PHENOMENOLOGY AND MODELLING OF FINANCIAL MARKETS

Petite Classe n. 2

Stylized facts in financial time-series

ECO 586 - J. Moran
M. Benzaquen

January 13, 2020

Introduction

We believe strongly that a good knowledge of empirical facts is paramount to, and should in fact precede, establishing good theory. In this session you will make acquaintance with financial data from the real world and study the stylized facts of the log-returns of the S&P 500 Index downloaded from Yahoo! Finance. A critical discussion of models "à la Black-Scholes" will follow.

Part 1 : Download and understand the data

1. Use the code you are given to download the S&P 500 Data from Yahoo! Finance using the `pandas_datareader.data` module. The ticker `^GSPC` is the ticker symbol for the S&P 500, but the code you are given can be used to look up any ticker.
2. "High" and "Low" represent the highest and lowest prices of the stock during the day. Look up the meaning of "Adj Close" in Yahoo! Finance. Why should we use this instead of "Close"? From now on, p_t refers to this variable for the day t.
3. Plot the "Adj Close" variable. Comments?

Part 2 : Log-return statistics

1. Define a column `df[''a_returns'']` for the daily additive returns, defined as $d_t^1 = p_t - p_{t-1}$, and do the same for a column `df[''l_returns_1'']` containing the log returns, defined as $r_t^1 = \log\left(\frac{p_t}{p_{t-1}}\right)$. Plot both returns through time. What do you notice? Which variable makes more sense to you and why?
2. Write a function to compute the variogram of the log-price, $V(\tau) = \langle\left(\log \frac{p_t}{p_{t-\tau}}\right)^2\rangle$. Plot $V(\tau)/V(1)$ for $0 \leq \tau < 500$ in log-log scale. What slope do you see?
3. Define the return at scale Δ (in days) as $r_t^\Delta = \log\left(\frac{p_t}{p_{t-\Delta}}\right)$ (for simplicity, $r_t^1 := r_t$). Noting that $r_t^\Delta = \sum_{i=1}^{\Delta} r_{t-i}^\Delta$, what simple model would explain the previous plot?
4. Write a function that computes the series for r_t^Δ .Compute the mean and std. of r_t^1, as well as that of $r_t^{250} = \log\left(\frac{p_t}{p_{t-250}}\right)$, the yearly returns. Comments?
   In mathematical finance, people often resort to the following model, "à la Black-Scholes " to study price dynamics:
   $$p_t = (1+\mu)p_{t-1} + \sigma\eta_t p_{t-1} \tag{1}$$
   where η_t is a gaussian random variable with $\langle\eta_t\eta_{t'}\rangle = \delta(t-t')$ and $\langle\eta_t\rangle = 0$ (white noise), and with $\sigma, \mu \ll 1$.
5. Interpret the terms μ and σ. How does one write r_t within this framework?
6. Use `np.random.randn` to draw a random gaussian numbers $\tilde{r}_t$ of the same length as r_t and multiply them by your estimation of σ to have the correct variance. Plot $\tilde{r}_t$. Comments?

We define the complementary cumulative density function (called also the **survival function** and abbreviated as sf) the returns as

$$P_{>}(x) := \int_x^{\infty} \mathrm{d}r \rho(r) \tag{2}$$

where ρ is the density function associated to the returns. It is used to see the nature of the "tails" of a distribution.

7. Plot the survival functions of the daily returns r_t on the right (use r_t on positive-valued bins) and left (-r_t on positive-valued bins) tails. Hint: look up how to do this on the **Numpy Cheatsheet**.
8. After importing the correct module with `import scipy.stats`, use the functions

```
y_normal = scipy.stats.norm.sf(x = x, scale = sigma)
y_student = scipy.stats.t.sf(x = x, scale = sigma, df = nu)
```

   to compute the survival functions of a normal distribution of std. σ and of a Student-t distribution of std. σ and tail parameter ν at x (can be a `numpy` array). Estimate σ from data, but play with ν. Which parameter fits best the right and left tails (try $\nu \in [2, 6]$)?
9. Do the same as the two previous functions, but for $\Delta \in \{30, 60, 90\}$. Qualitatively, what is happening? (Remember to adjust with the corresponding value of σ_Δ).
10. What can you comment about the model proposed in Eq. (1)?

Part 3 : Correlations

1. Before beginning, center the daily returns by removing the mean, i.e. $r_t \leftarrow r_t - \langle r_t \rangle$. This is standard when working with correlations.
2. Write a function to compute the correlation function $C_{r,r}(\tau) = \langle r_t r_{t+\tau} \rangle / \sqrt{\langle r_t^2 \rangle}$. Compute it for $-200 \leq \tau < 200$. What do you expect to see? Do you expect to see the same for $C_{|r|,|r|}(\tau) = \frac{\langle |r_t||r_{t+\tau}| \rangle}{\sqrt{\langle |r_t|^2 \rangle}}$ or $C_{r^2,r^2}(\tau) = \langle r_{t+\tau}^2 r_t^2 \rangle / \sqrt{\langle r_t^4 \rangle}$. (Hint: define a function `C(x,y,tau)` that computes the correlation $C_{x,y}(\tau) = \langle x_t y_{t+\tau} \rangle / \sqrt{\langle x_t^2 \rangle \langle y_t^2 \rangle}$).
3. Compute and plot $C_{|r|,|r|}(\tau)$ and $C_{r^2,r^2}(\tau)$ for $1 \leq \tau < 1000$. Choose a log-log scale when plotting. Comments?
4. What does this imply for the proper modeling of return dynamics?

Part 4 : Volatility and conclusion

1. Define now the volatility as $\sigma_t^2 = r_t^2$. Use a qualitative argument to predict the behaviour of $C_{\sigma^2,r} = \langle \sigma_t^2 r_{t+\tau} \rangle$ (use your intuition), then compute it and see for yourself by choosing $-100 \leq \tau < 300$. How can you interpret this? This is called the leverage effect.
2. In light of what you have seen, what arguments can you find against the modelling proposed in Eq. (1)?
3. You can re-execute all the cells by picking a different ticker. Try for instance the Euro Stoxx, with the ticker `^STOXX50E` or the CAC40 with ticker `^FCHI`. What do you notice?