

Stock Prices and Traded Quantities: Evidence from Ultra High Frequency Data ⁽¹⁾

Rocco Mosconi

Politecnico di Milano, e-mail: rocco.mosconi@polimi.it

Abstract: This paper provides some empirical evidence on the role of demand and supply mechanisms in the stock market, using ultra high frequency data on trades and the limit order book for two stocks traded on Borsa Italiana in February 2005. We conclude with some suggestion for modelling.

Keywords: Analysis of high frequency financial data, stock prices, order flow.

1. Introduction

The literature on efficient markets exclude that demand and supply dynamics play a role in setting the price of financial assets, which would be determined by news (Sholes, 1972). This is in contrast with the commonsense view that price falls are determined by pressure of sales. We analyze the issue using data on two stocks traded on Borsa Italiana in February 2005. We analyze the relationship among returns and the order flow (the difference between buyer initiated and seller initiated trades), trying to find regularities to be used to create an appropriate dynamic model. A similar analysis is carried on for exchange rates in Berger et. al. (2008), and for NYSE in Chordia et. al. (2007).

2. The data

In this study we analyze 2 stocks traded on Borsa Italiana (Mondadori and STM), from 1/2/2005 to 28/2/2005 (20 days, from 9:05 to 17:25, 10,000 minutes). The number of trades is 108499 for STM and 11902 for Mondadori. We have data on: $p_{t_i}^*$ = price of the i-th trade (occurring at time t_i), $v_{t_i}^*$ = value (€) of the i-th trade, L_{t_i} = best bid price at time t_i , U_{t_i} = best ask price at time t_i . Based on this, we can classify each trade as buy (buyer-initiated) or sell (seller-initiated) using the quote method (Lee-Ready, 1991):

$$d_{t_i}^B = 1 \text{ if } d_{t_i}^B > \frac{L_{t_i} + U_{t_i}}{2} \quad ; \quad d_{t_i}^S = 1 \text{ if } d_{t_i}^B < \frac{L_{t_i} + U_{t_i}}{2}$$

the trades whose price is exactly at the midpoint are not classified and excluded. Equally spaced data are then obtained by:

$$b_t = \sum_{t-1 < t_i < t} v_{t_i}^* d_{t_i}^B \quad ; \quad s_t = \sum_{t-1 < t_i < t} v_{t_i}^* d_{t_i}^S$$

We have considered different time aggregations, from 5" to 3 hours. Based on b_t and s_t we obtain: the order flow $o_t = b_t - s_t$ which measures the imbalance between buyer

⁽¹⁾ The present paper is financially supported by the PRIN 2006-13-1140 programme "Econometric analysis of interdependence, stabilization and contagion in real and financial markets".

initiated and seller initiated trades in the t -th time period; the volume $v_t = b_t + s_t$ which sums up all trades, irrespective of who is the initiator; price is measured by $p_t = \frac{L_t + U_t}{2}$, i.e. the last midpoint in the t -th period: notice that this measure might give a non zero return $r_t = \Delta \ln(p_t)$ even in the absence of trades.

3. Stylized facts

STYLIZED FACT 1: LOG PRICE AND CUMULATED ORDER FLOW ARE NOT COINTEGRATED: Despite the efficient markets theory, Figure 1 hints at some relationship among log price and cumulated order flow. However, both series seem non stationary, and spurious regression may be suspected. In fact, appropriate tests reject cointegration for both stocks and at any sampling frequency.

Figure 1: $\ln(p_t)$ and O_t for STM (correlation 0.75) and Mondadori (correlation 0.28)

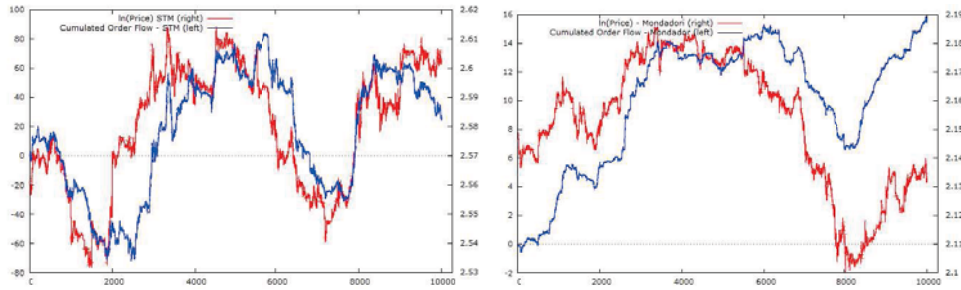
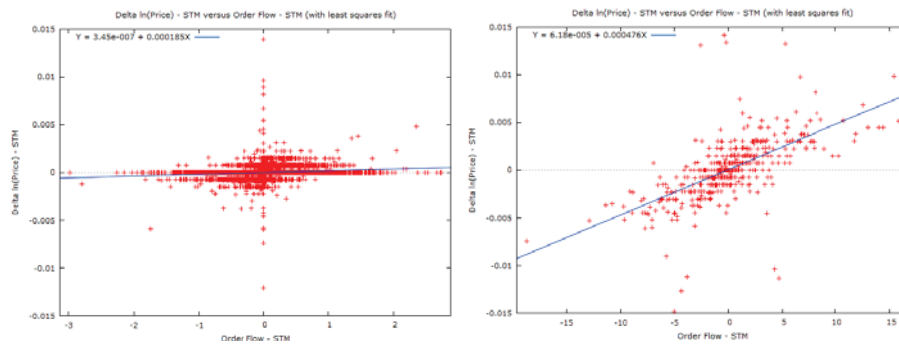


Table 1: p -values for cointegration tests (H_0 : cointegration rank = 0).

Stock	Δt	#lags	#obs	Trace test	Engle-Granger
STM	45"	60	13333	0.3576	0.8813
STM	60'45"	12	164	0.4650	0.6708
Mondadori	45"	60	13333	0.3741	0.0618
Mondadori	60'45"	12	164	0.1153	0.0918

Figure 2: STM: Scatter plot of r_t vs. o_t . $\Delta t=5''$ (left) and $\Delta t=20'$ (right)



STYLIZED FACT 2: RETURNS AND ORDER FLOW ARE POSITIVELY CORRELATED, AND THE CORRELATION INCREASES WITH TIME AGGREGATION. The absence of cointegration suggests to carry on the analysis in differences. Figure 2 shows scatter plots of r_t vs. o_t

for STM (Mondadori is quite similar). When $\Delta t=5''$ (left), we notice a probability mass on zero for both variables, and many outliers, mainly corresponding to large changes in price associated to zero order flow; the Kurtosis is 37.38 in o_t and 366.3 in r_t ; the correlation is significant, but quite low (0.09). When $\Delta t=20'$ (right), the evidence of non normality is less apparent: the Kurtosis drops to 2.83 for o_t and 5.41 for r_t ; the correlation grows to 0.63. The increase in correlation due to time aggregation suggest that the relationship is dynamic. Figure 3 shows the cross correlations for STM. For $\Delta t=5''$ (left), there is no symmetry: o_{t-h} is highly correlated with r_t for small h (up to 5), with correlation coefficients in the order of 0.15, but then the correlation vanishes; conversely r_{t-h} has a small, positive and very persistent correlations with o_t : out of the first 100 correlation coefficients, 29 are significant at the 10%, 67 are positive, with a maximum at 0.018. When $\Delta t=135''$ (right) the correlations between o_{t-h} and r_t are negligible, whereas the first few correlations between r_{t-h} and o_t remains the same.

Figure 3: STM: Cross Correlogram of r_t vs. o_{t-h} . $\Delta t=5''$ (left) and $\Delta t=135''$ (right)

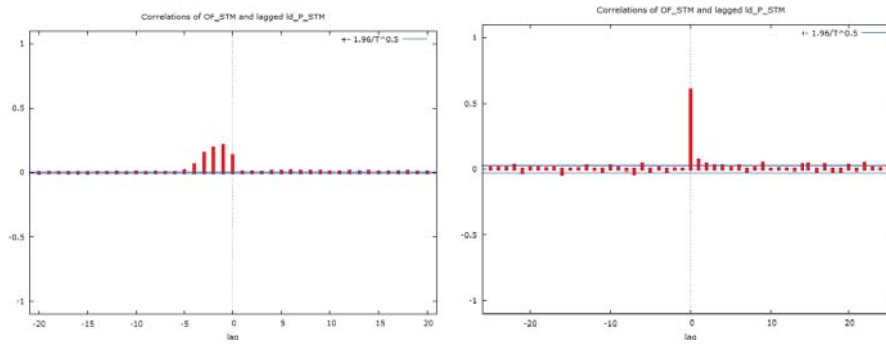


Figure 4: Correlation between order flow and returns based on $\Delta t=135''$ for STM and Mondadori. 40 overlapping windows of 400 observations - February 2005.

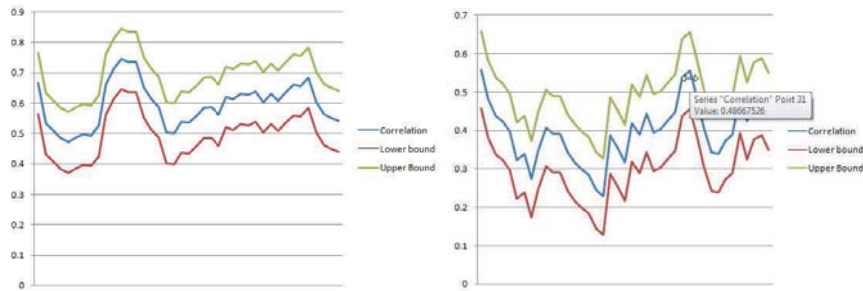
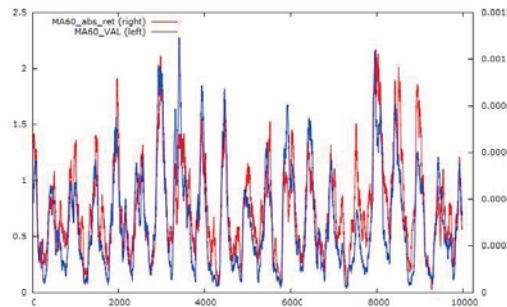


Figure 5: STM, Common seasonal pattern in $|r_t|$ and v_t . $\Delta t=60'$, MA of 60 obs.



STYLIZED FACT 3: THE CORRELATION BETWEEN RETURNS AND ORDER FLOW VARIES OVER TIME. Figure 4 show the correlation based on $\Delta t=135''$ for STM and Mondadori, with 95% bounds, in February 2005; correlations are computed in 40 overlapping windows of 400 observations [1;400],[101;500],...,[4000,4400]. The dynamic behaviour of the correlations seem have some similarities in the two stocks, but there are also specificities. Since correlation might reflect information asymmetries, a local increase in the correlation which is "stock specific" might indicate that informed traders are disclosing their information about the company through their activity on the market. Conversely, an increase in the correlation which is common to all stocks might reflect irrational behaviour like euphoria or panic.

STYLIZED FACT 4: VOLATILITY IS POSITIVELY CORRELATED WITH VOLUME. Coherently with the literature, evidence of a positive correlation between volatility (measured by $|r_t|$ or r_t^2) and traded volumes (v_t) is found in our data at any frequency. Also in this case, the correlation seems to increase with time aggregation. Figure 5 shows that the correlation might be to some extent due to the common seasonal pattern of volumes and volatility (both variables are lower in the central hours of the trading day).

4. Some preliminary ideas for modeling

The stylized facts illustrated above have implications for modeling. At high frequency ($\Delta t \leq 2'$) it seems possible to set up a forecasting model, since it is possible to recognize significant (although small) auto and cross correlations. However, this models should account for non normality (with "extra-heavy" tails and lots of zeros). Conversely, at lower frequency ($\Delta t \geq 20'$), most of the relationship between returns and order flow is simultaneous, with less striking non normality (possibly accommodated by modeling the volatility as a function of v_t). At this frequency, it might be interesting to model several stocks in a multivariate framework, using unobserved components and the Kalman Filter to investigate on the simultaneity across stocks of an increase or decrease of the correlation between order flow and return, which might be interpreted as related to irrationality of the market (euphoria or panic). Conversely, changes in the correlation which are "stock specific" might be related to informational asymmetries about that stock, cleared up through trading activity of informed traders.

References

- Berger D. W., Chaboud A. P., Chernenko S. V., Howorka E., Wright, J. H. (2008) Order Flow and Exchange Rate Dynamics in Electronic Brokerage System Data, *Journal of International Economics*, **75**, 93–109
- Chordia T., Roll R., Subrahmanyam A. (2007) Liquidity and Market Efficiency, Available at SSRN: <http://ssrn.com/abstract=794264>
- Lee C., Ready M. (1991) Inferring Trade Direction from Intraday Data, *The Journal of Finance*, **46**, 2, 733-746
- Sholes M.S. (1972) The Market for Securities: Substitution versus Price Pressure and the Effect of Information on Share Price, *Journal of Business*, **45**, 179-211.