Order book approach to price impact

P. WEBER and B. ROSENOW*

Institut für Theoretische Physik, Universität zu Köln, D-50923 Germany

(Received 16 January 2004; in final form 23 June 2005)

Buying and selling stocks causes price changes, which are described by the price impact function. To explain the shape of this function, we study the Island ECN orderbook. In addition to transaction data, the orderbook contains information about potential supply and demand for a stock. The virtual price impact calculated from this information is four times stronger than the actual one and explains it only partially. However, we find a strong anticorrelation between price changes and order flow, which strongly reduces the virtual price impact and provides for an explanation of the empirical price impact function.

Keywords: Order book; Price impact; Resiliency; Liquidity

In a perfectly efficient market, stock prices change due to the arrival of new information about the underlying company. From a mechanistic point of view, stock prices change if there is an imbalance between buy and sell orders for a stock. These ideas can be linked by assuming that someone who trades a large number of stocks might have private information about the underlying company, and that an imbalance between supply and demand transmits this information to the market. In this sense, volume imbalance and stock price changes should be connected causally, i.e. prices go up if demand exceeds supply and go down if supply exceeds demand. The analysis of huge financial data sets (Takayasu 2002) allows a detailed study of the price impact function (Hasbrouck 1991, Hausman et al. 1992, Kempf and Korn 1999, Evans and Lyons 2002, Hopman 2002, Plerou et al. 2002, Rosenow 2002, Bouchaud et al. 2003, Gabaix et al. 2003, Lillo et al. 2003, Potters and Bouchaud 2003), which quantifies the relation between volume imbalance and price changes.

In a series of previous studies (Hasbrouck 1991, Hausman et al. 1992, Kempf and Korn 1999, Evans and Lyons 2002, Plerou et al. 2002, Rosenow 2002, Bouchaud et al. 2003, Gabaix et al. 2003, Lillo et al. 2003, Potters and Bouchaud 2003), the average price impact of an imbalance between buy and sell market orders or of individual market orders was studied. Generally, this average price impact function was found to be a concave function of volume imbalance, which increases sublinearly for above average volume imbalance. However, there is evidence that this concave shape cannot be a property of the true price impact a trader in an actual market would observe. Firstly, such a concave price impact would be an incentive to do large trades in one step instead of breaking them up into many smaller ones as is done in practice. Secondly, in the presence of bluffers in the market, the enforcement of a strictly linear price impact is the only way in which a market maker or liquidity trader can protect herself against suffering losses.

One possible way to achieve a better understanding of the average price impact is the analysis of information about potential supply and demand stored in the limit order book. Using this information, one can calculate a virtual or instantaneous price impact, which would be caused by a market order matched with limit orders from the order book. This virtual price impact can be used as a reference point for an understanding of the average price impact of market orders. An analysis of the limit order book of the Stockholm Stock Exchange (Sandas 2001) suggests that the virtual price impact calculated from the order book is significantly larger than what is expected from a regression model. Similarly, a difference between hypothetical and actual price impact was found in Coppejans et al. (2001) and considered as evidence for discretionary trading, e.g. large trades are more likely to be executed when the order book has sufficient depth.

In this paper, we calculate both the average price impact of market orders and the virtual price impact

*Corresponding author. Email: rosenow@thp.uni-koeln.de
Table 1. Summary statistics for the ten stocks studied. All values are calculated for five minute intervals. Numbers in brackets are in units of the respective standard deviation.

<table>
<thead>
<tr>
<th>Returns</th>
<th>AMAT</th>
<th>BRCD</th>
<th>BRCM</th>
<th>CSCO</th>
<th>INTC</th>
<th>KLAC</th>
<th>MSFT</th>
<th>ORCL</th>
<th>QLGC</th>
<th>SEBL</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma )</td>
<td>0.004</td>
<td>0.0053</td>
<td>0.0049</td>
<td>0.0034</td>
<td>0.0034</td>
<td>0.0036</td>
<td>0.0025</td>
<td>0.0039</td>
<td>0.0041</td>
<td>0.005</td>
</tr>
<tr>
<td>(</td>
<td>G</td>
<td>_{\text{max}}</td>
<td>0.026</td>
<td>0.057</td>
<td>0.039</td>
<td>0.037</td>
<td>0.023</td>
<td>0.026</td>
<td>0.026</td>
<td>0.021</td>
</tr>
<tr>
<td>Median of (</td>
<td>G</td>
<td>)</td>
<td>(6.46)</td>
<td>(10.77)</td>
<td>(8.1)</td>
<td>(11.05)</td>
<td>(6.86)</td>
<td>(7.17)</td>
<td>(8.3)</td>
<td>(12.74)</td>
</tr>
<tr>
<td>Volume imbalance</td>
<td>( \sigma )</td>
<td>9551.78</td>
<td>9700.37</td>
<td>6196.3</td>
<td>22652.52</td>
<td>16660.35</td>
<td>5862.25</td>
<td>11029.98</td>
<td>16010.11</td>
<td>6513.98</td>
</tr>
<tr>
<td>(</td>
<td>Q</td>
<td>_{\text{max}}</td>
<td>344657</td>
<td>193633</td>
<td>127274</td>
<td>424923</td>
<td>323016</td>
<td>116790</td>
<td>376168</td>
<td>396978</td>
</tr>
<tr>
<td>Median of (</td>
<td>Q</td>
<td>)</td>
<td>(36.08)</td>
<td>(19.96)</td>
<td>(20.54)</td>
<td>(18.76)</td>
<td>(19.39)</td>
<td>(19.92)</td>
<td>(34.1)</td>
<td>(24.8)</td>
</tr>
<tr>
<td>Market orders</td>
<td>( \langle N_{\text{market}} \rangle )</td>
<td>52.38</td>
<td>45.31</td>
<td>44.97</td>
<td>58.34</td>
<td>72.62</td>
<td>58.89</td>
<td>83.59</td>
<td>47.41</td>
<td>56.82</td>
</tr>
<tr>
<td>( \sigma_{\text{market}} )</td>
<td>33.89</td>
<td>36.4</td>
<td>36.83</td>
<td>39.3</td>
<td>49.14</td>
<td>39.92</td>
<td>53.31</td>
<td>37.43</td>
<td>37.72</td>
<td>30.92</td>
</tr>
<tr>
<td>Limit orders</td>
<td>( \langle N_{\text{added}} \rangle )</td>
<td>321.27</td>
<td>247.06</td>
<td>233.54</td>
<td>302.6</td>
<td>391.91</td>
<td>344.59</td>
<td>429.8</td>
<td>227.17</td>
<td>303.68</td>
</tr>
<tr>
<td>( \sigma_{\text{added}} )</td>
<td>183.49</td>
<td>156.22</td>
<td>160.86</td>
<td>213.87</td>
<td>232.65</td>
<td>169.97</td>
<td>244.83</td>
<td>165</td>
<td>162.02</td>
<td>156.01</td>
</tr>
<tr>
<td>( \langle N_{\text{cancelled}} \rangle )</td>
<td>286.91</td>
<td>220.13</td>
<td>206.39</td>
<td>256.38</td>
<td>343.71</td>
<td>305.64</td>
<td>376.4</td>
<td>191.54</td>
<td>268.41</td>
<td>177.09</td>
</tr>
<tr>
<td>( \sigma_{\text{cancelled}} )</td>
<td>173.15</td>
<td>144.14</td>
<td>144.22</td>
<td>199.23</td>
<td>220.71</td>
<td>154.95</td>
<td>229.11</td>
<td>152.41</td>
<td>146.79</td>
<td>141.82</td>
</tr>
</tbody>
</table>

from information contained in the Island ECN order book. By comparing the two, we find that the virtual price impact is more than four times stronger than the actual one. In order to explain this surprising discrepancy, we study time-dependent correlations between order flow and returns and find strong evidence for resiliency: the market recovers from random uninformative price shocks as the flow of limit orders is anticorrelated with returns in contrast to the positive correlations between returns and market orders. Thus, limit orders placed in response to returns strongly reduce the virtual price impact and provide a link between virtual and actual price impact.

We include the possible influence of discretionary trading in our analysis by studying the average virtual price impact, the typical virtual price impact and the virtual price impact calculated from the average order book. The average virtual price impact is strongly influenced by low liquidity periods and hence not representative for the actual impact of market orders. In order to include the influence of discretionary trading in a semiquantitative way, we base our analysis on the virtual price impact as calculated from the average order book, which is weaker than the average and the typical price impact. Including the effect of resilience as apparent from the above-mentioned anticorrelations, the actual price impact can be explained quantitatively.

We analysed limit order data from the Island ECN, NASDAQ’s largest electronic communication network, which comprises about 20% of all trades. Island ECN is a limit order driven market which allows for placement and cancellation of limit orders. If a trader is willing to sell a certain volume (number of shares) of a stock at a given or higher price, she places a limit sell order. For buying at a given or lower price, a limit buy order is placed. Limit orders are stored in the order book at their respective price with the respective volume. The sell limit order with the lowest price defines the ask price \( S_{\text{ask}} \) for that stock. Similarly, the buy limit order with the highest price defines the bid price \( S_{\text{bid}} \). If a buy limit order with a price higher than the current ask price is placed, it is executed immediately against the sell limit orders at the ask price, vice versa for sell limit orders at or below the bid price. Such marketable limit orders will be called market orders in the following. Market orders are placed by impatient traders who want to trade immediately.

The time period we study is the entire year 2002. Each day, we disregard the first five minutes due to the build up of the order book. We checked that our results are robust with respect to this choice by removing e.g. the first fifteen minutes and the last five minutes per day. We consider 250 trading days and divide each trading day into 77 intervals of five minute length.

We chose the 10 most frequently traded stocks for the year 2002, ticker symbols and descriptive statistics can be found in Table 1. The volume of market buy orders is counted as positive and the volume of market sell orders as negative, and the volume imbalance \( Q(t) \) is defined as the sum of all signed market orders placed in the time interval \([t, t + \Delta t] \) with \( \Delta t = 5 \text{ min} \). For the definition of stock price changes, we use the same five

\( \dagger \) We do not include market orders executing ‘hidden’ limit orders in the definition of \( Q(t) \) as we want to compare with the order book containing ‘visible’ orders only.
minute intervals as for the volume imbalance and define
five minute returns
\[ G(t) = \ln S_M(t + \Delta t) - \ln S_M(t), \]
where the midquote price \( S_M(t) = \frac{1}{2}(S_{bid}(t) + S_{ask}(t)) \)
is the arithmetic mean of bid and ask price. We analyse
changes of the midquote price rather than transaction
price changes to avoid a distortion of our analysis
due to the bid–ask bounce. The midquote price changes
(i) when a market order fills the current bid or ask volume
and changes the bid or ask price in this way, (ii) when the
limit orders at the current bid or ask price are cancelled or
(iii) new limit orders are placed inside the spread. In the
first part of the paper we will concentrate on the relation
between the market order flow and changes of the
midquote price. In the second part, we will see that one has
to include the influence of limit orders in order to
understand the relation between market order flow and
price changes.

To make different stocks comparable, we normalize the
return time series \( G(t) \) by their standard deviation
\( \sigma_G \) and the volume imbalance time series \( Q \)
by its first centred moment \( \sigma_Q = \langle Q - <Q> \rangle \};
as for volumes the second moment is not well defined due to a slow decay of
the probability distribution. This normalization ensures that
the price impact functions for different stocks collapse
(Plerou et al. 2002).

**Price impact of market orders.** We define the price
impact of market orders as the conditional expectation value
\[ I_{market}(Q) = \langle G_{\Delta t}(t) \rangle_Q. \]
It describes the average relation between the return \( G \) in a
given five minute interval and the market order flow \( Q \) in
the same time interval. The functional form of \( I_{market}(Q) \)
is shown in figure 1, it is in general agreement with the
results (Kempf and Korn 1999, Plerou et al. 2002,
Rosenow 2002, Lillo et al. 2003). We find that \( I_{market}(Q) \)
is a concave function of volume imbalance (Hasbrouck 1991),
which can be well fitted by a power law
\( G = 0.48Q^{0.76} \). In principle, it would be desirable to
calculate the price impact function as a dynamical and stock
by stock function. Due to our averaging procedure
however, we obtain only one data point per five minute
interval, and we need to aggregate over time and different
stocks to obtain sufficient statistics for the calculation of
the conditional average in the region of large
volume imbalances. An analysis of the time dependence
of liquidity is presented in Weber and Rosenow (2004).

We note that the power law exponent characterizing
the price impact function depends both on the time
horizon and on the market studied. Generally, the exponent
tends to increase for an increasing time horizon
(Plerou et al. 2002). On a tick by tick basis, the exponent
is very small (Lillo et al. 2003) or the functional relation
can be characterized by a logarithm (Potters and
Bouchaud 2003). On an intermediate time scale of
15 min it was found to be 0.5 (Plerou et al. 2002,
based on a simple market maker model predicts a value of
0.5 as well. Analysing the TAQ data base† for the year
1997 instead of the years 1994 and 1995 as in Plerou et al.
(2002) and Gabaix et al. (2003) and for a time interval of
five minutes as compared to the fifteen minute interval in
Plerou et al. (2002) and Gabaix et al. (2003), we find an
exponent 0.58 for transaction price changes and 0.75
for midquote price changes. For the Island ECN data,
the exponent is 0.76 for midquote returns and 0.73 for
transaction returns, both calculated on a time scale of
five minutes. Intuitively, one might expect a stronger price
impact for the analysis of transaction prices as compared to
the analysis of midquote prices. This intuition is in
agreement with the results of the empirical analysis: as
most five minute volume imbalances \( Q \) are small with
\( Q < 1 \) (in units of \( \sigma_Q \)), there are many data points available in the region \( |Q|<1 \). Hence, on a logarithmic scale
one uses as many bins in this interval as in the region
\( |Q| > 1 \), and the region \( |Q| < 1 \) contributes significantly
in the result of a logarithmic fit. However, for \( |Q| < 1 \),
one has \( |Q|^\alpha > |Q|^\beta \) for \( \alpha < \beta \), and the price impact for
transaction prices is indeed stronger than the price impact
for midquote prices.

The concave shape of the function is very surprising:
this type of price impact would theoretically be an incentive
to make large trades as they would be less costly than
many small ones. In contrast, a convex price impact
would encourage a trader to brake up a large trade into
several smaller ones, which is what actually happens.
Having this in mind, we want to understand the mechanism
responsible for this concave shape and analyse
the trading information contained in the limit order
book. There are two mechanisms conceivable for the
explanation of the concave shape: (i) we will argue that
to a large extent the concave shape can be explained by

† We analysed the 44 most frequently traded NASDAQ stocks using data from the Trades and Quotes (TAQ) data base published by the New York Stock Exchange.
resiliency, i.e. by the interplay of limit orders and market orders, and (ii) one expects discretionary trading, i.e. large trades are made when the market is very liquid and when the price impact is small. In this paper, we will concentrate on mechanism (i) and will leave the empirical study of mechanism (ii) to a future publication.

**Order book and virtual price impact.** The whole market information of the Island ECN is stored in text files where each line represents a message of one of the four major types: add limit order, cancel limit order, execute limit order or trade message. While the last type announces the execution of hidden orders which are not visible in the order book, the first three types allow for a complete reconstruction of the order book at every instant of time. The placement of a market order is indicated by the execution of one or several limit orders. For each stock, the database contains about one to four million messages for the whole year. We do not include ‘hidden’ limit orders in our analysis as there is no information about their placement in our database. We have checked that on the level of the average price impact function \( I_{\text{market}}(Q) \) the inclusion of hidden limit orders do not change our result as the change in \( \sigma_Q \) accounts for the additional order volume.

Each message contains all necessary information: the ticker symbol of the respective stock, the time past midnight in milliseconds, the number of shares, the limit price, an indicator whether the order is of buy or sell type, and a unique order reference number. We use this number as a key to store and identify each order in our data structure and perform (partial) executions and cancellations until an order is completely executed or cancelled.

Since all open orders are purged from the book every evening, we can process the data day by day starting with an empty book each morning. In our analysis, every trading day is divided into intervals of five minutes length. To make the order book information amenable to a statistical analysis, we calculate at the beginning of each time interval and for each stock \( k \) the current order book as a density function \( \rho_k(y, t) \). Due to the complexity of this calculation, we use a discrete coordinate \( y_i \) to obtain the order book from the data structure by sorting orders with respect to their limit prices and aggregating the number of shares on a lattice with spacing \( \Delta y \). For each price \( S_{\text{limit}} \) at which a limit order is placed, the coordinate \( y_i \) is defined as

\[
y_i = \begin{cases} 
    \frac{\ln(S_{\text{limit}}) - \ln(S_{\text{bid}})}{\Delta y}, & \text{limit buy order}, \\
    \frac{\ln(S_{\text{limit}}) - \ln(S_{\text{ask}})}{\Delta y}, & \text{limit sell order}. 
\end{cases}
\]

(3)

Here, the function \( \lceil x \rceil \) denotes the smallest integer larger than \( x \). We define the density function such that \( \rho_k(i\Delta y, t)\Delta y \) is the total volume in the price interval \( [i\Delta y - 1, i\Delta y] \) in the order book, where \( i \) is an integer. In our analysis, we chose \( \Delta y = 0.3\sigma_G \) as a compromise between computational speed and accuracy. We note that throughout the paper \( y \) is measured in units of \( \sigma_G \).

For the calculation of the time-dependent density functions for all ten stocks from information about placement, cancellation, and execution of limit orders contained in the Island ECN database, we had to process about 60 GB of data.

Next, we study the average order book \( \rho_{\text{book}}(y_i) = \langle \rho_k(y_i, t) \rangle \), where \( \langle \ldots \rangle \) denotes an average over both time \( t \) and different stocks \( k \). It is characterized by a flat maximum at \( y_i \approx 1 \) and a slow decay for large \( y_i \) (see figure 2(a)). Its overall shape agrees with the results of Challet and Stinchcombe (2001), Maslov and Mills (2001) and Bouchaud et al. (2002).

We want to compare the actual price impact \( I_{\text{market}}(Q) \) to a virtual price impact function calculated from the average order book. To this end, we calculate the market depth for a given return and invert this relation. We imagine a trader who wants to buy a volume \( Q \) of stocks and has only offers from the order book available. Beginning at the ask price, she executes as many limit orders as necessary to match her market order, and changes the ask price by an amount \( G \). Traded volume (or market depth) \( Q_{\text{book}}(G) \) and return \( G \) are related by

\[
Q_{\text{book}}(G) = \sum_{y_i \in G} \rho_{\text{book}}(y_i)\Delta y. \quad (4)
\]
By inverting equation (4), we define the virtual price impact \( I_{\text{book}}(Q) \) with respect to the average order book. Here, we assume that the bid–ask spread remains constant and that the midquote price changes by the same amount as the ask price. According to the above definition, the virtual price impact \( I_{\text{book}}(Q) \) describes the price change due to a single market order of arbitrary size. Now, we want to compare the virtual price impact with the actual price impact \( I_{\text{market}}(Q) \), which is calculated as a function of the volume imbalance \( Q(t) \) aggregated over a five minute interval. Using the virtual price impact to predict a return due to a time aggregated volume imbalance is an approximation, which is justified if (i) the order book is symmetric with respect to the buy and sell side and if (ii) its nonlinearities do not influence the final result essentially. As far as the average price impact is concerned, the assumption of a buy–sell symmetry of the order book is justified and one sees from figure 2 that the nonlinearity of the virtual price impact is weak.

We find that the virtual price impact \( I_{\text{book}}(Q) \) is four times stronger than the price impact of actual market orders (see figure 2 (b)), a volume imbalance of \( 5\sigma_Q \) causes a virtual price change of \( 8\sigma_Q \) but only an actual price change of \( 2\sigma_Q \). In addition, \( I_{\text{book}}(Q) \) is a convex function that can be fitted by a power law \( I_{\text{book}}(Q) = 1.22Q^{1.19} \), and not a concave function as \( I_{\text{market}}(Q) \).

Strictly speaking, the average virtual price impact cannot be calculated from the average order book. Instead, the inversion of the depth as a function of return should be performed at each instant of time and for each stock, and the average should be performed afterwards. To this end, we define a time resolved and per stock depth

\[
Q_{\text{book}}(G,t,k) = \sum_{\gamma \geq G} n_k(\gamma, t, \Delta \gamma).
\]

By inverting this relation at each instant of time and for each stock, we obtain the virtual price impact \( I_{\text{book}}(Q, t, k) \). We find that this function fluctuates strongly in time and that its average over time and different stocks \( I_{\text{book}}(Q) \) is dominated by rare events with low liquidity.

For this reason, the calculation of \( I_{\text{book}}(Q, t, k) \) is somewhat subtle. In time intervals with very low liquidity, the domain of \( I_{\text{book}}(Q, t, k) \) does not even extend up to \( 0.5\sigma_Q \) since the amount of limit orders stored in the order book is too small. In this case, the return caused by an order with signed volume \( Q > 0.5\sigma_Q \) would be undefined and the average over all time intervals would be undefined as well. In order to expand the domain of \( I_{\text{book}}(Q) \) to at least \( 3.5\sigma_Q \), we extrapolate the depth linearly by connecting the last defined data point (with largest \( Q \) and \( G \)) with the origin. Since this procedure is necessary only for few time intervals, our extrapolation method does not disturb the final result. We checked this by using different methods, e.g. by continuing the depth function by a horizontal line instead of a linear extrapolation. The influence of the choice of a specific extrapolation method is clearly visible only for large volumes \( Q > 4\sigma_Q \). The average of \( I_{\text{book}}(Q) \) is calculated on an equidistant grid on the \( Q \) axis and the values of the individual functions \( I_{\text{book}}(Q, t, K) \) at these grid points are calculated by interpolation.

In doing so, one obtains \( I_{\text{book}}(Q) \) as a convex function of signed order volume which is much steeper than the average price impact, see figure 3. To reduce the influence of low liquidity periods on the virtual price impact, we have calculated a typical price impact \( I_{\text{book}}(\text{median}) \) by replacing the average over time and different stocks by the median. For large trading volumes, \( I_{\text{book}}(\text{median}) \) is considerably smaller than \( I_{\text{book}}(Q) \), see figure 3. \( I_{\text{book}}(\text{median}) \) is a convex function of signed volume and quite similar to \( I_{\text{book}}(Q) \).

As a basis for the explanation of the actual price impact \( I_{\text{market}}(Q) \), one should not consider \( I_{\text{book}}(Q) \) due to the influence of discretionary trading: in low liquidity periods, which dominate \( I_{\text{book}}(Q) \), one expects little trading activity. On the other hand, if the liquidity in the order book is average in the sense of the median, one expects discretionary trading to play a less important role. Therefore, \( I_{\text{book}}(\text{median}) \) is a better reference point for the explanation of the actual price impact \( I_{\text{market}}(Q) \). As \( I_{\text{book}}(Q) \) is again a little less steep than the typical price impact, one can argue that due to the influence of discretionary trading it is the best starting point for the explanation of the average price impact of market orders \( I_{\text{market}}(Q) \).

The exact influence of discretionary trading certainly needs to be analysed in more detail. However, such an analysis is beyond the scope of the present paper as the data analysis presented here is already quite involved.

The average order book and thus the virtual price impact can be described by ’zero intelligence models’ (Bouchaud et al. 2002, Daniels et al. 2003), in which orders are placed randomly. Due to the fact that we have calculated our virtual price impact from an average over time and different stocks, it is comparable to such zero intelligence models. The disagreement between our virtual price impact and the actual one is evidence for the necessity to include additional mechanisms describing ‘intelligent’ or collective behaviour.
Correlations between order flow and returns. Which effect is responsible for the pronounced difference between virtual and actual price impact? In the following, we will argue that a strong anticorrelation between returns and limit orders gives rise to resiliency and reduces the virtual price impact. Thus, virtual and actual price impact can be linked to each other. In our analysis, we calculate correlation functions between returns and the flow of market or limit orders. A technical alternative to the calculation of correlation functions would be Granger causality tests using a VAR analysis (Chen et al. 2001) or a semi-nonparametric estimate of the joint density of price change and volume (Gallant et al. 1992). These techniques were used to study the relation between volatility and trading volume.

In order to understand how order flow and price changes are related, we study the correlation functions

$$c_\alpha(t) = \frac{\langle Q_\alpha(t) G(t) \rangle - \langle Q_\alpha(t) \rangle \langle G(t) \rangle}{\sigma_{Q_\alpha} \sigma_G}$$  \hspace{1cm} (6)$$

between the volume imbalance of market orders ($\alpha = \text{market}$) or limit orders ($\alpha = \text{limit}$) and returns. In order to increase the time resolution of these correlators, the order volume is measured in intervals $[t, t + \delta t]$ with width $\delta t = 50$ s, and the returns are recorded for five minute intervals as before. For $\alpha = \text{market}$, $Q_{\text{market}}(t)$ is the volume of signed market orders, and for $\alpha = \text{limit}$

$$Q_{\text{limit}}(t) = \sum_{\gamma} \text{sign}(-\gamma) \left( Q_{\text{buy}}^{\text{add}}(\gamma) - Q_{\text{sell}}^{\text{ canc}}(\gamma) \right) \Delta \gamma$$  \hspace{1cm} (7)$$
is the net volume of limit buy orders minus the net volume of limit sell orders. In equation (7), $Q_{\text{buy}}^{\text{add}}(\gamma)$ is the volume of limit orders added to the book at a depth $\gamma$, and $Q_{\text{sell}}^{\text{ canc}}(\gamma)$ is the volume of orders cancelled from the book.

The correlation functions are plotted in figure 4. We find that $c_{\text{market}}(t)$ is zero for $t < -50$ s as required for an efficient market where returns cannot be predicted over extended periods of time. For times $t \geq -50$ s, we find positive correlations which are strongest when the time intervals for orders and returns overlap. For $t \geq 300$ s (non-overlapping time intervals), we observe a slow decay of the correlation function which is probably caused by the strong autocorrelations of the market order flow (Hasbrouck 1991, Hopman 2002, Lillo and Farmer 2003).

The correlation function between limit orders and returns vanishes for negative times $t < -50$ s and has a small positive value $c_{\text{limit}}(-50)$ s = 0.04. Surprisingly, for zero and positive time differences there is a significant anticorrelation between limit orders and returns, which is strongest for $t = 250$ s (overlapping time intervals) and decays slowly to zero for large positive times. We interpret this anticorrelation as an indication that rising prices cause an increased number of sell limit orders and vice versa for falling prices. Price changes seem to be counteracted by an orchestrated flow of limit orders.

**Figure 4.** Correlation functions between return and signed order flow (buy minus sell orders). (a) Market orders and returns show strong positive equal time correlations (shaded region) decaying slowly to zero. (b) Limit orders preceding returns have weak positive correlations with them, while equal time correlations (shaded region) are strongly negative.

**Limit order flow and feedback mechanism.** The anti-correlation between returns and limit order flow suggests that dynamical effects are responsible for the difference between virtual and actual price impact. We take into account the influence of discretionary trading in a semi-quantitative way by using $I_{\text{book}}(Q)$, the least steep of the different virtual price impact functions, as the starting point for the explanation of the average price impact of market orders. The rational behind this choice is the idea that trades will be made preferentially if the liquidity is above average.

In order to connect the virtual price impact to the actual one, one has to take into account both the average limit order flow within a five minute interval and the limit order flow in response to price changes. The density of the average limit order flow is described by

$$\rho_{\text{flow}}(\gamma) = \left( Q_{\Delta t}^{\text{add}}(\gamma) - Q_{\Delta t}^{\text{ canc}}(\gamma) \right)$$  \hspace{1cm} (8)$$

with $\Delta t = 5$ min. Near the ask price, the net volume of incoming limit orders is five times larger than the volume stored in the average order book. More than one $\sigma_G$ away from the bid and ask price, the order flow decreases rapidly. Summation of the order flow density up to a given return $G$ contributes the additional volume $Q_{\text{flow}}(G) = \sum_{\gamma < G} \rho_{\text{flow}}(\gamma) \Delta \gamma$, which is displayed in figure 5(a). It grows fast for small returns and saturates for larger returns.
Order book approach to price impact

Figure 5. (a) The average flow of limit orders integrated up to an order book depth $G$ (grey squares) changes rapidly at small returns and then stays constant. The additional volume of limit orders $Q_{\text{corr}}$ in response to a return $G$ (circles) increases linearly for large $G$. (b) Empirical price impact $I_{\text{market}}(Q)$ of market orders (open circles) compared to the theoretical price impact $I_{\text{theory}}(Q)$ (full circles), which takes into account orders from the order book, the average flow of limit orders and the additional flow $Q_{\text{corr}}$.

Furthermore, there is an additional volume of incoming limit orders generated by the returns $G$ due to the anticorrelation between returns and limit orders. The density of these additional orders is described by the conditional expectation value

$$\rho_c(\gamma, G) = (Q_{\text{corr}}^\text{add}(\gamma) - Q_{\text{corr}}^\text{unc}(\gamma))G - (Q_{\text{corr}}^\text{add}(\gamma) - Q_{\text{corr}}^\text{unc}(\gamma)).$$

Here, $Q_{\text{corr}}^\text{add}(\gamma)$ is the number of limit orders added to the book at a depth $\gamma$ within the time interval $[t, t + t_0]$. We find that $\rho_c(\gamma, G)$ approximately saturates for $t_0 \geq 30\text{ min}$.

In order to quantify the effect of this additional order flow, we must specify in which way to treat the influence of the extended time correlations of order flow. When taking into account only limit orders placed in the present time interval of five minutes, one underestimates the effect of correlated limit orders for the following reason: part of the market order volume placed in the current five minute interval is predictable in the sense that it can be explained by long correlations in the order flow (Bouchaud et al. 2003, Lillo and Farmer 2003). The correlations in the limit order flow which extend over more than five minutes cancel the price impact of the predictable part of the market order flow in such a way that returns become uncorrelated. In order to capture this complicated interplay partially, we integrate over all the correlated future limit order flow. This procedure would be exact in the (admittedly artificial) situation of ‘stationary price changes’ in which we assume that $G(t) \equiv G$ is constant in time. Then, the choice $t_0 = 30\text{ min}$ makes sure that also the additional limit order volume due to returns in past time intervals is taken into account. The correlation volume corresponding to a return $G$ is

$$Q_{\text{corr}}(G) = \sum_{|\gamma|\leq G} \rho_c(\gamma, G) \Delta \gamma.$$

$Q_{\text{corr}}(G)$ is slightly negative for small $G$ and increases then almost linearly for larger $G$ (see figure 4(a)). The total volume $Q(G)$ corresponding to a return $G$ is the sum

$$Q(G) = Q_{\text{book}}(G) + Q_{\text{flow}}(G) + Q_{\text{corr}}(G)$$

of the volume $Q_{\text{book}}(G)$ of orders stored in the limit order book up to a depth $G$, the volume $Q_{\text{flow}}(G)$ arriving within a five minute interval on average, and the correlation volume $Q_{\text{corr}}(G)$. The theoretical price impact function $I_{\text{theory}}(Q)$ calculated by inverting this relation is shown in figure 5(b).

The agreement between $I_{\text{theory}}(Q)$ and $I_{\text{market}}(Q)$ is quite good, up to $G = 10\sigma_G$ there are no deviations within the error bars of $I_{\text{market}}(Q)$. Market resiliency gives rise to additional liquidity, the influx of limit orders anti-correlated with past returns has a strong influence on the price impact of market orders. It strongly reduces the virtual price impact and is responsible for the empirically observed concave shape of the price impact function. When calculating the correlation volume $Q_{\text{corr}}(G)$ over time intervals of length $t_0 = 5\text{ min}$ instead of $t_0 = 30\text{ min}$, only half the discrepancy between virtual and actual price impact could be explained. Whether or not our original choice of $t_0 = 30\text{ min}$ is the correct one in light of the arguments related to long-range correlations in the market order flow, can only be examined by a more detailed analysis beyond the scope of this paper. Our qualitative conclusion that the virtual price impact is strongly reduced by limit orders is not affected by this quantitative issue.

We note that a reduction of ‘bare’ price impact by liquidity providers was recently postulated in Bouchaud et al. (2003) in order to reconcile the strong auto-correlations of market orders with the uncorrelated random walk of returns, and that Lillo and Farmer (2003) explain the uncorrelated nature of returns by liquidity fluctuations.

In summary, we find that the virtual price impact function as calculated from the average order book is convex and increases much faster than the concave price impact function for market orders. This difference can be explained by taking into account dynamical properties of the order book, i.e. the average net order flow and the strong anticorrelation between returns and limit order flow. This anticorrelation leads to an additional influx of limit orders as a reaction to price changes, which
reduces the price impact of market orders. Including these dynamical effects, we quantitatively model the price impact of market orders.

Acknowledgements

We thank J.-P. Bouchaud, M. Potters and D. Stauffer for helpful discussions.

References