Statistical Analysis of Investors' Behaviour: Different Temporal Scales

Mario Gutiérrez-Roig* and Josep Perelló
Departament de Física Fonamental
Universitat de Barcelona Diagonal 647, E-08028 (Barcelona)

The human collective behaviour in society is becoming more and more empirically studied by scientists. Nowadays, globalization, internet and our ICT society let us provide very large amount of data ready to study empirically. Most of the challenges can be summarized along the question how microscopic interactions between agents trigger macroscopic phenomena in a large scale level. Inside the context of financial markets, the role of agents' statistical properties, for instance the expectation time studied here, it is crucial to understand the price formation at the macro-level.

We study a dataset containing more than 7 million of individual recordings from 29,930 non-experts professional investors (clients) of a particular investment firm. All of them had traded between 2000 and 2007 in Spanish stock market IBEX. Price, date, and number of shares traded are recorded for each transaction, so that we can track the performance for every individual.

In order to guess an approximate investor's strategy it is reasonable to hypothesize that most influence over investors' response is the price of the asset which they are trading with. Therefore we can study the cross-correlation between what people do and what market does[1] through the coefficient

$$\rho^i_{Sr}\left(L\right) = \frac{1}{T^i_L}\sum_{k=1}^{T^i_L} \frac{\left(S^i_k\left(L\right) - \bar{S}^i\left(L\right)\right)\left(r_k\left(L\right) - \bar{r}\left(L\right)\right)}{\sigma_S \sigma_r}, \quad (1)$$

where T^i_L is the number of active periods of length L, $S^i_k\left(L\right)$ is the balance (shares bought minus shares sold) for investor i in the time-window of length L and $r_k\left(L\right)$ means the logarithmic price change in that time-window.

The mean of the distribution is found on negative region (FIG. 1), meaning that, on average, when market goes up investors sell and when the market falls down investors buy. One of the reasons to explain the dispersion of the points is because all agents does not base their strategy on the same time-horizon. This time-horizon can be heuristically determined by finding out which period L maximises $\rho^i_{S-r}\left(L\right)$ function (FIG. 2).

The implications of these preliminary results can be important for testing and estimating parameters in agent-based models, like the expectation horizon for the price formation of a single agent[2].

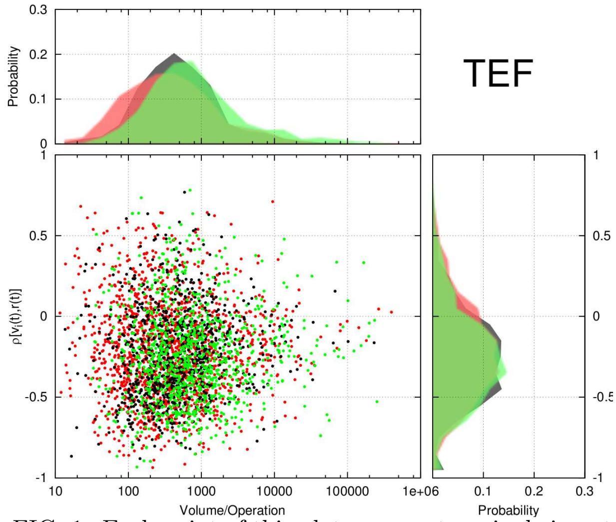

FIG. 1. Each point of this plot represents a single investor. The y component represents the correlation $\rho^{i,a}_{S-r}$ using time windows of $L = 1$days, while the x component represents total volume traded divided by the number of operations done. The color represents the "efficiency" (average income per operation); green for huge positive incomes, red for huge negative incomes and black for incomes in mid zone.

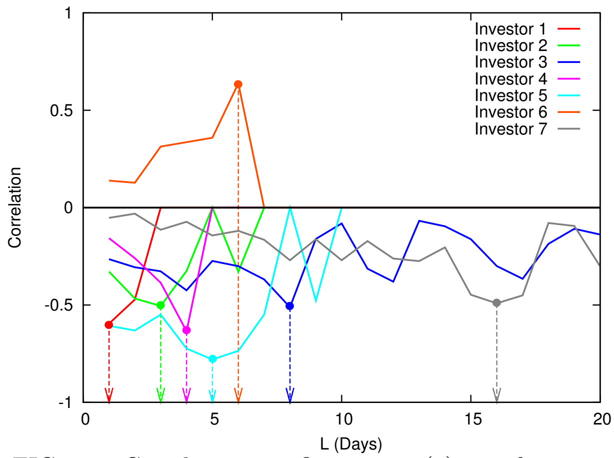

FIG. 2. Correlation coeficient in (1) in function of time-window length L for different invesors, each one represented by colored solid line. Dashed arrows point to the value of L which maximises the absolute value of the correlation for each investor and is expected to be the time-window length which the investor work with.

* mgutierrez@ffn.ub.es

[1] F. Lillo, E. Moro, G. Vaglica and R.N. Mantegna, *New Journal of Physics* **10**, 043019 (2008).

[2] C. Chiarella, G. Iori and J. Perelló, *Journal of Economic Dynamics and Control* **33**, 525-537 (2009).