

# Are Supply and Demand Driving Stock Prices?

Carl Hopman<sup>\*†</sup>

December 11, 2002

## ABSTRACT

This paper attempts to shed new light on price pressure in the stock market. I first define a rigorous measure of order flow imbalance using limit order data. It turns out that this imbalance is highly correlated with stock returns, with  $R^2$  around 50% for the average stock. This price impact of orders does not appear to be reversed later. In fact, the correlation between order flow and return is observed for micro time intervals of ten minutes all the way to macro time intervals of three months. I then attempt to distinguish between private information and uninformed price pressure by looking at the implications of a private information model. For idiosyncratic returns, where one would expect private information to be important, and the  $R^2$  to be high, the  $R^2$  is indeed around 41%. However, for the common market return, where one would expect private information to be minor, the  $R^2$  is even higher at 70%. The high  $R^2$  on the market suggests that private information does not explain well the observed co-movement of orders and prices. This points toward a bigger role for uninformed price pressure than is usually assumed, which, for example, could lead to the formation of stock market bubbles.

---

<sup>\*</sup>I would like to thank Cécile Boyer, Xavier Gabaix, Denis Gromb, S.P. Kothari, Jonathan Lewellen, Andrew Lo, Angelo Ranaldo, Dimitri Vayanos and Jiang Wang for their comments and suggestions. I am especially grateful to Xavier Gabaix for stimulating discussions and continuous encouragement for this project. I would also like to thank Mathias Auguy, Patrick Hazart and Bernard Perrot from the Paris Bourse for their help in making the data available.

<sup>†</sup>chopman@mit.edu, Massachusetts Institute of Technology.

# 1 Introduction

In *The General Theory of Employment, Interest and Money*, John Maynard Keynes concludes, “Thus certain classes of investment are governed by the average expectation of those who deal on the Stock Exchange as revealed in the price of shares, rather than by the genuine expectations of the professional entrepreneur.” In other words, prices will not be the rational valuation of a “professional entrepreneur,” but will reflect traders’ psychology through the mechanism of supply and demand. It seems that today’s financial press also has some sympathy for the price pressure argument, as it often explains price drops by heavy selling, or price increases by an excess of buyers, even in the absence of a change of fundamentals. However, the very elegant efficient market paradigm, and the work of Scholes (1972) in particular, has questioned this insight and argued that, except for short term temporary adjustments, prices will be driven only by information, either public or private.

An important problem for this paradigm has been raised by Roll (1988). His results suggest that news cannot explain more than 30% of the price changes, instead of 100% as one could at first expect. This  $R^2$  includes the regression on the industry and principal components of the return, which are assumed to reflect perfectly market-wide news. This means that only the relationship between company specific movements and news is really examined, and not assumed. To study these idiosyncratic movements, Roll distinguishes days with company news from days without. The idea is that on days without company specific news, the market wide factors should be the sole determinants and the  $R^2$  on these factors should be close to 100%. However, the difference in  $R^2$  between days without news and all days together is less than two percentage points (and both  $R^2$  are below 30%). It seems that a lot of the idiosyncratic variance exists without any idiosyncratic news. So for the idiosyncratic part of stock returns, where the effect of news is really studied, the news does not seem to explain returns very well. This result has been an important puzzle since its publication, and it is interesting to find variables that can account for price changes better than news do.

A related paper is French and Roll (1986), who show that public information explains only a small part of stock returns. They study stocks on days when the New York Stock Exchange is closed but the rest of the economy is active, and find that the variance on

these days<sup>1</sup> is on average only 14.5% of what it is when the exchange is open. Since the flow of public information is at least as big during these exchange holidays (which include Presidential elections), their results suggest that only 14.5% of a stock's daily variance can be attributed to public information. So public information<sup>2</sup> does not explain price changes very well.

In this paper I do not consider public information, concentrating instead on private information and mechanical price pressure. The importance of private information and the way it can affect prices as well as orders has been explored in the microstructure literature, with the work of Kyle (1985) in particular. In this context it is admitted that prices will move with supply and demand imbalance because the change in demand might reveal some private information. However, the quantitative importance of supply and demand as one of the driving forces of stock return has not been established, and it is often ignored altogether by mainstream finance, as it is taught, for instance, in MBA programs. Indeed, it is often argued that there is no imbalance, because the volume bought by some is equal to the volume sold by others.

In this paper, I hope to shed new light on the relationship between stock return and supply and demand imbalance. I first define rigorously the order flow imbalance. To that end, I do not rely on realized transactions, where the volume bought is equal to the volume sold, but on unrealized intentions, in the form of limit orders, where an imbalance can exist. This new measure of order flow imbalance turns out to be strongly correlated with stock movements, with an  $R^2$  around 50% for the average stock, higher than the  $R^2$  of Roll (1988). This impact of order imbalance on the price is not reversed later, and remains true not only for micro time horizons of 10-30 min., but also for macro time periods of three months. To address the causality issue, I show that there is no reverse causality, and argue that a common driving factor would be part of public information, ruled out by the work of French and Roll (1986). Finally, I attempt to disentangle two causal explanations:

---

<sup>1</sup>They estimate a two day variance, from the closing price before the holiday to the closing price on the day after the holiday, so that public news have a full trading day to affect prices. The two day variance is 1.145 a normal day's variance.

<sup>2</sup>In fact their argument also deals with a particular type of "private information": faster analysis of public information, as long as it is analyzed during the next day. So in this paper I do not consider this kind of "private information."

private information and mechanical price pressure. To that end, I distinguish the market return from the idiosyncratic return and study the implications of a private information model in the spirit of Kyle (1985), where order flow imbalance is a good measure of private information. Since the potential for leakage of company-specific information is greater than that of market-wide information, one would expect a relatively large fraction of idiosyncratic movements to be explained by the order flow imbalance, and indeed the  $R^2$  is about 41%. On the other hand, one would expect a relatively small fraction of the global market movements to be explained by the order flow imbalance, but the  $R^2$  is even higher at 70%. I, therefore, argue that the private information model is not well supported by the data and that it could be useful to look deeper into uninformed price pressure.

It is useful to see how uninformed orders could have a long term impact on the price. It is well known from the microstructure literature that uninformed orders can have a temporary impact: a buy order will push the price upwards for inventory reasons with a market maker, mechanically with a limit order market. However, information efficiency would suggest that this effect is only temporary if the order is uninformed: arbitrageurs will soon provide the necessary liquidity to bring the price back to its previous<sup>3</sup> level. But it is also possible that these arbitrageurs don't bring it completely back because the benefit is small and the risk is high (the price will indeed one day come back to its efficient value but the arbitrageur may need to wait a long time and face huge price changes in between). In sum, noise trades are not faced by infinite liquidity, and therefore have a long-term impact on the price.

If this long term impact accumulates over days and years, it can eventually create bubbles, such as the March 2000 Internet bubble. Of course, these bubbles will eventually burst, and prices will revert to fundamentals after ten years or so. However, the interim deviations are sufficiently important to warrant interest in their own right. Here I suggest that these bubbles could originate in the microstructure impact of the order flow imbalance when the orders are placed mainly by uninformed investors.

Several previous papers distinguish buyer initiated and seller initiated transactions to explain price changes. Although their measure is based on realized transactions, it is still related to my measure of order flow imbalance, since an excess of limit buy orders is likely to

---

<sup>3</sup>Modified for the information which has arrived in between.

generate aggressive buy orders that result in buyer initiated transactions. Hasbrouck (1991) uses NYSE transaction data and concludes that the impact of a trade is a positive, increasing and concave function of its size, with  $R^2$  at 10%. Hausman, Lo, and MacKinlay (1992) use an ordered probit model to take directly into account the discreteness of the tick size. Evans and Lyons (2000) also use signed transaction data on the Foreign Exchange dealer market rather than on the stock market, and they find  $R^2$  similar to mine, showing that the imbalance explains the Forex returns quite well. Compared to mine, their data only include realized transactions, without their volume or intraday information on a sample of only 89 trading days. They interpret their finding in a private information framework. However, they call private information the volume that traders are willing to buy or sell. This is something which noise traders with absolutely no information about the financial asset know as well. It is, therefore, not very different from a direct price pressure interpretation.

Other papers look at the impact of imbalances on stock prices by restricting themselves to large trades. This literature started with Scholes (1972). He finds that the impact of a trade does not increase with the block size, and concludes by rejecting the price pressure hypothesis. Yet in a later study of large trades, Holthausen, Leftwich, and Mayers (1990) use high frequency transaction data, which yield more precise estimates of the impact of large trades, and find that the impact does increase with the trade size (as Hasbrouck (1991) finds without restricting himself to large trades). This can cast doubt on the original conclusion of Scholes (1972), and the rejection of the price pressure hypothesis.

Other papers document a consistent link between volume and volatility, surveyed for example in Karpoff (1987). This link is a direct implication of the impact of trades and orders on asset prices, by taking the absolute value on both sides.

In the next section, I describe the data from the Paris Bourse and I define the order flow measure, taking into account the concavity of the price impact of an order as a function of its volume. I provide some summary statistics and time series properties of the order flow imbalance. The third section presents the relationship between the aggregated order flow imbalance and the stock return, and differentiates the impact of predictable versus unpredictable orders. I then show that this impact is not reversed subsequently, and that the correlation is true for very different time horizons. The fifth section attempts to distinguish

between private information and mechanical price pressure, and presents a simple model of mechanical price pressure.

## 2 Data, Definitions, and Summary Statistics

### 2.1 The Data

One of the most common arguments against the study of supply and demand for financial assets is that there is no imbalance. Indeed, the volume bought is equal to the volume sold when you look at realized transactions. To get around this problem, some researchers have distinguished between buyer and seller initiated transactions. However, this distinction does not solve the equality objection in a pure market maker setup where no limit orders are allowed. Indeed, suppose that only market orders are allowed, with a market maker who clears his inventory regularly,<sup>4</sup> then, even if the econometrician knows perfectly whether the market maker was on the buy side or the sell side, the total volume sold to the market maker is equal after each inventory clearing to the total volume bought from him. Therefore, there is never any imbalance in volume. This property limits the effectiveness of using transaction data to measure the order flow in a pure market maker setup, and probably extends to markets where limit orders are rare.

This is the main reason why I use the Paris Bourse data: limit orders are the norm not the exception, and their submission is available to the econometrician. Although in transaction terms the volume bought is still equal to the volume sold, in submission terms there can be many submitted orders that are never executed. There can, therefore, be an imbalance between submitted Buy and Sell orders (some are later executed, some are not) which I measure directly with this data set.<sup>5</sup>

---

<sup>4</sup>This condition can be weakened to having a bounded inventory with the equality between buy and sell volume true in the limit.

<sup>5</sup>If one goes deeper, one could ask what happens to the unexecuted limit orders. They of course get cancelled, most of them automatically at the end of the day or the month. If the impact of all the cancellations is equal to that of all the submissions, then again the total net (submitted minus cancelled) volume of orders is equal on the buy and sell side with a monthly time frame (and both are equal to the transaction volume). However, submissions are usually made relatively close to the current best quote, whereas cancellations often happen automatically, after the price has moved away and the submitted order has remained. If one considers a cancellation as a negative submission, then submitted orders, for a given volume, have a decreasing impact when submitted further away from the best quotes. As a first approximation, one can then argue that

Although market participants may feel the imbalance on markets that rely on market-makers instead of limit orders, it is not obvious how to measure<sup>6</sup> their unrealized wishes, i.e. the supply and demand imbalance. On the contrary, on the Paris Bourse limit orders are dominant and the imbalance is easy to measure. Besides, the Paris Bourse data is very clean and complete. Because the Bourse is a fully automated electronic exchange,<sup>7</sup> it is virtually free of errors.

The Paris Bourse is an order driven market, and there is no market maker, or any appointed liquidity provider. Traders give their orders to brokers who then pass them on to the central computer. It is then available on the traders' screen, usually within the next second. The Paris Bourse allows agents to place different types of orders. The most common is the limit order. Its main characteristic is to have a maximum price (for a buy order) at which the agent is ready to buy the stock (the buy and sell orders have exactly symmetric properties, so I will only describe buy orders). If a submitted buy order is higher than the current best ask, it is immediately executed. If not, it remains on the order book until either it is hit by a sell order, or it is cancelled, or it has a preassigned finite life<sup>8</sup>. If two different buy orders are at the same price, then a time priority is given to the order first entered in the book<sup>9</sup>.

There are two types of market orders. The first type is executed in full only if its volume is less than the available volume at the ask price. If not, the remaining volume is transformed into a limit buy order at this old ask price. The second type of market order is immediately executed in full, against all the available counterparts in the sell order book (and not only against the volume available at the ask price).

---

submissions are close enough to the best quotes to have an impact, whereas cancellations are not and can be ignored. This is the approximation I am forced to make, since I do not have the cancellation data. Having this data would allow me to have an even better estimation of the impact of supply and demand. As we'll see, the approximation already yields very good results.

<sup>6</sup>If one wanted to build an order flow imbalance measure on the NYSE for example, one could do the following. First identify buyer initiated trades from seller initiated, using the Lee and Ready or a similar algorithm. Because there are some limit orders on this market, there can be an imbalance in volume and one could use the net volume of trades. But because of the concavity described in section 2.4, using the SQRT aggregation defined in section 2.5 would be a better measure.

<sup>7</sup>Biais, Hillion, and Spatt (1995) provide a detailed description of the microstructure of the Paris Bourse.

<sup>8</sup>All the orders are automatically cancelled at the end of each Bourse month.

<sup>9</sup>When a limit order is submitted, it is possible to hide some part of it. The hidden part is not visible by any trader or broker until it gets executed. However, the impact of both parts are quite similar, and I do not distinguish between them in this paper.

The database goes from January 4<sup>th</sup> 1995 to October 22<sup>nd</sup> 1999 inclusive. I only look at the continuous trading session, which, until September 19<sup>th</sup> 1999, started just after 10AM and finished at 5PM. From September 20<sup>th</sup> 1999, it started at 9AM and finished at 5PM<sup>10</sup>.

The database includes all the transactions and all the orders that were submitted on the Paris Bourse, as well as the best quotes available at any time. In comparison, the TORQ (Trades, Orders, Reports and Quotes) database for the New-York Stock Exchange (NYSE) misses about half the total volume of submitted orders (Kavajecz (1999)).

The main French Index is the CAC40, which includes the 40 biggest stocks. I looked at the 40 stocks that were part of the CAC40 in January 1995. At the end of the sample, 34 of them were still quoted as independent companies, so the results of this paper are provided only<sup>11</sup> for these 34 stocks.

To make things more concrete, I sometimes present the results obtained for one company, Lafarge, which has average properties in many directions. But to show that the results are general, I prefer when possible to report the average results for the 34 stocks.

## 2.2 Variable Definitions

I calculate the (log) return using mid-quotes. I also performed robustness checks using transaction prices instead of the mid-quote and got nearly identical results.

I used different time horizons, from 10 min., 30 min., one day, one week, one month up to three months. At ten min., there is still quite a bit of microstructure noise (as measured by the bid-ask bounce or negative auto-correlation which disappears at 30 min.). On the other hand, at three months I have only 20 independent data points and, therefore, little statistical power. However, similar results were obtained for these widely different time intervals as reported in Table 12. For horizons longer than one day, the return is calculated from close to close. For one day, one can either calculate the return from opening to close (night excluded) or from close to close (night included) and I present results for both cases.

---

<sup>10</sup>There are also 2 call auctions, one just before the opening, the other at 5:05PM which was created on June 2 1998, but I remove all the order flow data they generated, because it is harder to define and measure order imbalance in these auctions.

<sup>11</sup>The 40 companies allow me to check for survivorship bias. However, since the results were similar for the 6 stocks that disappeared, and to have comparable results, I report results only for the 34 surviving stocks.

I distinguish the different buy orders (and similarly the sell orders) according to the level of urgency chosen by the trader submitting the order. This corresponds to the speed with which it is likely to be executed. The reason for this distinction is that one would expect more urgent orders (of similar size) to have a bigger<sup>12</sup> impact on the price, as observed in Figure 2.4. The basic distinction is between:

1. market orders, which are executed immediately;
2. spread orders, which are submitted between the best bid and ask (and thus change either the bid or the ask);
3. book orders, which are submitted inside the order book.<sup>13</sup>

To be more precise, after calculating the log-mid-quote  $p_0 = \frac{\ln(bid) + \ln(ask)}{2}$ , I call a buy order:

1. market if executed immediately, i.e. all the market orders and limit orders such that  $P \geq ask$ ;
2. spread if placed within the spread:  $ask > P > bid$ ;
3. book if  $\ln(bid) \geq \ln(P) > p_0 - 0.005$ .

To have consistent and symmetric definitions, I use the natural logarithm. However, one can consider the mid-quote as being roughly the arithmetic average between the bid<sup>14</sup> and the ask price. One can also think of book orders as being above the mid-quote minus 0.5% (for buy orders).

To define the order flow, I follow the work of Lo and Wang (2000) and use the share turnover (the number of shares in the order divided by the number of shares outstanding) as a measure of the volume of each submitted order. I call  $v_i$  the volume in share turnover

---

<sup>12</sup>This is predicted by the mechanical impact, which is direct for market and spread orders but indirect and less likely for book orders. It is also predicted by a private information setup, where a privately informed agent would want to use his information before others know it, so that urgent orders would on average be more informed.

<sup>13</sup>I discard orders too far away from the best quotes, as I have found that their impact is negligible.

<sup>14</sup>I use the best quotes available when the order is submitted, not outdated ones from the beginning of the time interval.

	Lafarge
Sample size	1202
Volatility (night excluded)	1.75%
Volatility (night included)	2.06%
Number of orders	
market buy	250
market sell	257
spread buy	101
spread sell	97
book buy	172
book sell	164
Average volume of one order $\times 10^6$	
market buy	7.3
market sell	7.0
spread buy	9.6
spread sell	9.7
book buy	11.5
book sell	12.4

Table 1: Summary statistics for Lafarge over one day.

The results are reported for the Lafarge stock over one day. Only the volatility measure changes when one includes or excludes the night.

	Mean	Std. Dev.
Sample size	1202	0
Volatility (night excluded)	1.73%	0.21%
Volatility (night included)	2.13%	0.46%
Number of orders		
market buy	241	126
market sell	282	190
spread buy	77	30
spread sell	73	28
book buy	166	90
book sell	158	86
Average volume of one order $\times 10^6$		
market buy	9.38	5.89
market sell	8.49	5.33
spread buy	12.3	8.90
spread sell	12.1	7.73
book buy	16.3	16.7
book sell	15.6	10.8

Table 2: Summary statistics over one day.

The results reported are the average of the results on the 34 stocks, and the cross-section standard deviation.

of each order  $i$ . Robustness checks confirm that using share volume or dollar volume gives very similar results.

To measure liquidity, I use the Weighted Average Spread (WAS). This information is also available from the database at each point in time. It consists of the weighted average bid (W.A.Bid) and ask (W.A.Ask). The W.A.Ask is the price which would be reached by a large buy market order fully executed against the current available book. The size of the large market order that is used to calculate the WAS is called the block size and is chosen by the Paris Bourse using liquidity criteria for that stock.<sup>15</sup> A large WAS means it is hard to place large orders and is a proxy for illiquidity. I define the:

- WAS facing buy orders as  $\ln(\text{W.A.Ask}) - \text{mid}$ ;
- WAS facing sell orders as  $-(\ln(\text{W.A.Bid}) - \text{mid})$ .

<sup>15</sup>For Lafarge for example, it is  $5 \times 10^{-5}$  of the shares outstanding and the average one sided WAS is 53 basis points.

## 2.3 Summary Statistics

Table 1 gives some summary statistics for Lafarge. Table 2 gives average summary statistics for the 34 stocks.

## 2.4 The Impact as a Concave Function of the Volume of each Order

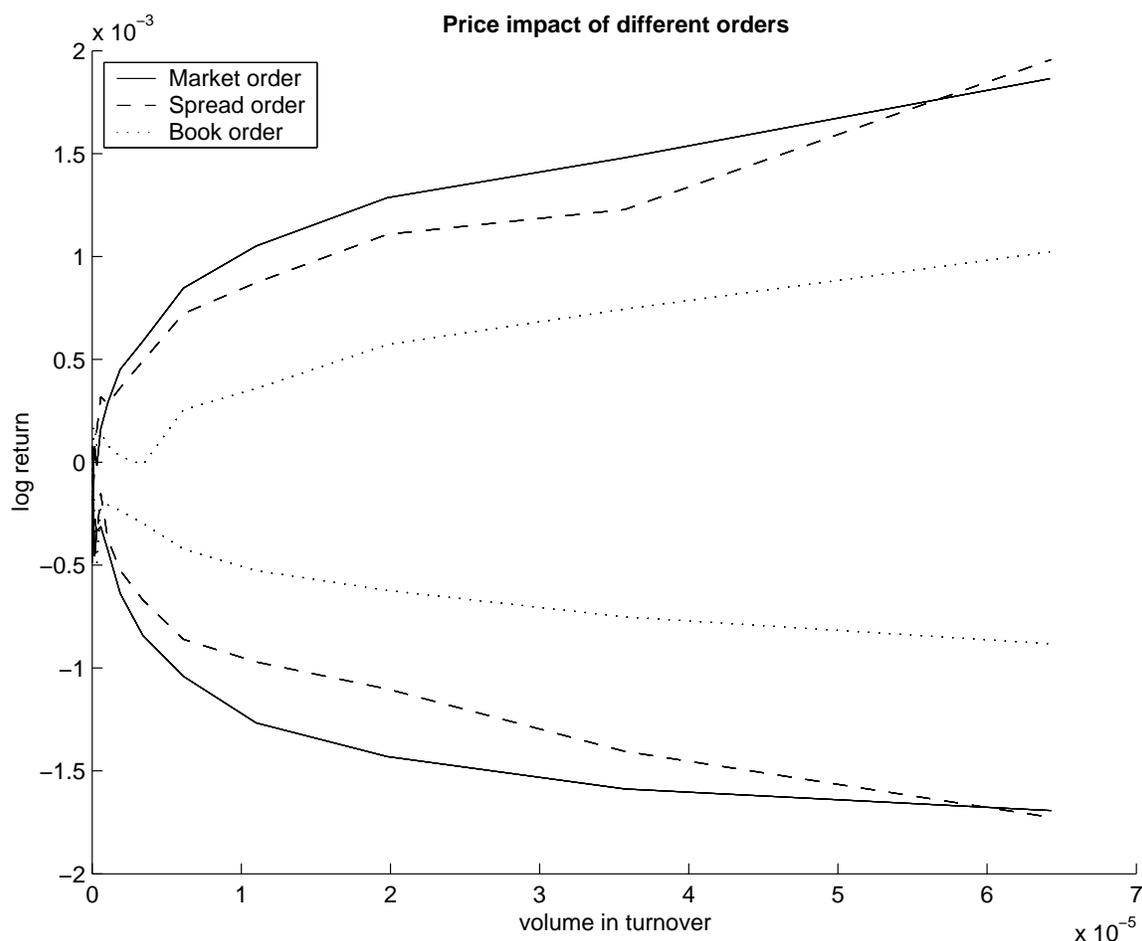


Figure 1: The 30 min. impact of one order on Lafarge's price, as a non-parametric function of its volume. The log-return is calculated from just before the order arrives to 30 min. after it has arrived. I use the Nadaraya-Watson kernel regression with Epanechnikow kernel. I distinguish the orders by their urgencies and between buy and sell orders. The results are reported for the Lafarge stock.

To have a better idea of how each order affects the price in the “long<sup>16</sup>” run, I study the

<sup>16</sup>The 30 min. impact is not reversed later. If anything, it tends to increase a little as I verify with a

change in log price from before the order arrives to 30 min.<sup>17</sup> after it has arrived. I use the non-parametric Nadaraya-Watson kernel<sup>18</sup> regression to find out about non-linearities in the price change as a function of the order's volume. The results are reported in Figure 2.4 for Lafarge. Similar results are obtained for the other stocks.

One can see that the price impact is a concave function of the volume of the order. It is also observed that more urgent orders have a larger impact, for a given volume. The curves that are obtained look similar to a power function  $r_t = \lambda v_i^\delta$ .

	Market Buy	Spread Buy	Book Buy
$\lambda \times 10^3$	84	85	142
(std. err. $\times 10^3$ )	(14)	(41)	(98)
$\delta$	.37	.38	.47
(std. err.)	(.03)	(.04)	(.05)

Table 3: Power function estimate for the 30 min. impact of an order as a function of its volume. The log-return is calculated from before the order arrives to 30 min. after it has arrived. I use non linear least squares to estimate the impact as a power function of the volume.  $r_t = \lambda v_i^\delta$ . I report the coefficients  $\lambda$  and  $\delta$ , and their standard errors estimated by block bootstrap. I report the estimates and standard errors averaged over the 34 stocks.

I use non-linear least squares to estimate  $\lambda$  and  $\delta$ , and report the results of this estimation for the three urgencies of buy orders in Table 3. The standard errors are estimated using block bootstrap, with a block size of one week, to take into account overlapping data, temporal dependence and heteroscedasticity. Although the power  $\delta$  appears to be slightly different between the different types of orders, I choose the approximation  $\delta = 0.5$  when doing the aggregations in the subsequent sections.

This concavity result has been known since at least Hasbrouck (1991). It can seem at first surprising. Someone could interpret this result as an advice to bundle orders instead of splitting them, in order to reduce the price impact of a given volume. This is contrary to what is observed in practice and would be a bad advice for two reasons. First, traders are

---

60 min. non parametric regression compared to the 30 min. reported in Figure 2.4. In addition, when regressing the 30 min. return on the lagged 30 min. order flow I also get a small but statistically significant positive coefficient which confirms the small continuation.

<sup>17</sup>The 30 min. interval is chosen because at shorter horizons the return is negatively autocorrelated (bid-ask bounce). On the other hand, the horizon is short enough to have as much statistical power as possible.

<sup>18</sup>I select the Epanechnikow kernel. I use a variable bandwidth to take into account the high density of small orders relative to large orders.

mostly concerned about execution costs, which are different from the 30 min. price impact, and even from the immediate price impact (which is the maximum price paid for a buy order, not the average price). Second, and perhaps more important, the observed concavity is obtained unconditionally. However, it is possible to have linear conditional impacts that become unconditionally concave. Indeed, suppose that when a buy order is faced by a lot of liquidity in the sell order book, it has a smaller impact. Suppose also that traders are willing to place larger orders in this condition because of the smaller impact. The result will be large orders with a relatively small impact. Conversely, when there is little liquidity, there will be more small orders and they will have a relatively larger impact. The two states bundled together will create concavity: large impact for small orders and small impact for large orders, compared to the conditional linearity. This variation with liquidity is exactly what I find in the data in Table 4.

Quintile	1 (most liquid)	2	3	4	5 (least liquid)	all quintiles
$\lambda \times 10^3$ (std. err. $\times 10^3$ )	145 (28)	191 (28)	238 (31)	305 (43)	541 (60)	255 (22)
$V \times 10^7$ (std. err. $\times 10^7$ )	133 (14)	107 (7)	100 (6)	94 (6)	83 (5)	105 (5)

Table 4: Variations in impact  $\lambda$  and average order volume  $\bar{V}$  according to liquidity quintile for the market buy orders. The five quintiles are constructed using the WAS facing buy orders:  $\ln(\text{W.A.Asks})$ -mid. A small spread (1st quintile) indicates high liquidity, and a large spread (5th quintile) small liquidity. The log-return is calculated from before the order arrives to 30 min. after it has arrived. The impact is estimated with the square root approximation:  $r_t = \lambda v_t^{0.5}$ . I report for each quintile the average order volume  $\bar{V}$  and the estimated  $\lambda$ , and their standard errors estimated by block bootstrap. I report the results, averaged across all 34 stocks, for the market buy orders.

Two different explanations for the concavity have been proposed in the literature. The first one, called stealth trading, is due to Barclay and Warner (1993). They argue that the price impact of orders increases with their private information content. They then propose<sup>19</sup> that informed traders prefer medium orders because large orders reveal their superior knowledge while small ones face high transactions cost. This explanation does not address

<sup>19</sup>This argument does not explain why small orders, considered to be uninformed, have a higher impact for a given total volume to be bought (or sold). It does not explain either why privately informed traders would not use larger orders, which have a smaller impact for a given volume to be bought (or sold).

the bundling/splitting problem, as investors could have incentives to bundle their orders in this setup. A more recent explanation is due to Gabaix, Gopikrishnan, Plerou, and Stanley (2002), who argue that large orders are placed by more patient traders, so that for a given volume they have a smaller impact than a bundle of small, impatient, orders. This could be related to the conditioning issue, as patient traders might wait for periods of higher liquidity.

Although it isn't possible to condition perfectly on liquidity, the Weighted Average Spread (WAS) is a reasonable proxy. I, therefore, divide the buy orders in five quintiles depending on the WAS that they are facing. The conditional impact I find inside each quintile isn't linear either, perhaps because the WAS is a noisy proxy for liquidity. However, the average volume and impact vary across the quintiles as I hypothesized above. Table 4 gives the average volume  $\bar{V}$  and the impact  $\lambda$  obtained for each quantile of buy market orders, using the square root approximation for the impact,  $r_t = \lambda v_i^{0.5}$ .

The pattern of decreasing impact<sup>20</sup> and increasing volume with increasing liquidity is found for the three urgencies, market, spread and book orders, and for both the buy and sell orders. It is a possible explanation for the unconditional concavity of the price impact as a function of an order's volume.

## 2.5 The Order Flow Measure.

This concavity, and a possible explanation, being established, I now take it into account to construct a measure of the order flow imbalance. Because I'm using fixed time intervals, I need to aggregate orders submitted during each time period. A first natural measure would have been to add the volume of each buy order and subtract the volume of sell orders:  $V = \sum_{i \in \text{buy orders}} (v_i)^1 - \sum_{i \in \text{sell orders}} (v_i)^1$ . Even if I used this volume measure, the fact that I have access to limit order data would ensure that there is an imbalance between submitted buy and submitted sell orders. I would thus measure the imbalance in investors' intention to trade. However, it turns out that, due to the observed concavity, the net volume is not the best aggregate order flow measure. An alternative, suggested by the work of Jones,

---

<sup>20</sup>This pattern is also found using kernel non parametric functions of the volume, instead of the sqrt approximation.

Kaul, and Lipson (1994), would have been to use only the net number of orders:  $N = \sum_{i \in \text{buy orders}} (v_i)^0 - \sum_{i \in \text{sell orders}} (v_i)^0$ . In fact, the impact of each order being well approximated by the square root function, I want to transform each order into something close to its own price impact, so as to obtain the “total price impact” when adding<sup>21</sup> up. So the aggregate measure that I use is the SQRT measure:  $SQRT = \sum_{i \in \text{buy orders}} (v_i)^{0.5} - \sum_{i \in \text{sell orders}} (v_i)^{0.5}$ . This last aggregate measure also turns out to be the one which is best correlated with price changes over fixed time intervals, as I report in section 3.2.

The three order flow variables I use are thus:

1. Market =  $\sum_{i \in \text{market buy}} (v_i)^\delta - \sum_{i \in \text{market sell}} (v_i)^\delta$
2. Spread =  $\sum_{i \in \text{spread buy}} (v_i)^\delta - \sum_{i \in \text{spread sell}} (v_i)^\delta$
3. Book =  $\sum_{i \in \text{book buy}} (v_i)^\delta - \sum_{i \in \text{book sell}} (v_i)^\delta$

where  $\delta = 0.5$ . In section 3.2, I also use  $\delta = 0$  (net number) and  $\delta = 1$  (net volume) which give qualitatively similar results but are not as good quantitatively.

## 2.6 Time Series Properties of the Order Flow.

To look at the dynamic properties of the order flow, I use the Vector Auto Regression (VAR) methodology. It turns out that orders are clustered: orders tend to be followed by orders in the same direction, and with similar characteristics (such as their urgency).

In Table 5, I observe that the order flow imbalance is “autocorrelated.” Orders placed one day and two days ago tend to be repeated today, in the same direction, and with the same urgency. It is not only an intraday phenomenon as it has sometimes been thought in the microstructure literature, since it remains significant at the horizon of two days.<sup>22</sup>

There are two possible explanations for this “autocorrelation,” which are possibly both true. The first is order splitting: institutions placing big orders will often split them into smaller orders, in the same direction and possibly of the same urgency. The second one is herd behavior: humans have a well-know psychological tendency to imitate each other

---

<sup>21</sup>The log return is additive.

<sup>22</sup>In my data, it is not significant at the three day horizon for most stocks.

	$Mkt_{t-1}$	$Sprd_{t-1}$	$Bk_{t-1}$	$Mkt_{t-2}$	$Sprd_{t-2}$	$Bk_{t-2}$	$\bar{R}^2$
$Mkt_t$ (z-stat)	<b>0.20</b> <b>(4.8)</b>	-0.10 (-0.9)	-0.02 (-0.4)	<b>0.07</b> <b>(2.2)</b>	-0.11 (-1.2)	-0.04 (-1.3)	6.8%
$Sprd_t$ (z-stat)	-0.01 (-0.6)	<b>0.21</b> <b>(5.5)</b>	0.01 (0.3)	0.01 (0.4)	<b>0.12</b> <b>(3.0)</b>	0.00 (0.0)	8.5%
$Bk_t$ (z-stat)	-0.02 (-0.7)	0.07 (1.0)	<b>0.18</b> <b>(5.0)</b>	-0.02 (-0.6)	0.06 (0.8)	0.07 (1.9)	5.8%

Table 5: The VAR of the daily order flows (SQRT), with 2 lags, averaged across all stocks. I regress the different daily order flows on past order flows. I distinguish between different urgencies, and aggregate using the SQRT function ( $SQRT = \sum_{i \in \text{buy orders}} (v_i)^{0.5} - \sum_{i \in \text{sell orders}} (v_i)^{0.5}$ ). I report the AR coefficients and the  $\bar{R}^2$  corrected for the degrees of freedom. I also report the z-stat obtained from the quantiles of block bootstrap replications. The coefficients,  $\bar{R}^2$  and z-stats are averaged across the 34 stocks.

(in crowd behavior or fashion following for example), which would also create the observed autocorrelation of orders. Since I do not have any information on who placed the order, I cannot distinguish the two here. However, similar results obtained for orders placed by individual investors in Jackson (2002) suggest that part of it is herd behavior.

Having defined rigorously the order flow imbalance, as an imbalance of submitted orders which takes into account the concavity of the impact of each order, and having mentioned the autocorrelation property of this order flow measure, I now study the relationship between this order flow measure and price changes over fixed time intervals.

### 3 High Correlation between Return and Order Flow Imbalance

#### 3.1 The Basic Return/Order Flow Regression over One Day

In Table 6, I regress the one day log-return (nights excluded) on the simultaneous order flow, distinguishing the 3 urgency levels and using the SQRT aggregation:

$$r_t = \alpha + \lambda_{market} SQRT_{,market_t} + \lambda_{spread} SQRT_{,spread_t} + \lambda_{book} SQRT_{,book_t} + \eta_t \quad (1)$$

	$\lambda_{Market} \times 10^3$	$\lambda_{Spread} \times 10^3$	$\lambda_{Book} \times 10^3$	$\bar{R}^2$
estimate	57	25	21	53.1%
z-stat	19	5	11	
	95% Confidence Interval			
Lower band	51	16	17	49.3%
Higher band	63	35	24	57.8%

Table 6: The return regressed on the simultaneous order flow (SQRT) for *Lafarge* over one day. I regress the one day log return (night excluded) on the simultaneous order flow imbalance, distinguishing between different urgencies, and aggregating using the square root function ( $SQRT = \sum_{i \in \text{buy orders}} (v_i)^{0.5} - \sum_{i \in \text{sell orders}} (v_i)^{0.5}$ ).  $r_t = \alpha + \lambda_{market}SQRT_{,market_t} + \lambda_{spread}SQRT_{,spread_t} + \lambda_{book}SQRT_{,book_t} + \eta_t$ . I report the  $\lambda$  coefficients and the  $\bar{R}^2$  corrected for the degrees of freedom. I also report the z-stat and the 95% confidence interval obtained from the quantiles of block bootstrap replications. The results are reported for the Lafarge stock.

I find a relatively high  $R^2$  of 52%, comparable to the results of Evans and Lyons (2000) on the foreign exchange. I also report the block bootstraps estimates of the 95% confidence interval and z-stat, obtained from the replication quantiles. I use block bootstrapping to take into account heteroscedasticity as well as any potential temporal<sup>23</sup> dependence. In fact, returns are nearly unpredictable except with the ten min. interval and simple bootstrapping gives the same confidence intervals for time intervals longer than ten min. Because the normalized regression coefficients are pivotal, bootstrap also provides a second order correction for the confidence interval. This can be useful since we know that high frequency returns are non normal and fat tailed. The confidence intervals obtained with White (or Newey-West) standard errors do not include this second order correction and are a little too narrow at intra-day frequency. Bootstrapping is also a simple way to get confidence intervals for the  $R^2$ , which is asymptotically normally distributed under the alternative H1:  $R^2 \neq 0$ .

In Table 7, I want to report the same results as in Table 6 for all the 34 stocks. For sake of brevity, I summarize the results and report the average and cross-section standard deviation of the estimates, as well as the average and standard deviation of the z-stat. Again, we notice the high  $R^2$  and the significance of the results.

These high  $R^2$  indicate that our measure of order flow imbalance is well correlated with

---

<sup>23</sup>The block size I use is one week.

	$\lambda_{Market} \times 10^3$	$\lambda_{Spread} \times 10^3$	$\lambda_{Book} \times 10^3$	$\bar{R}^2$
estimate: avrg.	54	49	25	47.7%
estimate: std. dev.	24	40	15	8%
	z-statistic			
z-stat: avrg.	14	6	8	
z-stat: std. dev.	5	2	3	

Table 7: The return regressed on the simultaneous order flow (SQRT) over one day: *average results for 34 stocks*. I regress the one day log return (night excluded) on the simultaneous order flow imbalance, distinguishing between different urgencies, and aggregating using the square root function ( $SQRT = \sum_{i \in \text{buy orders}} (v_i)^{0.5} - \sum_{i \in \text{sell orders}} (v_i)^{0.5}$ ).  $r_t = \alpha + \lambda_{market} SQRT_{market,t} + \lambda_{spread} SQRT_{spread,t} + \lambda_{book} SQRT_{book,t} + \eta_t$ . I report the  $\lambda$  coefficients and the  $\bar{R}^2$  corrected for the degrees of freedom. I also report the z-stat obtained from the quantiles of block bootstrap replications. The results reported are the average of the results on the 34 stocks, and the cross-section standard deviation.

price changes. In this sense, one can argue that it is a good measure of the order flow. In the next section, I look at two possible alternative measures, the net volume and the net number, to check that the SQRT is indeed a good measure. Regressing the return on these alternative measures also provides an economic interpretation of the impact coefficient.

### 3.2 The Return/Order Regression with Different Powers of the Volume.

	$\lambda_{Market} \times 10^5$	$\lambda_{Spread} \times 10^5$	$\lambda_{Book} \times 10^5$	$\bar{R}^2$
estimate	2.0	11.9	-2.6	10.5%
z-stat	3.6	9.4	-2.4	

Table 8: The return regressed on simultaneous *net number* of orders for Lafarge over one day. I regress the one day log return (night excluded) on the simultaneous order flow imbalance, distinguishing between different urgencies, and aggregating using the net number of orders ( $N = \sum_{i \in \text{buy orders}} (v_i)^0 - \sum_{i \in \text{sell orders}} (v_i)^0$ ).  $r_t = \alpha + \lambda_{market} N_{market,t} + \lambda_{spread} N_{spread,t} + \lambda_{book} N_{book,t} + \eta_t$ . I report the  $\lambda$  coefficients and the  $\bar{R}^2$  corrected for the degrees of freedom. I also report the z-stat obtained from the quantiles of block bootstrap replications. The results are reported for the Lafarge stock.

In Table 8, I report the same results as in Table 6 for Lafarge, but with the net number

of orders instead of the SQR.  $N = \sum_{i \in \text{buy orders}} (v_i)^0 - \sum_{i \in \text{sell orders}} (v_i)^0$ .

$$r_t = \alpha + \lambda_{\text{market}} N, \text{market}_t + \lambda_{\text{spread}} N, \text{spread}_t + \lambda_{\text{book}} N, \text{book}_t + \eta_t$$

I find that the  $\bar{R}^2$  is higher with the SQR. This is also true for the other stocks. The average  $\bar{R}^2$  across the 34 stocks is 47.7% for the SQR and 10.6% for the net number of orders.

The estimated  $\lambda$  also gives an economic estimate of the impact. All else equal, an imbalance of 100 orders submitted between the bid and the ask (spread orders have the largest average impact) will move the Lafarge stock price by 1.19%.

	$\lambda_{\text{Market}}$	$\lambda_{\text{Spread}}$	$\lambda_{\text{Book}}$	$\bar{R}^2$
estimate	12.8	6.7	-0.4	46.4%
z-stat	13.1	6.0	-1.7	

Table 9: The return regressed on the simultaneous *net volume* of orders for Lafarge over one day. I regress the one day log return (night excluded) on the simultaneous order flow imbalance, distinguishing between different urgencies, and aggregating using the net volume of orders ( $V = \sum_{i \in \text{buy orders}} (v_i)^1 - \sum_{i \in \text{sell orders}} (v_i)^1$ ).  $r_t = \alpha + \lambda_{\text{market}} V, \text{market}_t + \lambda_{\text{spread}} V, \text{spread}_t + \lambda_{\text{book}} V, \text{book}_t + \eta_t$ . I report the  $\lambda$  coefficients and the  $\bar{R}^2$  corrected for the degrees of freedom. I also report the z-stat obtained from the quantiles of block bootstrap replications. The results are reported for the Lafarge stock.

In Table 9, I report the same results as in Table 6 for Lafarge, but with the net volume of orders instead of the SQR.  $V = \sum_{i \in \text{buy orders}} (v_i)^1 - \sum_{i \in \text{sell orders}} (v_i)^1$ .

$$r_t = \alpha + \lambda_{\text{market}} V, \text{market}_t + \lambda_{\text{spread}} V, \text{spread}_t + \lambda_{\text{book}} V, \text{book}_t + \eta_t$$

I again find that the  $\bar{R}^2$  is higher with the SQR. This is also true for the other stocks. The average  $\bar{R}^2$  across the 34 stocks is 47.7% for the SQR and 35.8% for the net volume of orders.

The estimated  $\lambda$  also gives an economic estimate of the impact. All else equal, an imbalance in market orders of 0.1% of the shares outstanding will move the Lafarge stock price by 1.28%.

### 3.3 The Predictable Order Flow Imbalance has nearly No Impact on the Price.

We have seen that the order flow is autocorrelated, and that it is well correlated with the contemporaneous return. However, we do not expect that the return will be easily predictable. Otherwise, a simple statistical arbitrage would be available. So it should be the case that the fraction of the orders which is predictable does not have much impact on the price. This is what I verify in Table 10.

If a big fraction of the return were predictable, arbitrageurs would exploit it and remove most of the predictability. This strategy, diversifiable across time (and partly across stocks), would carry a low risk.

	$\lambda_{Mkt,pred}10^3$	$\lambda_{Mkt,res}10^3$	$\lambda_{Sprd,pred}10^3$	$\lambda_{Sprd,res}10^3$	$\lambda_{Bk,pred}10^3$	$\lambda_{Bk,res}10^3$	$\bar{R}^2$
estimate	15	<b>56</b>	-11	<b>51</b>	<b>45</b>	<b>24</b>	49.5%
(z-stat)	(0.8)	<b>(15.3)</b>	(-0.6)	<b>(6.4)</b>	<b>(2.2)</b>	<b>(8.0)</b>	

Table 10: The return regressed on *predicted (pred)* and *residual (res)* order flow (SQRT) over one day. I regress the one day log return (night excluded) on the order flow imbalance previously obtained from a VAR with 2 lags, distinguishing between the prediction obtained from the VAR (pred), and the residual from the VAR (res). I also distinguish the different urgencies, and aggregate the orders using the square root function ( $SQRT = \sum_{i \in \text{buy orders}} (v_i)^{0.5} - \sum_{i \in \text{sell orders}} (v_i)^{0.5}$ ). I report the average results across the 34 stocks.

It turns out to be nearly true. I distinguish the part of the order flow which is predicted (pred) using the VAR in Table 5, from the residual order flow (res) which is unpredicted by the VAR. The predicted part has usually an insignificant impact on the return, whereas the unpredicted order flow has a very significant impact. So the return is nearly unpredictable. However, the predicted book orders have a barely significant impact on the price. This also means, since the book orders are “autocorrelated,” that yesterday’s book orders will predict the return today. Although this might look like an opportunity for statistical arbitrage, it is more likely that the book orders needed to forecast the return were not known on the day they were submitted<sup>24</sup> so that arbitrageurs could not see and exploit this predictability in real time.

<sup>24</sup>The Paris Bourse allows hidden orders which become visible only gradually, when they are met by opposite market orders.

## 4 No Short-Term Reversal of the Price Impact.

We have seen in section 2.4 that each order has a price impact which lasts for at least 30 min. In section 3.1 the aggregated measure of the order flow is shown to be highly correlated with price changes over one day. But this impact could be only short term and be reversed within the next day or so, as is often assumed of mechanical price pressure.

	$Mkt_t$	$Sprd_t$	$Bk_t$	$\bar{R}^2$
$r_{t+1} \times 10^3$	6	-2	4	0.6%
(z-stat)	(1.3)	(-0.4)	(0.7)	

Table 11: The one day return regressed on lagged order flow. I regress the one day log return (night included) on lagged order flows. I distinguish between different urgencies, and aggregate using the SQRT function ( $SQRT = \sum_{i \in \text{buy orders}} (v_i)^{0.5} - \sum_{i \in \text{sell orders}} (v_i)^{0.5}$ ). I report the regression coefficients and the  $\bar{R}^2$  corrected for the degrees of freedom. I also report the z-stat obtained from the quantiles of block bootstrap replications. The results are averaged across the 34 stocks.

Table 11 checks if there is a reversal of the price impact during the next day. If there was, one would expect that a positive order flow imbalance today forecasts a negative return tomorrow, so as to remove part of today's impact on the price, and to find negative coefficients. This is not observed in Table 11, suggesting that the price impact is either permanent, or that it is only very slowly reversed, and that the regression of Table 11 cannot detect it.

This absence of short term reversal suggests that with time horizons longer than one day, one should also find a co-movement of the stock price with the order flow imbalance. This is what I report in the next section.

### 4.1 The Return/Order Regression with Different Time Periods.

In the previous sections I have used the daily time period as the reference. However, it is also interesting to look at different horizons. The results are similar at shorter horizons, implying that this co-movement appears in the microstructure and comes from the impact of each order, as was already suggested in section 2.4. The fact that the co-movement of orders and prices is also observed at longer horizons than one day suggests that this impact is not much reversed, at least for the next three months.

To have enough power at long horizons (only 20 data points with three month intervals), I only use one independent variable and do not distinguish between the different urgencies. I use the square root method of aggregation:  $SQRT = \sum_{i \in \text{buy orders}} (v_i)^{0.5} - \sum_{i \in \text{sell orders}} (v_i)^{0.5}$ . To have comparable results, I do the same regression:

$$r_t = \alpha + \lambda_{All} SQRT_{\text{all urgencies}_t} + \eta_t$$

with different time intervals: 10 min., 30 min., one day (night excluded), one day (night included),<sup>25</sup> one week, one month, three months. I report the estimate for  $\lambda$ , the z-stat and the  $\bar{R}^2$  of these regressions, averaged across the 34 stocks in Table 12.

	$\lambda_{All} \times 10^3$	z-stat	$\bar{R}^2$
10 min.	67	35	38.7%
30 min.	62	32	42.6%
one day (-night)	41	19	43.5%
one day (+night)	47	18	38.9%
one week	38	10.1	38.6%
one month	32	5.6	36.2%
three months	21	2.6	26.9%

Table 12: The return regressed on simultaneous order flow (SQRT) over *different time intervals*, average results for 34 stocks. I regress the log return on the simultaneous order flow imbalance, without distinguishing between different urgencies, and aggregating using the square root function ( $SQRT = \sum_{i \in \text{buy orders}} (v_i)^{0.5} - \sum_{i \in \text{sell orders}} (v_i)^{0.5}$ ).  $r_t = \alpha + \lambda_{All} SQRT_{\text{all urgencies}_t} + \eta_t$ . I report the  $\lambda$  coefficients and the  $\bar{R}^2$  corrected for the degrees of freedom. I also report the z-stat obtained from the quantiles of block bootstrap replications. The results reported are the average of the results on the 34 stocks.

As expected by the bigger sample sizes and more statistical power, the z-stats are very high for short time intervals and diminish all the way to three months. However, even at this horizon,  $\lambda$  is still statistically significantly positive for most stocks.

One also notices the diminishing  $R^2$  from one day to three months. This might suggest a partial reversal of the impact. However, when regressing future returns on past orders with various time horizon, I cannot find a statistically significant reversal of the price impact for most stocks and time intervals (there seems to be some economically important reversal after

<sup>25</sup>The night included is the previous one: returns are calculated from close to close.

six month horizon, but my short database does not yield statistically significant estimates). Another phenomenon which could better explain the decreasing  $R^2$  is that future orders are (slightly but significantly) negatively correlated with past returns, as reported in Table 13. When aggregated over long horizons, this negative lead-lag correlation can decrease the positive contemporaneous correlation. On the other hand, at very short horizons, the average  $R^2$  is also smaller, which can be explained by microstructure noise (discreteness of the tick size etc.).

The  $\lambda$  coefficient is also decreasing<sup>26</sup> from short to long horizons. This effect is stronger than for the  $R^2$  and can be explained by the “autocorrelation” of the order flow and the fact that predicted orders do not have an impact on the price as we have seen. These two effects combined generate<sup>27</sup> a decreasing<sup>28</sup>  $\lambda$ .

## 4.2 A Visual Impression of the Order Flow and Price

As a visual confirmation of the long term correlation of return and order flow, I report a graphical representation of their movements in Figure 2. The continuous line represents<sup>29</sup> the cumulative log return of Lafarge, using daily closing prices. It is thus the graph of (log) prices. The dashed line represents the cumulative order flow imbalance, that is, the sum of daily imbalances from date 0 to date  $t$ . The order flow indicator is the SQRT of orders. To take into account the different impacts of market, spread and book orders, I used the coefficients of a daily regression (nights included) when adding the three together.

The similarity of the two lines is striking. The ups and downs of the price level are also present in the cumulative order flow imbalance. This is true not only for the daily changes, but also for longer horizon of weeks and months, perhaps years.

---

<sup>26</sup>It is also smaller for ten min. than for each order separately as in Table 4, probably for the same reasons.

<sup>27</sup>It's easy to understand why with simplifying assumptions. Let's assume for now  $r_t = \lambda f_t + \eta_t$ ,  $f_{t+1} = \alpha f_t + 0 \times r_t + \epsilon_t$  with  $\alpha > 0$  and  $r_{t+1} = 0 \times f_t + 0 \times r_t + u_t$ . This gives  $r_{t+1} + r_t = \lambda^*(f_t + f_{t+1}) + v_t$  with  $\lambda^* = \lambda \frac{2+\alpha}{2+2\alpha}$ . So the impact coefficient  $\lambda$  is lower for longer horizons. Note that  $R^{2*} = \frac{(1+\alpha/2)^2}{1+\alpha} R^2$  is nearly constant, slightly bigger for longer horizons under these assumptions. Exactly the same results are obtained by refining these assumptions for the fact that predicted orders have no impact on the price.

<sup>28</sup>The same two assumptions also create the increase in  $\lambda$  from daily without night to daily with the previous night included in the return. Indeed, the orders that follow the night are probably correlated to the unobserved orders (placed on similar stocks in foreign markets) that happened during the night. So the night return is correlated with the following day orders, which increases the  $\lambda$ .

<sup>29</sup>Evans and Lyons (2000) produce a similar graph for the foreign exchange market.

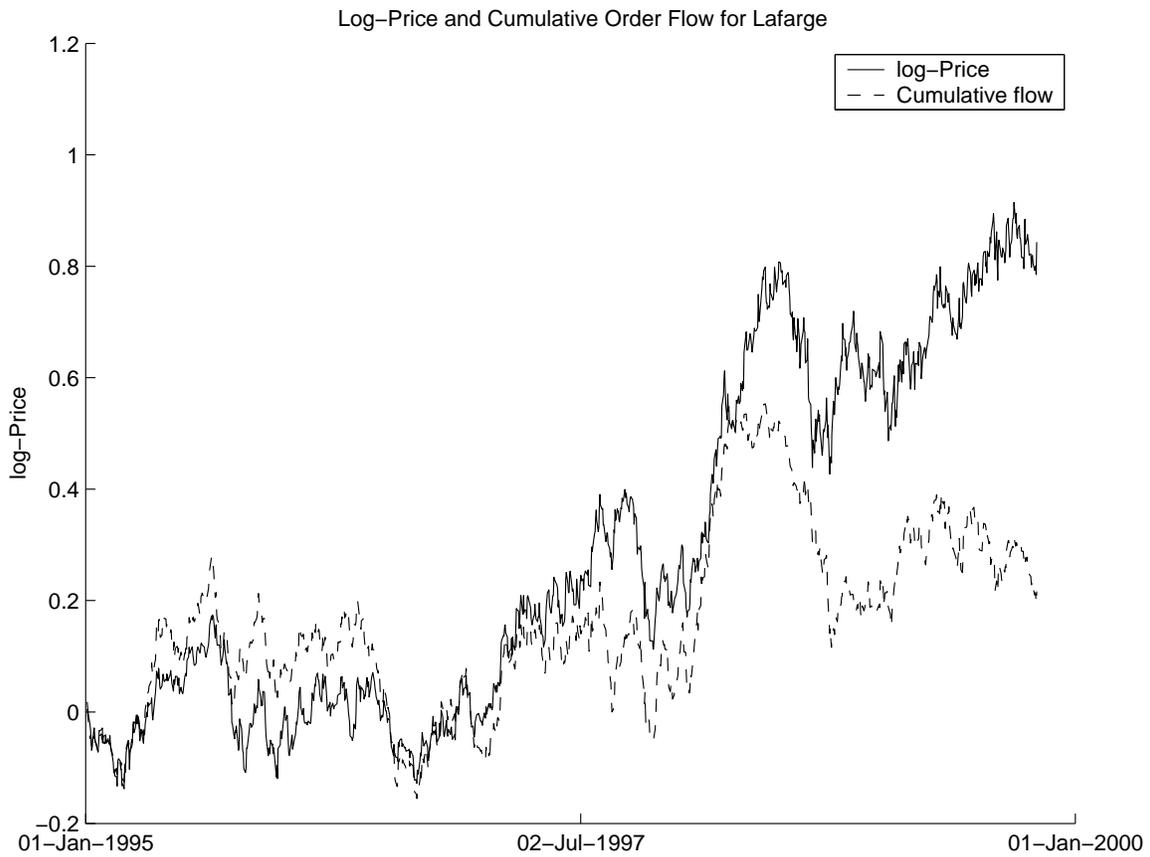


Figure 2: Cumulative return and cumulative order flow imbalance for Lafarge.

The continuous line is the cumulative return of Lafarge (using daily closing prices). The dashed line is the cumulative order flow imbalance. The order flow indicator is the Sqrt of orders. To take into account the various impacts of the three different urgencies, I used the three coefficients from daily regressions (night included) when adding together the different buy and sell orders.

## 5 Private Information or Mechanical Price Pressure?

After building a measure of the order flow that takes into account the concavity of the price impact and the possible inequality of submitted limit buy and sell orders, I have reported the strong co-movement of stock prices and order flow which appears at the microstructure horizons but remains at least until three months without being much reversed. The main question is: Why do they move together? I first look at the causality question: Do the orders cause the price to change? Then I attempt to disentangle two potential ways in which orders can cause price movements: private information and mechanical price pressure.

### 5.1 Causality

In this section I check that the causal interpretation is justified: Is it really the orders that cause price changes? Or is it the opposite: the return that causes traders to place orders in the same direction? Or is it a common factor that drives both? I first look at reverse causality and then address the common factor interpretation.

If there is reverse causality, and the price change stimulates traders to place orders in the same direction, the traders need a little time to observe the price change before they can trade on it. So by looking at high enough frequency, we should find that past returns are correlated positively with future orders. With one day horizon, the coefficients are insignificant. In Table 13, with a time interval of 30 min., the order flow is indeed correlated with past return, but with a negative coefficient: people provide liquidity and sell the stock when the price has previously moved up. This is the opposite of what reverse causality requires, and we can reject this interpretation.

Now suppose there was a common factor that prompted people to buy, as the same time as it triggered the “market makers<sup>30</sup>” to push the price upward. This factor is exactly what the literature usually labels public information: something that everyone knows at the same time, so that investors and market makers all react to it simultaneously, without some having an informational advantage over others.

But public information is studied in detail by French and Roll (1986) as I report in the

---

<sup>30</sup>There are no official market makers on the Paris Bourse but some brokers providing liquidity at the bid and ask price can play the same role.

	$r_{t-1}$	$Mkt_{t-1}$	$Sprd_{t-1}$	$Bk_{t-1}$	$\bar{R}^2$
$Mkt_t$ (z-stat)	<b>-0.66</b> <b>(-4.7)</b>	0.32 (14.4)	0.08 (3.6)	0.08 (5.4)	8.8%
$Sprd_t$ (z-stat)	<b>-0.32</b> <b>(-6.1)</b>	0.02 (3.3)	0.19 (12.6)	0.04 (4.8)	4.1%
$Bk_t$ (z-stat)	<b>-0.32</b> <b>(-3.6)</b>	-0.03 (-2.4)	0.03 (1.5)	0.31 (18.4)	8.9%

Table 13: No reverse causality, the 30 min. order flow regressed on lagged return and order flow. I regress the different order flows on past return and order flows. I distinguish between different urgencies, and aggregate using the SQRT function ( $SQRT = \sum_{i \in \text{buy orders}} (v_i)^{0.5} - \sum_{i \in \text{sell orders}} (v_i)^{0.5}$ ). I report the coefficients, the  $\bar{R}^2$  and the z-stat obtained from the quantiles of block bootstrap replications. The results are averaged across all 34 stocks.

Introduction. They show that public information explains only a small part of stock returns, less than 15% of their variance. But the order flow explains 50% of the return variance (the  $R^2$  of return on order flow). Therefore, the part of the return which is driven by the order flow cannot be entirely due to public information. For the same reason, the order flow itself cannot be entirely due to public information. In brief, public information, since it explains only little of the price changes, cannot explain both large price changes and the order flow which moves with them. So public information, the common factor that could have driven both the price and orders, does not appear to do so.

Now that I have verified that there is neither reverse causality from prices to orders, nor a common factor driving both, it is justified to think in causal terms from orders to price, and to speak of the “price impact of an order,” which I used anticipatively above. In the remainder I distinguish two sources for causal impacts, private information and mechanical price pressure.

## 5.2 Private Information

A causal impact of orders is what one would expect from a private information model, such as the model of Kyle (1985). A market maker<sup>31</sup> will adjust prices to any order, whether it is informed or not, because he cannot distinguish between them. This model also predicts

<sup>31</sup>Although there is no official market maker on the Paris Bourse, it is reasonable to assume that some rational “liquidity providers” play a similar role.

the observed direction of the impact.

The second causal interpretation which I consider is mechanical price pressure, described in detail in the next section. A third alternative type of explanation is proposed by Wang (1994), Vayanos (1999) and Evans and Lyons (2000). In these models, investors have private information on their own demand for shares, which vary due to an exogenous endowment, private investment opportunities, or risk aversion. These models are very similar to mechanical price pressure, except that they give justifications for the noise trades. Indeed, even noise traders with absolutely no information about the financial assets know before the others what order they're going to submit. So I do not distinguish these models from uninformed price pressure.

Although private information à la Kyle is certainly part of what is happening, it is possible that direct price pressure is important as well. Since mechanical price pressure is not as widely accepted as private information, I describe in the next section how it can exist in a well-arbitrated market, as well as the long run implications of price pressure, market bubbles.

### 5.3 Mechanical price pressure

Here I look in more detail at how orders could mechanically move the price, even if they do not contain private information. In the case of a market maker, Stoll (1978) has shown how inventory considerations could induce the market maker to move the price when he is faced with an order flow imbalance. However, as I have mentioned, with a market maker who regularly clears his inventory, the order flow<sup>32</sup> has to be balanced (since the market maker takes the opposite side of each trade and clears his inventory regularly). Although the imbalance probably exists from the point of view of participants, it is hard to measure their unrealized wishes, i.e. the true imbalance.

I therefore turn to the case of the limit order market, where it is possible to measure the imbalance, and where I reported it is highly correlated with price changes. For this type of market, it is clear that big market orders have a short term mechanical impact on the

---

<sup>32</sup>Defined here, with transactions instead of orders, as the volume of buyer-initiated transactions minus the volume of seller-initiated transactions.

Buy	Sell
€99 (10 shares)	€101 (10shares)
€98 (30 shares)	€102 (10 shares)
€97 (20 shares)	€103 (40 shares)

Table 14: Example, the order book before a market buy order arrives.

price, as we can see in the following example. Let's assume that the order book is given by Table 14, when a buy market order of 40 shares is submitted. It matches the book at €101, €102 and buys 20 shares at €103. The new ask price is €103. The mid-quote has gone up from €100 to €101. What a believer in information efficiency would argue is that this impact is only temporary, unless the order was informed. But this presupposes that some arbitrageurs will bring the price back to its "normal" value. The incentives for the arbitrageurs to do so may not be high enough: the price will indeed one day come back to its efficient value but the arbitrageur may need to wait a long time and face huge price changes in between. Therefore, the short term impact may take some time to disappear: Table 12 suggests that the impact has not disappeared at the three month horizon.

This price pressure framework explains naturally how market orders can move the price. As for the impact of limit orders, it is indirect: because sell limit orders provide additional liquidity on the sell side, a buy market order will have a smaller positive price impact. In my example, if someone places a limit sell order of 40 shares at €101, the market buy order will result in a transaction price of only €101 and a mid-quote of only €100. So it prevents the market order from moving the mid-quote up to €101. Therefore, sell limit orders have an indirect negative impact on the price.

I now propose a very simple model of how orders would affect the long term price with mechanical price pressure. Limit orders provide liquidity, whereas market orders demand liquidity. However, both can have a mechanical impact on the price as described above. To understand the implications of price pressure, I do not model the endogenous choice between liquidity demand and supply, that is market vs. limit orders<sup>33</sup>. Instead, I assume that all orders have the same impact on the price: buy orders push the log-price by  $+\lambda$  and sell

---

<sup>33</sup>Implicitly, I assume that there are enough sell limit orders to provide liquidity for the buy market orders to avoid market breakdowns, and the other way around for sell market orders.

orders by  $-\lambda$ . I also assume that the direction of the order is distributed randomly buy or sell, with an iid Bernoulli distribution, like flipping a coin.<sup>34</sup> If there are  $N_t$  orders between time 0 and  $t$ , the log-price change can be written:

$$p_t - p_0 = \lambda \sum_{i \leq N_t} \epsilon_i$$

where  $\epsilon_i = +1$  for a buy order and  $-1$  for a sell order.

This simple model predicts that the log price follows a random walk, thanks to the Central Limit Theorem. This result, which is often attributed to information, also ensues naturally from a price pressure model. Moreover, this price pressure model predicts that the log-price will follow a random-walk in transaction time<sup>35</sup> and not in physical time, as has been empirically documented by Ané and Geman (2000).

Finally, this random-walk result has important implications for the behavioral literature. It shows that behavioral traders can have an impact even if they are not systematically in the same direction: random orders will not perfectly cancel each other. Instead, this imperfect cancellation produces a random walk as I have described above. So one does not need a systematic crowd behavior to move stock prices. Random trades will do just as well<sup>36</sup>.

This model is very simplistic. Among other things,<sup>37</sup> it predicts that prices deviate

<sup>34</sup>Again, this is a simplification, because the order flow is autocorrelated. However, the part of the order flow that is predictable has no impact on the price, because of statistical arbitrage, as reported in Table 10. So what I model here is the unpredictable part, which is reasonably well described by the iid distribution.

<sup>35</sup>In this simple model, the distinction between market orders (which produce a transaction) and limit orders (which do not) is blurred. Empirically, however, the intensity of limit and market order submission are very correlated, so that the result of Ané and Geman (2000) would probably extend with order time instead of transaction time.

<sup>36</sup>In fact, since the fraction of the order flow which is predictable has nearly no impact on the price, a systematic and arbitrageable crowd behavior could have only a small impact when it happens, the rest being already taken into account, if profitable, by arbitrageurs.

<sup>37</sup>In this simple model, I have only considered one asset (the stock market). However, my empirical results show that the price pressure also works for each stock individually. Price pressure is harder to model in this case, because risk averse “arbitrageurs” can build portfolios with apparently very high Sharpe ratios, and therefore remove a big fraction of the mispricing even with a relatively small fraction of the global wealth. Indeed, an “arbitrageur” could invest in a long/short portfolio, which removes the market component of risk, and diversify the idiosyncratic risk. However, there are several difficulties in following this strategy. First, this long-short portfolio will be heavily loaded in the book-to-market factor of Fama and French (1993). So it is in fact risky. As a tentative explanation of where this factor comes from, it could be created by the trading of these “arbitrageurs” themselves when they get or lose money: as they invest in and out of their long/short portfolio, they move stocks by price pressure. This moves the price of the undervalued stocks together and in the opposite direction of the overvalued stocks (which they have in short position). Second, there is a lot of uncertainty in the distribution of future returns. The true fundamental value is difficult

infinitely from fundamentals.<sup>38</sup> To be more realistic, we need to assume that some rational arbitrageurs are ready to short the market when it is grossly overvalued and leverage their investment in the stock market when it is undervalued instead. This will create a dividend yield effect in the time series, as reported by Fama and French (1988), as well as the long term mean-reversion reported by Poterba and Summers (1988). When prices are high relative to fundamentals, they come back down. When they are low, they come back up. This pattern is consistent with observed stock market bubbles, such as the March 2000 Internet bubble, arguably driven by an irrational enthusiasm from uninformed investors for technology stocks.

## 5.4 Implications of Private Information for the Market Portfolio

Several results already reported in the paper suggest that price pressure might play a role in addition to that of private information. For example, mechanical price pressure would explain why book orders have an important role for price changes, although they are probably rarely used by privately informed traders. It would also explain why orders placed during periods of little liquidity have a larger impact<sup>39</sup> than when they are placed in periods of great liquidity, as reported in Table 4.

In this section I propose a more direct way to address the private information interpretation, by varying the level of private information that one expects to find in different assets or portfolios. To have very different levels of information asymmetry, I distinguish between company-specific returns and market-wide returns. Whereas there is a lot of potential for leakage at the company level (the CEO, key employees, managers, their family and friends, inquisitive analysts or fund managers etc.), it is difficult to find much potential for leakage at the market level. It therefore seems likely that only a small fraction of market movements

---

to estimate (a high price relative to book value could signal a growth company as well as an overvalued company). And the true time-varying covariance structure with many assets is also hard to estimate, which makes diversification harder. So even without the book-to-market factor, it would be difficult to build a very high Sharpe ratio portfolio without a good knowledge of the expected return and the covariance matrix.

<sup>38</sup>The cumulative imbalance between supply and demand can go to infinity over time, because it is an imbalance in submitted orders, not in realized transactions. So not even the total number of shares outstanding is a limit.

<sup>39</sup>Of course, a private information explanation would be that privately informed traders cannot delay their trades and have to trade at times of low liquidity, whereas uninformed traders can delay their trades. However, the frequency of order arrival hardly changes among the liquidity quintiles.

should be driven by private information.<sup>40</sup>

$$r_{mt} = k_m 1 + \lambda_m f_{mt} + \xi_{mt} \quad (2)$$

$$r_{it}^{idio} = k_i 1 + \lambda_i^{idio} f_{it}^{idio} + \xi_{it} \quad (3)$$

The first idea is that not all orders need to move one stock's price similarly. The notations are  $r_{mt}$  for the market portfolio's return,  $r_{it}^{idio}$  for the idiosyncratic part of stock  $i$ 's return,  $f_{mt}$  for the market order flow imbalance, and  $f_{it}^{idio}$  for the idiosyncratic order flow imbalance (I define these two flows empirically below). Equations 2 and 3 suggest that not all orders placed on stock  $i$  need to have the same impact  $\lambda_i$ .

To clarify the private information interpretation, I rely on Kyle (1985). In this model, there is a rational risk-neutral informed trader, a rational risk-neutral uninformed market maker, and some noise traders. The main result for our concern is that the market maker will move the price when he receives an order flow imbalance:

$$\Delta P = \lambda (v_{buy} - v_{sell}) \quad (4)$$

In the simplest setting of Kyle's model (single auction),

$$\lambda = 1/2 \frac{\sigma_{info}}{\sigma_{noise}}$$

where  $\sigma_{noise}$  measures the volume of noise trading and  $\sigma_{info}$  measures the information asymmetry between the informed trader and the uninformed market-maker.

Since the information asymmetry is higher on the idiosyncratic part ( $\sigma_{info}$  higher), one could imagine that orders placed on specific stocks,  $f_{it}^{idio}$ , would have a larger impact than orders placed indiscriminately on all stocks simultaneously, with  $\lambda_i^{idio} > \lambda_m$ . This inequality is not verified empirically as both types of orders have the same impacts, which are statis-

---

<sup>40</sup>One could argue that leakage is not the only type of private information. A professional trader could interpret public news better than other investors and have a temporary superior knowledge. However, the stock market is highly competitive and the professional trader would need to use his superior interpretation as quickly as possible. From the event study literature, it appears that full interpretation of public information happens within one day of the announcement. However, the work of French and Roll (1986) includes one day after the public information day. Therefore, their work excludes not only public information, but also this special kind of private information which comes from a superior interpretation of public information.

tically indistinguishable.<sup>41</sup> However, the theoretical higher impact of idiosyncratic orders depends on the importance of noise orders among idiosyncratic and market orders ( $\sigma_{noise}$ ). If there are a lot of noise traders taking bets on specific stocks, instead of rationally avoiding them (which they should do in order not to lose money to the informed traders), then  $\sigma_{noise}$  can be high for each stock and  $\lambda_i^{idio}$  low.

To avoid this ambiguity, I do not rely on the  $\lambda$  estimates but on the  $R^2$  of Equations 2 and 3. Within the strict Kyle model, there is no public information and all information arrives through orders. If the order flow were perfectly measured, the  $R^2$  when regressing the return on the order flow would be 100%. But this model is a very simplified one which we can extend to include public information. If this public information is incorporated directly into the price (without generating orders), the  $R^2$  on the order flow will not be 100%. Instead the  $R^2$  will correspond to the fraction of volatility due to private information and the rest will be due to public information.

$$R^2 = \frac{\sigma_{\text{private information}}^2}{\sigma_{\text{total information}}^2}$$

In Equations 2 and 3, company specific returns should be driven more by private information and have a large  $R^2$  whereas market-wide returns should be driven less by private information and have a small  $R^2$ .

I now build empirically the idiosyncratic and market order flows to be able to run Regressions 2 and 3. I use 30 min. intervals to have more statistical power (16,878 observations) and the SQRT aggregation.<sup>42</sup> I start by aggregating and normalizing the three types of orders for each company: I regress each stock's return on its market, spread and book order flow imbalance:

---

<sup>41</sup>Another way to find if market orders have a smaller impact than idiosyncratic orders is through the regression  $r_{it} = k_i + \lambda_i f_{it} + \lambda_m f_{mt} + \eta_{it}$ , which yields  $\lambda_i = 0.995$  (std. err. 0.04) and  $\lambda_m = 0.02$  (std. err. 0.03) suggesting that market orders have neither a bigger ( $\lambda_m > 0$ ) nor a smaller ( $\lambda_m < 0$ ) impact than other orders. However, the movements of other stocks have an impact on stock  $i$ . The regression  $r_{it} = k_i + \lambda_i f_{it} + \lambda_m f_{mt} + \beta_i r_{mt} + \eta_{it}$  yields  $\lambda_i = 0.995$  (std. err. 0.04),  $\lambda_m = -0.8$  (std. err. 0.05) and  $\beta_i = 0.8$  (std. err. 0.05) where  $f_{mt}$  and  $r_{mt}$  are equally weighted averages of the 33 other stocks. These coefficients imply that market orders do not move stock  $i$ 's price (when taking into account stock  $i$ 's orders), but that market movements unrelated to market orders do.

<sup>42</sup>Similar or even stronger results are obtained for the net number and the net volume of orders.

$$\begin{aligned}
r_{it} &= \underbrace{\alpha_i + \lambda_{i,market} \text{market}_{it} + \lambda_{i,spread} \text{spread}_{it} + \lambda_{i,book} \text{book}_{it}}_{f_{it}} + \eta_{it} && \text{I call this ag-} \\
r_{it} &= && + \eta_{it}
\end{aligned}$$

gregate  $f_{it}$  company  $i$ 's order flow imbalance. The reason for the aggregation is to simplify the rest of this section, by having only one order flow variable per stock. This regression also normalizes the  $\lambda_i$  coefficients to 1 for each stock, which allows simple comparisons between modified  $\lambda_i^{idio}$  for different stocks and between the market  $\lambda_m$  and the modified  $\lambda_i^{idio}$ .

I then define the market return as the equally weighted return for the 34 stocks. Similarly, I define the market order flow<sup>43</sup> as the equally weighted order flow:

$$r_{mt} = \frac{1}{N} \sum_{i=1}^N r_{it}$$

$$f_{mt} = \frac{1}{N} \sum_{i=1}^N f_{it}$$

I then define the idiosyncratic return for stock  $i$  as the residual of stock  $i$ 's return after regressing on the market return:

$$r_{it} = \theta_i + \beta_i r_{mt} + r_{it}^{idio} \tag{5}$$

Similarly, the idiosyncratic order flow is the residual of stock  $i$ 's order flow after regressing on the market flow:

$$f_{it} = \vartheta_i + b_i f_{mt} + f_{it}^{idio} \tag{6}$$

I then regress the return on the order flow, for the market as a whole, and for the idiosyncratic part of each stock, as described in Equations 2 and 3. Empirical results are reported in Tables 15 and 16. As described earlier, one would expect large  $R^2$  for the idiosyncratic return, where private information is important, and a smaller  $R^2$  for the market return, where it is not. The result I find empirically is exactly the opposite: for each of the 34 stocks, the idiosyncratic  $R^2$  is smaller than the market  $R^2$  and this difference is economically

---

<sup>43</sup>I also used other definitions of market return and market order flow. For instance, I extracted the principal component of the return and used the resulting eigenvector for both the return and order flow. This alternative definition gave very similar results, as the principal component from the order flow did.

	$\lambda_m$	$\bar{R}_m^2$
estimate	1.02	69.7%
(Std. Err.)	(0.02)	(0.9%)

Table 15: The market return regressed on the market order flow. I regress the 30 min. equally weighted market return on the equally weighted order flow imbalance:  $r_{mt} = k_m 1 + \lambda_m f_{mt} + \xi_{mt}$ . I report the  $\lambda_m$  coefficient and the  $\bar{R}^2$  corrected for the degrees of freedom. I also report the standard errors obtained from the quantiles of block bootstrap replications. The aggregation of orders is done using the square root function:  $\text{SQRT} = \sum_i (v_i)^{0.5}$ . The order flow of each stock is also normalized so that for each,  $\lambda_i = 1$ , before distinguishing idiosyncratic and market components.

	$\lambda_i^{idio}$	$\bar{R}_{i,idio}^2$
estimate	0.99	41.1%
(Std. Err.)	(0.04)	(1.7)%

Table 16: The idiosyncratic return regressed on the idiosyncratic order flow, averaged over 34 stocks. I regress the 30 min. idiosyncratic return (the residual after regressing on the equally weighted market return) on the idiosyncratic order flow imbalance (the residual after regressing on the equally weighted order flow imbalance):  $r_{it}^{idio} = k_i 1 + \lambda_i^{idio} f_{it}^{idio} + \xi_{it}$ . I report the  $\lambda_i^{idio}$  coefficient and the  $\bar{R}^2$  corrected for the degrees of freedom. I also report the 95% confidence interval obtained from the quantiles of block bootstrap replications. The aggregation of orders is done using the square root function:  $\text{SQRT} = \sum_i (v_i)^{0.5}$ . The order flow is also normalized so that for each stock,  $\lambda_i = 1$ , before distinguishing idiosyncratic and market components. The results are the average results over 34 stocks.

and statistically highly significant, using block-bootstrapping.<sup>44</sup>

The results of Tables 15 and 16 can at first be surprising. Indeed, the  $\lambda$  coefficient is the same for both regressions, but the  $R^2$  is higher for the market. In fact, the two results are compatible if the variance of the market order flow is large relative to the variance of the market return, which will happen if the market factor is more important for the order flow than it is for the return. This pattern is what I observe empirically when regressing Equations 5 and 6, i.e., the standard CAPM regression for the return and the equivalent for order flow. The  $R^2$  of these regressions is a measure of the importance of the market factor relative to the idiosyncratic component. In Equation 5 the average  $R^2$  for the return is 24.9%, whereas it is 34.2% for the order flow in Equation 6.

The most striking result in Table 15 is that for the market portfolio, the  $R^2$  of return on order flow is 70%, which, in absolute and economic terms, is extremely high. Economically, it seems far-fetched to argue that 70% of market-wide movements are due to private information. Evans and Lyons (2000) also find an  $R^2$  around 70% for foreign exchange data, where private information is similarly not well justified.

A useful benchmark to compare this 70% to can be found in Campbell (1991). There, he finds that only one third to one half of total market movements are due to fundamental news, whereas one half to two thirds are due to temporary, mean-reverting movements. This implies that the 70% driven by the order flow cannot all be permanent and driven by fundamental information about the asset ( $70\% > 50\%$ ). This suggests that orders are indeed generating mean-reverting price changes, more often called bubbles.

---

<sup>44</sup>The lower  $R^2$  for the idiosyncratic than the market returns suggest a lower fraction of private information movements for idiosyncratic returns. However, another possible reason for the small  $R^2$  on idiosyncratic orders could be that this regression is more misspecified. Indeed, let's assume for now that orders have widely time-varying impacts. Then a fixed  $\lambda$  will create a lower  $R^2$  than should be found with a perfect model. If, moreover, these time-varying  $\lambda$  average out for the market portfolio, and if a fixed  $\lambda$  is a better approximation for the market portfolio, then the  $R^2$  of Equation 2 will be less underestimated than for Equation 3. And the idiosyncratic  $R^2$  could be lower just due to model misspecification. For this reason, I emphasize not the low  $R^2$  of the idiosyncratic regression, but the high  $R^2$  of the market regression, which is a lower bound of what a perfect statistical model would provide.

## 6 Conclusion

In this paper, I first explain how there can be an imbalance in supply and demand for financial assets, as soon as one considers not only realized transactions, but also unrealized wishes using limit order data. Building on this observation, I construct a new measure of order flow imbalance that also takes into account the concavity of the price impact as a function of an order's volume. This order flow measure is highly correlated with contemporaneous price changes, with  $R^2$  around 50%. Besides, part of the order flow is predictable, but the predictable part has nearly no impact on the price, as one would expect from a well arbitrated market. I do not find any short term reversal of this price impact, which is observed for very different time horizons, from the micro-scale ten minutes to the macro-scale three months.

I then attempt to provide an economic interpretation of the co-movement of the order flow imbalance with price changes. I first establish the causality from orders to price changes. I refute the first alternative, reverse causality, by observing that orders follow price changes of the opposite direction instead of the same. In the second alternative, a common factor driving both orders and prices would be part of public information and is not compatible with the work of French and Roll (1986). I then stress two possible causal interpretations of the price impact, one based on private information, and the other based on mechanical price pressure. Although private information is certainly part of the reason why orders affect the price, I argue that price pressure could be present even for uninformed orders and propose a simple model for the implications of price pressure on the price, where stock prices follow a random walk in transaction time as empirically observed by Ané and Geman (2000). Uninformed price pressure would also produce bubbles driving the price away and back to its fundamental value, as was arguably observed in the March 2000 Internet bubble.

The main argument in favor of price pressure comes from the distinction between market return and idiosyncratic return. More precisely, one would expect only a small fraction of market-wide movements to be driven by private information, since there is little information asymmetry about the whole market. However, the  $R^2$  of return on orders is 70% for the market returns, significantly higher than the 41% obtained for idiosyncratic returns. Therefore, private information does not seem to be the only reason for the co-movement. Furthermore, 70% is higher than the upper bound (50%) of market movements that Campbell (1991) finds

can be attributed to fundamental news about the assets, the rest being driven by mean-reversion. This suggests that orders are indeed generating mean-reverting price changes, more often called bubbles.

This research hints at several possible directions for a better understanding of price pressure. A first one would be to get quantitative estimates of what is due to private information as opposed to uninformed price pressure. Another one would be to understand better bubbles and crashes, with behavioral explanations such as unrealistic optimism or infectious panic that could create for a long time an excess of demand or supply and move prices far away from fundamentals, as was arguably the case with the March 2000 Internet bubble.

## References

- Ané, T., and H. Geman, 2000, “Order Flow, Transaction Clock and Normality of Asset Returns,” *Journal of Finance*, 55, 2259–2284.
- Barclay, M. J., and J. B. Warner, 1993, “Stealth Trading and Volatility,” *Journal of Financial Economics*, 34, 281–305.
- Biais, B., P. Hillion, and C. Spatt, 1995, “An Empirical Analysis of the Limit Order Flow in the Paris Bourse,” *Journal of Finance*, 50, 1655–1689.
- Campbell, J. Y., 1991, “A Variance Decomposition for Stock Returns,” *Economic Journal*, 101, 157–179.
- Evans, M. D., and R. K. Lyons, 2000, “Order Flow and Exchange Rate Dynamics,” Working Paper, Berkeley.
- Fama, E. F., and K. R. French, 1988, “Dividend Yields and Expected Stock Returns,” *Journal of Financial Economics*, 22, 3–25.
- Fama, E. F., and K. R. French, 1993, “Common risk factors in the returns on stocks and bonds,” *Journal of Financial Economics*, 33, 3–56.
- French, K., and R. Roll, 1986, “Stock Return Variances: The Arrival of Information and the Reaction of Traders,” *Journal of Financial Economics*, 17, 5–26.
- Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley, 2002, “A theory of the cubic laws of stock market activity,” Working Paper, MIT.
- Hasbrouck, J., 1991, “Measuring the Information Content of Stock Trades,” *Journal of Finance*, 46, 179–207.
- Hausman, J., A. W. Lo, and A. C. MacKinlay, 1992, “An Ordered Probit Analysis of Transaction Stock Prices,” *Journal of Financial Economics*, 31, 319–379.

- Holthausen, R. W., R. W. Leftwich, and D. Mayers, 1990, "Large-block transactions, the speed of response, and temporary and permanent stock-price effects," *Journal of Financial Economics*, 26, 71–95.
- Jackson, A., 2002, "The aggregate behavior of individual investors," Working Paper, London Business School.
- Jones, C., G. Kaul, and M. Lipson, 1994, "Transactions, volume, and volatility," *Review of Financial Studies*, 7, 631–651.
- Karpoff, J., 1987, "The Relation between Price Changes and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis*, 22, 109–126.
- Kavajecz, K. A., 1999, "A specialist's quoted depth and the limit order book," *Journal of Finance*, 52, 747–771.
- Kyle, A. S., 1985, "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1335.
- Lo, A., and J. Wang, 2000, "Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory," *Review of Financial Studies*, 13, 257–300.
- Poterba, J. M., and L. H. Summers, 1988, "Mean Reversion in Stock Prices: Evidence and Implications," *Journal of Financial Economics*, 22, 27–59.
- Roll, R. W., 1988, "R-Squared," *Journal of Finance*, 43, 541–566.
- Scholes, M. S., 1972, "The market for securities: Substitution versus price pressure and the effects of information on share price," *Journal of Business*, 45, 179–211.
- Stoll, H., 1978, "The Supply of Dealer Services in Securities Markets," *Journal of Finance*, 33, 1133–1151.
- Vayanos, D., 1999, "Strategic Trading and Welfare in a Dynamic Market," *Review of Economic Studies*, 66, 219–254.

Wang, J., 1994, "A Model of Competitive Stock Trading Volume," *Journal of Political Economy*, 102, 127–168.