Stock price jumps: news and volume play a minor role

Armand Joulin[1], Augustin Lefevre[1]
Daniel Grunberg[1], Jean-Philippe Bouchaud[1]

December 2, 2008

[1] Science & Finance, Capital Fund Management, 6 Bd Haussmann, 75009 Paris, France.

Abstract

In order to understand the origin of stock price jumps, we cross-correlate high-frequency time series of stock returns with different news feeds. We find that neither idiosyncratic news nor market wide news can explain the frequency and amplitude of price jumps. We find that the volatility patterns around jumps and around news are quite different: jumps are followed by increased volatility, whereas news tend on average to be followed by lower volatility levels. The shape of the volatility relaxation is also markedly different in the two cases. Finally, we provide direct evidence that large transaction volumes are *not* responsible for large price jumps. We conjecture that most price jumps are induced by order flow fluctuations close to the point of vanishing liquidity.

Why do stock prices change? The traditional answer, within the theory of efficient markets, is that prices move because some new piece of information becomes available, leading to a revision of the expectations of market participants. If this picture was correct, and in the absence of "noise traders", the price should essentially be constant between two news items, and move suddenly around the release time of the news. Noise trading should add high frequency mean-reverting noise between news, that should not contribute to the long term volatility of the price. News release should be the main determinant of price volatility. There are, however, various pieces of evidence suggesting that this picture is incorrect. Volatility

is much too high to be explained only by changes in fundamentals [1]. ~~The volatility process itself is random~~, with highly non-trivial clustering and long-memory properties (for reviews, see e.g. [2, 3, 4]). Many of these properties look very similar to endogenous noise generated by complex, non-linear systems with feedback, such as turbulent flows [5, 6, 7, 8, 9], stick balancing dynamics [10], etc. [11]. On liquid stocks, there is in fact little sign of high frequency mean reversion that one could attribute to noise traders [12]. It rather appears that most of the volatility arises from trading itself, through the very impact of trades on prices. This was the conclusion reached by Cutler, Poterba and Summers in an early seminal paper [13] (see also [14]). More recently, the authors of [15] used high frequency data to decompose the volatility into an impact component and a news component, and found the former to be dominant (see also [12, 16]). Here, we want to confirm this conclusion directly, using different news feeds synchronized with price time series. Our main result is indeed that most large jumps (defined more precisely below) are not related to any broadcasted news, even if we extend the notion of 'news' to a (possibly endogenous) collective market or sector jump. We find that the volatility pattern around jumps and around news is quite different, confirming that these are distinct market phenomena [17]. We also provide direct evidence that large transaction volumes are *not* responsible for large price jumps, as also shown in [30]. We conjecture that most price jumps are in fact due to endogenous liquidity micro-crises [19], induced by order flow fluctuations in a situation close to vanishing outstanding liquidity.

We have studied the news feed from Dow Jones, that concerns 893 stocks from NASDAQ and NYSE, from Aug-02-2004 to Aug-31-2006, between 9:30 am to 4:00 pm. We identified a total of 93,698 relevant news items[1], or 105 news par stock on average, or one per stock every five days. The median stock has 62 news (see third plot of figure 1 for a distribution). We also had the Reuters news feed at our disposal: 32,000 news for the same 2-year period and using the same filter on the 893 stocks. Of these, only 8,000 coincide with Dow Jones news, i.e. there is a DJ news in a 15 min neighborhood provided both RT and DJ news are preceded by a news silence of 30 min. From Fig. 1 d) we infer that Dow Jones is slightly faster on US stocks; and including the Reuters news which do not seem to have a Dow Jones counterpart increases the total number of news by 26%. This does not change any of the qualitative conclusions below. We crossed this news feed with one-minute return data files. Some of these files are however incomplete, so we checked all our results on a restricted subset of 163 stocks in the period Nov-28-

---

[1]To separate the grain from the chaff, we took only the first news from an avalanche of news (same DJ_id), removed generic or automated news (imbalances at Close, etc), and required the name of the company (among the many tickers listed) to be present in the headline.

2005 to Jun-30-2006, for which we have a complete set of return time series. The number of news concerning these 163 stocks is 6,443.

We first compare the occurrence of price "jumps" with the occurrence of news on a given company. This requires to define jumps in a consistent, if slightly arbitrary fashion. We choose to compare the absolute size $|r(t)|$ of a one minute bin return to a short term (120 minutes) flat moving average $m(t)$ of the same quantity, in order to factor in slow modulations of the average volatility. An s-jump is such that $|r(t)| > s\,m(t)$. The number of s-jumps as a function of s is shown in Fig. 2; it is seen to decay as $\approx s^{-4}$, as expected from the well known approximately inverse cubic distribution of high frequency returns [20]. We note once again that this is a very broad distribution, meaning that the number of extreme events is in fact quite large. For example, for the already rather high value $s = 4$ and for only 166 stocks during 149 days, we found 177,674 jumps, which amounts to 1090 jumps per stock, or 7 to 8 jumps per stock per day! A threshold of $s = 8$ decreases this number by a factor ≈ 10, amounting to one jump every one day and a half per stock. On the same period and set of stocks we found 6443 news, or one news every 3 days for each stock. These numbers already suggest that a very large proportion of shocks (say, 4-jumps) cannot be attributed to idiosyncratic news (i.e. a news item containing the ticker of a given stock). The possibility of more collective news affecting macroeconomic variables will be discussed below.

In order to be more quantitative, we show in figure 3 a-b the conditional probability to observe a news (resp. a jump) at time t knowing that there was a news (resp. a jump) at time $t = 0$. One notices a substantial level of clustering in both cases. In figure 3 b, we also show the probability to observe a jump at time t conditioned to a news at time $t = 0$. Here, we see only a very small, but significant, asymmetric increase of the probability to see a jump induced by a news. This probability is only increased by a factor ≈ 2 in the two minutes following the news compared to the average background. So not only most jumps are not induced by news: most news do not induce any real jump at all! This decoupling means that most news are either expected (through rumors and leakage) or deemed insignificant by the market: genuine surprising intraday news are rare. Note that companies indeed tend to publish big surprises – like earnings – in overnight. Even earning forecasts by analysts are not taken very seriously by the market (see [21] for a discussion of this point).

Another clear-cut difference between jumps and news is the volatility pattern around the two types of events. In Fig. 4 a), we show the average absolute one-minute return $\sigma(t) = \langle |r(t)| \rangle$ conditioned to the presence of an $s = 4$ jump at $t = 0$, induced by a news or not. We can in fact determine the size distribution of news induced jumps. We find again a power law $s^{-\mu}$, but with an exponent $\mu \approx 2.7$ (see
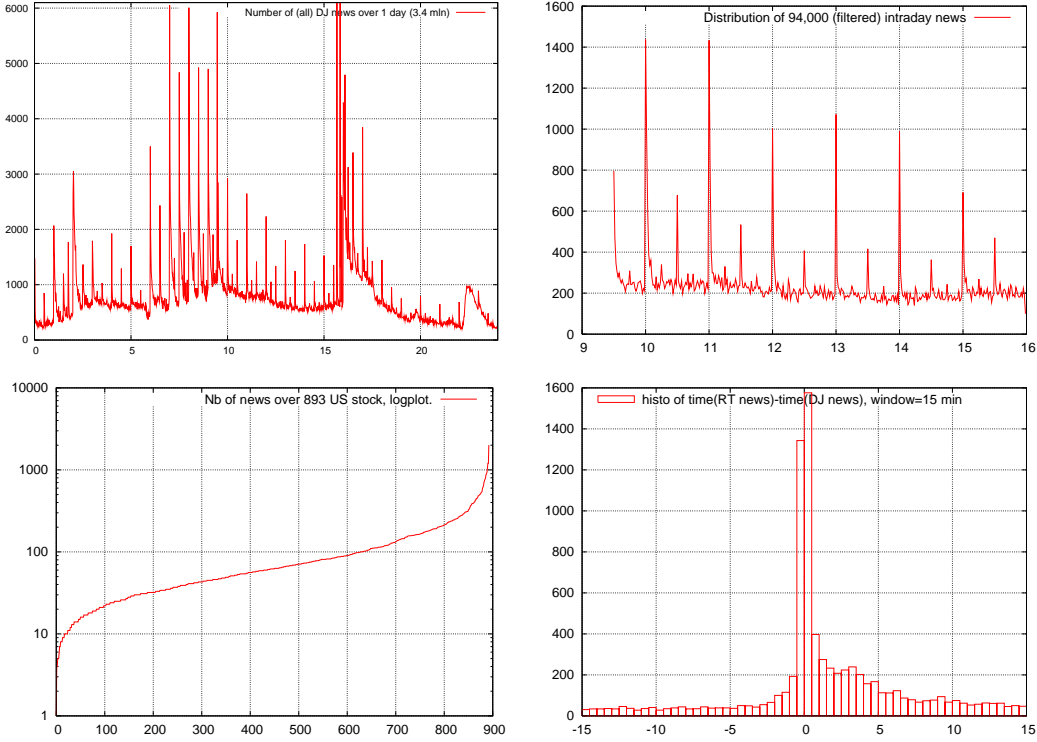
3

Figure 1: Statistics on Dow Jones news for 893 US stocks over 2 years.
a) Distribution of news throughout the day (551,000 news, first in series,
headline containing company's name). **b)** Idem for intraday, 'noisy' news
discarded (94,000 news). **c)** Distribution of number of news for each stock,
throughout 893 US stocks. **d)** Histogram of delays between DJ news and RT
news, expressed as time(RT-news)−time(DJ-news) in minutes. For 893 US
stocks and 2 years of news archive (intraday, only first news in a chain/story,
noisy news filtered out), we find one DJ news and one RT in the same window
of 15 minutes, and require each be preceded by a silence of 30 minutes. We
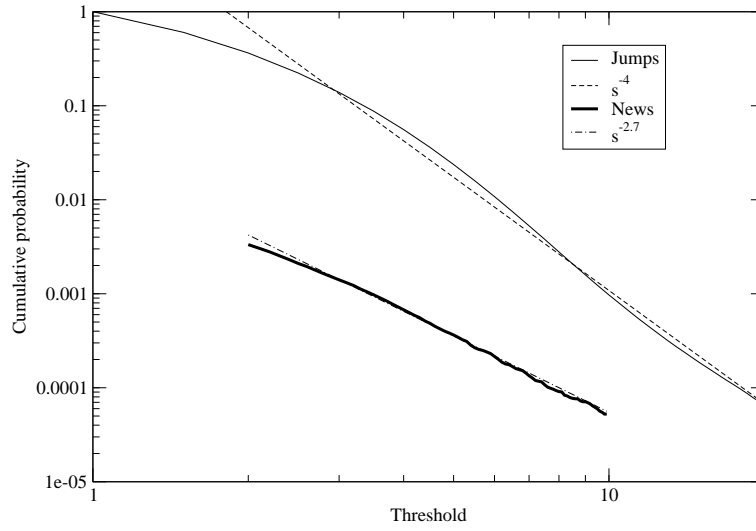see that DJ is quicker in 2/3 of cases (1.5 minutes ahead on avg).

4

Figure 2: Cumulative distribution of s-jumps as a function of s. The scale is log-log. The distribution decays as $\approx s^{-4}$. We also show the number of s-jumps associated with news. Interestingly, this distribution also decays as a power-law but with a smaller exponent, $\approx s^{-2.7}$.
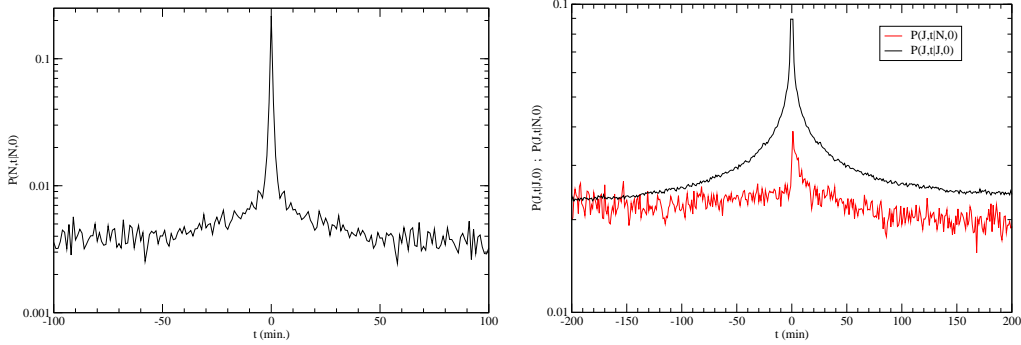
Figure 3: Clustering of news and jumps, measured by conditional probabilities **a)** Prob(news at time t | news at time 0), 200 minutes before and after $t = 0$, ignoring overnight news. **b)** Prob(jump at time t | jump at time 0) and Prob(jump at time t | news at time 0). Jumps here correspond to $s = 4$. Note that the probability to observe a jump after a news is actually decreased at intermediate times. For all densities, we divided each contribution by the average daily distribution of news or jumps, such as to remove intra-day seasonalities.

Fig. 1), clearly different from the value $\mu = 4$ mentioned above for jumps. The volatility pattern in the case of news is much wider than the rather narrow peak corresponding to endogenous jumps. In both cases, we find (Figure 5) that the relaxation of the excess-volatility follows a power-law in time $\sigma(t) - \sigma(\infty) \propto t^{-\beta}$ (see also [22, 23]). The exponent of the decay is, however, markedly different in the two cases: for news jumps, we find $\beta \approx 1$, whereas for endogenous jumps one has $\beta \approx 1/2$. Our results are compatible with those of [22], who find $\beta \approx 0.35$. The difference between endogenous and endogenous volatility relaxation has also been noted in [17], but on a very restricted set of news events. We found the value of β to be identical, within error bars, for the two threshold values $s = 4$ and $s = 8$ (see Fig. 5), although a three parameter fit is compatible with $\beta(s = 8) > \beta(s = 4)$ for jumps, as predicted by the multifractal random walk model [17] and the multiscale feedback model of [25].

If we now average the volatility pattern around *all* news, we find a rather modest peak height (the increase is only 30-40% of the baseline level), confirming the result of figure 3 b: news are often of no real importance and only mildly affect the price of a stock. The volatility decays with an exponent $\beta \approx 1/2$ towards a baseline value that is on average *lower* than before the news, as if news release had, on average, a sedating effect on markets. A similar effect can be noted in Fig. 3 b. This is however not true of jumps or of 'strong' news (corresponding

6

to $s \geq 4$ jumps): both endogenous jumps and unexpected news shake the market and induce extra uncertainty, that – as noted above – decays differently for jumps and for news.
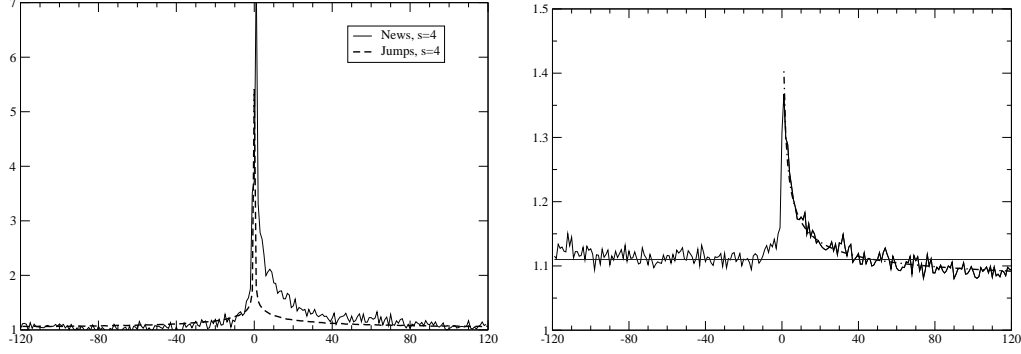

Figure 4: **a)** Volatility 120 minutes before and after the $s = 4$ news and $s = 4$ jumps. We divided each contribution by the U-shaped daily volatility pattern. Note that the news peak is much broader, but diverges faster for $t \to 0^+$ (see Fig. 5). **b)** Volatility 120 minutes before and after intraday news, averaged over *all* significant news. Note that the long time limit falls *below* the pre-news volatility.

The conclusion of this first part is that jumps and news appear to be, in general, distinct events, both concerning the amplitude of the volatility increase, and the shape of the volatility profile around these events. Jumps seem to occur for no identifiable reasons, and as a consequence spook the market; this perturbation is slow to decay (as $1/\sqrt{t}$). Although counter-intuitive at first, news seem (on average) to reduce any measure of uncertainty, since some previously unknown information becomes available. However, some news are really significant and also lead to an increased future volatility.

The above analysis focused on idiosyncratic company news only. However, some important macroeconomic news might affect the market (or a sector) as a whole and induce important price jumps that we fail to identify in our company news data set. Of course, the identification of possibly relevant macroeconomic news is at first sight much more difficult, since this would entail contextual word recognition. A short cut is to consider that any important macroeconomic news concerning e.g. interest rates, oil prices, inflation, etc. should induce a 'collective jump' in the market (see below for a precise definition). From the work of Cutler et al. [13], we already know that some of the market jumps are not explained by clearly identified macroeconomic events. But we could in fact *define* a market jump as being news, to which individual stocks might be sensitive to.
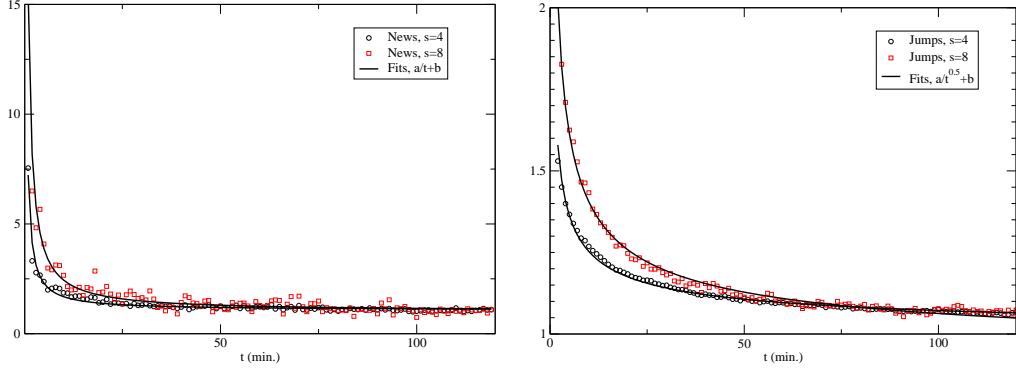
7

Figure 5: **a)** Relaxation of the volatility after $s = 4$ and $s = 8$ news jumps, and power law fit with an exponent $\beta = 1$. **b)** Relaxation of the volatility after $s = 4$ and $s = 8$ jumps, and power law fit with an exponent $\beta = 1/2$. Note the difference in the y-scale between the two curves (the first point, at $t = 0$, is suppressed in both cases).

There are different possible definitions of collective jumps. One is based on the correlation matrix of jump occurrences, $c_{ij} = T^{-1} \sum_t \theta_i^t \theta_j^t - p_i p_j$, where θ_i^t is one if stock i makes an s-jump in the one minute bin t and zero otherwise, p_i is the average of θ_i^t (i.e. the jump probability $p_i = T^{-1} \sum_t \theta_i^t$) and T is the number of bins. Most eigenvalues of c are seen to lie within the Marcenko-Pastur noise band [26, 27], but a few stand out, in particular the 'market' eigenvalue with a eigenvector v_i^1 close to uniform across all stocks: $v_i^1 \approx N^{-1/2}$, $\forall i$. A market jump can be defined such that the indicator $\chi^t = N^{1/2} \sum_i \theta_i^t v_i^1$ is larger than a certain threshold s'. For example $s' = 0.1$ means that roughly 10% of the stocks must jump to qualify as a market jump. Quite interestingly, we found a new 'stylized fact' concerning index jumps: the cumulative distribution of χ decays as a power-law $\chi^{-\nu}$ with exponent $\nu \approx 1.5$. In other words, the number of stocks involved in a market slide is very broadly distributed. Another definition of collective jumps is based on a standard sector classification to define a sector index; a sector jump is such that this index exceeds a certain threshold.

Focussing first on market jumps, and for a threshold $s' = 0.1$ that defines rather loose collective jumps, we find a total of 900 jumps, but only 13% of all individual jumps with $s = 4$ can be explained by these jumps, hardly more than the 10% expected from the very definition of these jumps. Extending the period around collective jumps to a five minute interval, one can increase this number to 20%, but this increase is the one expected from the unconditional probability of jumps, integrated over 5 minutes. If we include further meaningful sectors, we

8

increase the above 13% of explained jumps up to 21%. Increasing the threshold to $s' = 0.3$ further decreases these numbers to less than 5% even with a five minute interval. Changing the definition of collective jumps does not change the above picture.

The conclusion here is that even if one associates every collective jump to real macroeconomic news (which we know not be the case), a significant fraction of individual stock jumps are still unexplained. It therefore seems clear that the explanation of these jumps must lie elsewhere. One idea is that private information could play a role and trigger these jumps. But if an investor really had valuable private information he would trade as to reveal as little as possible of this information. This is best exemplified by Kyle's model of informed trading [28], where the trades of the insider are perfectly hidden in the uninformed flow. Nonetheless, it was recently claimed that large price jumps are due to large incoming volumes that destabilize the order book and lead to large swings [29]. This point of view was rather convincingly challenged in [18]. Here, we confirm explicitly that large price jumps are in fact *not* induced by large transaction volumes. We do this by studying tail correlations between absolute value of returns and volume. Define R_p and V_p as the p-quantiles of absolute returns and transaction volumes, i.e.:

$$P(|r| > R_p) = p \qquad P(V > V_p) = p. \tag{1}$$

Now, define the tail correlations as:

$$\mathcal{C}(p) = P(|r| > R_p | V > V_p) = P(V > V_p | |r| > R_p). \tag{2}$$

If $|r|$ and V are independent, then clearly $\mathcal{C}(p) = p$. Since both $|r|$ and V are power-law distributed, a relation of the type $|r| = \lambda V^\alpha + \epsilon$, as postulated in [29], leads to a finite limit for $\mathcal{C}(p)$ when $p \to 0$ whenever the residual ϵ is asymptotically subdominant (see e.g. [4], Ch. 11). This would be true even if only a finite fraction of the events followed such a relation. If $\mathcal{C}(p \to 0) \to 0$, on the other hand, this would mean that extreme returns and extreme volumes are asymptotically independent. We have studied $\mathcal{C}(p)$ both at the trade by trade level and for one-minute aggregate returns and volumes. The results are given in Fig. 6. Clearly, at the trade by trade level, there is *no* extreme correlations at all since $\mathcal{C}(p) \approx p$ to a good approximation. For one minute bins, there are some significant correlations ($\mathcal{C}(p) > p$) that are anyway expected since large price changes tend to be followed by large volumes of activity. But still, it is very clear that $\mathcal{C}(p)$ tends to zero when $p \to 0$, meaning that large jumps are *not* induced by large trading volumes.

So what is left to explain the seemingly spontaneous large price jumps? We believe that the explanation comes from the fact that markets, even when they are 'liquid', operate in a regime of *vanishing liquidity*, and therefore are in a *self-organized critical state* [31]. On electronic markets, the total volume available in
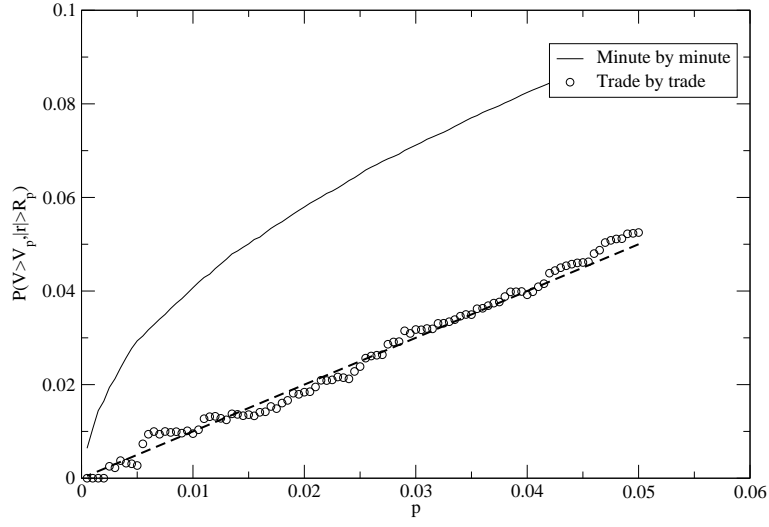
Figure 6: Tail correlations between returns and volumes, both at the trade by trade level (for 196 NASDAQ stocks between June 12,2006 and August 3, 2006) and at the one-minute bin level (averaged over all stocks of the above data set). Clearly, in both cases, $\mathcal{C}(p \to 0) \to 0$.

the order book is, at any instant of time, a tiny fraction of the stock capitalisation, say $10^{-5} - 10^{-4}$ (see e.g. [15]).[2] Liquidity providers take the risk of being "picked off", i.e. selling just before a big upwards move or vice versa, and therefore place limit orders quite cautiously, and tend to cancel these orders as soon as uncertainty signals appear. Such signals may simply be due to natural fluctuations in the order flow, which may lead, in some cases, to a catastrophic decay in liquidity, and therefore price jumps. There is indeed evidence that large price jumps are due to local liquidity dry outs. It would be very interesting to find a simple order flow model that captures these spontaneous critical liquidity fluctuations. Such a model would usefully shed light on the controversial nature of price jumps and help understand their origin.

References

[1] R. J. Shiller, *Do Stock Prices move too much to be justified by subsequent changes in dividends?*, American Economic Review, **71**, 421 (1981); R. J. Shiller, *Irrational Exuberance*, Princeton University Press (2000).

[2] R. Cont, *Empirical properties of asset returns: stylized facts and statistical issues.* Quantitative Finance **1**, 223 (2001)

[3] L. Borland, J. P. Bouchaud, J.-F. Muzy, G. Zumbach, *The dynamics of Financial Markets: Mandelbrot's multifractal cascades, and beyond*, Wilmott Magazine, March 2005.

[4] J.-P. Bouchaud and M. Potters, *Theory of Financial Risks and Derivative Pricing*, Cambridge University Press, 2004.

[5] U. Frisch, *Turbulence: the Kolmogorov legacy*, Cambridge University Press (1997).

[6] A. Fisher, L. Calvet, B.B. Mandelbrot, *Multifractality of DEM/$ rates*, Cowles Foundation Discussion Paper 1165.

---

[2]This explain why liquidity takers have to cut their orders in small pieces, thereby creating long range correlations in order flow [12, 30].

[7] J.-F. Muzy, J. Delour, E. Bacry, *Modelling fluctuations of financial time series: from cascade process to stochastic volatility model*, Eur. Phys. J. B **17**, 537-548 (2000); E. Bacry, J. Delour and J.F. Muzy, *Multifractal random walk*, Phys. Rev. E **64**, 026103 (2001).

[8] L. Calvet, A. Fisher, *Forecasting multifractal volatility*, Journal of Econometrics, **105**, 27, (2001); L. Calvet, A. Fisher, *Multifractality in Asset Returns: Theory and Evidence*, Review of Economics and Statistics **84**, 381-406 (2002).

[9] T. Lux, *Turbulence in financial markets: the surprising explanatory power of simple cascade models*, Quantitative Finance **1**, 632, (2001); T. Lux, *Multi-Fractal Processes as a Model for Financial Returns: Simple Moment and GMM Estimation*, in revision for Journal of Business and Economic Statistics

[10] J. L. Cabrera and J. G. Milton, *On-off intermittency in a human balancing task* Phys. Rev. Lett. **89** 158702 (2002).

[11] A. Krawiecki, J. A. Hoyst and D. Helbing, *Volatility clustering and scaling for financial time series due to attractor bubbling* Phys. Rev. Lett. **89** 158701 (2002)

[12] J. P. Bouchaud, Y. Gefen, M. Potters, M. Wyart, *Fluctuations and response in financial markets: The subtle nature of 'random' price changes*, Quantitative Finance **4**, 176 (2004); J. P. Bouchaud, J. Kockelkoren, M. Potters, *Random walks, liquidity molasses and critical response in financial markets*, Quantitative Finance, **6**, 115 (2006).

[13] D. M. Cutler, J. M. Poterba, and L. H. Summers. *What moves stock prices?* The Journal of Portfolio Management, 15(3):412, 1989.

[14] Ray C. Fair, *Events That Shook the Market*, The Journal of Business, 2002, vol. 75, no. 4

[15] M. Wyart, J. P. Bouchaud, J. Kockelkoren, M. Potters, M. Vettorazzo, *Relation between Bid-ask spread, impact and volatility in order driven markets*

[16] C. Hopman, *Are supply and demand driving stock prices?*, Quantitative Finance.

[17] D. Sornette, Y. Malevergne, J.F. Muzy, *What causes crashes*, Risk Magazine, 67 (Feb. 2003).

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

[19] A. Ponzi, F. Lillo, R. N. Mantegna, *Market reaction to temporary liquidity crises and the permanent market impact*, physics/0608032

[20] V. Plerou, P. Gopikrishnan, L.A. Amaral, M. Meyer, H.E. Stanley, *Scaling of the distribution of price fluctuations of individual companies*, Phys. Rev. **E60** 6519 (1999); P. Gopikrishnan, V. Plerou, L. A. Amaral, M. Meyer, H. E. Stanley, *Scaling of the distribution of fluctuations of financial market indices*, Phys. Rev. **E 60** 5305 (1999)

[21] O. Guedj, J. P. Bouchaud, *Experts earning forecasts: bias, herding and gossamer information*, International Journal of Theoretical and Applied Finance, **8**, 933 (2005).

[22] A. G. Zawadowski, J. Kertesz, G. Andor, *Short-term market reaction after extreme price changes of liquid stocks* Quantitative Finance **6**, 283-295 (2006)

[23] F. Lillo and R. N. Mantegna, *Power law relaxation in a complex system: Omori Law After a Financial Market Crash*, Physical Review E **68**, 016119 (2003)

[24] G. Zumbach, *Volatility processes and volatility forecast with long memory*, Quantitative Finance **4** 70 (2004).

[25] L. Borland, J.P. Bouchaud, *On a multi-timescale statistical feedback model for volatility fluctuations*, e-print cond-mat/0507073 (unpublished).

[26] V. A. Marčenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Math. USSR-Sb, **1**, 457-483 (1967)

[27] L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters, *Noise Dressing of. Financial Correlation Matrices*, Phys. Rev. Lett. **83**, 1467 (1999); M. Potters, J.-P. Bouchaud and L. Laloux, *Financial Applications of Random Matrix Theory: Old laces and new pieces*, Acta Physica Polonica B, **36** 2767 (2005).

[28] A. S. Kyle, *Continuous auctions and insider trading*, Econometrica **53**, 1315 (1985).

[29] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. *A theory of power-law distributions in financial market fluctuations*, Nature, **423**, 267270, (2003); X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. *Institutional Investors and Stock Market Volatility*, Quarterly Journal of Economics, **121** 461-504 (2006).

[30] F. Lillo, J. D. Farmer, *The long memory of efficient markets*, Studies in Nonlinear Dynamics and Econometrics, **8**, 1 (2004); F. Lillo, S. Mike, and J. D. Farmer, *Theory for long memory in supply and demand*, Physical Review E, 7106, 287 (2005).

[31] P. Bak, *How Nature Works: The Science of Self-Organized Criticality.* Copernicus, Springer, New York, 1996.