Stylized Facts

Jim Gatheral, Merrill Lynch

September 20, 2007

1 Rama Cont

Cont (2001) lists the following stylized facts:

- Absence of autocorrelation
- Heavy tails
- Gain/loss asymmetry
- Aggregational Gaussianity
- Intermittency
- Volatility clustering
- Conditional heavy tails
- Slow decay of autocorrelation in absolute returns
- Leverage effect
- Volatility/volume correlation
- Asymmetry in time scales

To quote Rama Cont "these stylized facts are so constraining that it is not easy to exhibit even an (*ad hoc*) stochastic process which possesses the same set of properties and one has to go to great lengths to reproduce them with a model".

2 Power laws

Consider the distribution of papers with "QCD" in the text. There are 16,597 of them in total. These are distributed as follows:

Number of citations	Number of papers
1,000+	8
500+	43
100+	686
50+	1,561
0+	16,591

Just eye-balling the data, you can see that if you double the cutoff, you divide the number of papers by approximately 4 (for large cutoffs).

Repeating this experiment for papers with "exponentiation" in the text, I got

Number of citations	Number of papers
1,000+	0
500+	0
100+	3
50+	11
0+	86

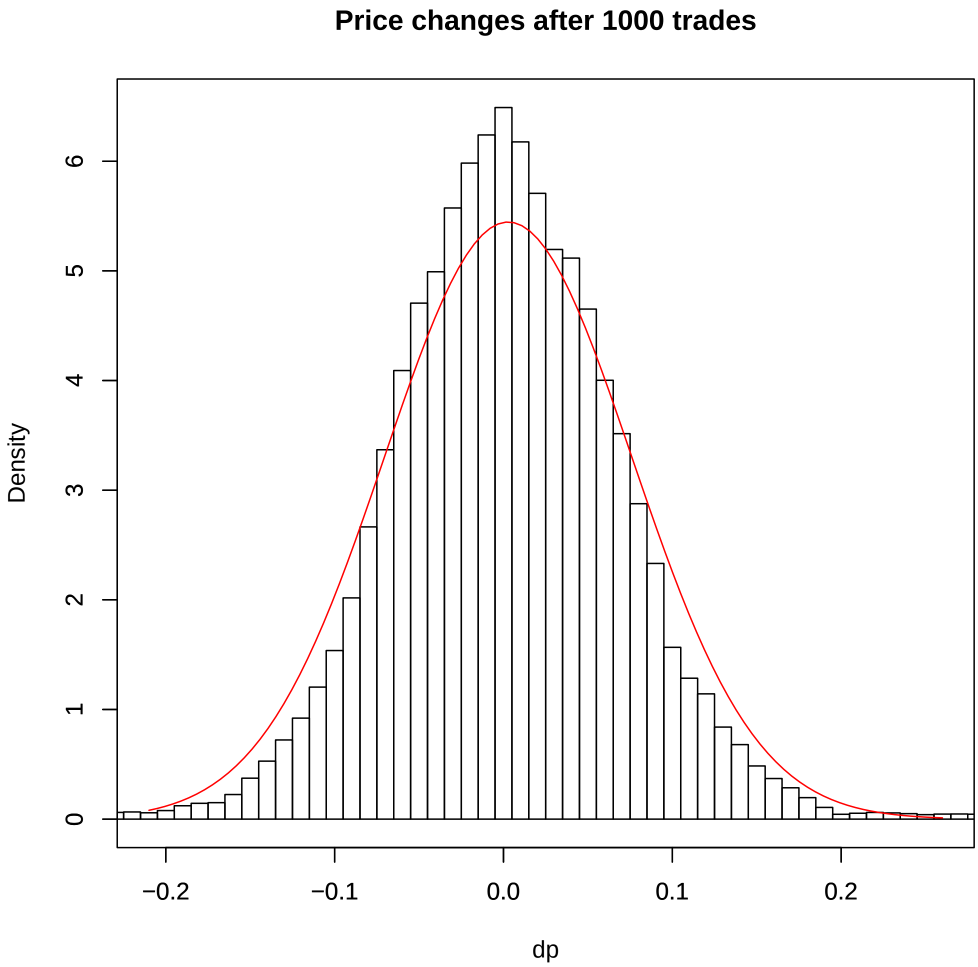
This kind of power-law distribution occurs all the time in social sciences. Pareto for example noted that 80% of wealth was owned by 20% of the population, 64% of wealth was owned by 4% of the population, and so on.

2.1 Power laws in stock returns

For a good introduction to the analysis of empirical financial data, see Bouchaud and Potters (2003). We start with 17 days of CSCO trade data from 20-Aug-2007 to 12-Sep-2007 (over 700,000 trades after cleaning). We compute price changes over intervals of a given number of trades. Consider the histogram and log-log plot of price changes in Figures 1 and 2.

We see from Figure 1 that the empirical distribution has a high peak and fat tails. Although our first reaction might be that this is to be expected, the idea of stochastic volatility is that returns in trading time should be normally distributed. Here we see that returns in trading time have fat tails.

Figure 1: Histogram of CSCO price changes after 1,000 trades. The red line is a fit of the normal density.

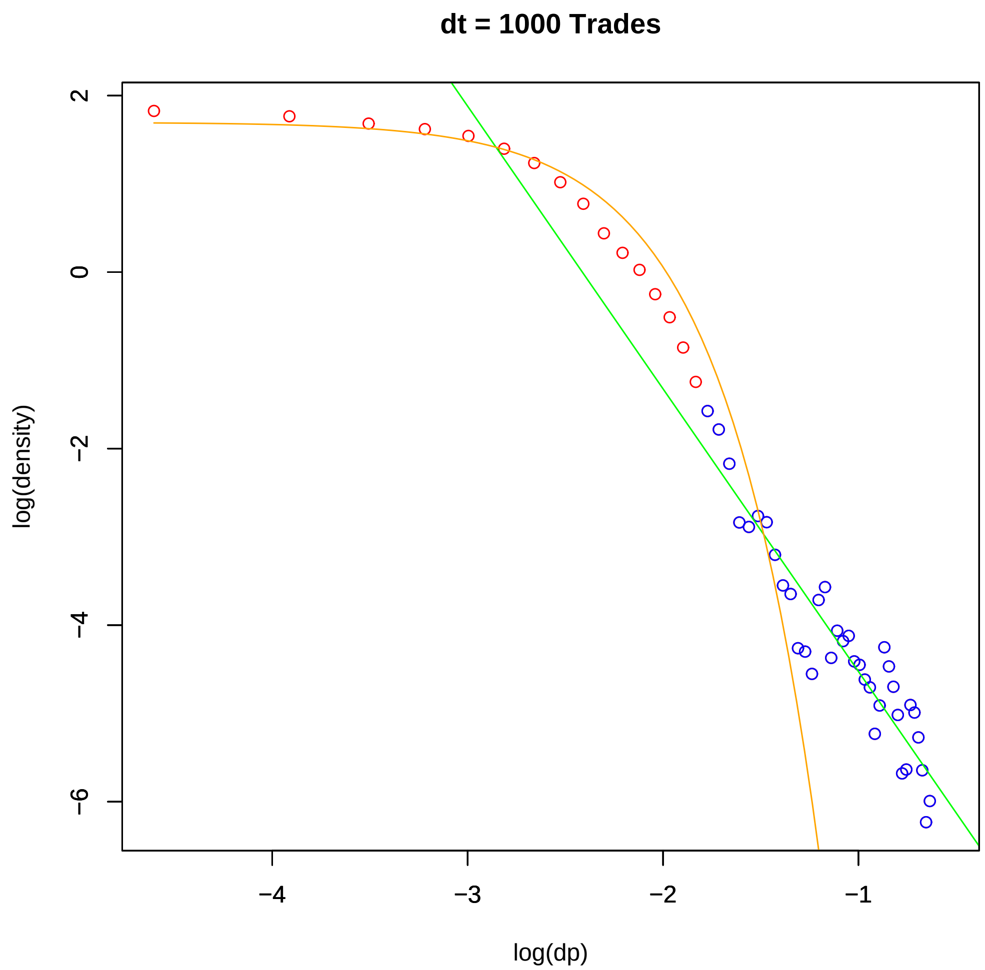


2.2 Autocorrelation of absolute returns

Perelló, Masoliver, and Bouchaud (2004) point out that the autocorrelation of squared returns decays as a power law. They go on to show that a model with two timescales has similar behavior. See also Gatheral (2007).

In Figure 3, we graph the decay of the autocorrelation CSCO absolute

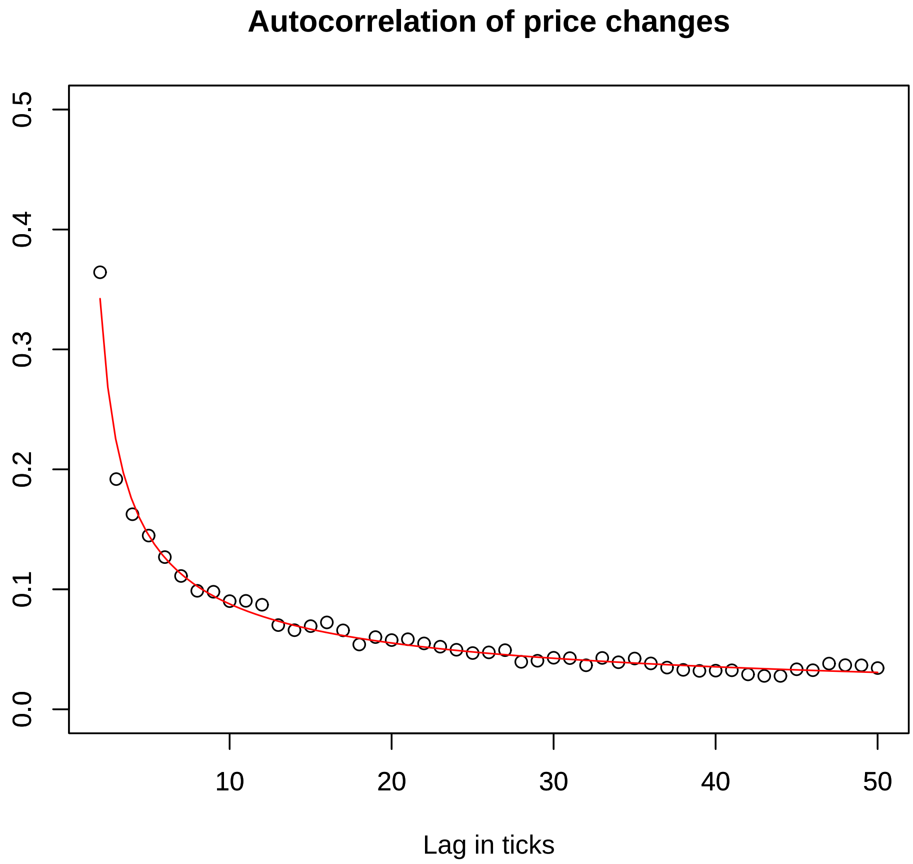
Figure 2: Log-log plot of CSCO prices changes. The green line is a (power-law) fit to the blue points. The orange curve is the fitted normal density.



tick-by-tick returns as a function of lag in ticks:

We see that the autocorrelation of absolute returns is indeed well-approximated by a power-law function of lag.

Figure 3: Autocorrelation of absolute tick returns of CSCO on 07-Sep-2006. The red line is the power-law fit $C/\text{lag}^{0.61}$.

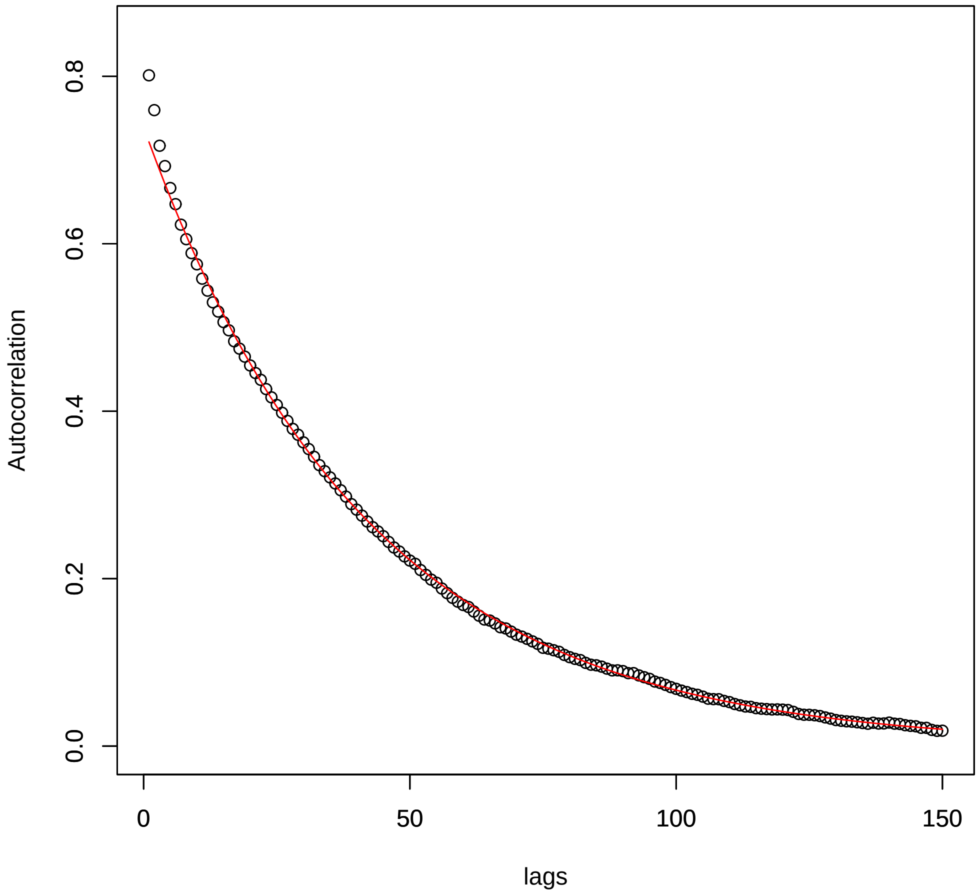


2.3 Autocorrelation of trade signs

In Figure 4, we see that the signs of trades are highly autocorrelated. This can be understood both in terms of order splitting and terms of rational optimal limit order strategies.

Once again, as we can see from Figure 5, the autocorrelation of trade signs is well-approximated by a power law.

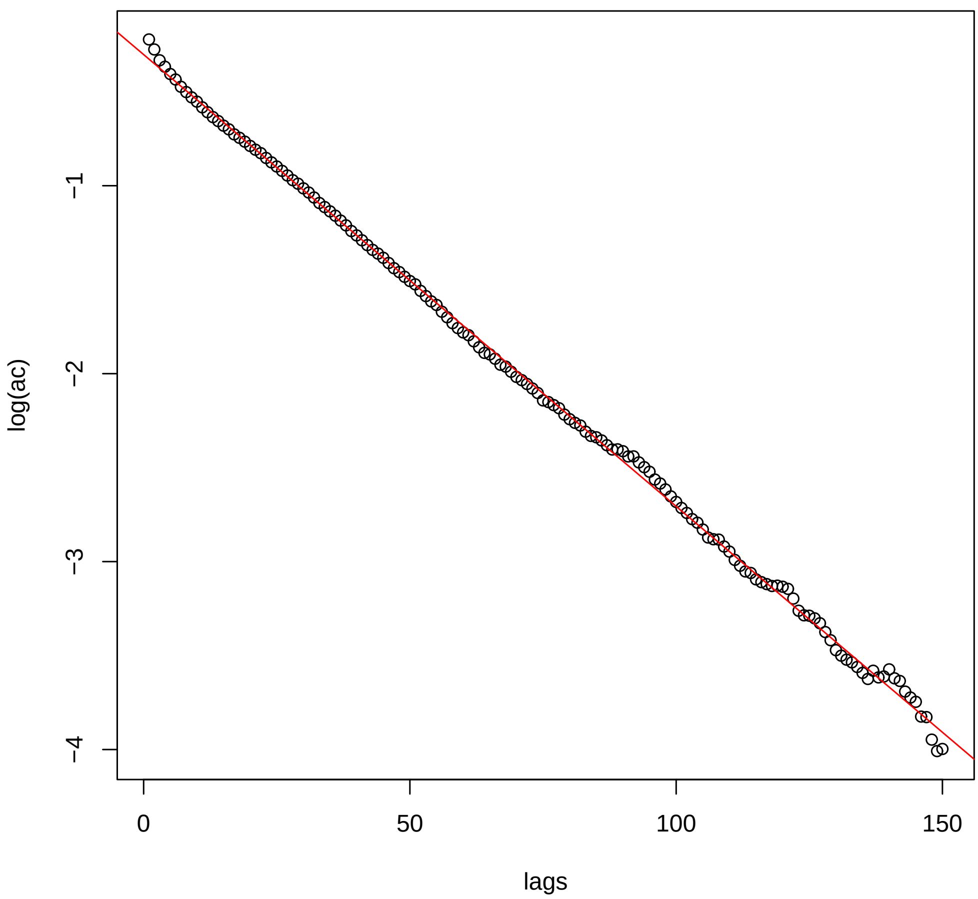
Figure 4: CSCO on 07-Sep-2006: Autocorrelation of trade signs.



2.4 Distribution of trade sizes

From Figure 6, we see that the distribution of trade sizes also has power-law tails with an exponent of roughly 5/2. This of course corresponds to a power-law tail in the CDF of approximately 3/2.

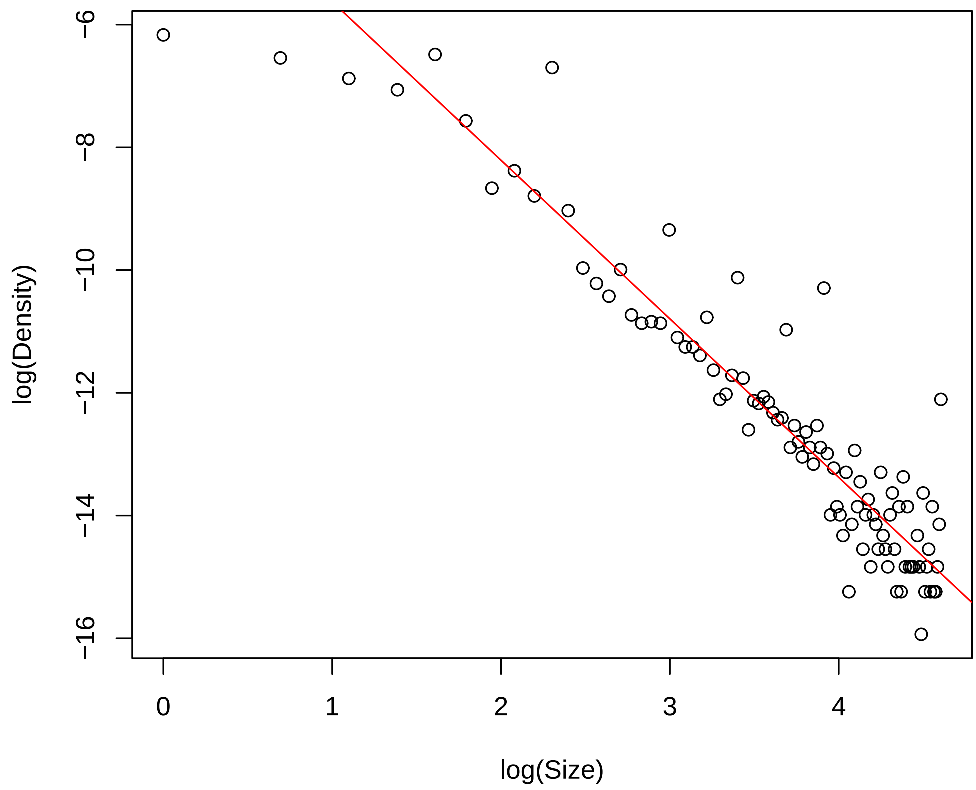
Figure 5: CSCO on 07-Sep-2006: Log-log plot of trade-sign autocorrelation.



3 Various other observations

Gillemot, Farmer, and Lillo (2005) point out that the analysis of Ané and Geman (2000) must be wrong because empirically observed returns have approximately the same distribution no matter what timescale is selected: clock time, volume time or transaction time. That's a major problem for stochastic volatility models. A Yale PhD student called Yang Li apparently

Figure 6: CSCO on 07-Sep-2006: Distribution of trade sizes.



showed where Ané and Geman went wrong but I haven't been able to find a copy of her thesis.

Plerou, Stanley, Gabaix, and Gopikrishnan (2004) claim to find a relationship between the power 3 in the tail of the distribution of returns with the power 3/2 in the tail of the distribution of trade sizes. The relationship is through the well-known approximate formula for market impact. We see from Figure 2 however that the cubic law is not exact; the power-law exponent on the 1000-tick timescale is nearer 4 than 3.

References

Ané, Thierry, and Hélyette Geman, 2000, Order flow, transaction clock, and normality of asset returns, *The Journal of Finance* 55, 2259–2284.

Bouchaud, Jean-Philippe, and Marc Potters, 2003, *Theory of Financial Risk and Derivative Pricing: From Statististical Physics to Risk Management* (Cambridge University Press: Cambridge) 2 edn.

Cont, Rama, 2001, Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance* 1, 223–236.

Gatheral, Jim, 2007, Developments in volatility derivatives pricing, http://www.math.nyu.edu/fellows_fin_math/gatheral/VolDerivatives2007_Paris_Handouts.pd

Gillemot, László, J. Doyne Farmer, and Fabrizio Lillo, 2005, There's more to volatility than volume, Discussion paper Santa Fe Institute.

Perelló, Josep, Jaume Masoliver, and Jean-Philippe Bouchaud, 2004, Multiple time scales in volatility and leverage correlations: a stochastic volatility model, *Applied Mathematical Finance* 11, 25–50.

Plerou, Vasiliki, H Eugene Stanley, Xavier Gabaix, and Parameswaran Gopikrishnan, 2004, On the origin of power-law fluctuations in stock prices, *Quantitative Finance* 4, 11–15.