of the PSE. The relative values of the spread and the average impact also varies by a factor 5 between the different stocks, which enables to test the linear relation (7). Once again we find a good agreement with the predicted bound, and the linear fit gives a slope of 2.86, or  $\lambda \approx 1.43$ , corresponding to a market making horizon of roughly 20 trades (see Fig. 1).

It is also interesting to analyze small tick Futures markets, for which the typical spread is ten times smaller than on stock markets. We have studied a series of small tick Index Futures in 2005 (except the MIB for which the data is 2004), again both as a function of time and across the 7 indexes of our set. Results are shown in Fig. 4; the bound is again very well obeyed both across contracts and across time, even when the time averaging is restricted to only 1000 consecutive trades. This shows that on these highly liquid contracts, where the transaction rate is as high as a few per second, the equilibrium between limit and market orders is reached very quickly.

## 2.4 Comparison with empirical data: NYSE stocks

The case of the NYSE is quite interesting since the market is still ruled by specialists, who however compete to provide liquidity with other market participants

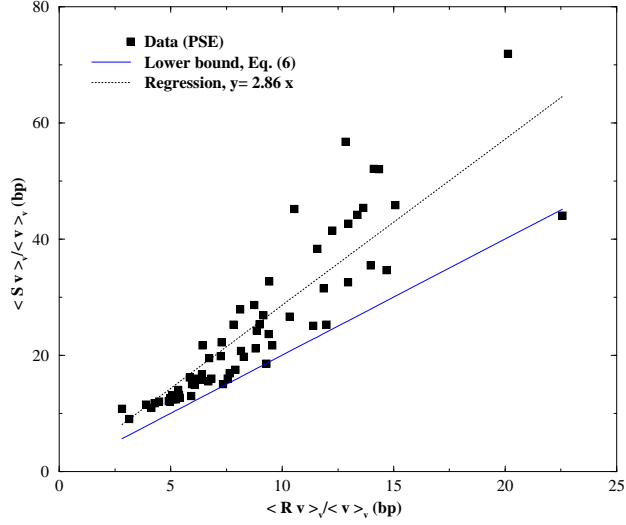


Figure 3: Test of Eq. (7) for 68 stocks of the Paris Stock Exchange in 2002. Each point corresponds to a pair  $(\langle Sv \rangle_v / \langle v \rangle_v, \langle \mathcal{R}v \rangle_v / \langle v \rangle_v)$ , computed by averaging over the year. Both quantities are expressed in basis points. We also show our bound, Eq. (7), with  $\lambda = 1$ , and a linear fit that gives an effective value of  $\lambda \approx 1.43$ , corresponding to a market making horizon of roughly 20 trades (see Fig. 1).

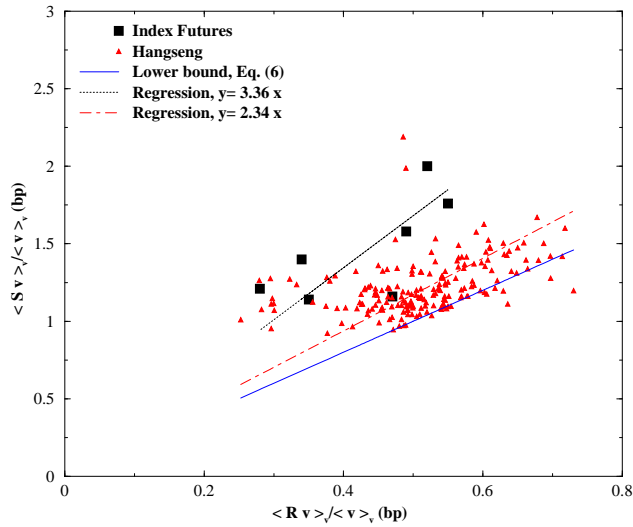


Figure 4: Test of Eq. (7) for small tick Index Futures in 2005: CAC, DAX, FTSE, IBEX, MIB, SMI, HANGSENG. Each black square corresponds to a pair  $(\langle Sv \rangle_v / \langle v \rangle_v, \langle \mathcal{R}v \rangle_v / \langle v \rangle_v)$ , computed by averaging over the year, while small triangles are computed by averaging over 1000 non overlapping trades on the HANGSENG futures. Both quantities are expressed in basis points. We also show our bound, Eq. (7), with  $\lambda = 1$ . A linear fit that gives an effective value of  $\lambda \approx 1.65$  across Index Futures, and  $\lambda \approx 1.17$  for the HANGSENG across time.

placing limit orders. We again test Eq.(7) cross sectionally, using the set of the 155 most actively traded stocks on the NYSE in 2005. We use the bid-ask quote posted by the specialist. Once again we find a good agreement with the predicted bound, although in this case, the gap between the bound and the empirical points is substantially larger than in the case of the PSE. This translates into a significantly larger slope of 3.96, which suggests that, perhaps not surprisingly, market makers on the NYSE post spreads that are systematically over-estimated compared to the situation in electronic markets. This result is in agreement with the study of Harris and Hasbrouck performed in the nineties on the NYSE [37], which showed that limit orders were more favorable than market orders. Quite interestingly, our analysis allows us to compare in a simple and meaningful way the spread in different markets.

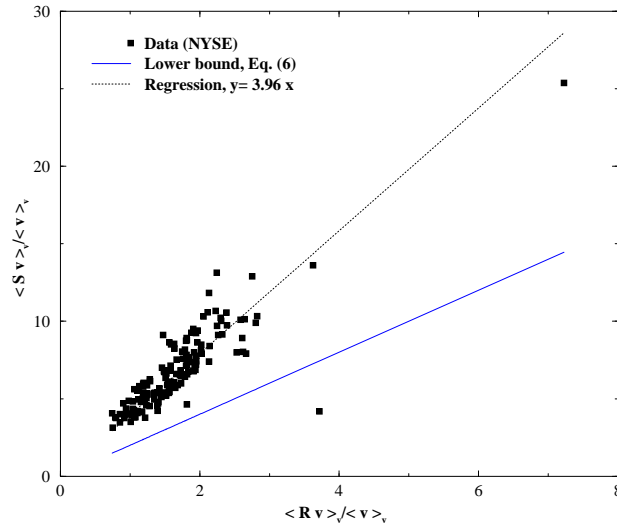


Figure 5: Test of Eq. (7) for stocks of the NYSE 2005. Each point corresponds to a pair  $(\langle Sv \rangle_v / \langle v \rangle_v, \langle Rv \rangle_v / \langle v \rangle_v)$ , computed by averaging over the year. Both quantities are expressed in basis points. We also show our bound, Eq. (7), with  $\lambda = 1$ , and a linear fit that gives an effective value of  $\lambda \approx 1.97$ , quite significantly larger than for the PSE.

## 2.5 The case of large tick electronic markets

The string of arguments leading to Eq. (7) does not directly apply in the case where the tick size is large. In that case the spread  $S$  is most of the time stuck to its minimum value, i.e. one tick, while the size of the queue  $q$  at the bid and at the ask tends to be extremely large (see e.g. Appendix, Table 3). Because of the large value of the spread, limit orders are *a priori* favorable, but huge limit order volumes accumulate as liquidity providers attempt to take advantage of the

spread. The size of the queue  $q$  at the bid or at the ask is thus much larger than the typical value of the traded volume at each transaction  $v$ :  $v/q \approx 0.01$  (see Table 3), to be compared with  $v/q \approx 0.2 - 0.3$  (see Appendix, Table 2) for smaller tick stocks. Therefore, the simple market making strategy considered above, which assumes that one can participate to a small fraction of *all* transactions, cannot be implemented. We thus expect that the spread on these markets will be substantially larger than predicted by the bound Eq.(7), because the competition between liquidity providers, that acts to reduce the spread, cannot operate. We indeed find that the ratio between  $\langle Sv \rangle_v$  and  $\langle \mathcal{R}v \rangle_v$  is large for large tick stocks. For example, in the case of Ericsson, during the period March-November 2004, for which the tick size is  $\sim 50$  bp, we find  $\langle Sv \rangle_v / \langle \mathcal{R}v \rangle_v \approx 4.5$ , corresponding to an effective value of  $\lambda \approx 2.25$ , even larger than on the NYSE.

## 2.6 Comparison with empirical data: conclusion

Our empirical analysis shows that on small tick electronic markets, an approximate symmetry between limit and market orders indeed hold, in the sense that the spread is very nearly equal to minimum value needed for market making to be profitable. On the NYSE, spreads appears to be significantly larger but the predicted linear relation between spread and impact still holds. Similarly, a large value of the tick size thwarts the full expression of the symmetry between market and limit orders.

## 3 Liquidity *vs.* volatility

### 3.1 A simple model

Consider again the simple model discussed above where all market orders have the same size and are not correlated in time. In this case, the impact function  $\mathcal{R}$  is constant over time [13], and the volatility *per trade*  $\sigma_1$  is therefore easily found to be  $\sigma_1 = \mathcal{R}$ . Using Eq.(8) one obtains:

$$\sigma_1 = \frac{S}{2} \tag{9}$$

In what follows we argue that this simple idea is still valid in the more realistic case where market orders are distributed in volume and, more importantly, have long term correlations in sign.

### 3.2 The situation in real markets

It is now a well established fact that order flow has long term memory, in the sense that the correlation function of the market order signs  $C(\ell) = \langle \epsilon_i \epsilon_{i+\ell} \rangle$  decays for

large  $\ell$  as  $\ell^{-\gamma}$ , with an exponent  $\gamma$  less than unity, i.e. the sign correlation function is in fact *non summable* [13, 14].<sup>9</sup> Naively, this should lead to strong correlations in the price dynamics. However, these correlations are nearly perfectly offset by market resiliency, i.e. an adapted reaction of the limit order flow and the dynamics of quotes to the flow of market orders [13, 32, 14]. As a result, the price is nearly perfectly diffusive.

By definition of the instantaneous impact and of the volatility per trade  $\sigma_1^2 = \langle (P_{\ell+1} - P_\ell)^2 \rangle$ , one has:

$$\sigma_1^2 \equiv \langle \mathcal{R}_1^2(v) \rangle_v. \quad (10)$$

The instantaneous impact  $\mathcal{R}_1(v)$  at fixed  $v$  is expected to fluctuate over time for two reasons. First, the impact appears to follow the liquidity of the market measured by the spread (see section 2). On the PSE, the spread has a distribution close to an exponential, hence one has  $\langle S^2 \rangle = a \langle S \rangle^2$  with  $a \approx 2$  (see Table 2, Appendix).<sup>10</sup> Second, recall that the impact is measured between two market orders, in between which a lot may happen. Large impact fluctuations could arise from external news, which may induce large quote revisions due to cancellation of limit orders but with no transactions, resulting in large jumps of the price, unrelated to the trading activity. In order to account for this possibility, we write:

$$\sigma_1^2 = \langle \mathcal{R}_1^2(v) \rangle_v = a' \mathcal{R}_1^2 + \Sigma^2; \quad \text{with} \quad \mathcal{R}_1 \equiv \langle \mathcal{R}_1(v) \rangle_v. \quad (11)$$

A specific model for this Einstein-like relation was worked out in [13], and tested on France-Telecom (see also [38]). Here, we establish that this relation holds quite precisely across different stocks: see Fig. (6). Perhaps surprisingly, the exogenous quote revision contribution  $\Sigma^2$  appears to be small. This might be related to the observation made in Farmer et al. [35] that for most price jumps, some limit orders are too slowly cancelled and get ‘grabbed’ by fast market orders, which means that these events in fact contribute to  $\mathcal{R}_1$ . In the following, we will therefore neglect  $\Sigma^2$ , as suggested by Fig. (6): in this sense the volatility of the stock can be fully explained by market impact.

Our final assumption is that of *universality*, i.e. when the tick size is small enough and the typical number of shares traded is large enough, all stocks should behave identically up to a rescaling of the average spread and the average volume. The universality of the shape of the order book was indeed checked to hold rather well in [27]. This means that, for example,

$$\langle vS \rangle_v = b \langle v \rangle_v \langle S \rangle_v, \quad (12)$$

where  $b$  is stock independent. Similarly,

$$\langle v \mathcal{R}_1(v) \rangle_v = b' \langle v \rangle_v \mathcal{R}_1, \quad (13)$$

---

<sup>9</sup>These long ranged correlations were also noted in e.g. [19, 36], but the detailed shape of  $C(\ell)$  was not investigated.

<sup>10</sup> $a$  may take different values on other markets [31], but the value of  $a$  is irrelevant for the following discussion.

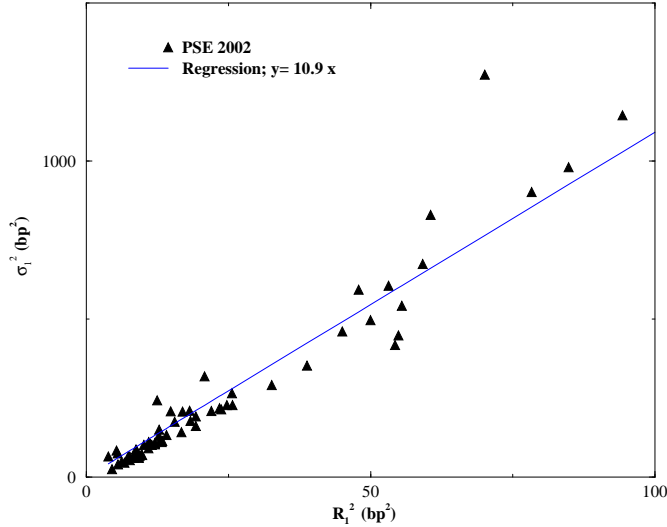


Figure 6: Plot of  $\sigma_1^2$  vs.  $\mathcal{R}_1^2$ , showing that the linear relation Eq. (11) holds quite precisely with  $\Sigma^2 \approx 0$  and  $a' \approx 10.9$ . (The intercept of the best affine regression is even found to be slightly negative). Data here corresponds to the 68 stocks of the PSE in 2002.

where  $b'$  is also stock independent. Note that this assumption is consistent with the empirical observation of [39], where the impact function  $\mathcal{R}_1(v)$  for different US stocks could be rescaled onto a unique Master curve. We test Eqs. (12,13) in Fig. 7 in the case of the Paris Stock Exchange, from which we extract  $b \approx 1.02$  and  $b' \approx 1.80$ . Interestingly, we find that the volume and the spread are nearly uncorrelated, whereas the volume traded and the impact are correlated ( $b' > 1$ ).

Therefore, using Eq. (7) as an equality (as suggested by the empirical results of Section 2, and Eqs. (11,12,13)), we obtain the main result of this paper:

$$\langle S \rangle_v = c \sigma_1, \quad (14)$$

where  $c$  is a stock independent numerical constant, which can be expressed using the constants introduced above as  $c = 2\lambda b' / \sqrt{a'b}$ . This very simple relation between volatility *per trade* and average spread was noted in [13, 22], and we present further data in the next section to support this conjecture. Therefore, the constraints that (i) high frequency execution strategies impose that the price is diffusive [Eq. (10)], and (ii) the cost of limit and market orders are equal [Eq.(7)], lead to a simple relation between liquidity and volatility. As an important remark, note that the above relation is not expected to hold for the volatility *per unit time*  $\sigma$ , since it involves an extra stock-dependent and time-dependent quantity, namely the the trading frequency  $\nu$ , through:

$$\sigma = \sigma_1 \sqrt{\nu}. \quad (15)$$

We will discuss this issue further in section 4.

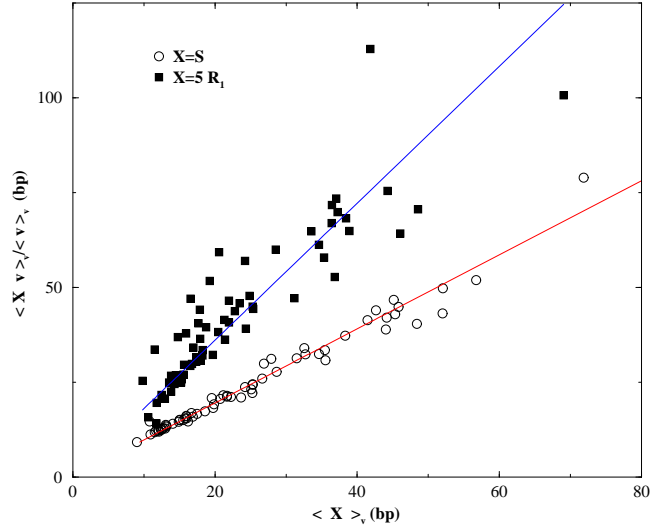


Figure 7: Plot of  $\langle vX \rangle_v / \langle v \rangle_v$  vs.  $\langle X \rangle_v$ , where  $X$  is either the spread  $S$  or the instantaneous impact  $\mathcal{R}_1$  (multiplied by a factor 5 for clarity). The quality of the linear regression tests our universality assumption, which is good for  $S$  and fair for  $\mathcal{R}_1$ . The value of  $b \approx 1.02$  and  $b' \approx 1.80$  are given by the slope of these regressions. Data here corresponds to the 68 stocks of the PSE in 2002.

### 3.3 Comparison with empirical data

Using the same data sets as in Sections 2.3, 2.4 and 2.5, we now test empirically the predicted linear relation between spread and volatility per trade, Eq. (14). The average spread  $\langle S \rangle$  is defined as the average distance between bid and ask immediately before each trade (and not as the average over all posted quotes). The volatility per trade is defined as the root mean square of the trade by trade return.<sup>11</sup> Our results for the Paris Stock Exchange are shown in Figs 8 and 9. We see that Eq. (14) describes the data very well. Interestingly, using the results obtained above across the PSE stocks, we have  $a' \approx 10.9$ ,  $b \approx 1.02$ ,  $b' \approx 0.53$ ,  $\lambda \approx 1.43$ , leading to  $c \approx 1.53$ , in close correspondence with the direct regression result  $c \approx 1.58$ . Similar results are obtained for Index futures (Figs. 10-a & b) or for the NYSE (Fig. 11), with values of  $c$  which are all very similar  $c \sim 1.2 - 1.6$ . We have also checked that there is an average intra-day pattern which is followed in close correspondence both by  $\langle S \rangle$  and  $\sigma_1$ : spreads are larger at the opening of the market and decline throughout the day. Note that the trading frequency  $\nu$  increases as time elapses, which, using Eq. (15), explains the familiar U-shaped pattern of the volatility per unit time.

<sup>11</sup>Since prices are very close to random walks, defining the volatility from returns defined on a longer time scale gives very similar results. On our set of PSE stocks, we find that  $\sigma_{128} / \sqrt{128} \approx 0.84\sigma_1$ , indicating a small anti-correlation of returns ( $\sim 15\%$ ) of short time scales.

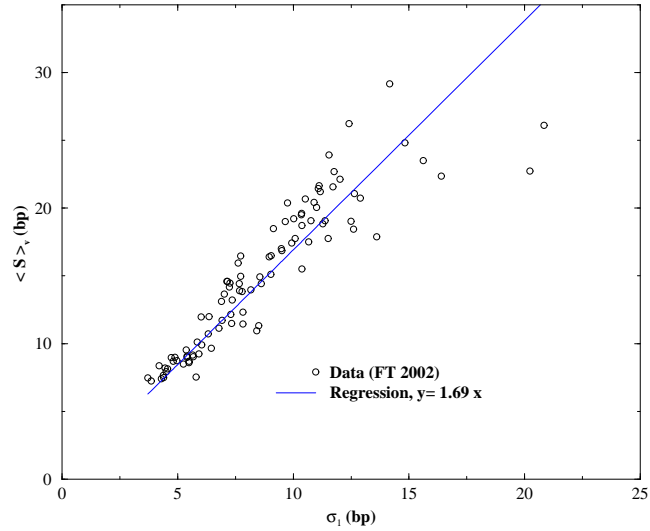


Figure 8: Test of Eq. (14) for France Telecom in 2002. Each point corresponds to a pair  $(\langle S \rangle, \sigma_1)$ , computed by averaging over 10000 non overlapping trades ( $\sim$  two trading days). Both quantities are expressed in basis points. From a linear fit, we find  $c \approx 1.69$ .

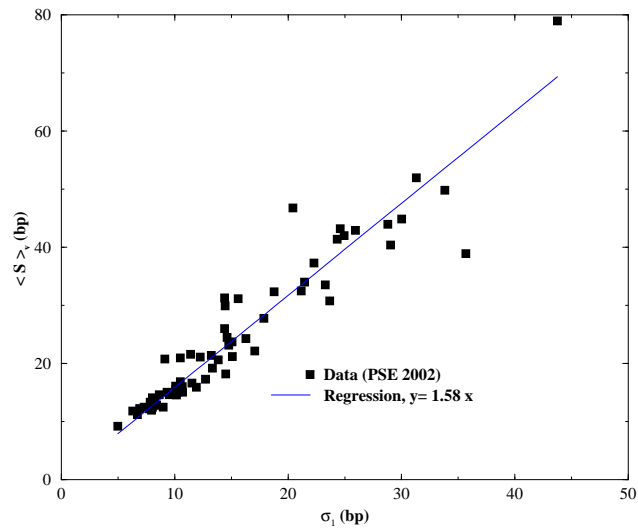


Figure 9: Test of Eq. (14) for 68 stocks from the Paris Stock Exchange in 2002, averaged over the entire year. The value of the linear regression slope is  $c \approx 1.58$ .



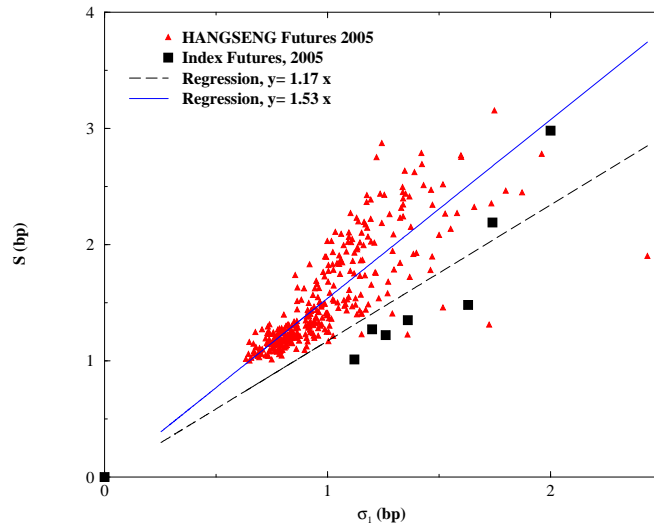


Figure 10: Test of Eq. (14) for the HANGSENG futures contract (triangles), and across small tick Index Futures in 2005: CAC, DAX, FTSE, IBEX, MIB, SMI, HANGSENG (squares). Each point corresponds to a pair  $(\langle S \rangle, \sigma_1)$ , computed by averaging either over 1000 non overlapping trades (triangles) or over the whole year (squares). From a linear fit, we find  $c \approx 1.53$  for the HANGSENG across time and  $c \approx 1.17$  across Index Futures.

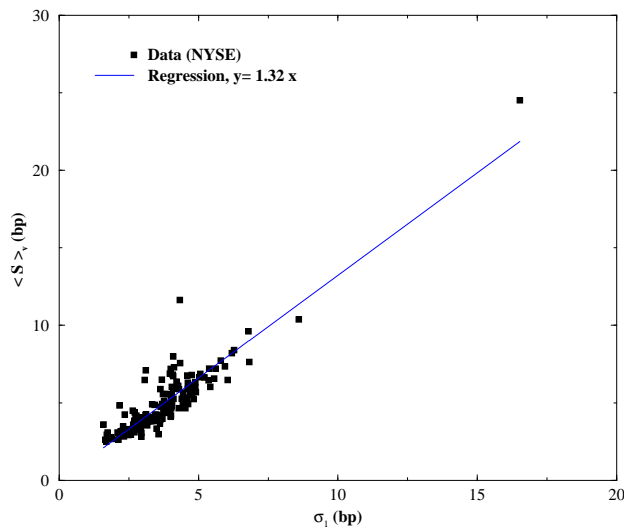


Figure 11: Test of Eq. (14) for stocks from the NYSE in 2005. Each point corresponds to a pair  $(\langle S \rangle, \sigma_1)$ , computed by averaging over the entire year. Both quantities are expressed in basis points. From a linear fit, we find  $c \approx 1.32$ .

## 4 Discussion and conclusion

The main theoretical result of this paper is the possibility to express the profit of simple market-making strategies directly in terms of the spread and the impact function. This relation does not require any assumption on the fraction of informed trades; in fact our result holds even if trades are all uninformed but impact the price. Imposing that the strategy leads to a zero average gain allows one to derive a linear equation between spread and impact. This relation is in good agreement with empirical data on small tick contracts, with a slope compatible with the zero gain hypothesis. This implies that limit and market orders have similar costs and therefore play symmetric roles in electronic markets, at least when the tick size is small. This result is in fact expected since all market participants can choose to use one type of order or the other; any persistent asymmetry would be easily detected and be corrected (see below). This means, quite interestingly, that minimization of execution costs leads to self-organization of liquidity in electronic markets (at least in normal conditions). Our analysis allows us to compare in an objective way the spreads in different markets and suggests that spreads are distinctly too large on the NYSE.

Making reasonable further assumptions, we have then shown that spread  $S$  and volatility per trade  $\sigma_1$  are also proportional, a result that we confirm empirically. This very simple relation sheds light on the debate in the literature about the determinants of the bid-ask spread, and suggests that the bid-ask spread is dominated by adverse selection, if one considers that the volatility per trade is a measure of the amount of ‘information’ included in prices at each transaction. There are two complementary economic interpretations of the relation  $\sigma_1 \sim S$  in small tick markets: (i) since the typical available liquidity in the order book is quite small, market orders tend to grab a significant fraction of the volume at the best price; furthermore, the size of the ‘gap’ above the ask or below the bid is observed to be on the same order of magnitude as the bid-ask spread which therefore sets a natural scale for price variations. Hence both the impact and the volatility per trade are expected to be of the order of  $S$ , as observed; (ii) the relation can also be read backwards as  $S \sim \sigma_1$ : when the volatility per trade is large, the risk of placing limit orders is large and therefore the spread widens until limit orders become favorable. Therefore, there is a clear two-way feedback that imposes the relation  $\sigma_1 \sim S$ , valid on average; any significant deviation tends to be corrected by the resulting relative flow of limit and market orders. Our result therefore appears as a fundamental property of the markets organization, which should be satisfied within any theoretical description of the micro-structure. Zero intelligence models [30], for example, fail to predict any universal relation between  $S$  and  $\sigma_1$ .

Our relation involves the volatility per trade whereas most of the econometric work has instead focused on the volatility per unit time  $\sigma$ . The relation between the two involves the trading frequency  $\nu$ , which is itself time- and stock-

dependent. As a function of time, we find, in agreement with [42], that volatility per trade and trading frequency are positively correlated; the volatility  $\sigma = \sigma_1 \sqrt{\nu}$  therefore increases because both  $\sigma_1$  and  $\nu$  increase.<sup>12</sup> Across stocks, on the other hand, the volatility per unit time does not exhibit systematic variations with capitalization  $\mathcal{C}$ :  $\sigma \sim \mathcal{C}^\varphi$ , whereas the trading frequency increases with capitalization as  $\nu \sim \mathcal{C}^\zeta$ . For stocks belonging to the FTSE-100, Zumbach finds  $\varphi \approx 0$  and  $\zeta \approx 0.44$  [22], while for US stocks the scaling for  $\nu$  is less clear [43]. Interestingly, our result then leads to a result between average spread and capitalization of the form  $S \sim \mathcal{C}^{\phi-\zeta/2}$ , in good agreement with Zumbach's data [22] and with the impact data of Lillo et al. [39].

The fundamental question at this stage is to know what fixes the volatility  $\sigma$  and the trading frequency  $\nu$ . Clearly, the trading frequency has to do with the available liquidity and the way large volumes have to be cut in small pieces. But is the volatility per unit time *the* primary object, driven by a fundamental process such as the arrival of news, to which the volatility per trade and therefore the spread is slaved? Or on the contrary is the market micro-structure and trading activity imposing, in a bottom-up way, the value of the volatility? Understanding these coupled dynamical problems appears to be a major challenge for the theory of financial markets, and an unavoidable step to understand the interrelation between liquidity and market efficiency [9, 44, 18, 19, 36, 13, 14, 45].

We want to thank J. D. Farmer, Th. Foucault and G. Zumbach for useful discussions.

## Appendix: Summary statistics

---

<sup>12</sup>The long-memory property of  $\sigma$  is actually argued to be related to long range correlation in the trading frequency rather than in the volatility per trade, see [41].

Code	Name	Code	Name
ACA	Credit Agricole	IFG	Infogrames Entertainment
AC	Accor	LG	Lafarge
AF	Air France-KLM	LI	Klepierre
AGF	Assurances Generales de France	LY	Suez
AI	Air Liquide	MC	LVMH
ALS	Alstom RGPT	ML	Michelin
ALT	Altran	MMB	Lagardere
AVE	Aventis	MMT	M6-Metropole Television
BB	Societe BIC	NAD	Wanadoo
BN	Groupe Danone	NK	Imerys
CAP	Cap Gemini	OGE	Orange
CA	Carrefour	OR	L Oreal
CDI	Christian Dior	PEC	Pechiney
CGE	Alcatel	PP	PPR
CK	Casino Guichard (pref.)	PUB	Publicis Groupe
CL	Credit Lyonnais	RF	Eurazeo
CNP	CNP Assurances	RHA	Rhodia
CO	Casino Guichard	RI	Pernod-Ricard
CS	AXA	RNO	Renault
CU	Club Mediterranee	RXL	Rexel
CY	Castorama Dubois	SAG	Safran
DEC	JC Decaux	SAN	Sanofi-Aventis
DG	Vinci	SAX	Atos Origin
DSY	Dassault Systemes	SCO	SCOR
EF	Essilor International	SC	Simco
EN	Bouygues	SU	Schneider Electric
FP	Total	SW	Sodexo Alliance
FR	Valeo	TEC	Technip
FTE	France Telecom	TFI	Television Francaise 1
GFC	Gecina	TMM	Thomson
GLE	Societe Generale	UG	Peugeot
GL	Galleries Lafayette	UL	Unibail
HAV	Havas	VIE	Veolia Environnement
HO	Thales	ZC	Zodiac

Table 1: Codes and names of the PSE stocks analyzed in Table 2.

Code	Turnover	$\langle q_t \rangle$	$\langle v \rangle$	# trade	$\sigma_1$	$\langle S \rangle$	$\sigma_S$	$\mathcal{R}_1$	Tick
ACA	15.4	78	11.1	347	9.12	20.76	18.28	2.30	5.13
AC	22.1	87	26.3	211	8.97	12.45	13.00	2.99	2.68
AF	4.0	40	7.2	139	14.40	25.96	25.57	3.85	7.01
AGF	9.6	80	16.9	142	13.32	19.17	18.88	4.28	5.46
AI	34.0	154	26.9	316	8.21	13.72	11.07	3.06	6.87
ALS	14.2	57	13.0	274	14.75	23.11	20.80	4.84	12.02
ALT	12.7	53	14.6	217	23.66	30.75	27.88	7.15	11.04
AVE	134.1	362	62.6	536	6.70	11.18	7.31	2.59	7.54
BB	0.8	30	11.9	16	28.80	43.94	47.51	7.78	2.55
BN	59.5	310	45.9	324	6.30	11.80	7.79	2.37	7.62
CAP	34.7	94	28.9	300	12.71	17.27	17.03	4.39	5.17
CA	79.8	229	40.1	497	6.92	12.20	9.16	2.48	5.35
CDI	5.1	59	14.9	86	18.76	32.32	29.08	6.23	2.79
CGE	79.8	152	21.3	936	10.49	20.94	14.85	3.31	15.37
CK	0.3	37	11.4	6	20.43	46.72	32.24	7.37	8.06
CL	30.1	163	23.0	328	8.05	13.60	13.04	1.96	3.43
CNP	1.7	41	9.3	45	15.58	31.12	33.74	3.53	2.67
CO	13.8	98	22.9	151	10.38	15.43	14.42	3.47	6.58
CS	96.1	144	36.6	657	8.32	12.78	10.17	3.13	6.39
CU	1.0	22	7.7	32	30.02	44.83	47.65	8.85	3.97
CY	15.4	2148	77.1	50	8.64	14.61	14.49	2.35	8.03
DEC	0.7	27	12.7	13	43.75	78.92	77.84	13.81	8.15
DG	16.8	134	25.4	165	8.02	14.05	10.90	2.76	7.76
DSY	10.4	65	19.6	133	17.06	22.11	23.83	5.71	5.28
EF	5.6	59	20.4	69	13.21	21.38	23.01	3.94	2.52
EN	17.8	66	21.3	210	10.15	14.55	14.35	3.47	3.46
FP	322.4	662	114.4	705	4.96	9.17	4.86	2.12	6.71
FR	8.3	80	24.1	86	13.84	20.60	22.46	4.39	3.50
FTE	112.7	137	29.5	956	9.12	15.04	12.57	2.93	6.39
GFC	0.3	35	9.9	7	17.85	27.74	31.55	4.56	5.62
GLE	82.7	239	46.0	449	8.07	12.70	9.27	3.04	7.07
GL	0.9	43	13.6	17	24.34	41.34	39.96	6.92	7.24
HAV	8.1	57	14.6	139	21.15	32.44	26.53	7.41	17.15
HO	11.3	61	19.7	143	10.68	15.02	15.86	3.65	2.85
IFG	3.2	24	4.9	163	29.05	40.35	32.95	8.51	21.79
LG	38.2	193	36.6	261	7.82	13.31	9.99	2.90	7.33
LI	0.2	50	15.8	3	14.43	29.90	23.18	4.69	8.54

Code	Turnover	$\langle q_t \rangle$	$\langle v \rangle$	# trade	$\sigma_1$	$\langle S \rangle$	$\sigma_S$	$\mathcal{R}_1$	Tick
LY	58.4	111	28.8	507	8.40	12.84	11.94	3.11	4.31
MC	52.9	143	34.7	381	8.01	12.41	10.75	3.01	4.18
ML	13.2	71	21.2	156	10.23	14.98	15.36	3.35	2.71
MMB	13.3	76	18.9	177	10.67	15.96	15.84	3.60	3.50
MMT	0.7	33	10.1	17	33.82	49.77	48.42	9.71	3.57
NAD	4.6	47	6.0	188	14.39	31.27	20.91	4.11	20.36
NK	1.1	43	13.4	21	25.95	42.89	39.34	7.69	8.07
OGE	36.4	182	21.2	429	11.40	21.53	13.35	3.57	16.03
OR	68.0	211	41.9	406	7.30	12.45	9.45	2.76	6.59
PEC	9.5	110	31.9	75	15.06	23.71	24.89	5.07	5.45
PP	36.8	154	31.5	292	10.08	15.44	13.25	3.39	7.23
PUB	11.9	78	27.3	109	15.08	21.19	22.92	4.97	3.85
RF	0.1	25	7.0	3	22.27	37.26	32.69	7.07	8.25
RHA	2.1	32	9.7	55	21.45	33.99	32.97	6.71	10.83
RI	12.6	138	39.1	80	10.49	16.82	16.01	3.64	6.42
RNO	35.8	158	34.4	260	8.01	12.80	11.13	2.71	4.38
RXL	0.7	34	13.0	14	31.30	51.91	50.49	9.21	6.77
SAG	1.5	36	10.3	35	24.59	43.15	42.89	7.29	7.48
SAN	94.2	301	56.4	417	7.76	12.18	8.60	3.04	7.92
SAX	6.0	57	20.4	73	23.28	33.48	33.79	7.45	5.23
SCO	2.0	25	9.3	55	35.67	38.88	40.00	8.37	7.40
SC	0.5	55	13.9	10	12.24	21.08	19.01	3.58	6.10
SU	26.2	129	33.0	198	9.52	14.60	13.43	3.30	5.63
SW	11.2	67	19.5	144	14.48	18.19	20.40	4.26	3.22
TEC	9.4	123	31.6	74	16.27	24.27	25.77	5.06	7.31
TFI	17.4	63	21.0	207	11.91	15.87	16.19	4.09	3.71
TMM	28.0	78	20.8	338	10.08	16.08	15.59	3.18	4.29
UG	33.2	141	36.2	229	7.95	11.91	10.80	2.86	4.43
UL	3.0	64	22.9	33	14.61	24.45	23.47	4.87	8.07
VIE	19.8	77	24.8	199	11.52	16.60	17.81	3.75	3.57
ZC	0.9	33	8.7	26	24.95	41.99	42.40	7.28	4.21

Table 2: Pool of the 68 stocks of the PSE studied in this paper, with their summary statistics for 2002. The daily turnover is in million Euros,  $\langle q_t \rangle$  is the average amount in book (bid+ask) in thousand Euros,  $\langle v \rangle$  is the average size of market order (in thousand Euros). The total number of trades (in thousand) corresponds to the whole year 2002. The volatility per trade  $\sigma_1$ , the average spread  $\langle S \rangle$ , the spread standard deviation  $\sigma_S$ , the response  $\mathcal{R}_1$  and the average tick size are all in basis points. Note that  $\sigma_S \approx \langle S \rangle$ , characteristic of an exponential distribution of the spread. Note also that the volume available at the best prices is  $\sim 10^{-3}$  of the daily turnover.

Code	Turnover	$\langle q_t \rangle$	$\langle v \rangle$	# trade	$\sigma_1$	$\langle S \rangle$	$\sigma_S$	$\mathcal{R}_1$	Tick
LMEB	262	21	199	211	8.4	46.7	4.4	1.5	46.6

Table 3: Summary statistics for Ericsson in the period March 2004-November 2004. Units are the same as in Table 2, except  $\langle q_t \rangle$  which is now in million Euros. Note that  $\langle v \rangle / \langle q_t \rangle \approx 10^{-2}$ .

## References

- [1] A. Orléan, *Le pouvoir de la finance*, Odile Jacob, Paris (1999); *A quoi servent les marchés financiers ?*, in *Qu'est-ce que la Culture ?*, Odile Jacob, Paris (2001).
- [2] M. O'Hara, *Market microstructure theory*, Cambridge, Blackwell, (1995).
- [3] B. Biais, Th. Foucault, P. Hillion, *Microstructure des marchés financiers*, PUF (1997)
- [4] A. Madhavan, *Market microstructure: A survey*, Journal Financial Markets, **3**, 205 (2000)
- [5] L. R. Glosten, P. Milgrom, *Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders*, Journal of Financial Economics **14** 71 (1985)
- [6] L. R. Glosten, *Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices*, Journal of Finance, **42** 293 (1987).
- [7] J. Hasbrouck, *Measuring the information content of stock trades*, Journal of Finance, **XLVI**, 179 (1991).
- [8] R. D. Huang, H. R. Stoll, *The components of the Bid-Ask spread: A general approach*, Review of Financial Studies, **4** 995 (1997).
- [9] A. Madhavan, M. Richardson, M. Roomans, *Why do security prices fluctuate? a transaction-level analysis of NYSE stocks*, Review of Financial Studies, **10**, 1035 (1997).
- [10] F. Black, *Towards a fully automated exchange*, Review Financial Analysts, **27**, 29 (1971).
- [11] A. S. Kyle, *Continuous auctions and insider trading*, Econometrica **53**, 1315 (1985).
- [12] W.-M. Liu, *Monitoring and Limit Order Submission Risks*, University of New South Wales Working Paper, 2005.

- [13] J. P. Bouchaud, Y. Gefen, M. Potters, M. Wyart, *Fluctuations and response in financial markets: The subtle nature of ‘random’ price changes*, Quantitative Finance **4**, 176 (2004).
- [14] F. Lillo, J. D. Farmer, *The long memory of efficient markets*, Studies in Nonlinear Dynamics and Econometrics, **8**, 1 (2004).
- [15] J. P. Bouchaud, J. Kockelkoren, M. Potters, *Random walks, liquidity molasses and critical response in financial markets*, Quantitative Finance, xxx, xxx (2006).
- [16] H. Bessembinder, *Bid-Ask Spread in the Interbank Foreign Exchange Markets*, Journal of Financial Economics, **35**, 317 (1994)
- [17] M. Coppejans, I. Domowitz, A. Madhavan, *Liquidity in Automated Auction*, Working paper, 2001.
- [18] T. Chordia, R. Roll, A. Subrahmanyam, *Order Imbalance, Liquidity and Market Returns*, Journal of Financial Economics, **65**, 111 (2001).
- [19] T. Chordia, A. Subrahmanyam, *Order imbalance and individual stock returns*, Journal of Financial Economics, **72** 485 (2004)
- [20] T. Chordia, L. Shivakumar, A. Subrahmanyam, *Liquidity dynamics across small and large firms*, Economic Notes, **33** 111 (2004)
- [21] L. Gillemot, J. D. Farmer, F. Lillo, *There’s more to volatility than volume*, physics/0510007
- [22] G. Zumbach, *How the trading activity scales with the company sizes in the FTSE 100*, Quantitative Finance, **4**, 441 (2004)
- [23] Th. Foucault, *Order flow composition and trading costs in a dynamic limit order market*, Journal of Financial Market, **2**, 99 (1999)
- [24] Th. Foucault, O. Kadan, E. Kandel, *Limit order book as a market for liquidity*, working paper (2003).
- [25] M. G. Daniels, J. D. Farmer, G. Iori, E. Smith, *Quantitative model of price diffusion and market friction based on trading as a mechanistic random process*, Phys. Rev. Lett. **90**, 108102 (2003).
- [26] E. Smith, J. D. Farmer, L. Gillemot, S. Krishnamurthy, *Statistical theory of the continuous double auction*, Quantitative Finance **3**, 481 (2003).
- [27] J.P. Bouchaud, M. Mézard, M. Potters, *Statistical properties of stock order books: empirical results and models*, Quantitative Finance **2**, 251 (2002).



- [28] H. Luckock, *A steady-state model of the continuous double auction*, Quantitative Finance, **3**, 385 (2003)
- [29] I. Rosu, *A dynamic model of the limit order book*, Working paper, 2005.
- [30] J. D. Farmer, P. Patelli and I. Zokvo, *The predictive power of zero intelligence in financial markets*, PNAS **102**, 2254-2259 (2005)
- [31] S. Mike, J. Doynne Farmer, *An empirical behavioral model of price formation*, physics/0509194.
- [32] P. Weber, B. Rosenow, *Order book approach to price impact*, Quantitative Finance, **5**, 357 (2005)
- [33] F. Lillo, S. Mike, and J. D. Farmer, *Theory for long memory in supply and demand*, Physical Review E, 7106, 287 (2005).
- [34] R. Almgren, C. Thum, E. Hauptmann, H. Li, *Direct estimation of equity market impact*, working paper (May 2005).
- [35] J. D. Farmer, L. Gillemot, F. Lillo, S. Mike, A. Sen, *What really causes large price changes?*, Quantitative Finance, **4**, 383 (2004).
- [36] C. Hopman, *Are supply and demand driving stock prices?*, MIT working paper, Dec. 2002, to appear in Quantitative Finance.
- [37] L. Harris, J. Hasbrouck, *Market vs. limit orders: the SuperDOT evidence on order submission strategy*, Journal of Financial and Quantitative Analysis **31**, 213-31 (1996).
- [38] B. Rosenow, *Fluctuations and market friction in financial trading*, Int. J. Mod. Phys. C, **13**, 419 (2002).
- [39] F. Lillo, R. Mantegna, J. D. Farmer, *Master curve for price-impact function*, Nature, **421**, 129 (2003).
- [40] V. Plerou, P. Gopikrishnan, X. Gabaix, H. E. Stanley, *Quantifying Stock Price Response to Demand Fluctuations*, Phys. Rev. E **66**, 027104 (2002).
- [41] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, X. Gabaix, and H. E. Stanley, *Diffusion and Economic Fluctuations*, Phys. Rev. E **62**, 3023 (2000).
- [42] R. F. Engle, *The econometrics of ultra-high frequency data*, Discussion paper 96-15, University of California, San Diego, 1996.
- [43] Z. Eisler, J. Kertecz, *Size matters, some stylized facts of the market revisited*, xxx.lanl.gov/ physics/0508156.

- [44] M. D. Evans, R. K. Lyons, *Order Flow and Exchange Rate Dynamics*, Journal of Political Economy, **110** 170 (2002)
- [45] T. Chordia, R. Roll, A. Subrahmanyam, *Liquidity and Market Efficiency*, Working paper 2005.