Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

Relation between bid–ask spread, impact and volatility in order-driven markets

Matthieu Wyart , Jean-Philippe Bouchaud , Julien Kockelkoren , Marc Potters & Michele Vettorazzo

To cite this article: Matthieu Wyart , Jean-Philippe Bouchaud , Julien Kockelkoren , Marc Potters & Michele Vettorazzo (2008) Relation between bid–ask spread, impact and volatility in order-driven markets, Quantitative Finance, 8:1, 41-57, DOI: 10.1080/14697680701344515

To link to this article: https://doi.org/10.1080/14697680701344515

Published online: 14 Dec 2007.

Quantitative Finance, Vol. 8, No. 1, February 2008, 41–57

Routledge
Taylor & Francis Group

Relation between bid–ask spread, impact and volatility in order-driven markets

MATTHIEU WYART, JEAN-PHILIPPE BOUCHAUD*, JULIEN KOCKELKOREN, MARC POTTERS and MICHELE VETTORAZZO

Science & Finance, Capital Fund Management, 6 Bvd Haussmann, 75009 Paris, France

(*Received 5 May 2006; in final form 16 March 2007*)

We show that the cost of market orders and the profit of infinitesimal market-making or -taking strategies can be expressed in terms of directly observable quantities, namely the spread and the lag-dependent impact function. Imposing that any market taking or liquidity providing strategies is at best marginally profitable, we obtain a linear relation between the bid–ask spread and the instantaneous impact of market orders, in good agreement with our empirical observations on *electronic* markets. We then use this relation to justify a strong, and hitherto unnoticed, empirical correlation between the spread and the volatility *per trade*, with R^2s exceeding 0.9. This correlation suggests both that the main determinant of the bid–ask spread is adverse selection, and that most of the volatility comes from trade impact. We argue that the role of the time-horizon appearing in the definition of costs is crucial and that long-range correlations in the order flow, overlooked in previous studies, must be carefully factored in. We find that the spread is significantly larger on the NYSE, a liquid market with specialists, where monopoly rents appear to be present.

Keywords: Microstructure; Bid-ask spread; Impact; Liquidity

1. Introduction and review of the literature

One of the most important attributes of financial markets is to provide immediate liquidity to investors (Orléan 1999, 2001), who are able to 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, to favour steady and orderly trading and avoid 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 from sellers, and get compensated by 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 (Glosten and Milgrom 1985, Glosten 1987, Hasbrouck 1991, O'Hara 1995, Biais *et al*. 1997, Huang and Stoll 1997, Madhavan *et al*. 1997, Madhavan 2000, Sandas 2000, Easley *et al*. 2002, Hollifield *et al*. 2007), and see Stoll (2000) for a recent review.

However, most financial markets have nowadays become fully electronic (with the notable exception of the New York Stock Exchange, NYSE—although this will soon change). In these markets, liquidity is *self-organized*, in the sense that any agent can choose, at any instant of time, to either provide liquidity or consume liquidity. 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 fixed minimum (maximum) price. 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') so no transaction takes place. For a transaction

*Corresponding author. Email: jean-philippe.bouchaud@cfm.fr
†Now at: Division of Engineering and Applied Sciences, Harvard University, Pierce Hall, 29 Oxford Street, Cambridge, Massachusetts 02138, USA.

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online © 2008 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/14697680701344515

to occur, an agent must consume liquidity by issuing 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, or vice versa)†. 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, and the relative merit of limit versus market orders. In the present work, we argue that on electronic markets, profitable high frequency strategies using either limit or market orders should not exist, imposing a linear relation between the bid–ask spread S and the average impact of market orders. This, in turn justifies a simple, but to our knowledge hitherto unnoticed, proportionality relation between the spread and the volatility *per trade*.

In a large fraction of the economics literature (Glosten and Milgrom 1985, O'Hara 1995, Biais *et al.* 1997, Madhavan 2000), liquidity providers are described as market makers who earn their profit from the spread. The value of the spread is non-zero because this market making strategy has costs. Three types of cost are discussed in the literature (Stoll 2000):

(i) order processing costs (which includes sheer profit for the market maker);
(ii) adverse selection costs: liquidity takers may have superior information on the future price of the stock, in which case the market maker loses money‡;
(iii) inventory risk: market makers may temporarily accumulate large long or short positions which are risky. If agents are risk-sensitive and have to limit their exposure, this adds extra costs.

Theoretical models that account for these costs typically introduce a rather large amount of free parameters (such as risk-aversion, fraction of informed trades, fraction of patient/impatient traders, etc.) most of which cannot be measured directly. In order to extract the different determinants of the spread from empirical data, some drastic assumptions must be made. For example, assuming the order flow to be short-ranged correlated, Huang and Stoll (1997) argue (using data from 1992) that 90% of the spread is associated with order processing costs, and not to adverse selection§. This would mean rather comfortable profits for market making¶, and is a somewhat surprising conclusion since the spread on purely electronic markets is found comparable to the spread in markets with specialists. A related approach is that of Madhavan *et al.* (1997), where the ratio of adverse selection to processing costs was estimated to be in the range 35–50% on the NYSE in 1990 (see also Stoll (2000) for similar numbers). We will review this theoretical framework below and detail the similarities and differences with our own analysis; one particularly crucial difference is the assumption that the order imbalance has short-ranged correlations (Huang and Stoll 1997, Madhavan *et al.* 1997), and therefore that market impact of a single trade is *permanent*, in striking disagreement with empirical data, where the order flow is instead found to be a long-memory process (Bouchaud *et al.* 2004, Lillo and Farmer 2004), and single trade impact *transient*, but decaying very slowly (Bouchaud *et al.* 2004, 2006). The long-range correlation between trades, and the corresponding temporal dependence of market impact will turn out to play an important role in the following discussion.

On general grounds, both adverse selection and inventory risk imply 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 documented empirically (Bessembinder 1994, Madhavan 2000, Stoll 2000, Chordia *et al.* 2001, Coppejans *et al.* 2001, Chordia and Subrahmanyam 2004, Chordia *et al.* 2004, Gillemot *et al.* 2005), but is not particularly spectacular and stands as one among other reported correlations, e.g. with traded volume, flow of limit orders, market capitalization, etc. (Stoll 2000). 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 strong relation was first noted for the case of France-Telecom (Bouchaud *et al.* 2004), and independently on the stocks of the FTSE-100 (Zumbach 2004), but no theoretical argument was given in favour of this relation.

From a theoretical point of view, several statistical models of limit and market order flows have been analysed to understand the distribution of the bid–ask spread, and relate its average value to flow and cancellation rates (Biais *et al.* 1997, Foucault 1999, Bouchaud *et al.* 2002, Daniels *et al.* 2003, Foucault *et al.* 2003, Luckock 2003, Smith *et al.* 2003, Rosu 2005). 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 (Foucault *et al.* 2003), therefore neglecting the long term volatility of the price (see also Luckock (2003) and Rosu (2005)). 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

†Other determinants of liquidity discussed in the literature are the depth of the order book and market resiliency, see Black (1971) and Kyle (1985).
‡This is also discussed as the free option trading problem in the literature, see e.g. Liu (2005) and references therein.
§Adverse selection is even found to have, within this framework, a negative contribution to the spread!
¶Direct processing costs can be estimated to be at least ten times smaller than the spread, in particular on electronic markets.

limit orders, market orders and cancellation (Daniels *et al.* 2003, Bouchaud *et al.* 2002, Smith *et al.* 2003)†. 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 (Farmer *et al.* 2005). The problem with such models is that although the order flow itself 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 off by high frequency automated execution machines. These programs search to optimize execution costs (Almgren *et al.* 2005, Hollifield *et al.* 2007) by adequately conditioning the order flow (proportion of limit and market orders, timing, aggressivity) and use any short-term predictability to do so. As a result 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 within the order book (Bouchaud *et al.* 2004, Lillo and Farmer 2004, Weber and Rosenow 2005). A key observation is that 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, or 10^{-3}–10^{-2} of the daily volume (see table 2 in appendix B). Clearly, the reason for such a small outstanding liquidity is that liquidity providers want to avoid giving a free-trading option to informed traders. As a consequence, liquidity takers must cut their total order in small chunks; this creates the long term correlation in order flow (Lillo *et al.* 2005). But since on electronic markets sophisticated buyers and sellers can trade using at their best convenience either limit or market orders, the average cost of limit and market orders should be very similar. If, say, market orders were on average significantly more expensive than limit orders, more limit orders would be issued, thereby reducing the spread and the cost of market orders, until an equilibrium is reached‡, That a competitive ecology between limit and market orders should exist on order-driven markets was emphasized by Handa *et al.* (1998), Bouchaud *et al.* (2004) and Bouchaud *et al.* (2006). However, as our analysis reveals, this ecology turns out to be considerably more intricate than anticipated by Handa *et al.* (1998).

In the following, we introduce the idea of infinitesimal strategies, participating to a vanishing fraction of market or limit orders. Imposing that such strategies lead at best to marginal profits motivates a linear relation between the instantaneous price impact of a market order and the bid–ask spread, which we check empirically. Interestingly, we find that the profitability of these strategies depend in a non-trivial way on the time horizon over which they are implemented. We show in particular that fast market making strategies can be profitable even though the long-term average cost of limit orders is positive, a rather paradoxical situation brought about by the presence of long-range correlations in the order flow and the temporal structure of the impact function.

The linear relation between spread and impact in turn 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, on electronic markets and on the NYSE. This result shows that in a competitive electronic market the bid–ask spread in fact mostly comes from 'adverse selection', provided one extends this notion to account for the fact that trades 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, as we now explain in detail.

2. Limit orders versus market orders and market impact

2.1. A simple theoretical framework

We start by reviewing the theoretical framework proposed by Madhavan *et al.* (1997) (hereafter MRR), which helps define various quantities and hone in on relevant questions. We will call v_i the volume of the ith market order, and ε_i the sign of that market order ($\varepsilon = +1$ for a buy and $\varepsilon = -1$ for a sell). The assumptions of the model are (i) that all trades have the same volume $v_i = v$ and (ii) the ε_i's are generated by a Markov process with correlation ρ, which means that the average value of ε_i conditioned on the past only depends on ε_{i-1} and is given by

$$\langle \varepsilon_i \rangle \big|_{\varepsilon_{i-1}} = \rho \varepsilon_{i-1}, \tag{1}$$

where $\langle \ldots \rangle$ denotes averaging. The case $\rho = 0$ corresponds to independent trade signs, whereas $\rho > 0$ describes positive autocorrelations of trades. Note that in this model, correlations decay exponentially:

$$C(\ell) = \langle \varepsilon_i \varepsilon_{i+\ell} \rangle = \rho^{\ell}. \tag{2}$$

The MRR model assumes that the 'true' price p_i evolves both because of random external shocks (or news) and because of trade impact. It is natural to postulate that both external news and surprise in order flow should move the price. Since the surprise at the ith trade is given by $\varepsilon_i - \rho\varepsilon_{i-1}$, MRR write the following evolution equation for the price:

$$p_{i+1} - p_i = \xi_i + \theta[\varepsilon_i - \rho\varepsilon_{i-1}], \tag{3}$$

where ξ is the shock component, with variance $\langle \xi_i^2 \rangle = \Sigma^2$, and θ measures trade impact, assumed to be constant (all trades are assumed to have the same volume). Since market makers cannot guess the surprise of the

†More elaborated 'weak intelligence' models have been studied recently, see Mike and Doyne Farmer (2005).
‡Data from brokers VWAP machines indeed show that the fraction of issued market orders is close to 50%.

next trade, they post a bid price b_i and an ask price a_i given by

$$a_i = p_i + \theta[1 - \rho\varepsilon_{i-1}] + \phi; \quad b_i = p_i + \theta[-1 - \rho\varepsilon_{i-1}] - \phi, \tag{4}$$

where ϕ is the extra compensation claimed by the market maker, covering processing costs and the shock component risk. The above rule ensures no *ex-post* regrets for the market maker. The spread is therefore $S \equiv a - b = 2(\theta + \phi)$, whereas the midpoint $m \equiv (a + b)/2$ immediately before the ith trade is given by

$$m_i = p_i - \theta\rho\varepsilon_{i-1}. \tag{5}$$

These equations allow one to compute several important quantities for the following discussion, although not explicitly considered by MRR. The first one is the lagged impact function introduced in Bouchaud *et al.* (2004, 2006)

$$\mathcal{R}_\ell = \langle \varepsilon_i \cdot (m_{\ell+i} - m_i) \rangle, \tag{6}$$

which is found, within the MRR model, to increase from $\mathcal{R}_1 = \theta(1 - \rho)$ to $\mathcal{R}_\infty = \theta$ (see appendix A and figure 1). Due to correlations between trades, the long time impact is therefore enhanced compared to the short term impact by a factor:

$$\lambda_\infty = \frac{1}{1 - C_1}, \tag{7}$$

where $C_1 = C(\ell = 1) = \rho$ in the MRR model, but the above relation is more general (see appendix A).

The second quantity is the mid-point volatility, defined as

$$\sigma_\ell^2 = \frac{1}{\ell} \langle (m_{\ell+i} - m_i)^2 \rangle, \tag{8}$$

which is easily computed to be†

$$\sigma_1^2 = \mathcal{R}_1^2 + \Sigma^2; \quad \sigma_\infty^2 = \frac{1+\rho}{1-\rho}\mathcal{R}_1^2 + \Sigma^2; \quad \langle \xi_i^2 \rangle = \Sigma^2. \tag{9}$$

Within the above interpretation, the MRR model leads to the following simple relations between spread, impact and volatility per trade:

$$S = 2\lambda_\infty \mathcal{R}_1 + 2\phi, \quad \sigma_1^2 = \mathcal{R}_1^2 + \Sigma^2, \tag{10}$$

relations which we generalize and test empirically in the following. From the data presented in MRR, one observes that ϕ was rather large on the NYSE in 1990: $\phi/\lambda_\infty \mathcal{R}_1 \sim 1 - 2$.

Note that in the simplest case of independent trade signs ($\rho = 0$), the impact function is time independent. In the absence of extra compensation for the market makers, $\phi = 0$ and the above equation reduces to $\mathcal{R}_1 = S/2$. In economical terms, this last equality has a very simple meaning: it indicates that on average, the new mid-price after the transaction $m_{i+1} = m_i + \varepsilon_i \mathcal{R}$ is equal to the last transaction price $m_i + \varepsilon_i S/2$, and therefore that $\mathcal{R}_1 = S/2$ is precisely the condition where both market orders and limit orders have zero ex-post cost. This is more generally the meaning of the MRR relation, equation (10): the transaction price is exactly equal to the expected long term value of the mid-point.

It is interesting to discuss the cost of limit orders $\mathcal{C}_\mathrm{L}$ slightly differently. Suppose one wants to trade at a random instant in time. Compared to the initial mid-point value, the average execution cost of an infinitesimal buy limit order is given by

$$\mathcal{C}_L = \frac{1}{2}\left(-\frac{S}{2}\right) + \frac{1}{2}(\mathcal{R}_1 + \mathcal{C}_L^+); \tag{11}$$

with probability 1/2, the order is executed right away, $S/2$ below the mid-point; otherwise, the mid-point moves on average by a quantity $\mathcal{R}_1$, to which must be added the cost of a limit order conditioned to the last trade being a buy, $\mathcal{C}_\mathrm{L}^+$, for which a similar equation can be obtained:

$$\mathcal{C}_L^+ = \frac{1-\rho}{2}\left(-\frac{S}{2}\right) + \frac{1+\rho}{2}(\mathcal{R}_1^+ + \mathcal{C}_L^{++}), \tag{12}$$

with obvious notations. Since the MRR model is Markovian, one has $\mathcal{R}_1^+ = \mathcal{R}_1$ and $\mathcal{C}_\mathrm{L}^{++} = \mathcal{C}_\mathrm{L}^+$, so that

$$\mathcal{C}_L^{++} = -\frac{S}{2} + \frac{1+\rho}{1-\rho}\mathcal{R}_1. \tag{13}$$

Plugging this last relation in equation (11), we finally find

$$\mathcal{C}_L = -\frac{S}{2} + \frac{2}{1-\rho}\mathcal{R}_1. \tag{14}$$

Imposing that $\mathcal{C}_\mathrm{L} \equiv 0$, one recovers the MRR relation between the spread and the asymptotic impact (equation (10) with $\phi = 0$). Note, however, the cost of a market order *compared to the initial mid-point value* is $S/2$ within the MRR model—but of course the order is still executed at the 'right' long term value of the stock.

2.2. *Real markets are more complicated*

The above model, although suggestive and capturing the essence of the correlation between spread, impact and volatility, is however not fully satisfactory since it completely neglects the very broad distribution of traded volumes (often found to be log-normal, or power-law tailed) and, perhaps more importantly, the non-Markovian, long ranged correlation of the trade

†The is an extra contribution to σ_1^2 coming from any high-frequency noise component that we neglect here, coming from decimalization, small volumes at bid/ask, etc. See Madhavan *et al.* (1997) and Bouchaud *et al.* (2004) and footnote on p. 52.

signs, which is found to decay as (Bouchaud *et al.* 2004, Lillo and Farmer 2004)†

$$C(\ell) = \langle \varepsilon_i \varepsilon_{i+\ell} \rangle \approx \frac{c_0}{\ell^\gamma}, \quad \gamma < 1, \tag{15}$$

instead of the fast, exponential decay assumed in the MRR model. Because the exponent γ is found to be less than unity, the correlation function is not integrable, which technically makes the series of trade signs a long-memory process. As emphasized in Bouchaud *et al.* (2004, 2006), this imposes a number of non-trivial constraints on price impact for the returns to remain uncorrelated while the order flow is strongly auto-correlated. In particular, simple models (such as Huang and Stoll's (1997) and Stoll's (2000)) where price changes include a term proportional to ε_i would lead to strong super-diffusion (trends) of prices in the long run (Bouchaud *et al.* 2004), in disagreement with empirical data.

The volume-dependent lagged impact is now defined as‡

$$\mathcal{R}_\ell(v) = \langle \varepsilon_i \cdot (m_{\ell+i} - m_i) \rangle \big|_{v_i = v}. \tag{16}$$

In the MRR model, v takes a single value and $\mathcal{R}_\ell(v)$ reduces to the previously defined quantity. The function $\mathcal{R}_\ell(v)$ was studied in detail in Bouchaud *et al.* (2004). To a good level of approximation, the following factorization property is found to hold: $\mathcal{R}_\ell(v) \approx R(\ell) f(v)$, where $f(v)$ is a strongly concave function, and $R(\ell)$ is an increasing function of ℓ that varies by a factor of $\sim$2 when ℓ increases from 1 to several thousands (corresponding to a few days of trading)§. The shape of $R(\ell)$, averaged over a collection of different stocks of the PSE, is shown in figure 1, and compared with the simple form assumed in the MRR model (see caption for more details). Perhaps more importantly, the enhancement factor λ_∞ is found empirically to be substantially larger than predicted by equation (7). For example, on the pool of 68 PSE stocks studied below, we find, averaged over all stocks, $\lambda_\infty \approx 1.75$ whereas $1/(1 - C_1) \approx 1.32$ (see table 2 of appendix B). The difference between the two will turn out to play a crucial role in the following.

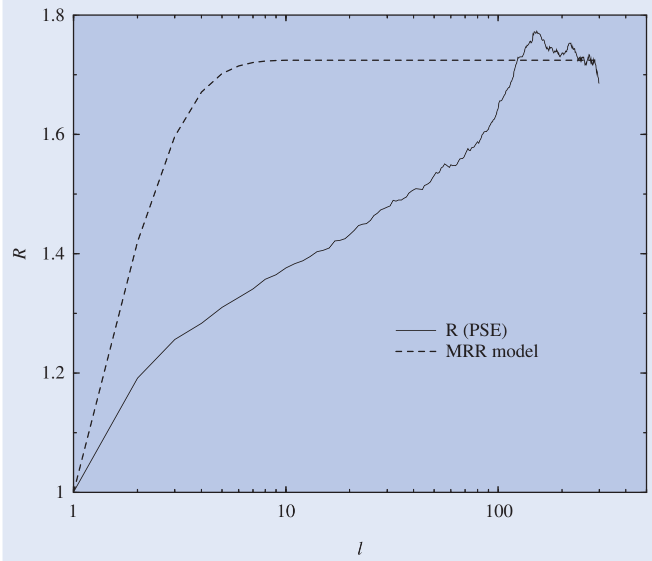

Figure 1. Average over 68 PSE stocks of the impact function $R(\ell)$ as a function of ℓ (plain line). The average is performed by rescaling the individual $R(\ell)$ such that $R(\ell = 1) \equiv 1$, and by rescaling ℓ by the average daily number of trades and multiplying by 100. Dotted line: prediction of the MRR model with $\rho = 3/7$, such that $\lambda_\infty = 1.75$. The discrepancy with empirical data shows the importance of correctly accounting for long-range correlations in order flow.

2.3. *Market order strategies*

In this section and below, we want to show how the simple relations derived in the MRR model can be extended and tested in the general case of fluctuating volumes and long-ranged correlation of trade signs. A first idea is to measure empirically the average execution cost of market orders. One can define the *ex-post* cost $\mathcal{C}_\mathrm{M}(T)$ the difference between the transaction price at (trade-)time i and the mid-point price at time $i + T$ later, with $T \gg 1$ but still much smaller than the typical horizon of the trading strategy itself (a few days or more), in order not to mix in the quality of the decision to trade. The above definition of execution cost marks the trade to market after T and is referred to as the *realized spread* in the literature (Bessembinder 2003, Stoll 2000). The volume weighted averaged cost (over N trades) of a single market order over horizon T is therefore¶

$$\mathcal{C}_M(T) = \frac{1}{N\langle v \rangle} \sum_{i=1}^{N} \varepsilon_i v_i \left(m_i + \varepsilon_i \frac{S_i}{2} - m_{i+T} \right) \equiv \frac{\langle vS \rangle}{2\langle v \rangle} - \frac{\langle v \mathcal{R}_T(v) \rangle}{\langle v \rangle}. \tag{17}$$

The choice $T \gg 1$ allows us to use the asymptotic value of R, $R(\ell \gg 1) \approx \lambda_\infty R(1)$, where we have introduced a factor λ_∞ in conformity with the notation of the previous section. Using the factorization property of $\mathcal{R}_\ell(v)$, we finally obtain for the average cost of a single market order:

$$\mathcal{C}_M(T \gg 1) = \frac{\langle vS \rangle}{2\langle v \rangle} - \lambda_\infty \frac{\langle v \mathcal{R}_1(v) \rangle}{\langle v \rangle}, \tag{18}$$

†These long ranged correlations were also noted in e.g. Chordia and Subrahmanyam (2004) and Hopman (2002), but the detailed shape of the tail of $C(\ell)$ was not investigated, and its long-memory nature not discussed.

‡In the definition of $\mathcal{R}_1$ care has been taken to remove any long term trend of the midpoint. In any case, since $\langle \varepsilon \rangle$ is close to zero, this trend contribution would very nearly vanish.

§The true asymptotic behaviour of $R(\ell)$ for longer horizons is difficult to determine empirically due to statistical noise, and might in fact be stock dependent, see Bouchaud *et al.* (2006) for a discussion of this point.

¶Note that this definition neglects the fact that one single large market order may trigger transactions at several different prices, up the order book ladder, and pay more than the nominal spread. Nevertheless this situation is empirically quite rare on the markets we are concerned with, and corresponds to only a few percent of all cases (Farmer *et al.* 2004).

meaning, as intuitively clear, that this cost is positive when spreads are large, but may become negative if the total price impact $\lambda_\infty \mathcal{R}_1$ is large. In the plane $x = \langle v\mathcal{R}_1(v)\rangle/\langle v\rangle$, $y = \langle vS\rangle/\langle v\rangle$ (which will repeatedly be used below to represent empirical data) the condition $\mathcal{C}(T \gg 1) = 0$ defines a straight line of slope $2\lambda_\infty$ separating an upper region where market orders are on average costly, from a region where single market orders are favoured: see figure 2.

The above computation suggests an upper bound on the spread, which we establish more rigorously in the next section. For larger spreads, the positive average cost of market order would deter their use; limit orders would then pile up and reduce the spread. What would happen if the spread was below the red line of slope $2\lambda_\infty$ in figure 2? Naively, market orders have a negative cost in that region, and one might be able to devise profitable strategies based solely on market orders. The idea would be to try to benefit from the impact term $\mathcal{R}_\infty$ in the above balance equation. The growth of $\mathcal{R}_\ell$ ultimately comes from the correlation between trades, i.e. the succession of buy (sell) trades that typically follow a given buy (sell) market order. The simplest 'copy-cat' strategy which one can rigorously test on empirical data is to place a market order with vanishing volume fraction (not to affect the subsequent history of quotes and trades), immediately following another market order. This strategy suffers on average from the impact of the initial trade, used as a guide to guess the direction of the market. Therefore, the profit $\mathcal{G}_{CC}$ of such a copy-cat strategy, marked to market after a long time and neglecting further unwinding costs, is reduced to†

$$\mathcal{G}_{CC} = [\lambda_\infty - 1]\frac{\langle v\mathcal{R}_1(v)\rangle}{\langle v\rangle} - \frac{\langle vS\rangle}{2\langle v\rangle}. \tag{19}$$

Imposing that this gain is non-positive, one obtains a lower line in the plane x, y, of slope $2(\lambda_\infty - 1)$. Only below this green line can the above infinitesimal copy-cat strategy be profitable. We therefore expect markets to operate above this line and below the red line of slope $2\lambda_\infty$. Note however that market orders below the $2\lambda_\infty$ line are not necessarily favourable in practice, since the cost for executing a *series* of market orders (which is the typical situation faced by large investors, since the outstanding liquidity is, as noted above, always quite small) must include the impact of past trades and this increases their average cost. Hence, the slope of the effective zero-cost line for a series of market orders is indeed smaller than $2\lambda_\infty$. Similarly, the long-time impact of an isolated market order, uncorrelated with the order flow, is in fact very small (Bouchaud *et al.* 2004). These isolated market orders thus also have a positive cost, equal to half the spread. The only way to benefit from the average impact $\mathcal{R}_\ell$ is to free-ride on a wave of orders launched by others, as in the above copy-cat strategy. Let us now take the complementary point of view of limit orders and determine the region of profitable market making strategies.

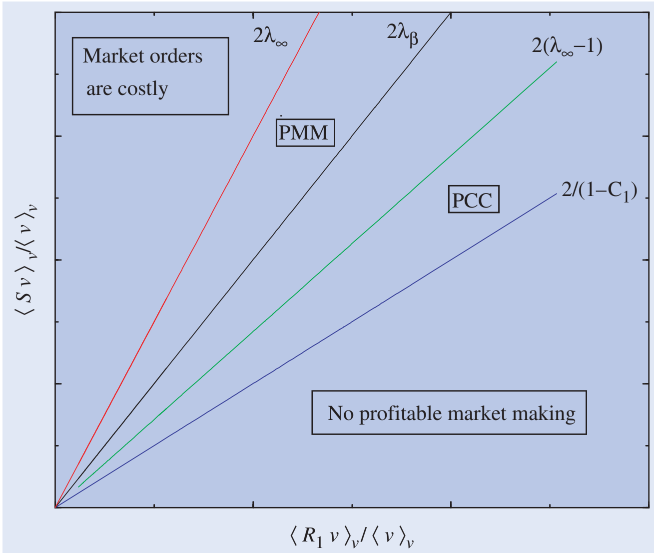

Figure 2. General 'phase diagram' in the plane $x = \langle v\mathcal{R}_1(v)\rangle/\langle v\rangle, y = \langle vS\rangle/\langle v\rangle$ showing several regions: (i) above the reel line of slope $2\lambda_\infty$, market orders are costly (on average) and market making is profitable; (ii) below the blue line of slope $\approx 2/(1-C_1)$, limit orders are costly and no market-making strategy is profitable; (iii) above the black line of slope $2\bar{\lambda}_\beta$, market making on time scale T (or faster) is profitable (PMM); (iv) below the green line of slope $2(\lambda_\infty - 1)$, copy-cat strategies can be profitable (PCC). Since neither market orders nor liquidity providing should be systematically penalized for markets to ensure steady trading, we expect that markets should operate in the 'neutral wedge' in between the blue and the red line. Competition between liquidity providers should push the market towards the blue line. Since copy-cat strategies should not be profitable either, the PCC green line cannot lie above this blue line. Note that the blue, red and black lines all coincide within the MRR model.

2.4. *An infinitesimal market making strategy*

Our aim is to discuss the profitability of providing liquidity to the market formalizing the idea of infinitesimal strategies used in the previous section. To do so we compute the gain of a simple market making strategy which consists in participating to a vanishing fraction of all trades through limit orders. The simplest strategy is to consider a market maker with a certain time horizon T who provides an infinitesimal fraction ϕ of the total available liquidity. As illustrated by equation (17), the cost incurred by the market maker comes from market impact: the price move between 0 and T is anti-correlated with the accumulated position. When the crowd buys, the price goes up while the market making strategy accumulates a short position which would be costly to buy back at time T, and vice versa. More precisely, we consider a *steady-state* market making strategy (which avoids explicit unwinding costs). The strategy is such that

†A more rigorous estimate of the gain of a copy-cat strategy participating to all the trades can be obtained following the method outlined in the next section.

volume offered dynamically depends on the accumulated position, which insures that the inventory is always bounded. We choose the tendered fraction ϕ to be given by $\phi_i = \phi_0(1 + \alpha V_i \varepsilon)$, where V_i is the (signed) position accumulated up to time i^-, and $\varepsilon = +1$ for orders placed at the ask and $\varepsilon = -1$ for orders placed at the bid. This mean-reverting strategy insures that the typical position is always bounded. One can now use this strategy for an arbitrary long time T; its profit and loss is simply given by

$$\mathcal{G}_L = \sum_{i=0}^{T-1} \varphi_i \varepsilon_i v_i \left(m_i + \varepsilon_i \frac{S_i}{2} \right). \tag{20}$$

For large T, one can replace this expression by

$$\mathcal{G}_L = T \left\langle \varphi_i \varepsilon_i v_i \left(m_i + \varepsilon_i \frac{S_i}{2} \right) \right\rangle \tag{21}$$

with $O(T^0)$ corrections due to the residual position at T. Discarding the constraint $\phi_i \geq 0$ and neglecting volume–volume correlations, which are much smaller than sign–sign correlations (Bouchaud *et al.* 2004, Lillo and Farmer 2004), we finally find

$$\frac{\mathcal{G}_L(\beta)}{T\varphi_0\langle v\rangle} = \frac{\langle vS\rangle}{2\langle v\rangle}\left[1 - \frac{1-\beta}{\beta}\sum_{\ell=1}^{\infty}\beta^\ell C(\ell)\right] - \frac{1-\beta}{\beta}\sum_{\ell=1}^{\infty}\beta^\ell \frac{\langle v\mathcal{R}_\ell(v)\rangle}{\langle v\rangle}, \tag{22}$$

where $\beta = 1 - \alpha\varphi_0\langle v\rangle$ fixes the typical time scale of the market making strategy. The above expression is exact in the limit $\alpha \to 0$, and only approximate otherwise. When $\beta \to 0$ (fast market making), equation (22) reduces to

$$\frac{\mathcal{G}_L(\beta \to 0)}{T\varphi_0\langle v\rangle} \approx \frac{\langle vS\rangle}{2\langle v\rangle}[1 - C_1] - \frac{\langle v\mathcal{R}_1(v)\rangle}{\langle v\rangle}, \tag{23}$$

whereas $\beta \to 1$, corresponding to slow market making, yields

$$\frac{\mathcal{G}_L(\beta \to 1)}{T\varphi_0\langle v\rangle} = \frac{\langle vS\rangle}{2\langle v\rangle} - \frac{\langle v\mathcal{R}_\infty(v)\rangle}{\langle v\rangle}. \tag{24}$$

Setting $\mathcal{G}_L(\beta)$ to zero leads to a linear relation between spread and impact:

$$\frac{\langle vS\rangle}{\langle v\rangle} = 2\bar{\lambda}_\beta \frac{\langle v\mathcal{R}_1(v)\rangle}{\langle v\rangle}. \tag{25}$$

Using the empirical shape of $\mathcal{R}_\ell$ and $C(\ell)$, the slope $2\bar{\lambda}_\beta$ is found to increase between $\approx 2/(1 - C_1)$ and $2\lambda_\infty$ when β increases. Contrarily to market orders which benefit from the growth of the impact $\mathcal{R}_\ell$ with time, slow market making is suboptimal. When $\beta \to 1$, $\bar{\lambda}_\beta \to \lambda_\infty$ and the lower limit of profitability of very slow market making is precisely the red line of figure 2 where market orders become profitable. Faster strategies correspond to smaller values of $\bar{\lambda}_\beta$, closer to $1/(1 - C_1)$, leading to an extended region of profitability for market making. From the assumption that the above market making strategy for any value of β should be at best marginally profitable (since one might find more sophisticated strategies, which take full advantage of the correlations between signs and volumes), we finally obtain the following bound between spread and impact:

$$\frac{\langle vS\rangle}{\langle v\rangle} \leq \frac{2}{1 - C_1}\frac{\langle v\mathcal{R}_1(v)\rangle}{\langle v\rangle}, \tag{26}$$

defining the blue line of slope $2/(1 - C_1)$ in the x, y plane of figure 2. Consistently with the MRR model, when $\lambda_\infty = 1/(1 - C_1)$, the blue and red line of figure 2 exactly coincide. Using that fact that $\mathcal{R}_1^{n+} \leq \mathcal{R}_1^{(n-1)+}$, a simple generalization of the argument presented at the end of section 2.1 allows one to show that the cost of limit orders is indeed negative above the blue line.

2.5. Theoretical analysis: conclusions

Equations (18), (19) and (26) and the resulting microstructural 'phase diagram' of figure 2 are our central results. These equations show that the cost or profitability of an infinitesimal market and limit order strategies can be estimated from empirical data alone, without having to make any further assumption on the fraction of informed trades, the correlation between trades, etc. In order to proceed, we made two approximations. Firstly, we assumed that these strategies could be made infinitesimal, which allows us to neglect their impact on the price dynamics. In practice, trades occur in discrete volume, and strictly speaking the assumption of infinitely small volumes does not hold. However, the volume of typical trades is much larger than the minimum size, which suggests that this approximation is accurate. Secondly, we neglected all direct transaction costs, which obviously affect profitability. These costs are in general very small compared to the spread, and can therefore also reasonably be neglected.

Our main result is that profitability, perhaps surprisingly, depends on the *frequency* of these strategies, a result closely related to the anomalous time dependence of the impact function. Market orders are favoured at low frequencies, when impact has fully developed, whereas limit orders are favoured at high frequencies, where impact is still limited and the execution probability significant.

Our analysis delineates, in the impact-spread plane, a central wedge bounded from above by a slope $2\lambda_\infty$ and from below by a slope $\approx 2/(1 - C_1)$, within which both market orders and limit orders are viable. In the upper wedge, market orders would always be costly and would be substituted by limit orders. In the lower wedge, market making strategies, even at high frequencies, would never eke out any profit. Such a market would not be sustainable in the absence of any incentive to provide liquidity. But if the spread happened to fall in this region, the enhanced flow of market orders would soon reopen the gap between bid and ask.

Our next assumption is that simple statistical strategies must have marginal profit. This is quite reasonable since high-frequency strategies carry relatively small risks. Applying this idea to market making strategies, we conclude that competition between liquidity providers will push the spreads close to the lower limit, corresponding to the blue line of slope $\approx 2/(1-C_1)$ in figure 2. Now, since market taking (copy-cat) strategies should not be profitable either, the green line of slope $2(\lambda_\infty - 1)$ should necessarily lie below the blue line, leading to the following inequality on the asymptotic impact enhancement factor λ_∞:

$$1 \le \lambda_\infty \le 1 + \frac{1}{1-C_1}, \tag{27}$$

where the lower bound comes from the existence of correlation between trades (see equation (7)). In other words, the impact function cannot grow more than to roughly twice its initial value, otherwise statistical arbitrage would set in. Interestingly, our data is compatible with the above bound; in practice the blue and green lines turn out to be not very far from each other.

Finally, we note that market microstructure studies insist on large inventory risks being an important determinant of the bid–ask spread. However, large inventories correspond to long horizons and slow market making. Our analysis above shows that accumulating inventories on a long horizon is not only risky, but may also be extremely costly on average. When $\lambda_\infty > 1/(1-C_1)$, market making on large horizons is significantly more costly than on short horizons, by an amount proportional to the spread itself. This is a very strong effect, which makes the existence of low-frequency market makers very unlikely. Therefore, inventory risk by itself should not be important in determining the value of the spread, at least on electronic markets.

In conclusion, we expect that electronic markets should operate in the vicinity of the blue line of figure 2, imposing a linear relation between spread and market impact of slope close to $2/(1-C_1)$. This is what we test on empirical data in the following section.

3. Comparison with empirical data

3.1. 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 appendix B. From the Trades and Quotes data, one has access to 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 $\mathcal{R}_1$, the one-lag correlation C_1, the spread S and λ_∞. 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. Clearly, these block trades are outside the scope of the above arguments; in any case they represent typically a 5–10% fraction of the total number of trades and do not significantly affect the following results.

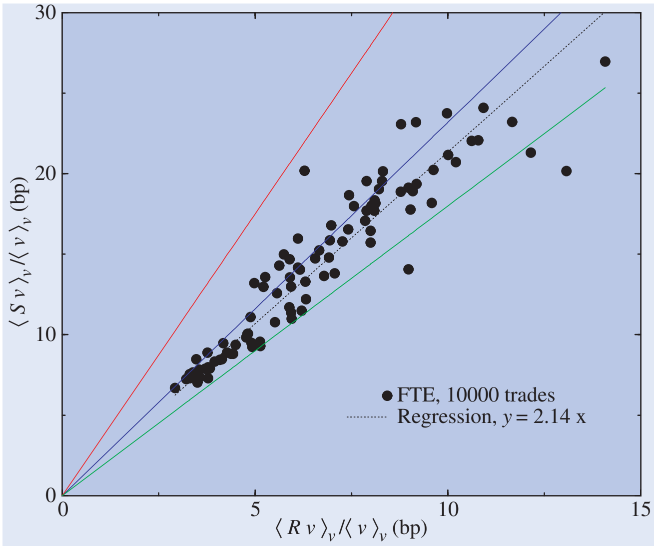

Figure 3. Franco Telecom in 2002, Each point corresponds to a pair ($y = \langle vS \rangle / \langle v \rangle$, $x = \langle v\mathcal{R}_1 \rangle / \langle v \rangle$), computed by averaging over 10 000 non-overlapping trades ($\sim$ two trading days). Both quantities are expressed in basis points. We also show the different bounds, equations (18), (19) and (26), and a linear fit that gives a slope of 2.14. The correlation is $R^2 = 0.93$.

We test the above ideas 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 $\langle Sv \rangle / \langle v \rangle$ and $\langle v\mathcal{R}_1(v) \rangle / \langle v \rangle$ over a number of successive trades. In the example of figure 3, each point corresponds to an average over 10 000 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 equations (19) and (26). For France Telecom, we find that λ_∞ is close to the average value 1.85 shown in figure 1. Therefore $2(\lambda_\infty - 1) \approx 2$ in this case, meaning that copy-cat market making strategies are impossible, as expected for highly liquid stocks. We also find that $C_1 \approx 0.14$ (see appendix B). Our results shown in figure 3 are in good agreement with the above theoretical bounds, even for averages over rather short time scales. A linear fit with zero intercept gives a slope equal to 2.14, to be compared with $2/(1-C_1) \approx 2.32$, meaning that providing liquidity is hardly rewarded at all for this very liquid, small tick stock. In fact, if the intercept of the linear fit is left free, its value (which should equal the 'processing costs' 2ϕ in the MRR model) is found to be slightly negative.

We also test equation (26) cross-sectionally in figure 4, 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

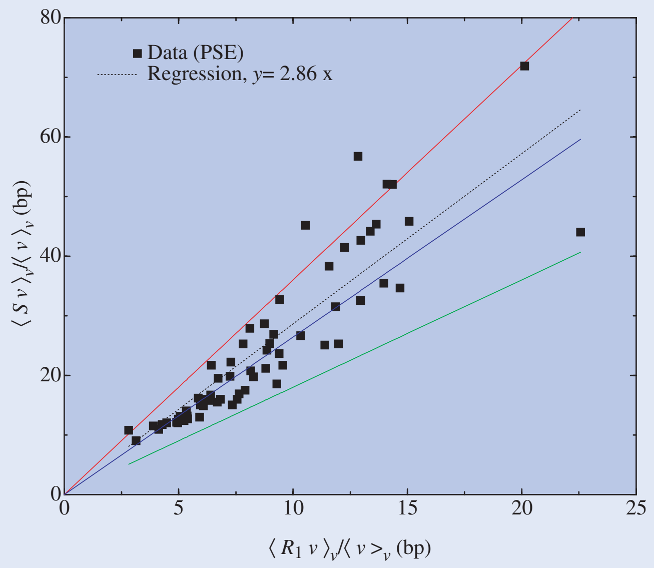

Figure 4. 68 stocks of the Paris Stock Exchange in 2002. Each point corresponds to a pair $(y = \langle vS\rangle/\langle v\rangle, x = \langle v\mathcal{R}_1\rangle/\langle v\rangle)$, computed by averaging over the year. Both quantities are expressed in basis points. We also show the different bounds, equations (18), (19) and (26), and a linear fit that gives a slope of 2.86, while $\langle 2/(1 - C_1)\rangle \approx 2.64$. The correlation is $R^2 = 0.90$.

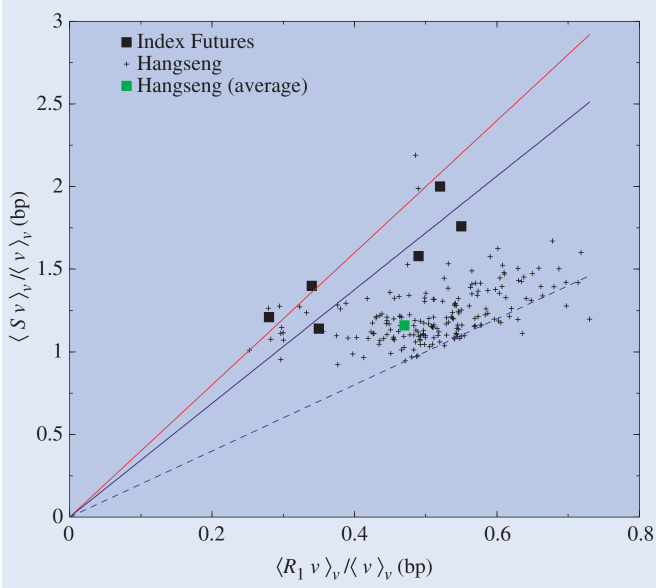

Figure 5. Small tick Index Futures in 2005: CAC, DAX, FTSE, IBEX, MIB, SMI, HANGSENG. Each black square corresponds to a pair $(y = \langle vS\rangle/\langle v\rangle, x = \langle \mathcal{R}_1 v\rangle/\langle v\rangle)$, computed by averaging over the year, while small crosses are computed by averaging over 1000 non overlapping trades on the HANGSENG futures. Both quantities are expressed in basis points. We also show the bounds, equations (26) and (18), with $1/(1 - C_1) \approx 1$ (dotted blue line), corresponding to the HANGSENG, and $1/(1 - C_1) \approx 1.72$ (full blue line), corresponding to the average over all other futures.

enables to test the linear relations (19) and (26). Once again we find a good agreement with the predicted bound, and the linear fit with zero intercept gives a slope of 2.86, while $\langle 2/(1 - C_1)\rangle \approx 2.64$. Hence, fast market making strategies are on average weakly profitable on the PSE. However, the intercept of a two-parameter regression is very slightly negative, showing that no order processing costs component can be detected on these fully electronic markets.

It is also interesting to analyse 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. For most contracts, the value of C_1 is quite large ($\langle C_1\rangle \approx 0.42$) except for the HANGSENG where $C_1 \approx 0.035$. Results are shown in figure 5; the bounds are again quite 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 as high as a few per second, the equilibrium between spread and impact is reached very quickly.

3.2. *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 placing limit orders. We again test equations (18) and (26) cross-sectionally, using the set of the 155 most actively traded stocks on the NYSE in 2005†. We use the quoted bid–ask posted by the specialist. We have first determined the average impact function $R(\ell)$, which has a shape roughly similar to figure 1, although the asymptotic plateau value is slightly larger, leading to $\lambda_\infty \approx 2.1$. On the other hand, $1/(1 - C_1)$ is also slightly larger, equal to 1.39.

Plotting the data in the spread-impact plane, we now find (see figure 6) that the empirical results cluster around to the upper red line limit where market orders become costly. The regression has a significantly larger slope of 3.3 and now a positive intercept $2\phi \approx 1.3$ basis points‡. This suggests that, perhaps not surprisingly, the existence of monopoly rents on NYSE: market makers post spreads that are systematically overestimated compared to the situation in electronic markets, with a non-zero extrapolated spread 2ϕ for zero market impact. This result is in agreement with the study of Harris and Hasbrouck (1996) performed in the early 1990s on the NYSE, which showed that limit orders were more favourable than market orders, and also with Handa and Schwartz (1996), who showed that pure limit order strategies were indeed profitable. On the other hand, the value of the regression slope on the purely electronic PSE show that pure limit order strategies can only be marginally profitable.

†The list of the 155 names is available on request.
‡This is five times smaller than the average spread, leading to $\phi/\theta \sim 0.25$, much smaller than the result $\phi/\theta \sim 1 - 2$ found within the MRR model in 1990, or a similar value reported in Stoll (2000).

We have checked that using the traded spread instead of the quoted spread does not change appreciably the above conclusions.

3.3. *The case of large tick electronic markets*

A priori, the string of arguments leading to equation (26) 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 B, table 3). Because of the large value of the spread, limit orders appear to be favourable, 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 B, table 2) for smaller tick stocks. Therefore, the simple market making strategy considered above, which assumes that one can participate in 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 equation (26), because the competition between liquidity providers, that acts to reduce the spread, cannot fully operate. We indeed find that the ratio between $\langle vS \rangle$ and $\langle v\mathcal{R}_1 \rangle$ 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 $\sim 50\,\text{bp}$, we find $\langle vS \rangle / \langle v\mathcal{R}_1 \rangle \approx 4.5$. However, we also find on the same data that $\lambda_\infty \approx 4.5 \pm 1$, meaning that market orders are in fact not systematically unfavoured in these large tick electronic markets. In fact, all data points are found to lie between our bounds, equations (18) and (26), but indeed significantly higher than the blue line of figure 2 in this case.

3.4. *Comparison with empirical data: conclusion*

Our empirical analysis shows that on liquid markets, an approximate symmetry between limit and market orders indeed holds, in the sense that neither market orders nor limit orders are systematically unfavourable. Markets operate in the 'neutral wedge' of figure 2.

For fully electronic markets, competition for providing liquidity is efficient in keeping the spread close to its lowest value, marginally compensating impact cost. There is therefore hardly any room for market making strategies. Although the cost of *isolated* market orders is found to be negative, the empirically established proximity of the blue and green line in figure 2 means that there is no room for simple market *taking* strategies either. In this discussion, time horizon and long range correlations in the order flow play an important role, overlooked in previous studies (Huang and Stoll 1997, Madhavan *et al.* 1997, Stoll 2000): somewhat paradoxically, liquidity providers as a whole offer average negative costs to market orders but high frequency market making strategies still manage to get (marginally) compensated. Our analysis shows that the ecology between liquidity takers and liquidity providers turns out to be considerably more complex than anticipated by Handa and Schwartz (1998): when costs are computed on large time scale, limit orders are on average costly. This implies that a significant fraction of limit orders cannot be due to market makers, since limit orders as a whole are in arrears. The common assumption that limit orders can be attributed to liquidity providers compensated by the spread cannot be correct in electronic markets. This argument can only concern a small fraction of high-frequency market makers, whose existence is nevertheless crucial to prevent liquidity crises.

On the NYSE, spreads appears to be significantly larger: isolated market orders are now marginally costly. A linear relation between spread and impact still applies, albeit with a larger slope and a residual intercept, corresponding to market maker monopoly rents, which are absent in electronic markets.

4. Liquidity versus volatility

4.1. *Theoretical considerations*

Consider again the MRR model discussed above, which predicts a simple relation between volatility and impact, equation (9). Using the relation between spread and impact established above, this suggests a direct link between volatility per trade and spread, which we motivate and test in this section.

By definition of the volatility per trade $\sigma_1^2 = \langle (m_{\ell+1} - m_\ell)^2 \rangle$ and of the instantaneous impact $r_{1,i} \equiv (m_{i+1} - m_i)\varepsilon_i$, one has as an identity,

$$\sigma_1^2 \equiv \left\langle r_{1,i}^2 \right\rangle. \tag{28}$$

The instantaneous impact $r_{1,i}$ is expected to fluctuate over time for several reasons. First, the volume of the trade, the volume in the book and the spread strongly fluctuate with time. For example, on the PSE, the spread has a distribution close to an exponential, hence one has $\langle S^2 \rangle \approx 2\langle S^2 \rangle$ (see table 2, appendix B)†. Large impact fluctuations may also arise from quote revisions due to addition or cancellation of some limit orders. Second, there might also be important news affecting the 'fundamental price' of the stock. These result in large, instantaneous jumps of the mid-point, unrelated to the trading activity itself. In order to account for both effects, we write, generalizing the above MRR relation:

$$\sigma_1^2 = a\overline{\mathcal{R}}_1^2 + \Sigma^2 \tag{29}$$

where $\overline{\mathcal{R}}_1 \equiv \langle \mathcal{R}_1(v) \rangle$ is the average impact after one trade, a is a coefficient measuring the variance of impact

†The distribution appears to be a power-law on other markets (Mike and Doyne Farmer 2005), but this is irrelevant for the following discussion.

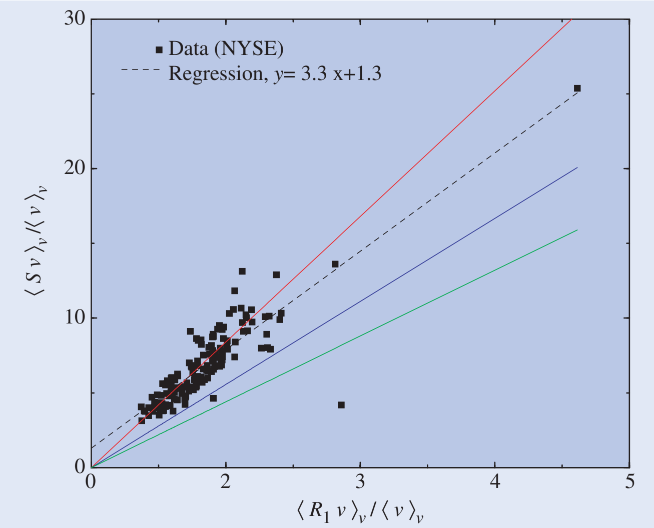

Figure 6. 155 stocks of the NYSE 2005. Each point corresponds to a pair ($y = \langle vS\rangle/\langle v\rangle$, $x = \langle v\mathcal{R}_1\rangle/\langle v\rangle$), computed by averaging over the year. Both quantities are expressed in basis points. We also show our bounds, equations (18), (19) and (26). The data shows clearly that market orders are less favourable than in the electronic Paris Bourse. The regression now has a positive intercept of 1.3 bp with an $R^2 = 0.87$.

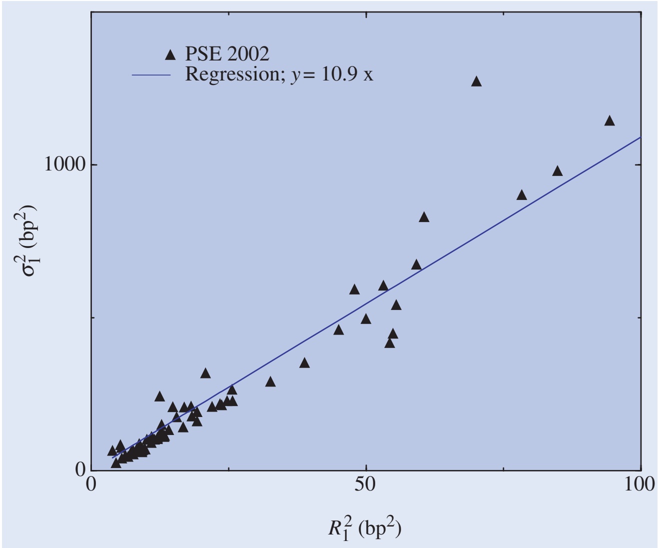

Figure 7. Plot of σ_1^2 versus $\bar{\mathcal{R}}_1^2$, showing that the linear relation equation (29) holds quite precisely with $\Sigma^2 = 0$ and $a \approx 10.9$. (The intercept of the best affine regression is even found to be slightly negative). Data here corresponds to the 68 stocks of the PSE in 2002. The correlation is very high: $R^2 = 0.96$.

fluctuations and Σ^2 is the news component of the volatility (see section 2.1). A specific model for equation (29) was worked out in Bouchaud et al, (2004), and tested on France Telecom (see also Rosenow (2002)). Here, we establish that this relation holds quite precisely across different stocks of the PSE, with a correlation of $R^2 = 0.96$ (see figure 7). Perhaps surprisingly, the exogenous 'news volatility' contribution Σ^2 is found to be small. (The intercept of the best affine regression is even found to be slightly negative). This could be related to the observation made in Farmer et al, (2004) that for most price jumps, some limit orders are cancelled too slowly and get 'grabbed' by fast market orders, which means that most of these events are already included in $\overline{\mathcal{R}}_1$, in line with our general statements on the approximate symmetry between limit and market orders†. In the following, we will therefore neglect Σ^2, as suggested by figure 7: in this sense the volatility of the stocks can be mostly attributed to market activity and trade impact. This is in agreement with the conclusions of Evans and Lyons (2002) on currency markets; see also the discussion in Bouchaud *et al*. (2004) and Hopman (2002).

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 within the same market should behave identically up to a rescaling of the average spread and the average volume. In particular we assume that the statistics of (i) the volume of market orders, (ii) the spread S and (iii) the impact $\mathcal{R}$, and the correlations between these quantities are independent on the stock when these quantities are normalized by their average value‡. This universality implies that

$$\langle vS\rangle = b\langle v\rangle\langle S\rangle, \tag{30}$$

where b is stock independent. Similarly,

$$\left\langle v\mathcal{R}_1(v)\right\rangle_v \equiv b'\langle v\rangle\overline{\mathcal{R}}_1, \tag{31}$$

where b' is also stock independent. Note that this assumption is consistent with the empirical observation of Lillo *et al*. (2003), where the impact function $\mathcal{R}_1(v)$ for different US stocks can indeed be rescaled onto a unique Master curve by a proper scaling of both the x and y axis. We test equations (30) and (31) in figure 8 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 ($b = 1$), whereas the volume traded and the impact are correlated ($b' > 1$), as expected.

Therefore, using equation (26) as an equality (as suggested by the empirical results of section 3), and equations (29)–(31), we obtain the main result of this section:

$$\langle S\rangle = c\sigma_1, \tag{32}$$

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 Bouchaud *et al*. (2004) and Zumbach (2004), and we present further data

†One could argue that our results simply show that the news volatility Σ itself is proportional to $\overline{\mathcal{R}}_1$ and thus to the spread S. However, there is no reason why this should *a priori* be the case. For example, a model where jumps of typical amplitude J have a small probability per trade p leads to $\Sigma = \sqrt{p}J$, whereas the cost of such jumps, contributing to S, is $pJ \ll \Sigma$.
‡The universality of the shape of the order book was indeed checked to hold rather well in Bouchaud *et al*. (2002).

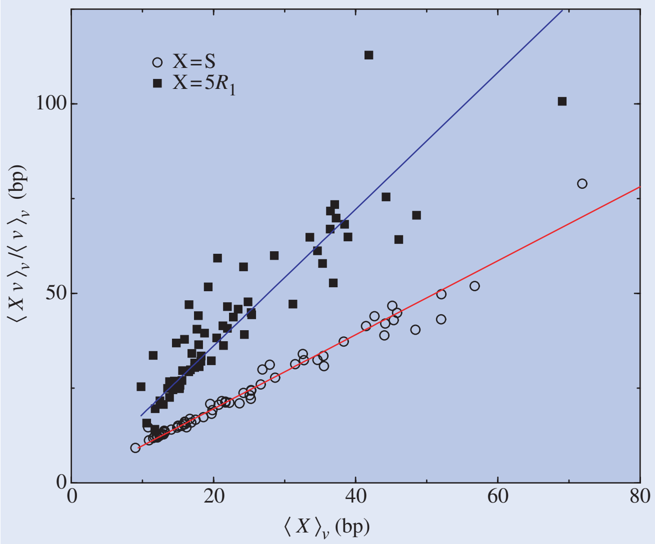

Figure 8. Plot of $\langle vX\rangle_v/\langle v\rangle$ versus $\langle X\rangle$, where X is either the spread S or the instantaneous impact $\mathcal{R}_1(v)$ (multiplied by a factor 5 for clarity). The quality of the linear regression tests our universality assumption, which is excellent for S ($R^2=0.98$) and satisfactory for $\mathcal{R}_1$ $\left(R^2=0.9\right)$. 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.

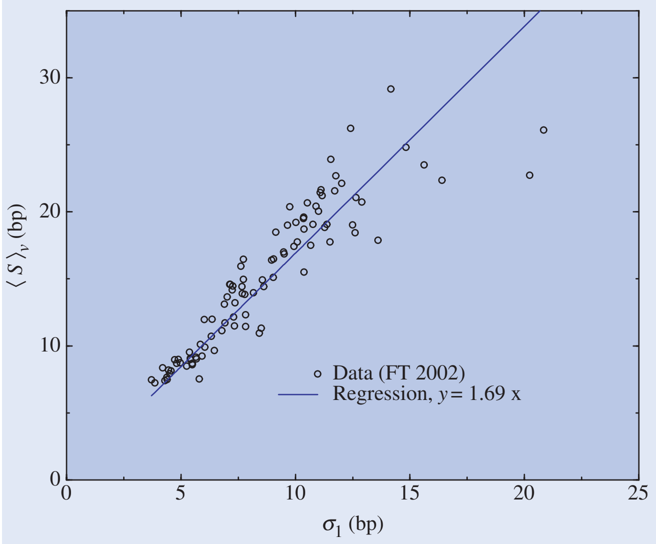

Figure 9. Test of equation (32) for France Telecom in 2002. Each point corresponds to a pair $(\langle S\rangle,\sigma_1)$, computed by averaging over 10 000 non-overlapping trades ($\sim$ two trading days). Both quantities are expressed in basis points. From a linear fit, we find $c\approx 1.69$ with $R^2=0.90$.

in the next section to support this conjecture. Therefore, the constraints that (i) optimized high frequency execution strategies impose that the price is diffusive (see Bouchaud *et al*. (2004, 2006)), and (ii) the cost of limit and market orders are nearly equal (equations (18) and (26)), 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* σ, since it involves an extra stock-dependent and time-dependent quantity, namely the trading frequency ν, through:

$$\sigma=\sigma_1\sqrt{\nu}. \tag{33}$$

We will discuss this issue further in section 5.

4.2. *Comparison with empirical data*

Using the same data sets as in sections 3.1 and 3.2, we now test empirically the predicted linear relation between spread and volatility per trade, equation (32). 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†. Our results for the Paris Stock Exchange are shown in figures 9 and 10. We see that equation (32) describes the data very well, with R^2s over 0.9. 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 (figures 11(a) and (b)) or for the NYSE (figure 12), 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 σ_1: spreads are larger at the opening of the market and decline throughout the day. Note that the trading frequency ν increases as time elapses, which, using equation (33), explains the familiar U-shaped pattern of the volatility per unit time.

5. Discussion and conclusion

The main theoretical result of this paper is the possibility to express the cost of market orders and the profit of infinitesimal market-making/taking strategies in terms of directly observable quantities, namely the spread and the lag-dependent impact function. Imposing that any market taking or liquidity providing strategies is at best marginally profitable allows one to define viable regions of the microstructural 'phase-diagram' (figure 2) where electronic markets should operate, and suggest a linear relation between spread and instantaneous impact. This relation is in good agreement with empirical data on small tick contracts, with a slope compatible with marginal profitability of both fast market making and copy-cat, market taking strategies. Somewhat paradoxically, we find that liquidity providers as a whole offer average negative costs to market orders although high frequency market making

†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%) on short time scales.

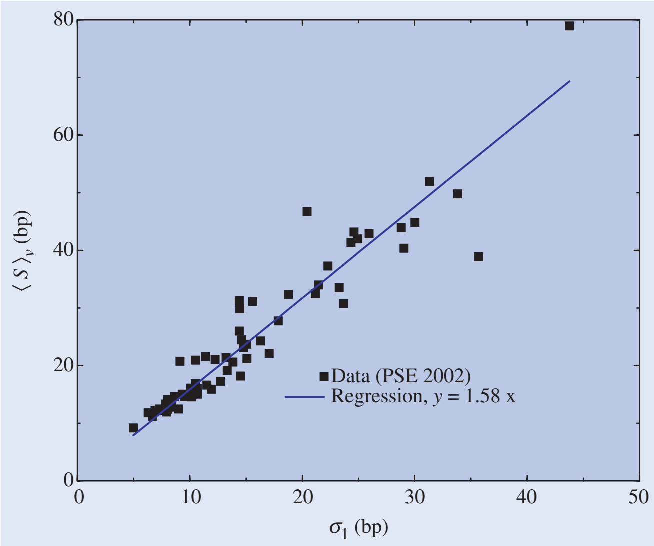

Figure 10. Test of equation (32) for 68 stocks from the Paris Stock Exchange in 2002, averaged over the entire year. The value of the linear regression slope is $c \approx 1.58$, with $R^2 = 0.96$.

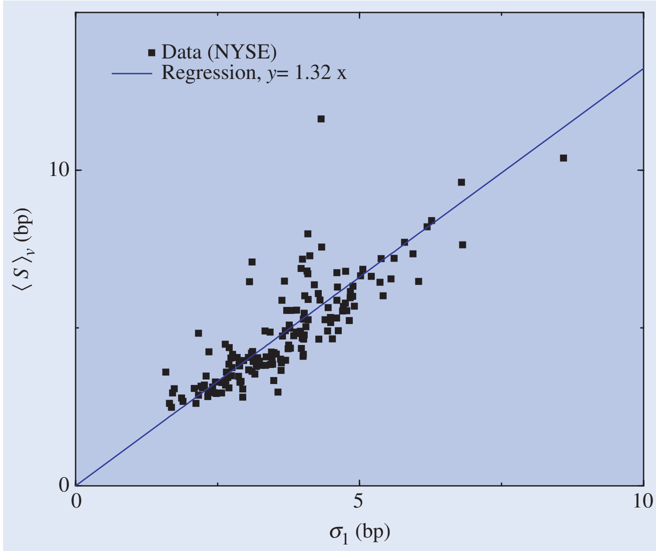

Figure 12. Test of equation (32) for stocks from the NYSE in 2005. Each point corresponds to a pair $(\langle S\rangle, \sigma_1)$, computed by averaging over the entire year. Both quantities are expressed in basis points. From a linear fit, we find $c \approx 1.32$, with $R^2 = 0.91$.

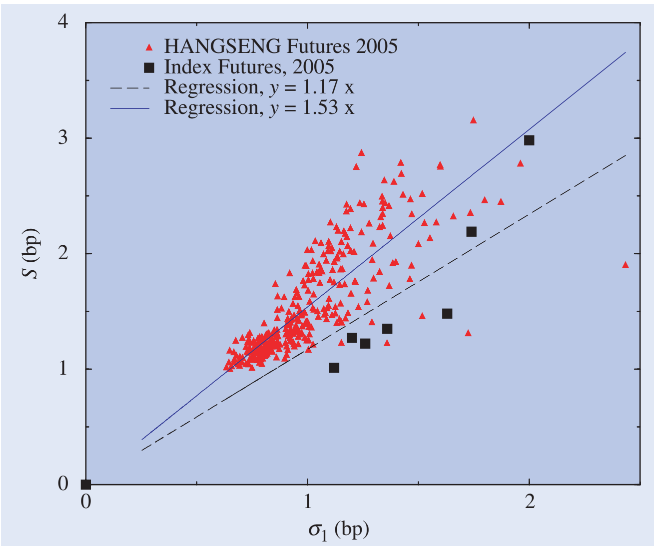

Figure 11. Test of equation (32) for the HANGSENG futures contract (triangles), and across small tick Index Futures in 2005: CAC, DAX, FTSE, IBEX, MIB, SMI, HANGSENG (squares). Each point corresponds to a pair $(\langle S\rangle, \sigma_1)$, computed by averaging either over 1000 non-overlapping trades (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.

strategies still manage to get (marginally) compensated. Our analysis allows us to compare in an objective way the spreads in different markets and suggests that spreads are distinctly larger on the NYSE. Note that our analysis does not require any model specific assumptions such as the nature of order flow correlations or the fraction of informed trades. In fact our results hold even if trades were all uninformed but still mechanically impact the price.

Making reasonable further assumptions, we have then shown that spread S and volatility per trade σ_1 are also proportional, a result that we confirm empirically with correlations above 0.9. This very simple relation means that most of the volatility comes from trading alone, and suggests that the bid–ask spread is dominated by adverse selection, provided one considers the volatility per trade as a measure of the amount of 'information' included in prices at each transaction. There are indeed 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 itself 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 backward 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 favourable.

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 (Farmer *et al.* 2005), or bounded-range models (Foucault *et al.* 2003, Luckock 2003, Rosu 2005) fail to predict any universal relation between S and σ_1.

Our relation involves the volatility per trade whereas most of the econometric work has instead focused on the volatility per unit time σ. The relation between the two involves the trading frequency ν, which is itself both time- and stock-dependent. As a function of time, we find, in agreement with Engle (1996), that volatility per trade and trading frequency are positively correlated; the volatility $\sigma = \sigma_1\sqrt{\nu}$ therefore increases because both σ_1 and ν increase†. Across stocks, on the other hand, the volatility per unit time exhibits only weak systematic variations with capitalization C: $\sigma \sim C^{\phi}$ with $\phi \approx 0$, whereas the trading frequency increases with capitalization as $\nu \sim C^{\zeta}$. For stocks belonging to the FTSE-100, Zumbach (2004) finds $\zeta \approx 0.44$, while for US stocks the scaling for ν is less clear (Eisler and Kertecz 2005). Interestingly, our result then leads to a result between average spread and capitalization of the form $S \sim C^{\phi-\zeta/2} \sim C^{-0.22}$, in good agreement with Zumbach's (2004) data, with the impact data of Lillo *et al.* (2003) and with our own data on the PSE.

The fundamental question at this stage is to know what fixes the volatility σ and the trading frequency ν. Clearly, the trading frequency has to do with the available liquidity and the way large volumes have to be cut into 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 is the market microstructure 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 order flow and price changes, and liquidity and market efficiency (Madhavan *et al.* 1997, Chordia *et al.* 2001, Evans and Lyons 2002, Hopman 2002, Chordia and Subrahmanyam 2004, Bouchaud *et al.* 2004, Lillo and Farmer 2004, Chordia *et al.* 2005).

Acknowledgments

We want to warmly thank S. Bogner, J.D. Farmer, Th. Foucault and G. Zumbach for important and useful discussions. We also thank the referees for very constructive remarks, which helped improving the manuscript.

References

Almgren, R., Thum, C., Hauptmann, E. and Li, H., Direct estimation of equity market impact. *Working Paper*, 2005 (University of Toronto).

Bessembinder, H., Bid–ask spread in the interbank foreign exchange markets. *J. Financ. Econ.*, 1994, **35**, 317–348.

Bessembinder, H., Issues in assessing trade execution costs. *J. Financ. Markets*, 2003, **6**, 233–257.

Biais, B., Foucault, Th. and Hillion, P., *Microstructure des Marchés Financiers*, 1997 (PUF: Paris).

Black, F., Towards a fully automated exchange. *Rev. Financ. Analysts*, 1971, **27**, 29–35.

Bouchand, J.P., Gefen, Y., Potters, M. and Wyart, M., Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quant. Finance*, 2004, **4**, 176–190.

Bouchaud, J.P., Kockelkoren, J. and Potters, M., Random walks, liquidity molasses and critical response in financial markets. *Quant. Finance*, 2006, **6**, 115–123.

Bouchaud, J.P., Mézard, M. and Potters, M., Statistical properties of stock order books: empirical results and models. *Quant. Finance*, 2002, **2**, 251–256.

Coppejans, M., Domowitz, I., Madhavan, A., Liquidity in automated auction. *Working Paper*, 2001 (Duke University).

Chordia, T., Roll, R. and Subrahmanyam, A., Order imbalance, liquidity and market returns. *J. Financ. Econ.*, 2001, **65**, 111–130.

Chordia, T., Roll, R. and Subrahmanyam, A., Liquidity and market efficiency. *Working Paper*, 2005 (Emory University).

Chordia, T. and Subrahmanyam, A., Order imbalance and individual stock returns. *J. Financ. Econ.*, 2004, **72**, 485–518.

Chordia, T., Shivakumar, L. and Subrahmanyam, A., Liquidity dynamics across small and large firms. *Econ. Notes*, 2004, **33**, 111–143.

Daniels, M.G., Farmer, J.D., Iori, G. and Smith, E., Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. *Phys. Rev. Lett.*, 2003, **90**, 108102.

Easley, D., Hvidkjaer, V. and O'Hara, M., Is information risk a determinant of asset returns?. *J. Finan.*, 2002, **57**, 2185–2222.

Eisler, Z. and Kertecz, J., Size matters, some stylized facts of the market revisited. xxx.lanl.gov/physics/0508156, 2005.

Engle, R.F., The econometrics of ultra-high frequency data. *Discussion Paper 96-15*, 1996 (University of California, San Diego).

Evans, M.D. and Lyons, R.K., Order flow and exchange rate dynamics. *J. Polit. Econ.*, 2002, **110**, 170–180.

Farmer, J.D., Gillemot, L., Lillo, F., Mike, S. and Sen, A., What really causes large price changes? *Quant. Finance*, 2004, **4**, 383–397.

Farmer, J.D., Patelli, P. and Zokvo, I., The predictive power of zero intelligence in financial markets. *PNAS*, 2005, **102**, 2254–2259.

Foucault, Th., Order flow composition and trading costs in a dynamic limit order market. *J. Financ. Market*, 1999, **2**, 99–134.

Foucault, Th., Kadan, O. and Kandel, E., Limit order book as a market for liquidity. *Rev. Financ. Stud.*, 2003, **18**, 1171–1217.

Gillemot, L., Farmer, J.D., Lillo, F., There's more to volatility than volume. physics/0510007, 2005.

Glosten, L.R., Components of the Bid–ask spread and the statistical properties of transaction prices. *J. Finan.*, 1987, **42**, 1293–1307.

Glosten, L.R. and Milgrom, P., Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *J. Financ. Econ.*, 1985, **14**, 71–100.

Handa, P. and Schwartz, R.A., Limit order trading. *J. Finan.*, 1996, **51**, 1835–1861.

Handa, P., Schwartz, R.A. and Tiwari, A., The ecology of an order-driven market. *J. Portfolio Manag.*, 1998, **Winter**, 47–56.

Harris, L. and Hasbrouck, J., Market versus limit orders: the SuperDOT evidence on order submission strategy. *J. Finan. Quant. Anal.*, 1996, **31**, 213–231.

Hasbrouck, J., Measuring the information content of stock trades. *J. Finan.*, 1991, **XLVI**, 179–207.

†The long-memory property of σ is argued in Plerou *et al.* (2000) to be related to long range correlation in the trading frequency rather than in the volatility per trade, but see Gillemot *et al.* (2005).

Hollifield, B., Miller, A. and Sandas, P., Empirical analysis of limit order markets. *Rev. Econ. Stud.*, 2004, **71**, 1027–1063.
Hopman, C., Are supply and demand driving stock prices? *MIT Working Paper*, 2002 (MIT), to appear in *Quant. Finan.*
Huang, R.D. and Stoll, H.R., The components of the bid–ask spread: a general approach. *Rev. Financ. Stud.*, 1997, **4**, 995–1034.
Kyle, A.S., Continuous auctions and insider trading. *Econometrica*, 1985, **53**, 1315–1335.
Lillo, F. and Farmer, J.D., The long memory of efficient markets. *Stud. Nonlinear Dyn. Econometrics*, 2004, **8**, 1–26.
Lillo, F., Farmer, J.D. and Mantegna, R., Master curve for price-impact function. *Nature*, 2003, **421**, 129–130.
Lillo, F., Mike, S. and Farmer, J.D., Theory for long memory in supply and demand. *Phys. Rev. E*, 2005, **71**, 066122.
W.-M. Liu, Monitoring and Limit Order Submission Risks. *Working Paper*, 2005 (University of New South Wales).
Luckock, H., A steady-state model of the continuous double auction. *Quant. Finance*, 2003, **3**, 385–404.
Madhavan, A., Market microstructure: a survey. *J. Financ. Markets*, 2000, **3**, 205–268.
Madhavan, A., Richardson, M. and Roomans, M., Why do security prices fluctuate? A transaction-level analysis of NYSE stocks. *Rev. Financ. Stud.*, 1997, **10**, 1035–1064.
Mike, S. and Doyne Farmer, J., An empirical behavioral model of price formation. physics/0509194, 2005.
Orléan, A., *Le Pouvoir de la Finance*, 1999 (Odile Jacob: Paris).
Orléan, A., A quoi servent les marchés financiers? in, *Qu'est-ce que la Culture?*, 2001 (Odile Jacob: Paris).
O'Hara, M., *Market Microstructure Theory*, 1995 (Blackwell: Cambridge, MA).
Plerou, V., Gopikrishnan, P., Amaral, L.A.N., Gabaix, X. and Stanley, H.E., Diffusion and economic fluctuations. *Phys. Rev. E*, 2000, **62**, 3023–3026.
Rosenow, B., Fluctuations and market friction in financial trading. *Int. J. Mod, Phys. C*, 2002, **13**, 419–425.
Rosu, I., A dynamic model of the limit order book. *Working Paper*, 2005 (University of Chicago).
Sandas, P., Adverse selection and competitive market making: empirical evidence from a limit order market. *Rev. Financ. Stud.*, 2001, **14**, 705–734.
Smith, E., Farmer, J.D., Gillemot, L. and Krishnamurthy, S., Statistical theory of the continuous double auction. *Quant. Finance*, 2003, **3**, 481–514.
Stoll, H., Friction. *J. Finance*, 2000, **55**, 1479–1514.
Weber, P. and Rosenow, B., Order book approach to price impact. *Quant. Finance*, 2005, **5**, 357–364.
Zumbach, G., How the trading activity scales with the company sizes in the FTSE 100. *Quant. Finance*, 2004, **4**, 441–456.

Appendix A: Impact and volatility in the MRR model

From the basic equation determining the dynamics of the mid-point,

$$m_{i+1} - m_i = p_{i+1} - p_i - \theta\rho(\varepsilon_i - \varepsilon_{i-1}) = \xi_i + \theta(1-\rho)\varepsilon_i, \tag{A1}$$

one gets

$$m_{i+\ell} - m_i = \sum_{j=i}^{i+\ell-1} \xi_j + \theta(1-\rho)\sum_{j=i}^{i+\ell-1} \varepsilon_j. \tag{A2}$$

Therefore, taking into account the correlation between εs, and the assumption that external shocks are uncorrected with the order flow, the impact function is

$$\mathcal{R}_\ell = \langle \varepsilon_i (m_{i+\ell} - m_i)\rangle = \theta(1-\rho)\sum_{\ell'=0}^{\ell-1} \rho^{\ell'} = \theta(1-\rho^\ell). \tag{A3}$$

Note that in this model, the 'bare' impact function $G_0(\ell)$ defined in Bouchaud *et al.* (2004, 2006) through

$$m_i = \sum_{j=-\infty}^{i-1} \xi_j + \sum_{j=-\infty}^{i-1} \xi_j + \sum_{j=-\infty}^{i-1} G_0(i-j-1)\varepsilon_j, \tag{A4}$$

is here found to be constant, equal to $G_0(\ell) = \theta(1-\rho)$. Finally, one finds

$$\sigma_1^2 = \langle (m_{i+1} - m_i)^2 \rangle = \Sigma^2 + \theta^2(1-\rho)^2 \tag{A5}$$

and

$$\sigma_\infty^2 = \Sigma^2 + \theta^2(1-\rho)^2\left(1 + 2\frac{\rho}{1-\rho}\right) = \Sigma^2 + \theta^2(1-\rho^2). \tag{A6}$$

More generally, assuming that only the sign surprise matters, one can write, for arbitrary correlations between signs:

$$m_{i+\ell} - m_i = \sum_{j=i}^{i+\ell-1} \xi_j + \theta \sum_{j=i}^{i+\ell-1} \varepsilon_j - \langle \varepsilon_{j+1} \rangle_j, \tag{A7}$$

where the last term is the conditional expectation of the next sign. The impact function now generalizes to

$$\mathcal{R}_\ell = \theta[1 - C(\ell)], \tag{A8}$$

and therefore $\lambda_\infty = 1/(1 - C_1)$.

Appendix B: summary statistics

This appendix gives the summary statistics in tabular form, see tables 1–3.

Table 1. Codes and names of the PSE stocks analysed in table 2.

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	Sodexho Alliance
FR	Valeo	TEC	Technip
FTE	France Telecom	TFI	Television Francaise 1
GFC	Gecina	TMM	Thomson
GLE	Societe Generale	UG	Peugeot
GL	Galeries Lafayette	UL	Unibail
HAV	Havas	VIE	Veolia Environnement
HO	Thales	ZC	Zodiac

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 σ_1, the average spread $\langle S \rangle$, the spread standard deviation σ_S, the average response $\overline{\mathcal{R}}_1$ 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	Tnv	$\langle q_t \rangle$	$\langle v \rangle$	#	σ_1	$\langle S \rangle$	σ_S	$\overline{\mathcal{R}}_1$	C_1	λ_∞	Tick
ACA	15.4	78	11.1	347	9.12	20.76	18.28	2.30	0.28	1.70	5.13
AC	22.1	87	26.3	211	8.97	12.45	13.00	2.99	0.26	2.16	2.68
AF	4.0	40	7.2	139	14.40	25.96	25.57	3.85	0.23	2.30	7.01
AGF	9.6	80	16.9	142	13.32	19.17	18.88	4.28	0.28	2.28	5.46
AI	34.0	154	26.9	316	8.21	13.72	11.07	3.06	0.26	1.42	6.87
ALS	14.2	57	13.0	274	14.75	23.11	20.80	4.84	0.14	1.84	12.02
ALT	12.7	53	14.6	217	23.66	30.75	27.88	7.15	0.17	2.71	11.04
AVE	134.1	362	62.6	536	6.70	11.18	7.31	2.59	0.21	1.72	7.54
BB	0.8	30	11.9	16	28.80	43.94	47.51	7.78	0.33	1.66	2.55
BN	59.5	310	45.9	324	6.30	11.80	7.79	2.37	0.28	1.03	7.62
CAP	34.7	94	28.9	300	12.71	17.27	17.03	4.39	0.20	2.55	5.17
CA	79.8	229	40.1	497	6.92	12.20	9.16	2.48	0.22	1.85	5.35
CDI	5.1	59	14.9	86	18.76	32.32	29.08	6.23	0.29	1.74	2.79
CGE	79.8	152	21.3	936	10.49	20.94	14.85	3.31	0.13	1.70	15.37
CK	0.3	37	11.4	6	20.43	46.72	32.24	7.37	0.22	0.93	8.06
CL	30.1	163	23.0	328	8.05	13.60	13.04	1.96	0.36	1.77	3.43
CNP	1.7	41	9.3	45	15.58	31.12	33.74	3.53	0.31	1.43	2.67
CO	13.8	98	22.9	151	10.38	15.43	14.42	3.47	0.27	1.50	6.58
CS	96.1	144	36.6	657	8.32	12.78	10.17	3.13	0.17	1.97	6.39

(*continued*)

Table 2. Continued.

Code	Tnv	$\langle q_t \rangle$	$\langle v \rangle$	#	σ_1	$\langle S \rangle$	σ_S	$\overline{\mathcal{R}}_1$	C_1	λ_∞	Tick
CU	1.0	22	7.7	32	30.02	44.83	47.65	8.85	0.23	1.84	3.97
CY	15.4	2148	77.1	50	8.64	14.61	14.49	2.35	0.26	1.67	8.03
DEC	0.7	27	12.7	13	43.75	78.92	77.84	13.81	0.33	1.53	8.15
DG	16.8	134	25.4	165	8.02	14.05	10.90	2.76	0.22	1.24	7.76
DSY	10.4	65	19.6	133	17.06	22.11	23.83	5.71	0.27	2.30	5.28
EF	5.6	59	20.4	69	13.21	21.38	23.01	3.94	0.25	1.90	2.52
EN	17.8	66	21.3	210	10.15	14.55	14.35	3.47	0.25	1.85	3.46
FP	322.4	662	114.4	705	4.96	9.17	4.86	2.12	0.20	1.49	6.71
FR	8.3	80	24.1	86	13.84	20.60	22.46	4.39	0.25	1.99	3.50
FTE	112.7	137	29.5	956	9.12	15.04	12.57	2.93	0.14	1.85	6.39
GFC	0.3	35	9.9	7	17.85	27.74	31.55	4.56	0.27	0.98	5.62
GLE	82.7	239	46.0	449	8.07	12.70	9.27	3.04	0.23	2.26	7.07
GL	0.9	43	13.6	17	24.34	41.34	39.96	6.92	0.29	1.38	7.24
HAV	8.1	57	14.6	139	21.15	32.44	26.53	7.41	0.18	1.91	17.15
HO	11.3	61	19.7	143	10.68	15.02	15.86	3.65	0.29	1.94	2.85
IFG	3.2	24	4.9	163	29.05	40.35	32.95	8.51	0.15	2.31	21.79
LG	38.2	193	36.6	261	7.82	13.31	9.99	2.90	0.24	1.67	7.33
LI	0.2	50	15.8	3	14.43	29.90	23.18	4.69	0.20	0.70	8.54
LY	58.4	111	28.8	507	8.40	12.84	11.94	3.11	0.20	1.97	4.31
MC	52.9	143	34.7	381	8.01	12.41	10.75	3.01	0.26	2.20	4.18
ML	13.2	71	21.2	156	10.23	14.98	15.36	3.35	0.30	1.93	2.71
MMB	13.3	76	18.9	177	10.67	15.96	15.84	3.60	0.25	1.94	3.50
MMT	0.7	33	10.1	17	33.82	49.77	48.42	9.71	0.32	1.66	3.57
NAD	4.6	47	6.0	188	14.39	31.27	20.91	4.11	0.17	1.89	20.36
NK	1.1	43	13.4	21	25.95	42.89	39.34	7.69	0.29	1.40	8.07
OGE	36.4	182	21.2	429	11.40	21.53	13.35	3.57	0.16	1.82	16.03
OR	68.0	211	41.9	406	7.30	12.45	9.45	2.76	0.25	1.39	6.59
PEC	9.5	110	31.9	75	15.06	23.71	24.89	5.07	0.24	2.41	5.45
PP	36.8	154	31.5	292	10.08	15.44	13.25	3.39	0.25	2.23	7.23
PUB	11.9	78	27.3	109	15.08	21.19	22.92	4.97	0.23	2.46	3.85
RF	0.1	25	7.0	3	22.27	37.26	32.69	7.07	0.23	0.72	8.25
RHA	2.1	32	9.7	55	21.45	33.99	32.97	6.71	0.22	1.87	10.83
RI	12.6	138	39.1	80	10.49	16.82	16.01	3.64	0.22	1.49	6.42
RNO	35.8	158	34.4	260	8.01	12.80	11.13	2.71	0.25	2.27	4.38
RXL	0.7	34	13.0	14	31.30	51.91	50.49	9.21	0.26	1.50	6.77
SAG	1.5	36	10.3	35	24.59	43.15	42.89	7.29	0.22	1.73	7.48
SAN	94.2	301	56.4	417	7.76	12.18	8.60	3.04	0.25	1.49	7.92
SAX	6.0	57	20.4	73	23.28	33.48	33.79	7.45	0.27	2.58	5.23
SCO	2.0	25	9.3	55	35.67	38.88	40.00	8.37	0.23	2.16	7.40
SC	0.5	55	13.9	10	12.24	21.08	19.01	3.58	0.22	1.28	6.10
SU	26.2	129	33.0	198	9.52	14.60	13.43	3.30	0.26	2.15	5.63
SW	11.2	67	19.5	144	14.48	18.19	20.40	4.26	0.32	1.92	3.22
TEC	9.4	123	31.6	74	16.27	24.27	25.77	5.06	0.24	2.18	7.31
TFI	17.4	63	21.0	207	11.91	15.87	16.19	4.09	0.25	2.07	3.71
TMM	28.0	78	20.8	338	10.08	16.08	15.59	3.18	0.19	2.34	4.29
UG	33.2	141	36.2	229	7.95	11.91	10.80	2.86	0.26	2.11	4.43
UL	3.0	64	22.9	33	14.61	24.45	23.47	4.87	0.27	1.41	8.07
VIE	19.8	77	24.8	199	11.52	16.60	17.81	3.75	0.23	2.23	3.57
ZC	0.9	33	8.7	26	24.95	41.99	42.40	7.28	0.24	1.64	4.21

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 $\langle v \rangle / \langle q_t \rangle \approx 10^{-2}$.

Code	Turnover	$\langle q_t \rangle$	$\langle v \rangle$	# trade	σ_1	$\langle S \rangle$	σ_S	$\overline{\mathcal{R}_1}$	Tick
LMEB	262	21	199	211	8.4	46.7	4.4	11.5	46.6