# Microscopic models of Financial Markets

Nicolas MEILLER, Alexandre ELKRIEF, Maxime GENIN, Thomas AYOUL

June 6, 2016

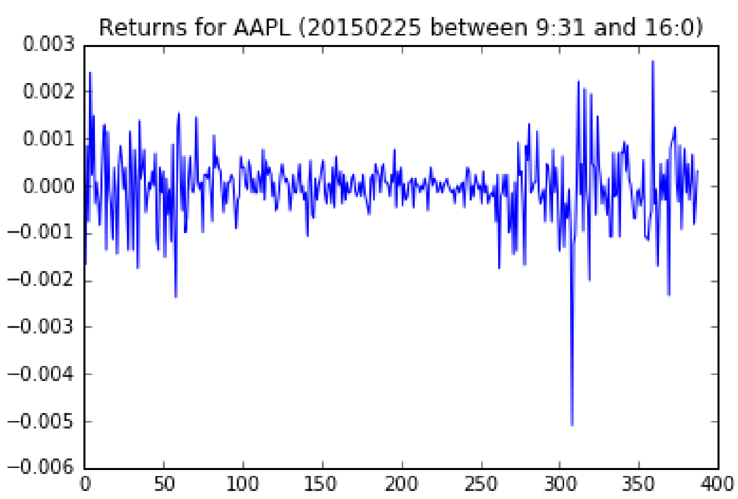
## Abstract

Most famous models in finance assume that the stock returns are independant and gaussian. In practice, at every time scale, we can observe that it is not the case : the returns are fat-tailed and volatility clustering proves their auto-correlation. Lux and Marchesi proposed a model that can reproduce these stylized facts. It divides the traders in two categories : the fundamentalists and the chartists. The latter can be either optimistic or pessimistic. Every trader can change category depending on the comparison between his past performance and another trader's when they meet with some probability. The price process is then determined by the number of chartists (optimistic and pessimistic) and fundamentalists who drive the supply and demand for a given stock. Our goal is to fit the parameters to intraday financial data in order to predict these price movements.

## 1 Overview of the problem

### Intraday stylized fact

Lux-Marchesi work focuses on daily data but they did not work on the estimation on their parameters to realistic data. With the development of algorithmic trading and the possibility to work on high-frequency data, we decided to focus to intraday data. On these financial data, we can observe the common stylized facts such fat tails. For instance, we can observe volatility clustering on the stock price of Apple. The volatility is high just after the opening, quite small at the middle of the day and increases before the close, as we can see in the following figure. As a result, it is relevant to try to fit Lux-Marchesi model on intraday data.



### The model

The total number of traders will be noted $N$. We make the assumption that $N$ is constant on any given day. The number of chartists and fundamentalists will be respectively $n_c$ and $n_f$. The optimistic traders are $n_+$ and the pessimistics $n_-$. We then have : $n_+ + n_- = n_c$ and $n_c + n_f = N$. As the traders may change their view depending on the the majority view, we define the opinion index $x = (n_+ - n_-)/n_c$.

Among chartists, the transition probabilities between optimistics and pessimistics traders are :

$$\begin{cases} \pi_{+-} = \mu_1 \frac{n_c}{N} \exp(U_1) \\ \pi_{-+} = \mu_1 \frac{n_c}{N} \exp(-U_1) \end{cases}$$

with $U_1 = \alpha_1 x + \alpha_2 \dot{p}$ where $\dot{p}$ is the time derivative of the price. Please note we changed a little bit here the definition of $U_1$ in order to make our task easier.

The probabilities to go from chartists to fundamentalists are:

$$\begin{cases} \pi_{+f} = \mu_2 \frac{n_+}{N} \exp\left[\alpha_3 * (\frac{r+\dot{p}/\mu_2}{p} - R - |\frac{p_f - p}{p})\right] \\ \pi_{-f} = \mu_2 \frac{n_+}{N} \exp\left[\alpha_3 * (-\frac{r+\dot{p}/\mu_2}{p} + R - |\frac{p_f - p}{p})\right] \\ \pi_{f-} = \mu_2 \frac{n_+}{N} \exp\left[-\alpha_3 * (-\frac{r+\dot{p}/\mu_2}{p} + R - |\frac{p_f - p}{p})\right] \\ \pi_{f+} = \mu_2 \frac{n_+}{N} \exp\left[-\alpha_3 * (\frac{r+\dot{p}/\mu_2}{p} - R - |\frac{p_f - p}{p})\right] \end{cases}$$

From there, we can determine the excess demand from chartists and fundamentalists. The price process is then determined as follows:

$$\begin{cases} ED_c = (n_+ - n_-) * t_c \\ ED_f = n_f \gamma (p_f - p) \\ ED_{tot} = ED_c + ED_f \end{cases}$$

Elementary price price movement:

$$\begin{cases} \pi_{\uparrow p} = \max(0, \beta(ED + \mu) \\ \pi_{\downarrow p} = \min(0, \beta(ED + \mu) \end{cases}$$

### Our assumptions

We will assume that the numbers of optimistics and pessimistics traders are linked respectively with the bid and the ask in the order book. We will treat as fundamentalists the orders that are in the order book before the opening. The intuition there is that they are holding longer term positions.

## 2 Parameters estimation

Our first goal is to estimate the transition matrix, that we will note :

$$A_t = \begin{pmatrix} \pi_{++} & \pi_{+-} & \pi_{+f} \\ \pi_{-+} & \pi_{--} & \pi_{-f} \\ \pi_{f+} & \pi_{f-} & \pi_{ff} \end{pmatrix} \tag{1}$$

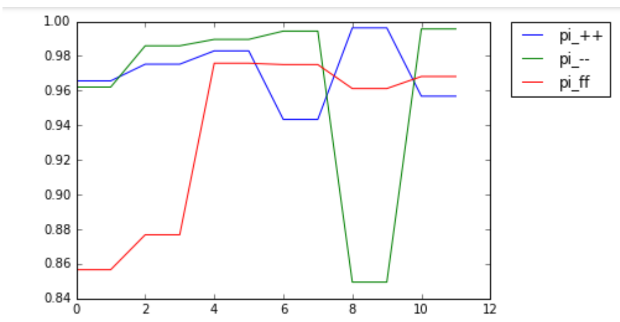
It must satisfy the dynamics $n_{t+1} = A_t n_t$ where $n_t = \begin{pmatrix} n_+(t) \\ n_-(t) \\ n_f(t) \end{pmatrix}$

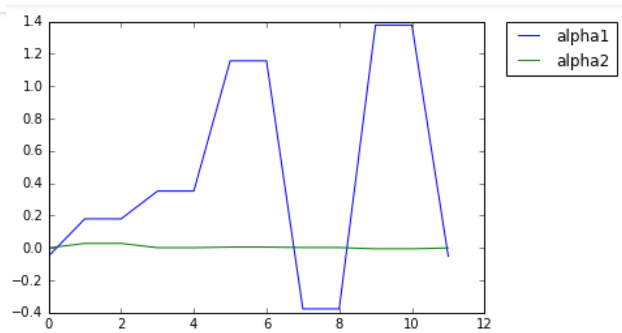
There are 3 equations and 6 unknowns (using the fact that $A_t$ has to be a stochastic matrix) which can not be solved.

## Model 1: Linear regression

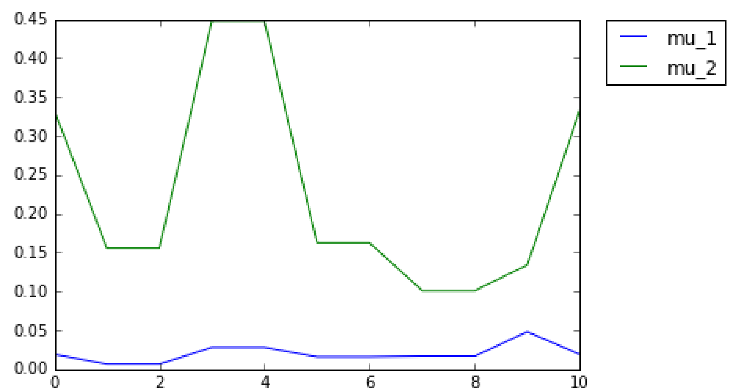
Our first model assumes that $A$ is constant on a given window. This approach is obviously wrong given the particular forms we gave to our probabilities (they depend on $x$ and $\dot{p}$) but is justifiable considering that empirically, the transition probabilities change very slowly relative to n. We therefore consider it constant and attempt to fit a linear regression.

Given the parameters described above, we find ourselves with 3 equations adn 6 unknowns. We therefore use the fact that $N = n_+ + n_- + n_f$ to consider the sub-2x2 matrices of the transition matrix A. We run the regression on each of those and average results. This method proved to be more robust than a direct brute force linear regression on all parameters.




We made the assumption that $\pi_{++}$, $\pi_{--}$,$\pi_{ff}$ were constant but the figure above shows that although they remain in a relatively tight range, they are very volatile. The figure on the right, shows that $\alpha_2$ is practically null, which implies that the transition probabilities depend very little on $\dot{p}$.




The figure above representing $\mu_1$ and $\mu_2$ shows that $\mu_2$ is very difficult to estimate and use for the purpose of predictions because of its volatility. On the contrary, it is is very interesting to see that $\mu_1$ is practically constant ($= 0.03$) and therefore be used with much precision. Similarly, we can estimate $\alpha 3$ which determines the transition probabilities with the fundamentalists and $\beta$ and $\gamma$, which will later be used to compute the excess demand.

### Trading strategy

Our trading strategy is built with the following parameters:

- Starting time
- Regression window (30 min)

- Trading window (10 min)
- Total trading time (200 min)

For each time step t, we regress on the prior 30 mns to estimate all parameters. We then keep those parameters constant during the trading window (10 min). We use the parameters to compute the number of each type of traders and then derive the excess demand giving the price movement for the next time step. According this price movement we great a signal to to enter and exit positions.

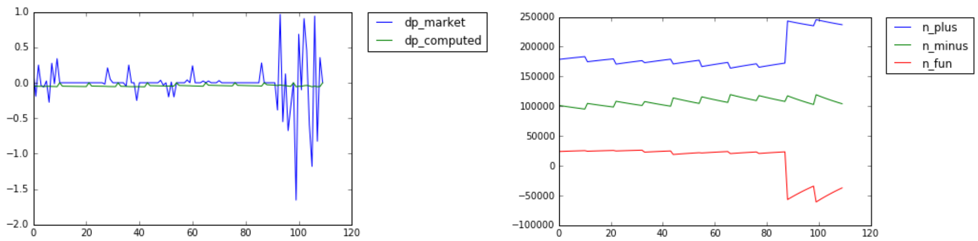



Figure 2: Apple

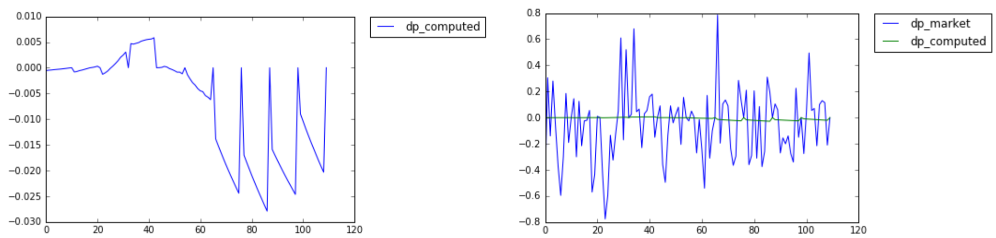



Figure 3: Google

**Trading results**

For Apple we find a relatively constant and therefore well estimated n. We can howver observe jumps between different days. We also have the issue that the market's $\dot{p}$ is more volatile than the $\dot{p}$ we compute. Similary, for Google, the market's $\dot{p}$ is very volatile. Therefore, despite having a positive $\dot{p}$-estimate, the model doesn't capture enough of the realised swings in price. For both stocks, our strategy yields a positive PnL but the lack of robustness, of the signals leads us to believe that this performance is not very sustainable.

## Model 2 : Least-squares estimation

To get a more precise estimation we can use the particular forms of our model. For instance we can see that:

$$\pi_{+-}\pi_{-+} = \mu_1^2 \frac{n_c^2}{N^2}$$

We have also :

$$\begin{aligned} \pi_{+f}\pi_{f+} &= \mu_2^2 \frac{n_+ n_f}{N^2} \\ \pi_{-f}\pi_{f-} &= \mu_2^2 \frac{n_- n_f}{N^2} \end{aligned} \tag{2}$$

Assuming we know $\mu_1$ and $\mu_2$, we have left 6 equations and 6 unknowns which seems to be easier to solve despite the non linearities.

We will solve the equation $n_{t+1} = A_t n_t$ for each time $t$ by minimizing the norm $||n_{t+1} - A_t n_t||^2$. This problem has no reason to be convex so we will add a regularization term to help the minimization. This term will ensure the continuity through time of the probabilities. We will thus be minimizing :

$$||n_{t+1} - A_t n_t||^2 + \lambda ||A_t - \tilde{A}||$$

where $\tilde{A}$ is a matrix we want not to deviate too far from, like for instance $A_{t-1}$.

### Transition probabilities derivatives estimation

The resulting estimated matrix $\hat{A}$ is a function of the intensity parameters $\mu_1$ and $\mu_2$. Let note $A^{(t)}_{\mu_1,\mu_2}$ the resulting matrix. According to our model, the dependence of $A^{(t)}_{\mu_1,\mu_2}$ with respect to $\mu_1$ and $\mu_2$ has to be linear. Recall that, for instance, $\pi_{+-} = \mu_1 \frac{n_c}{N} \exp(U_1)$, that leads to $\frac{\partial \pi_{+-}}{\partial \mu_1} = \frac{n_c}{N} \exp(U_1)$.

Therefore, varying the parameters $\mu_1$ and $\mu_2$ and processing a new estimate matrix $\hat{A}^{(t)}_{\mu_1,\mu_2}$ gives a good estimate of the the dependence of $A$ with those parameters. It is now possible to estimate the derivative of each entries of the matrix with respect to $\mu_1$ and $\mu_2$ and therefore get rid of these unknowns.

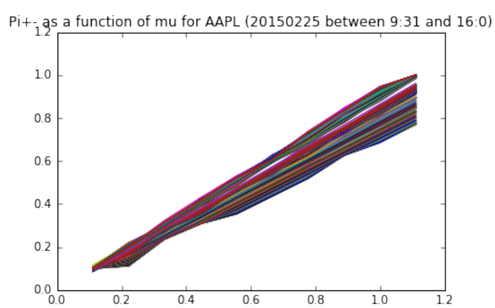
Let thus note $\hat{\pi}^{(t)}_{+-}(\mu_1, \mu_2)$ the value we estimated, i.e. the corresponding entry in the matrix $A^{(t)}_{\mu_1,\mu_2}$. By inverting the previous formula, we get our estimation of $U_1$ :

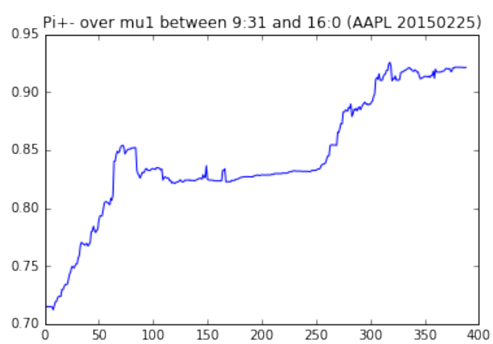
$$\hat{U_1}^{(t)} = \log \frac{N}{n_c^{(t)}} \frac{\partial \hat{\pi}^{(t)}_{+-}}{\partial \mu_1}$$

### Numerical results

In practice, we can observe that linearity with a big precision which is good. It shows that the way we solve for the matrix A was consistent with our model. Indeed, here is the plot of $\hat{\pi}^{(t)}_{+-}(\mu_1, \mu_2)$ as a function of $\mu_1$, for several values of $t$.

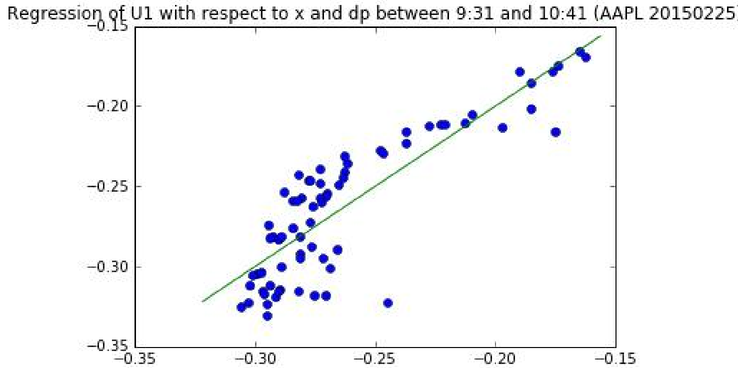
Moreover, we can see that the slopes does not change a lot during a day and that they are quite continuous. We also remark that the periods where it changes are correlated with high volatility (just after the open and just before the close).




Finally, we can do the regression of $\hat{U_1}^{(t)}$ with respect to $x^{(t)}$ and $\dot{p}(t)$. Here is the plot of the estimated versus the true values. The $R^2$ score is around 0.85 which means that the regression is quite accurate.

However, we find that the coefficients $\alpha_1$ and $\alpha_2$ have completely different impacts on the movement of $\hat{U_1}^{(t)}$ and of $\pi_{+-}$ as a result. Indeed, the regression attributes much more weight to $x^{(t)}$ than to $\dot{p}(t)$ as $\alpha_1$ is in the order of magnitude of 1 and $\alpha_2$ is in the order of magnitude of 0.01. After taking a closer look at the data, we noticed that $\dot{p}(t)$ varies much faster than $\pi_{+-}$ and $x^{(t)}$. We therefore tried to smooth $\dot{p}(t)$ by taking the moving average of prices. After running the regression again, this increased the $\alpha_2$ by one order




of magnitude to approximately 0.1 which is still extremely small. This would suggest that $\dot{p}(t)$ has very little impact on the short term transition probabilities $\pi_{+-}$ and $\pi_{-+}$ relative to $x^{(t)}$. We also noticed that these seem to be much more correlated to the variation of $p(t)$, or the volatility, than that of $\dot{p}(t)$.

# 3 Hypothesis test

In this section, we will focus on testing the various hypothesis we have made.

### Frequency of revaluation of opinion $\mu_1$

One of the strongest hypothesis we have done in this model is supposing that the frequency of revaluation of opinion $\mu_1$ is a constant over time. Therefore, in this case, we can suppose that over a short period of time, the transition probabilities will only depend on the number of chartists (optimists and pessimists) and the number of fundamentalists. In order to check this hypothesis, we have reverse engineered $\mu_1$ based on forward looking data. In order to simplify the calculus, we focus only on chartists ( this hypothesis is not too strong as $\mu_1$ should only depend on $n_+$ and $n_-$ cf initial equations). Using transition probabilities, we have the following equations :
Let's write $d = \frac{n_c}{N}\exp(U_1)$ which we have estimated previously.

$$\begin{cases} \pi_{+-} = \mu_1 d \\ \pi_{-+} = \mu_1 (\frac{n_c}{N})^2 (\frac{1}{d}) \\ n_-(t+1) = \pi_{-+} n_+(t) + (1 - \pi_{+-}) n_-(t) \end{cases}$$

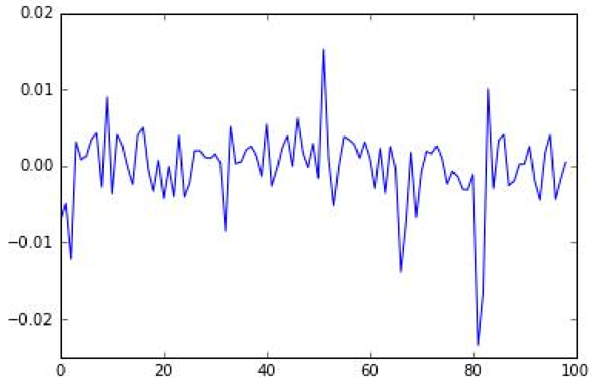
Which gives us the following equation for $\mu_1(t)$

$$\mu_1(t) = \frac{n_-(t+1) - n_-(t)}{(\frac{n_c}{N})^2 (\frac{1}{d}) n_+(t) - d n_-(t)}$$

The following graph represents the result obtained

The hypothesis we have done on $\mu_1$ being a constant is clearly wrong. It can be explained by the fact that our model considers the whole number of traders $N$ to be a constant, i.e that the traders are living in a closed environment. However in reality, the number of traders varies with new traders on the market and traders leaving the market every minutes.
Another explanation is that the quantity

$$\frac{n_-(t+1) - n_-(t)}{n_+(t+1) - n_+(t)}$$




Back engineering of $\mu_1$ as a function of $t$

is not constant (should be equal to 1 under the above hypothesis). Therefore, in our model, it is not possible to estimate $\mu_1$ independently of the number of fundamentalists.

# Conclusion

In this paper, we successfully estimated parameters based on Lux-Marchesi's equations. We verified that the transition probabilities ($\pi_{+-}$ and $\pi_{-+}$) are linear in $\mu_1$ and $\mu_2$ as postulated by the paper. However, we empirically realised that these transition probabilities are very volatile, particularly on an intraday scale. Our results also show that the transition probabilities are not very dependent on *dotp*. As a result, this model appears to be too theoretic and complicated to fit to real life price processes, specifically on a short term time scale. We also found ourselves needing to make too many assumptions on the data. The inability to track individual orders put by individual traders results in imprecisions. For example, if 10 traders enter the market and 10 traders exit, our model will not factor in anything when in reality the market has changed. We did notice some interesting behaviours in that the variation of the slope of the function of the transition probabilities with respect to $\mu_1$ seems to be correlated with volatility. This relationship could be investigated in future research.

# Bibliography


Lux-Marchesi (2000). *Volatility clustering in financial markets : A microsimulation of interacting agents*