Quantitative Finance, Vol. 6, No. 1, February 2006, 7–14

Large stock price changes: volume or liquidity?

PHILIPP WEBER and BERND ROSENOW*

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

(*Received 22 January 2004; in final form 14 April 2005*)

We analyse large stock price changes of more than five standard deviations for (i) TAQ data for the year 1997 and (ii) order book data from the Island ECN for the year 2002. We argue that a large trading volume alone is not a sufficient explanation for large price changes. Instead, we find that a low density of limit orders in the order book, i.e. a small liquidity, is a necessary prerequisite for the occurrence of extreme price fluctuations. Taking into account both order flow and liquidity, large stock price fluctuations can be explained quantitatively.

Keywords: Large stock price changes; Trading volume; Liquidity

1. Introduction

The existence of fat tails in the distribution of commodity (Mandelbrot 1963) and stock price changes (Lux 1996, Campbell *et al*. 1997, Gopikrishnan *et al*. 1998), together with the relevance of these fat tails for the practical problem of risk management, has spurred a large amount of interest in the price process in financial markets. There is evidence that the cumulative distribution function of returns for both individual stocks and stock indices decays as a power law with exponent around three (Lux 1996, Gopikrishnan *et al*. 1998).

The idea of power law distributed returns is very appealing as the appearance of a power law is reminiscent of universality and critical phenomena, thus suggesting that there might be a basic and universal mechanism behind the distribution of price changes. Many phenomenological as well as microscopic models have been developed that are able to explain the main stylized facts concerning financial time series (Campbell *et al*. 1997, Takayasu 2002). However, so far there is no general consensus concerning the mechanism behind extreme price fluctuations in the tail of the distribution.

In order to understand the mechanism underlying the empirically observed return distribution in detail, one needs to study the price impact of trades. Besides the influence of breaking news, stock prices change if there is an imbalance between supply and demand. If more people want to buy than to sell, stock prices will move up, and if more people want to sell than to buy, they will move down. This relation is quantified by the price impact function (Hasbrouck 1991, Hausmann *et al*. 1992, Kempf and Korn 1999, Evans and Lyons 2002, Hopman 2002, Plerou *et al*. 2002, Rosenow 2002, Gabaix *et al*. 2003, Lillo *et al*. 2003, Potters and Bouchaud 2003, Bouchaud *et al*. 2004), which describes stock price changes as a conditional expectation value of the order flow. The order flow is the difference between the number of shares bought and the number of shares sold in a given time interval.

Recently, Gabaix *et al*. (2003) suggested a quantitative explanation for the power law distribution of returns. They describe the cumulative distribution of order flows within time intervals of a fixed length Δt by a power law with exponent $\zeta_V = 1.5$. They model the price impact function by a time-independent square root function and use it as a predictor for stock price changes. According to this argument, the exponent ζ_G of the cumulative return distribution is twice the exponent ζ_V of the order flow distribution. In this model, large returns are exclusively caused by large order flows.

This approach was criticized by Farmer and Lillo (2004) because the test for the square root relation between order flow and returns presented by Gabaix *et al*. (2003) lacks power in the presence of correlations in the order flow and because the functional form used to describe the price impact of large orders seems to vary for different stock markets. Instead, Farmer *et al*. (2004) conclude from a tick by tick analysis that large price changes are due to the granularity of the order book, which gives rise to a time varying liquidity.

The present study is a contribution to the discussion concerning the origin of extreme returns. We present an

*Corresponding author. Email: rosenow@thp.uni-koeln.de

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online
http://www.tandf.co.uk/journals
DOI: 10.1080/14697680500168008

empirical study of extreme stock price changes within time intervals of length $\Delta t = 5\,\mathrm{min}$. We analyse data for the 44 most frequently traded NASDAQ stocks contained in the Trades and Quotes (TAQ) data base for the year 1997, which is published by the New York Stock Exchange. In addition, we analyse the year 2002 of order book data from the Island ECN for the ten most frequently traded stocks†. For both data bases, we find little evidence that price changes larger than five standard deviations can be explained by an extreme order flow alone. For the order book data, we are able to reconstruct the price impact function for time intervals with large price changes and find that large returns occur only in times where the liquidity is below average. Large returns are explained quantitatively when taking into account both the order flow and the time-dependent slope of the price impact function. Our analysis suggests that an unusually small slope of the price impact function is a necessary ingredient for the explanation of extreme stock price changes.

The concept of market liquidity encompasses various transactional properties of markets (Kyle 1985). Market depth denotes the amount of order flow innovation which is required to change prices a given amount. Resiliency describes the speed with which prices recover from a random uninformative shock, and tightness is the cost for turning around a certain amount of shares within a short period of time. Within a dynamic model of insider trading and sequential auctions, Kyle (1985) finds that both depth and volatility are constant in time. Glosten (1994) discusses the equilibrium price schedule in an open limit order book and its robustness against destabilizing strategies. Intraday patterns in price discovery and transaction cost are discussed in an empirical study (Madhavan *et al.* 1997). Using a linear parameterization for the price impact of individual trades, a sharp drop of price impact after the first half trading hour and a slight increase at the end of the trading day are observed. A weak seasonality of the bid ask spread and the average quote depth at the bid and ask price was found by Chordia *et al.* (2001) when these liquidity measures were averaged over a daily window. In addition, liquidity is found to be influenced by contemporaneous market returns, market trends, and market volatility. An analysis of the limit order book of the Stockholm Stock Exchange (Sandas 2001) shows that the depth calculated from the order book is significantly smaller than what is expected from a regression model. Similarly, a difference between hypothetical and actual price impact (Coppejans *et al.* 2002) is considered as evidence for discretionary trading, i.e. large trades are more likely to be executed when the order book has sufficient depth. Returns are correlated with depth in the sense that rising prices imply a liquidity increase on the offer side of the book and *vice versa* for falling prices. Very recently, the relation between volatility and liquidity was studied for the Euronext trading platform (Beltran *et al.* 2004). While in the framework of a two-state Markov switching process the virtual price impact was found to increase in the high volatility state, a dynamical analysis based on a VAR model was not conclusive with respect to the influence of volatility on liquidity.

In a study of the average price impact for the same data set studied here, Weber and Rosenow (2003) found that the virtual price impact calculated from the order book depth is four times stronger than the actual one. The difference between the two is accounted for by resiliency. Due to the important contribution of resiliency, price changes within time intervals of a finite length $\Delta t = 5\,\mathrm{min}$ are more difficult to interpret than a study on a tick by tick basis. On the other hand, the answer to the question "how do stock prices change in a certain time interval in response to a given order flow" is relevant for the understanding of stock markets, which, after all, operate in real time. Since the distribution of order sizes follows a power law (Gopikrishnan *et al.* 2000), large order flows are not the result of many small orders adding up, but are rather due to a single large transaction. For this reason, an analysis of intervals with fixed length should be asymptotically correct when we are interested in large orders.

In order to have a theoretical framework in which the origin of price changes can be discussed, we set up a price equation,

$$S_i = S_{i-1} + c_i + \lambda_i Q_i + u_i, \tag{1}$$

in the spirit of Glosten and Harris (1988). Here, the index i labels successive transactions at times t_i, S_i is the transaction price, c_i the transitory spread component, λ_t the slope of the virtual price impact at time t_i, and u_i is a white noise which describes the fact that prices change not only due to trading, but also due to the arrival of new public information. As we will mostly be concerned with the analysis of midquote price changes in the empirical section, we let $c_i \equiv 0$ in the following. For the price change in an interval with a fixed length Δt, one finds (Foster and Viswanathan 1993)

$$S(t+\Delta t) - S(t) = \sum_{t_i \in [t,\, t+\Delta t]} \lambda_{t_i} Q_{t_i} + \sum_{t_i \in [t,\, t+\Delta t]} u_{t_i}. \tag{2}$$

We would like to mention that the price impact of an order volume Q_{t_i} according to equations (1) and (2) is permanent and hence not easy to reconcile with the fact that long-range correlations in the order flow are observed empirically (Lillo and Farmer 2004, Bouchaud *et al.* 2004).

In light of equations (1) and (2) there are three possible causes for large price changes: (i) large order flows Q_{t_i}; (ii) large price impacts (small liquidities) λ_{t_i}; and (iii) public information u_{t_i}. We will present empirical evidence that, in contrast to the theory (Gabaix *et al.*

†We analysed the following companies (ticker symbols): AMAT, BRCD, BRCM, CSCO, INTC, KLAC, MSFT, ORCL, QLGC, SEBL.

2003), the order flow alone cannot explain extreme stock price changes and that a low liquidity is a necessary condition for the occurrence of such events. To this end, we perform an event study of returns larger than five standard deviations. Specifically, we focus on the question of whether the order flow is a good predictor for these returns under the assumption of a time-independent average price impact. We obtain a negative answer to this question and argue instead that extreme returns are observed only in times where the market liquidity is below average. The magnitude of large returns can be explained quantitatively when taking into account both order flow and market liquidity.

The TAQ data base contains information about transaction data such as the number of shares traded and transaction price as well as information about quotes, i.e. the lowest sell offer (ask price $S_{\mathrm{ask}}(t)$) and the highest buy offer (bid price $S_{\mathrm{bid}}(t)$). The stock price change or return in a time interval Δt is defined as

$$G_{\Delta t}(t) = \ln S(t+\Delta t) - \ln S(t). \tag{3}$$

For the analysis of TAQ data, $S(t)$ is the price at which the last transaction before time t took place. For the analysis of order book data, $S(t)$ is chosen as the midquote price $S_{\mathrm{M}}(t) = \frac{1}{2}(S_{\mathrm{bid}}(t) + S_{\mathrm{ask}}(t))$ as we want to make comparisons with hypothetical price impacts calculated from the order book. The market order flow Q in a time interval is the sum of all signed market order volumes executed between t and $t+\Delta t$. For the TAQ data, the sign of a transaction is determined by the Lee and Ready algorithm (Lee and Ready 1991), which compares the transaction price with the midquote price. The sign is positive for buy orders (transaction price larger than the midquote price) and negative for sell orders (transaction price smaller than the midquote price). For the order book data, the data base contains information about the direction of a trade. With this data set we were able to test the Lee and Ready algorithm by first computing the results using the algorithm and then performing the same analysis with respect to the buy and sell information contained in the order book data base. On the level of single events, the transaction directions from the Lee and Ready algorithm deviate from the exact ones, but, upon averaging, both methods yield a nearly identical price impact function.

Returns G are normalized by their standard deviation σ_G, which is well defined because the cumulative distribution function of returns follows a power law with exponent around three (Lux 1996, Gopikrishnan *et al.* 1998). Since trading volume is described by a cumulative distribution with power law exponent $\zeta_V = 1.5$ (Gopikrishnan *et al.* 2000), its standard deviation is not well defined. Hence, the order flow Q is normalized by its first centred moment $\sigma_Q = \langle |Q - \langle Q\rangle|\rangle$.

2. Average price impact and large events

The relation between price changes and market order flow is described by the price impact function,

$$I_{\mathrm{market}}(Q) = \langle G_{\Delta t}(t)\rangle_Q. \tag{4}$$

It describes the average price change G caused by an order flow Q† in the same time interval. We ask whether the average price impact function $I_{\mathrm{market}}(Q)$ is able to describe extremely strong price changes $G > 5\sigma_G$. We determined all time intervals with price changes larger than five standard deviations and checked carefully that these large price changes are not due to errors in the data base but correspond to 'real' events. While the order book data seem to be free of errors, some errors are contained in the TAQ data. We filtered the raw TAQ data against recording errors and apparent price changes due to the combination of data from different ECNs (electronic communications networks). We used the algorithm of Chordia *et al.* (2001), which discards all trades for which the difference between trade price and midquote price is larger than four times the spread, which is defined as $S_{\mathrm{ask}} - S_{\mathrm{bid}}$. In addition, we checked visually the return and trading volume time series surrounding the largest price changes on a tick by tick basis and did not find evidence for data errors after applying the filtering algorithm. The data filtering removes about 1% of all transactions and has a significant effect on the cumulative distribution function $P(G > x)$. It is a common assumption that returns follow asymptotically a power law with $P(G > x) \sim x^{-\alpha}$. The tail exponent α was initially considered to be smaller than two. As a consequence, the standard deviation would not be well defined, and the time aggregation of independent returns would be described by a Levy stable distribution (Mandelbrot 1963). However, subsequent works favour α around three (Lux 1996, Gopikrishnan *et al.* 1998). As there is still no final consensus about whether a power law is the best description for the return distribution function, the present paper does not want to make a new contribution to this discussion. From a methodological point of view, we use power law fits as a simple descriptive method to discuss differences between data sets and to characterize nonlinearities. For the raw data without any filtering, we find $\alpha = 2.1$, and after applying the filter we find $\alpha = 3.9$ by fitting a straight line in a double logarithmic diagram. We note that the filtering algorithm (Chordia *et al.* 2001) is very restrictive in the sense that it discards quite a few events where the TAQ data set reports erratic and strong oscillations (of several σ_G) of the price which are probably due to the combination of data from different ECNs. While the price has already reached its new 'true' value in the leading ECN, there may still be limit orders at the old price in some smaller ECNs which are exploited by arbitrage traders. While these oscillations are 'true' price

†We do not include market orders executing 'hidden' limit orders in the definition of $Q(t)$ as we want to make a comparison with the order book containing 'visible' orders only.

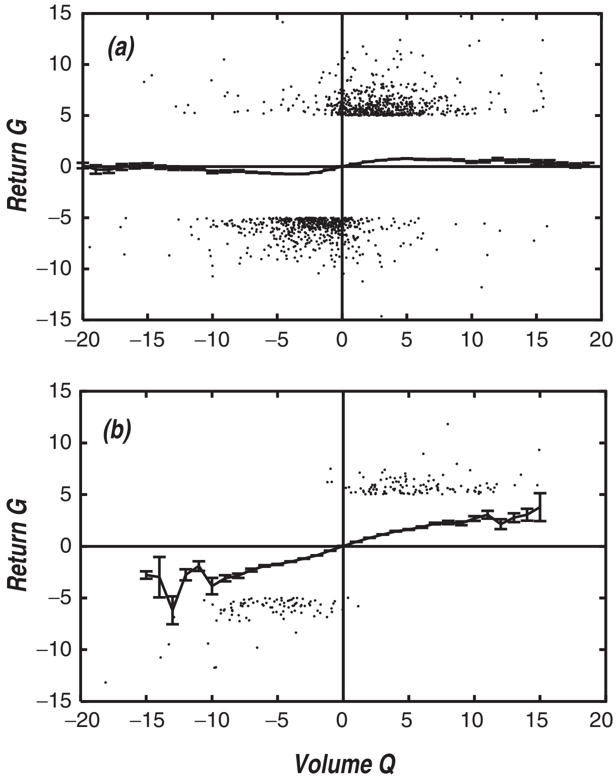

Figure 1. (*a*) Average price impact function for the 44 most frequently traded NASDAQ stocks in the year 1997 with standard deviation of the mean. Price changes larger than five standard deviations cluster in the region of small volume imbalance, and all of them are clearly outside the error bars. (*b*) Same as (*a*) but for 2002 data from the Island ECN order book for the ten most frequently traded stocks.

changes in the sense that they are not due to recording errors, they are an artifact of the trading system and were not included in our analysis.

Figure 1 shows both the price impact function and those events with price changes larger than five standard deviations σ_G. We find 1198 such events for the TAQ data base and 210 for the Island ECN data. The large events cluster at quite small values of Q where the price impact function is significantly below $G = 5\sigma_G$. For some of these events the signs of Q and G do not agree. We believe that this disagreement is (i) caused by the inaccuracy of the Lee and Ready algorithm, as such situations are less frequent for the order book data, and (ii) due to the analysis of intervals with a fixed length rather than the analysis of individual transactions.

We note that even for large order flows the average price impact function is several standard deviations (measured by the statistical error of the mean) below the line $G \equiv 5\sigma_G$ for the TAQ data and at least one standard deviation of the mean below the line $G \equiv 5\sigma_G$ for the order book data. We conclude that order flow alone cannot explain the occurrence of large returns but must be accompanied by another effect. An obvious explanation is that price impact is stronger than average during the occurrence of large returns. In the following, we will argue that this is indeed the correct explanation.

3. Time varying price impact

As the average price impact function does not provide for a satisfactory explanation of large returns, we study the time dependence of price impact. In order to achieve this goal, it is insufficient to calculate the price impact function as a conditional average over many time intervals. Instead, one needs to estimate the strength of price impact within short time intervals. This goal can only be achieved when using additional information about limit orders. In an electronic market place, market orders are matched with limit orders stored in the order book. A buy limit order indicates that a trader is willing to buy a specified number of shares at a given or lower price, while a sell limit order signals that a trader wants to sell a certain number of shares at a given or higher price. The buy limit order with the highest price determines the bid price, and the sell limit order with the lowest price the ask price. The price change due to a given market order is determined by the limit orders stored in the order book. If a trader places a buy market order with volume q, it executes as many limit orders as necessary to fill that volume. In this way, the order book determines the price change due to a single market order. We describe limit orders by their density $\rho_{\mathrm{book}}(\gamma_i, t)$ as a function of time t and the discrete coordinate

$$\gamma_i = \begin{cases} [(\ln(S_{\mathrm{limit}}) - \ln(S_{\mathrm{bid}}))/\Delta\gamma]\Delta\gamma & \text{limit buy order,} \\ [(\ln(S_{\mathrm{limit}}) - \ln(S_{\mathrm{ask}}))/\Delta\gamma]\Delta\gamma & \text{limit sell order,} \end{cases} \tag{5}$$

which describes the position of orders in the order book. Here, the function $[x]$ denotes the smallest integer larger than x. The total volume of limit orders placed in the interval $[(i-1)\Delta\gamma, i\Delta\gamma]$ is given by $\rho_{\mathrm{book}}(i\Delta\gamma, t)\Delta\gamma$ with integer i. The use of a discrete grid improves computational efficiency, reduces the amount of computer memory needed for data analysis, and speeds up calculations by more than one order of magnitude. In our analysis, we chose $\Delta\gamma = 0.3\sigma_G$ as a compromise between computational speed and accuracy.

A market buy order with volume q executes limit sell orders stored in the order book beginning at the ask price until the whole volume q is traded. The lowest remaining sell limit order forms the new ask price. Assuming a constant spread, the relation between return G and order volume q is given by

$$q(G) = \sum_{\gamma_i \le G} \rho_{\mathrm{book}}(\gamma_i, t)\Delta\gamma, \tag{6}$$

which is just the market depth for a given return G. The return G in equation (6) is denoted as the instantaneous or virtual price impact of the order q. In the framework of the model equation (1), the order book density is approximated as constant and the depth would be just $q(G) = G/\lambda_t$. From equation (6) one sees that the same order volume q can be related to quite different returns G depending on the function $\rho_{\mathrm{book}}(\gamma_i, t)$. In the following, we will argue that it is this time dependence of the order book which is important for the occurrence of large price changes. We note that, from the order book, one obtains only information about the price change as a function of buy or sell volume. Order book information can be related to time aggregated signed order volumes only under the assumption (i) that the order book is symmetric around the midquote price and that (ii) nonlinearities can be neglected. Both assumptions are generally not satisfied. For this reason, we will consider either the buy or the sell volume $\tilde{Q}$ in a given 5-min interval, depending on the direction of the return in that interval. In this way, $\tilde{Q}$ is equal to the volume of buy market orders if $G_{\Delta t} > 0$ in the 5-min interval. For $G_{\Delta t} < 0$, on the other hand, $\tilde{Q}$ is equal to the volume of sell market orders and has a negative sign. We have recalculated I_{market} as a function of $\tilde{Q}$ by averaging with respect to either the sell or the buy volume. This new $\tilde{I}_{\text{market}}$ is quite similar to the original one.

We try to find a quantitative explanation of extreme price changes by taking into account not only order flow but also market liquidity as described by market depth and market tightness in the beginning of a given time interval. The depth D is the size of the market order required to change the price by a given amount $5\sigma_G$ and is obtained from equation (6). The tightness T is the cost of a round trip (buying and selling a volume of $2\sigma_Q$ over a short period of time). To determine the tightness for a given time interval, we calculated the virtual price impact $I_{\text{book}}(\tilde{Q})$ by inverting the relation in equation (6), and define the tightness as

$$T = \frac{1}{|I_{\text{book}}(2\sigma_Q)| + |I_{\text{book}}(-2\sigma_Q)|}. \tag{7}$$

In the framework of the model equation (1), the order book density is approximated as constant and the tightness would be just $T = 1/4\sigma_Q\lambda_t$.

We compare the ratio of the actual price change $G_{\Delta t}(t)$ and the predicted price change,

$$G_{\text{pred}}(t) = \tilde{I}_{\text{market}}(\tilde{Q}(t)), \tag{8}$$

with the inverse liquidity as described by the inverse depth and the inverse tightness. We analyse 5-min returns larger than five standard deviations for the ten most liquid stocks† contained in the 2002 Island ECN data set. The liquidity measures depth and tightness are normalized by the average depth $\overline{D}$ and the average tightness $\overline{T}$ calculated from the average order book. In order to calculate G_{pred}, we computed $\tilde{I}_{\text{market}}(\tilde{Q}(t))$ up to $|\tilde{Q}| = 18\sigma_Q$ as the statistics is insufficient for $|\tilde{Q}| > 18\sigma_Q$. Therefore, we had to discard 11 events with $|\tilde{Q}| > 18\sigma_Q$ from this analysis. In addition, for eight events the order book did not contain enough limit orders to trade a volume of $2\sigma_Q$, so we were not able to compute the tightness T. In the analysis of the inverse tightness as liquidity measure, these events are excluded. For reasons of consistency, we also removed two events with $\overline{D}/D > 30$ in

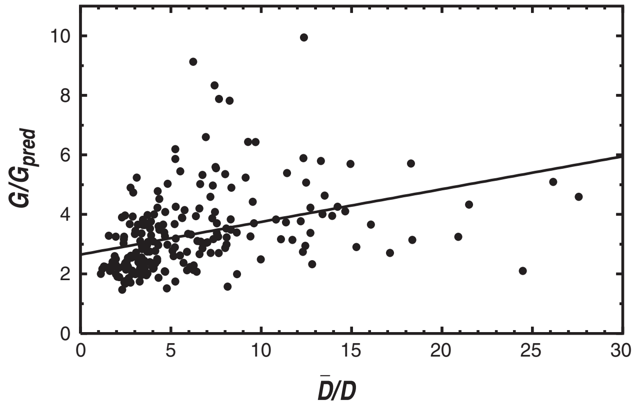

Figure 2. Ratio of actual price change to predicted price change plotted against the inverse market depth for large 5-min returns contained in the 2002 Island data. A linear regression (full line) has a correlation coefficient $R^2 = 0.14$.

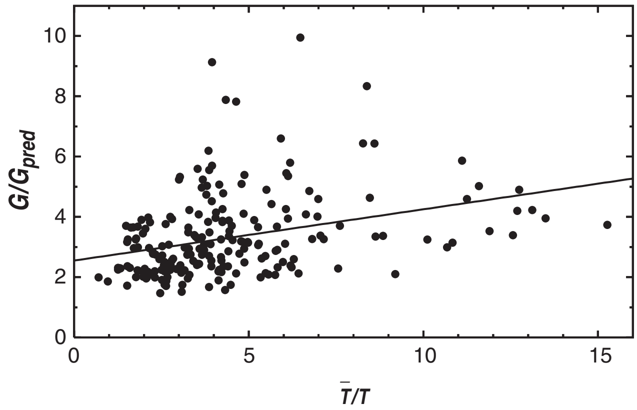

Figure 3. Ratio of actual price change to predicted price change plotted against the inverse market tightness for large 5-min returns contained in the 2002 Island data. A linear regression (full line) has a correlation coefficient $R^2 = 0.11$.

figure 2. A scatter plot for events with $|G_{\Delta t}| > 5\sigma_G$ is shown in figures 2 and 3. Contrary to the expectation that the ratio of actual and predicted price change is explained by a small liquidity, there is only a moderate correlation between liquidity and returns for both depth and tightness. This visual impression is confirmed by correlation coefficients $R^2 = 0.14$ and $R^2 = 0.11$ for depth and tightness, respectively.

When studying the price impact of the order flow in a given time interval, it is not sufficient to invoke the order book density $\rho_{\text{book}}(\gamma_i, t)$ at one instant of time. In addition, one has to consider changes in the order book which occur within a given time interval. Weber and Rosenow (2003) showed that the virtual price impact of a given order volume is roughly four times stronger than the actual one. This difference is due to additional limit orders placed in reaction to a price change. Hence, the inclusion of dynamical effects is crucial for calculating the correct price impact.

†We analysed the following companies (ticker symbols): AMAT, BRCD, BRCM, CSCO, INTC, KLAC, MSFT, ORCL, QLGC, SEBL.

In order to calculate the density of limit orders arriving in a given time interval, we fix a reference frame by the bid and ask price in the beginning of the interval. Sell limit orders arriving at a price lower than this ask price are counted as if they were arriving at the ask price, and *vice versa* for buy limit orders. While $\rho_{\text{book}}(\gamma_i, t)$ describes the density of limit order volume at a depth γ_i in the beginning of the time interval $[t, t+\Delta t]$, we define another density function $\rho_{\text{flow}}(\gamma_i, t, \Delta t)$ describing the density of limit order volume placed at a depth γ_i minus the limit order volume removed during this time interval with

$$\rho_{\text{flow}}(\gamma_i) = \langle Q^{\text{add}}_{\Delta t}(\gamma_i) - Q^{\text{canc}}_{\Delta t}(\gamma_i)\rangle, \tag{9}$$

where $Q^{\text{add}}_{\Delta t}(\gamma_i)$ is the volume of limit orders added to the book at a depth γ_i, and $Q^{\text{canc}}_{\Delta t}(\gamma_i)$ is the volume of orders canceled from the book. Thus, $\rho_{\text{flow}}(\gamma_i, t)\Delta\gamma$ is the net limit order volume arriving in the time interval $[t, t+\Delta t]$ and in the price interval $[(i-1)\Delta\gamma, i\Delta\gamma]$. The total limit order density available for transactions is then given by

$$\rho(\gamma_i, t) = \rho_{\text{book}}(\gamma_i, t) + \rho_{\text{flow}}(\gamma_i, t, \Delta t). \tag{10}$$

The relation between $\rho(\gamma_i, t)$ and the order flow $\tilde{Q}$ is

$$\tilde{Q}(G) = \sum_{\gamma_i \le G} \rho(\gamma_i, t)\Delta\gamma. \tag{11}$$

By inverting this relation we calculate a price impact function $I_{\text{actual}}(\tilde{Q})$. The sell order side of this function for ten events with price changes larger than $5\sigma_G$ is shown in figure 4. In figure 5, the average over all such events is compared with the average price impact function $\tilde{I}_{\text{market}}(\tilde{Q})$. One sees that the slope of $I_{\text{actual}}(\tilde{Q})$ is much larger than the slope of $\tilde{I}_{\text{market}}(\tilde{Q})$. As a consequence, in time intervals with large price changes there are less limit orders available than on average. Hence, we suggest the use of the slope of the actual price impact function as a measure of market liquidity.

The curves displayed in figure 4 look quite linear, and also the average of the I_{actual} for all large events (see figure 5) is approximately linear†. Accordingly, we expect a linear fit to the actual price impact functions to be a good description for the strength of price impact. For each time interval with $|G_{\Delta t}| > 5\sigma_G$, we define a susceptibility $\chi(t)$ by a linear fit through the origin to the actual price impact function $I_{\text{actual}}(\tilde{Q})$ up to a return $G = 5\sigma_G$ or $G = -5\sigma_G$, depending on the sign of $G_{\Delta t}$. Although in the simple model equation (1) the order book density is approximated as constant and dynamical effects are not included, one can formally identify $\chi(t) = \lambda_t$.

Liquidity is measured by the inverse $1/\chi(t)$. In this way a large slope of the price impact function corresponds to a low liquidity.

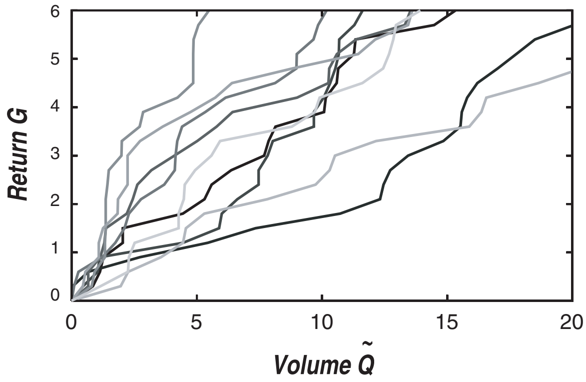

Figure 4. Price change as a function of buy or sell volume for ten of the largest price changes in the Island ECN data.

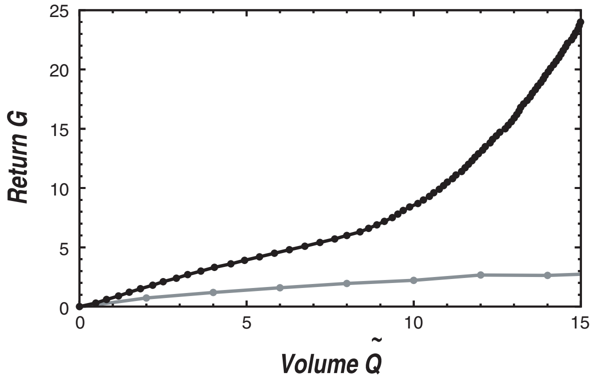

Figure 5. Price change as a function of buy or sell volume averaged over all time intervals with returns larger than $5\sigma_G$ (connected black circles). The price change averaged over all transactions (connected grey circles) is much smaller than that for the extreme events.

In figure 6 the ratio of G_{pred} and $G_{\Delta t}$ is plotted against the susceptibility $\chi/\overline{\chi}$ for all events with $|G_{\Delta t}| > 5\sigma_G$. We have normalized χ by $\overline{\chi}$, the slope of a linear fit to the average price impact function $\tilde{I}_{\text{market}}$ up to $|G_{\Delta t}| = 5\sigma_G$. To make this analysis consistent with the analysis of tightness and depth we removed one event with extremely small liquidity ($\chi/\overline{\chi} > 60$). The data points in figure 6 cluster in the vicinity of a linear fit with $R^2 = 0.79$. In comparison with the two liquidity measures used above, this result is a considerable improvement. We believe that the improved description of large returns with the help of χ is due to the fact that the susceptibility χ takes into account the dynamics of the order book, which is important for describing liquidity. From this analysis, we conclude that the time-dependent slope of the price impact function has a large explanatory power for the occurrence of extreme price changes.

As additional evidence for the idea that the return in a given time interval is caused by a combination of the order flow and the time varying liquidity, we discuss

†We tested $I_{\text{actual}}(\tilde{Q})$ for each time interval with price change larger than $5\sigma_G$ for nonlinearities. As a simple descriptive method we fitted these curves with power laws. The exponents we found vary between 0.15 and 2.35 with a mean of 1.32 and they scatter with a standard deviation of 0.41. On the other hand, a power law fit to the average of I_{actual} for all such events yields an exponent of 1.03, which is approximately linear.

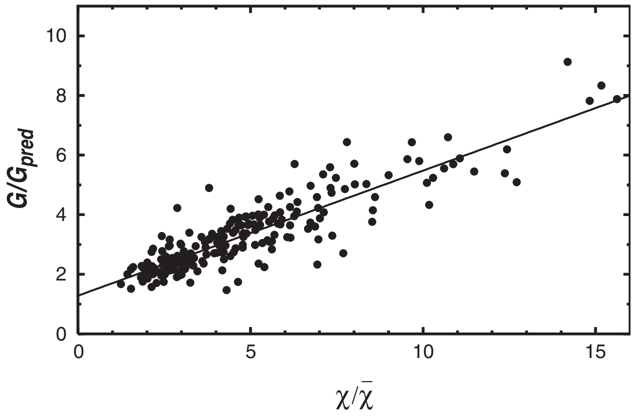

Figure 6. Ratio of actual price change to predicted price change plotted against the slope of the actual price impact function normalized by the slope of the average price impact function. The data points cluster in the vicinity of a linear fit with $R^2 = 0.79$.

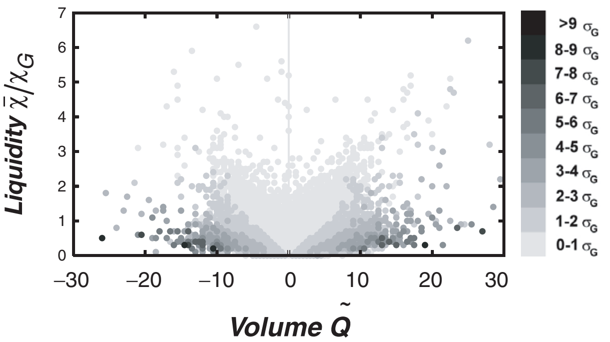

Figure 7. Expected return $\langle G_{\Delta t}(t)\rangle_{\tilde{Q},\overline{\chi}/\chi}$ as a function of order flow $\tilde{Q}$ and liquidity $\overline{\chi}/\chi$. For every combination of $\tilde{Q}$ and $\overline{\chi}/\chi$ we plotted (i) the average return if there is more than one matching time interval, (ii) the return if there is only one event or (iii) nothing if the combination never occurred. The magnitude of the return is coded from bright grey for small returns to black for the largest ones.

returns as a function of both order flow $\tilde{Q}$ in the direction of the price change and the susceptibility $\chi/\overline{\chi}$. For this analysis, the susceptibility χ needs to be redefined because the order book density at a depth $\gamma_i > G_{\Delta t}$ does not affect the price dynamics. As this effect would weaken the explanatory power of χ for time intervals with small returns $G_{\Delta t} < 5\sigma_G$, we define a new susceptibility χ_G by a linear fit through the origin to the actual price impact function $I_{\text{actual}}(\tilde{Q})$ up to the actual return $G_{\Delta t}$. In time intervals with $|G_{\Delta t}| < 1\sigma_G$, the linear fit extends up to $\sigma_G \text{sgn}(G_{\Delta t})$ in order to include enough data points for a reliable fit.

In figure 7, the average return is plotted as a function of both market order flow and liquidity measured by $\overline{\chi}/\chi_G$. The magnitude of returns is coded in a grey scale from bright grey for small returns to black for the largest ones. One observes quite sharp borders between regimes of different expected returns, again demonstrating that, for a given order flow, the magnitude of the return depends on liquidity. In addition, one sees that large returns occur only for liquidities below average, while small returns can be found even for very large volumes.

4. Discussion

In the last section, we argued that a lack of liquidity is a necessary condition for the occurrence of large returns and that a combination of low liquidity and order flow allows for a quantitative explanation of returns. So far, we have not yet discussed the possible influence of public information on large stock price changes. In an earlier empirical study using a structural model of intraday price formation (Madhavan *et al.* 1997), public information was found to account for 35 to 46% of the volatility of transaction price movements. As public information seems to influence price volatility, it is conceivable that public information also influences unusually large price changes. In the standard pricing model equations (1) and (2), the white noise term u_t describes price changes due to the arrival of new public information. In a limit order market, this type of price change takes place via cancelation and placement of limit orders. To be specific, if there is good news about a given company in the beginning of a 5-min interval, e.g. a positive earnings report, prices might go up due to the cancelation of sell limit orders in the vicinity of the current ask price and the placement of new sell limit orders at a higher price, and *vice versa* for buy limit orders. In such a situation, we would observe a *negative* ρ_{flow} in the vicinity of the ask price and a positive ρ_{flow} some distance away from it. Thus, the total density of limit orders $\rho_{\text{book}} + \rho_{\text{flow}}$ is reduced in the vicinity of the ask price and the slope χ of a linear fit to the order book is reduced as well. In other words, our liquidity measure $1/\chi(t)$ reflects price movements due to public information. In this sense, our liquidity measure describes the combined influence of order book depth, resiliency, and public information.

In our analysis, we have explained large price movements by a reduction in liquidity, i.e. we have argued that there is a causal relation between low liquidity and large price movements. In a recent study of the turbulent October 1987 period (Goldstein and Kavajecz 2004) the limit order book spread was shown to have increased significantly on October 28, the day after the market drop on October 27. Although the size of the spread does not affect midquote price changes, this result supports the possibility that causality may go from large price movements to a reduced liquidity as well. While we cannot make general statements about the possibility that large price movements may lead to a liquidity reduction, it is clear from our analysis that *during the 5-min interval with the large price drop* a reduced liquidity is causally responsible for the large price movement in the sense that an average order volume causes an above average price change. We did not study the reason for the below average liquidity in such a time interval, i.e. we do not address the possibility that a price movement at an earlier time may have caused the reduction of liquidity.

In summary, we have studied two alternative approaches to explain large stock price changes: large fluctuations in trading volume and time varying liquidity.

We find little evidence that extreme stock price changes are caused exclusively by a large trading volume. Using order book data, we have reconstructed the price impact of trading volume for all time intervals with returns larger than five standard deviations. We find that the price impact in these time intervals is much stronger than the average one and that such an anomalously large price impact is a necessary prerequisite for the occurrence of extreme price changes. The combined effect of trading volume and time varying liquidity can account for extreme price changes quantitatively.

References

Beltran, H., Durré, A. and Giot, P., How does liquidity react to stress periods on a limit order market? NBB Working Paper No. 49, 2004.

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

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C., *The Econometrics of Financial Markets*, 1997 (Princeton University Press: Princeton, NJ).

Chordia, T., Roll, R. and Subrahmanyam, A., Market liquidity and trading activity. *Journal of Finance*, 2001, **56**, 501–530.

Coppejans, M.T., Domowitz, I.H. and Madhavan, A., Liquidity in an automated auction (December 21, 2001). AFA 2002 Atlanta Meetings.

Evans, M.D. and Lyons, R.K., Order flow and exchange rate dynamics. *Journal of Political Economy*, 2002, **110**, 170–180.

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

Farmer, J.D. and Lillo, F., On the origin of power law tails in price fluctuations. *Quantitative Finance*, 2004, **4**, C7–C11.

Foster, F. and Viswanathan, S., Variations in trading volume, return volatility, and trading costs: evidence on recent price formation models. *Journal of Finance*, 1993, **48**, 187–211.

Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H.E., A theory of power-law distributions in financial market fluctuations. *Nature*, 2003, **423**, 267–270.

Glosten, L.R., Is the electronic open limit order book inevitable? *Journal of Finance*, 1994, **49**, 1127–1161.

Glosten, L. and Harris, L., Estimating the components of the bid–ask spread. *Journal of Financial Economics*, 1988, **21**, 123–142.

Goldstein, M. and Kavajecz, K., Trading strategies during circuit breakers and extreme market movements. *Journal of Financial Markets*, 2004, **7**, 301–333.

Gopikrishnan, P., Meyer, M., Amaral, L.A.N. and Stanley, H.E., Inverse cubic law for the probability distribution of stock price variations. *European Physics Journal* B, 1998, **3**, 139.

Gopikrishnan, P., Plerou, V., Gabaix, X. and Stanley, H.E., Statistical properties of share volume traded in financial markets. *Physical Review* E, 2000, **62**, 4493–4496.

Hasbrouck, J., Measuring the information content of stock trades. *Journal of Finance*, 1991, **46**, 179–207.

Hausmann, J., Lo, A. and MacKinlay, C., An ordered probit analysis of transaction stock prices. *Journal of Finance and Economics*, 1992, **31**, 319–379.

Hopman, C., Are supply and demand driving stock prices? MIT Working Paper, 2002.

Kempf, A. and Korn, O., Market depth and order size. *Journal of Financial Markets*, 1999, **2**, 29–48.

Kyle, A.S., Continuous auctions and insider trading. *Econometrica*, 1985, **53**, 1315–1335.

Lee, C.M. and Ready, M.J., Inferring trade direction from intraday data. *Journal of Finance*, 1991, **46**, 733–746.

Lillo, F. and Farmer, J.D., The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 2004, **8**(3), Article 1.

Lillo, F., Farmer, J.D. and Mantegna, R.N., Master curve for price-impact function. *Nature*, 2003, **421**, 129.

Loretan, M. and Phillips, P.C.B., Testing the covariance stationarity of heavy-tailed time series: an overview of the theory with applications to several financial datasets. *Journal of Empirical Finance*, 1994, **1**, 211.

Lux, T., The stable Paretian hypothesis and the frequency of large returns: an examination of major German stocks. *Applied Financial Economics*, 1996, **6**, 463–475.

Madhavan, A., Richardson, M. and Roomans, M., Why do security prices change? A transaction-level analysis of NYSE stocks. *Review of Financial Studies*, 1997, **10**, 1035–1064.

Mandelbrot, B.B., The variation of certain speculative prices. *Journal of Business*, 1963, **36**, 394–419.

Plerou, V., Gopikrishnan, P., Gabaix, X. and Stanley, H.E., Quantifying stock price response to demand fluctuations. *Physical Review* E, 2002, **66**, 027104[1]–027104[4].

Potters, M. and Bouchaud, J.-P., More statistical properties of order books and price impact. *Physica* A, 2003, **324**, 133–140.

Rosenow, B., Fluctuations and market friction in financial trading. *International Journal of Modern Physics* C, 2002, **13**, 419–425.

Sandas, P., Adverse selection and competitiv market making: empirical evidence from a limit order market. *Review of Financial Studies*, 2001, **14**, 705–734.

Takayasu, H., editor, *Empirical Analysis of Financial Fluctuations*, 2002 (Springer: Berlin).

Weber, P. and Rosenow, B., Order book approach to price impact. Eprint cond-mat/0311457, 2003.