

# Financial bubbles analysis with a cross-sectional estimator

Frédéric Abergel \*

Nicolas Huth<sup>†</sup> \*

Ioane Muni Toke \*

September 16, 2009

## Abstract

We highlight a very simple statistical tool for the analysis of financial bubbles, which has already been studied in [1]. We provide extensive empirical tests of this statistical tool and investigate analytically its link with stocks correlation structure.

## Introduction

Forecasting the burst of financial bubbles would be incredibly useful for many players in stock exchanges, including regulators, portfolio managers and investment banks. Fundamental indicators relying on economic analysis can be monitored. But is it possible to find some statistical regularity in market crashes? Several authors, including [2, 3, 4], have already attempted to answer this question. In this paper, we focus on a very simple statistical tool, first introduced in [1], and study analytically its link with stocks correlation structure. This approach is similar to the one studied in [2, 4] although different through the statistical object under consideration.

## 1 A spatial survival function

In [1], an unusual and interesting statistical tool is introduced in order to study market crashes. Given  $N$  stocks on a market place, a reference date<sup>1</sup>  $t_{ref}$  and the current date  $t$ , we set

$$S_N(z) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{X_i(t_{ref}, t) > z\}}$$

where  $X_i(t_{ref}, t) := \frac{S_i(t)}{S_i(t_{ref})}$  is the performance of stock  $i$  over  $[t_{ref}, t]$ . In the following, we shall leave the time argument for notational simplicity. Intuitively,  $S_N(z)$  can be seen as the proportion of stocks displaying a greater performance than  $z$ , i.e. the survival function of stocks on day  $t$ . From this point of view, it is a measure of stocks dispersion: a slow decreasing  $S_N(z)$  indicates broadly distributed performances, thus reflecting an important dispersion. Since  $t_{ref}$  is supposed to be as close as possible to the onset of the bubble,  $X_i(t_{ref}, t)$  might be of the order of monthly or yearly returns. Short notes on similar statistical objects can be found in [5, 6]. This is very different from looking at daily returns as in [2, 4] and might be more relevant for bubble detection since it often takes long time for a bubble to build up and for bubbling stocks to disperse.

Statistical properties of  $S_N(z)$  are interestingly robust. The main features of this quantity are:

- for  $z \sim +\infty$ ,  $S_N(z) \sim z^{-\alpha}$ ;
- the variance of the  $X_i$ 's increases dramatically before crashes.

These two features are robust with respect to the choice of the arbitrary reference date  $t_{ref}$  and are valid over several financial crashes. We illustrate these two facts on figures 1 and 2. We use daily close prices of three different sets of stocks: the stocks composing the Australian Stock Exchange All Ordinaries index (AORD, 500 stocks); the stocks composing the New York Stock Exchange Composite index (NYA, 1800 stocks); the stocks composing the Shanghai Composite index (SSE, 900 stocks). We choose the first trading day of 2003 as our date  $t_{ref}$ . As for the first fact, figure 1 shows three examples of distributions of  $X_i$ 's at random dates. It appears that the power-law tail is indeed a good fit as the normalized prices grow. As for the second fact, we plot on figure 2 the timeseries of

\*BNP Paribas Chair of Quantitative Finance, Ecole Centrale Paris, MAS Laboratory.

<sup>†</sup>Natixis, Equity Derivatives and Arbitrage. E-mail: nicolas.huth@natixis.com. The authors would like to thank the members of Natixis' equity derivatives quantitative R&D team for fruitful discussions.

<sup>1</sup>supposed to be close to the onset of a financial bubble

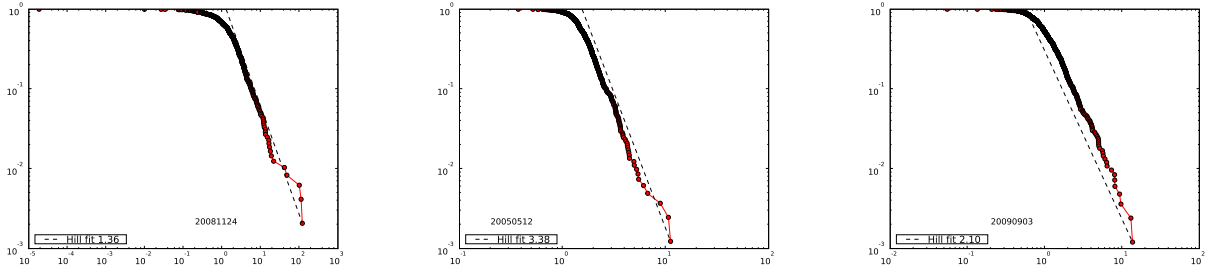


Figure 1: Survival functions of cross-sectional distributions of normalized prices on three sets of stocks at random dates, in log-log scale. (left) AORD stocks on Nov 24th, 2008; (center) NYA stocks on Dec 12th, 2005; (right) SSE stocks on Sep 3rd, 2009

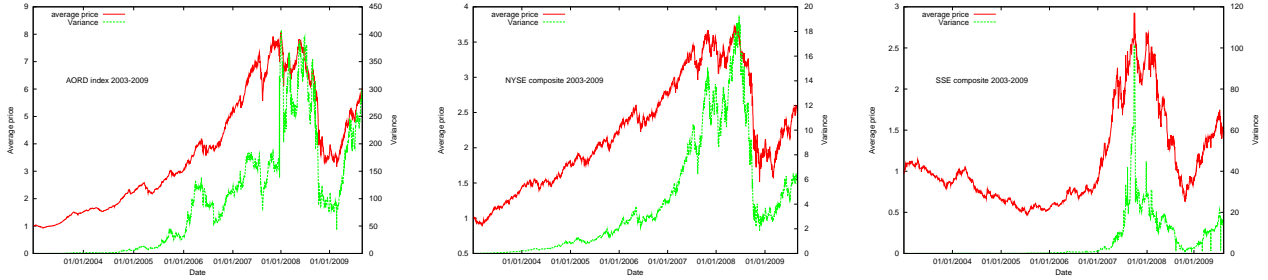


Figure 2: Evolutions of the mean (“average price”) and variance of the  $X_i$ ’s on three sets of stocks from Jan 1st, 2003 to Sep 3rd, 2009.(left) AORD stocks; (center) NYA stocks ; (right) SSE stocks

the average and variance of the  $X_i$ ’s. It appears clearly that on all stocks, the variance grows dramatically as the bubble inflates, especially prior to crashes.

We also observe that these properties are robust with respect to the choice of  $t_{ref}$ : on figure 3 (left), timeseries of the variance of the  $X_i$ ’s are plotted for eleven different reference dates (first trading day of the year from 1998 to 2008). It appears that the bursts of variance exists whatever the date  $t_{ref}$ . For the series that begin before 2000, the Internet bubble is clearly visible. The recent 2007 bubble is also visible for all series, even for the most recent reference dates (see zoom in inset).

In [1], the author argues that the power-law exponent  $\alpha$  dives towards 1 before crashes, leading to the divergence of the mean performance and thus to the burst of the financial bubble. We find this threshold to be not robust with respect to  $t_{ref}$ : on figure 3 (right), the Hill estimator of the Pareto exponent of the tail of the distributions of the  $X_i$  is plotted, and the value of the index clearly depends on the reference date  $t_{ref}$ . However, the decrease in  $\alpha$  and the rise of the variance are linked since the variance of a random variable distributed as a power-law of index  $\alpha$  is given by  $\frac{\alpha}{(\alpha-1)^2(\alpha-2)}$ . On figure 3 (right), we observe that the Pareto exponent shows local minima and sharp peaks prior crashes. For example, first minimum is obtained on March 6th 2000 for all (started) time series, and the maximum of the Internet bubble is observed on March 9th 2000. More recently, all times series exhibit a sharp minimum on June 30th-July 1st 2008, prior to the large market dive of September 2008 (market height in 2008 is attained on June 6th).

## 2 Link with the covariance structure

On the one hand, financial bubbles happen when market quotes of certain sectors of the economy are booming, thus increasing the dispersion over whole stocks. On the other hand, it is common knowledge that market crashes are associated to bursts of volatility and correlation. How can we relate these two phenomena? Empirically, the rise in dispersion is given by the increase in the variance of performances  $V_N$ , which can be computed as follows

$$V_N = \int_0^{+\infty} z^2 s_N(z) dz - \left( \int_0^{+\infty} z s_N(z) dz \right)^2$$

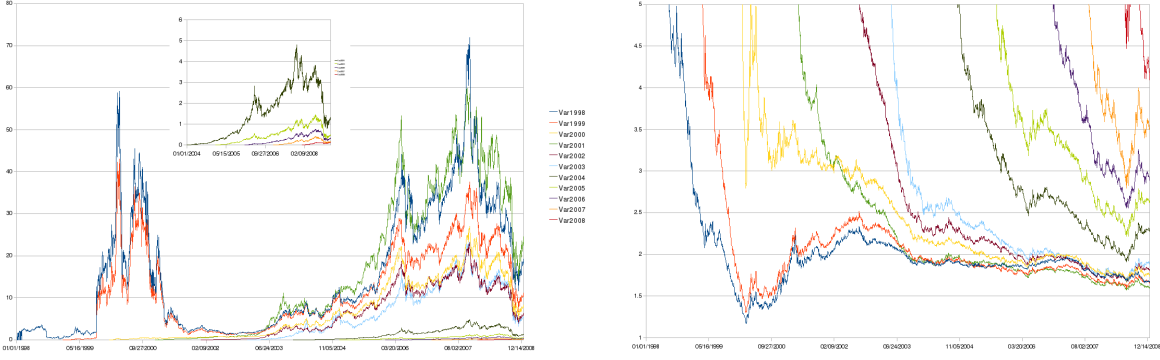


Figure 3: (left) Variance of the normalized prices  $X_i$  for eleven different reference dates  $t_{ref}$ . (right) Hill estimator of the Pareto exponent of the tail of distribution of the normalized prices  $X_i$  for eleven different reference dates  $t_{ref}$ . (all) These graphs are computed using daily close prices of the stocks composing the US Russell3000 index (3000 stocks).

where  $s_N(z) = -\frac{d}{dz}S_N(z) = \frac{1}{N} \sum_{i=1}^N \delta(X_i - z)$  is the empirical probability density function associated to performances. It is easily seen that

$$V_N = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X}_N)^2 = \frac{1}{2N^2} \sum_{i,j=1}^N (X_i - X_j)^2$$

with  $\bar{X}_N := \frac{1}{N} \sum_{i=1}^N X_i$ . Since  $V_N$  is a sum of random variables, we would like to apply a convergence theorem such as the law of large numbers. However, we must be cautious since the  $X_i$ 's are correlated and not identically distributed.  $V_N$  will converge towards its mean if its variance is asymptotically nil. We set  $m_i := \mathbb{E}(X_i)$ ,  $\sigma_i^2 := \text{Var}(X_i)$ ,  $\rho_{ij} := \text{Corr}(X_i, X_j)$ . Straightforward computations show that

$$\begin{aligned} \mathbb{E}(V_N) &= \frac{1}{N} \sum_{i=1}^N \sigma_i^2 - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \rho_{ij} \sigma_i \sigma_j + \frac{1}{N} \sum_{i=1}^N m_i^2 - \left( \frac{1}{N} \sum_{i=1}^N m_i \right)^2 \\ \mathbb{E}(V_N^2) &= \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}((X_i - \bar{X}_N)^4) + \frac{2}{N^2} \sum_{i=1}^N \sum_{j>i}^N \mathbb{E}((X_i - \bar{X}_N)^2 (X_j - \bar{X}_N)^2) \end{aligned}$$

These two quantities exist provided that  $c_{ij}^{\alpha,\beta} := \mathbb{E}(|X_i^\alpha X_j^\beta|) < \infty$  for all  $(i, j)$  and  $(\alpha, \beta) \in \mathcal{E} := \{(\gamma, \delta) \in \{0, 1, 2, 3, 4\}^2 \mid \gamma + \delta \leq 4\}$ . If the sequences  $\{c_{ij}^{\alpha,\beta}, i, j = 1, \dots, N, (\alpha, \beta) \in \mathcal{E}\}$  are such that  $\lim_{N \rightarrow +\infty} \text{Var}(V_N) = 0$ , then

$$\mathbb{P} - \lim_{N \rightarrow +\infty} V_N = \lim_{N \rightarrow +\infty} \left( \frac{1}{N} \sum_{i=1}^N \sigma_i^2 - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \rho_{ij} \sigma_i \sigma_j + \frac{1}{N} \sum_{i=1}^N m_i^2 - \left( \frac{1}{N} \sum_{i=1}^N m_i \right)^2 \right)$$

The above equation relates the dispersion of stocks  $V_N$  to their variance-covariance (and mean) structure. The dispersion  $V_N$ :

- increases as individual variances increase: if stocks are individually very unstable, then it is very likely that the whole market will be so;
- increases as covariances, in particular as correlations, decrease: the more anti-correlated stocks are, the more distant each pair will be, so that the whole dispersion increases.

Both effects are illustrated on table 1. We simulate a Gaussian vector with mean zero in dimension  $N = 1000$  a  $M = 100$  times, then compute the mean of  $V_N$  over these  $M$  simulations for different levels of correlation or standard deviation.

Assume that the initial  $X_i$ 's are centered and normalized so that  $m_i = 0$  and  $\sigma_i = 1$ . We are then able to give bounds on  $\lim V_N$  depending on correlation. Setting  $\rho_{ij} = 1$  for all  $i, j$  leads to  $\mathbb{E}(V_N) = 0$  and  $\rho_{ij} = -1$  for all  $i, j$  leads to  $\mathbb{E}(V_N) = 2$ , as shown in table 1.

$\rho$	Mean $V_N$	$\sigma$	Mean $V_N (\times 10^{13})$
-1	1.997	0.1	0.056
-0.8	1.808	0.3	0.377
-0.6	1.595	0.5	1.654
-0.4	1.393	0.7	3.052
-0.2	1.199	0.9	3.866
0	0.999	1.1	5.895
0.2	0.797	1.3	8.765
0.4	0.603	1.5	14.249
0.6	0.400	1.7	13.104
0.8	0.199	1.9	22.164
1	$7.430 \times 10^{-13}$		

Table 1: Simulation of a Gaussian vector with different levels of correlation and standard deviation. As correlation decreases, the dispersion is larger. Standard deviations (resp. correlations) are set to 1 (resp. 0.5) in the left (resp. right) table.

### 3 Conclusion and further research

We have studied empirically and analytically an indicator of bubbles build up and burst on financial markets. This statistical tool is the variance of fixed starting date performances over the stocks universe. It is quite robust with respect to the choice of the market place and the starting date. The .com and subprime bubbles are well identified by this methodology. Fundamentally, we establish a link between the building up of a bubble and anti-correlation between stocks as well as individual variances bursts.

Regarding further research, we have two things in mind:

- test empirically the link we establish between bubbles and stocks variance-covariance structure;
- suggest an agent-based model to explain the fundamental microscopic mechanisms underlying this link between bubbles and stocks variance-covariance structure.

The first direction requires a statistical measure for the variance and correlation of  $X_i(t_{ref}, t)$  over  $i$  at each time  $t$ . This problem is quite complex in practice since, unless one goes 30 years back, one does not have historical daily data to compute them. A solution might be found in the use of high frequency data by computing the variances and correlations needed over the returns of the previous day. Furthermore, having access to these statistical quantities would be useful for normalizing returns in order to bound the market variance  $V_N$ , thus making it an indicator with fundamental thresholds.

Some exchange models for wealth distribution, such as [7], have striking similarities with our approach and could be therefore used to understand the power law distribution of stocks ensemble and global variance burst during bubbles from a microscopic point of view towards interactions between individual stocks.

## References

- [1] Kaizoji, T. A precursor of market crashes: Empirical laws of japan’s internet bubble. *The European Physical Journal B* **50**, 5 pages (2006).
- [2] Lillo, F. & Mantegna, R. N. Variety and volatility in financial markets. *Physical Review E* **62**, 6126 (2000).
- [3] Sornette, D. & Johansen, A. Significance of log-periodic precursors to financial crashes. *Quantitative Finance* **1**, 452–471 (2001).
- [4] Borland, L. Statistical signatures in times of panic: Markets as a Self-Organizing system. *Arxiv preprint arXiv:0908.0111* (2009).
- [5] Kaizoji, T. & Kaizoji, M. Power law for ensembles of stock prices. *Physica A: Statistical Mechanics and its Applications* **344**, 240–243 (2004).
- [6] Men, D., Wang, J. & Shao, J. The statistical analysis of stock prices and trading volumes for the chinese stock markets. In *Computing, Communication, Control, and Management, 2008. CCCM ’08.*, vol. 3, 37–41 (2008).
- [7] Bouchaud, J. P. & Mezard, M. Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications* **282**, 536–545 (2000).