Econophysics MU5PYC31

Financial Data Analysis

Electronic format (pdf) to be returned by Friday 19 February 2021
can be done in pairs (two students for a report)

The main goal of this study is to statistically analyze real financial data and to compare the obtained results with the predictions of the standard model of asset pricing. Here, we will examine the daily time series of the S&P500 index from 1950 to 2021 (a total of 17 880 trading days). The time horizon is then $\Delta t = 1$ day $= 1/252$ year (there are about 252 trading days per year).

The evaluation of this short report will take into account the quality of the presentation (appearance, figures, spelling, rigor…) and the interest of comments and conclusions draw from this study.

First, **download the file** "SP500_1950_2021.txt" from my web page at LPTMC:

`https://www.lptmc.jussieu.fr/files/SP500_1950_2021.txt`

The data file consists of 7 columns:

1. Trading day number ($1 \leq t \leq 17\,880$).
2. Date (from January 3rd 1950 to January 22nd 2021).
3. Open price: the price at which the asset is first traded at the opening of the market.
4. Today's high: the highest price at which the asset is traded during a trading day.
5. Today's low: the lowest price at which the asset is traded during a trading day.
6. Closing price: the final price at which the asset is traded at the closure of the market. **This is the price S_t we will use in this study**.
7. Volume of transactions: the number of assets traded during a trading day.

1 S&P500 index time series

1.1 Daily changes from 1950 to 2021

1.1 Plot the closing price S_t as a function of the trading day number t for the whole period (this is Figure {**Fig1**}).

1.2 The log return being defined by $g_t = \delta S_t \equiv \ln \frac{S_{t+1}}{S_t}$, calculate its mean $\langle g \rangle$ and standard deviation σ_g.

1.3 Plot the log return (in σ_g unit) as a function of the trading day number t {**Fig2**}. Compare qualitatively with a Gaussian random walk.

1.4 According to the standard model of asset pricing, the log return should be normally distributed. Plot, in the same Figure, i) the probability distribution function $P(g_t)$ of the log return as a function of g_t/σ_g and ii) the normal distribution with mean $\langle g \rangle$ and deviation σ_g found above. Use first a linear y-axis ({**Fig3**}) and then a logarithmic y-axis ({**Fig4**}).

1.5 Calculate the excess kurtosis κ of $P(g_t)$ and discuss its value.

1.2 Monthly changes from 1950 to 2021

Let's now consider the monthly changes of the index. The time horizon is then $\Delta t = 1$ month (1/12 year). In this section we are going to reexamine the previous study by taking only one daily index value out of 22 (there are about 22 trading days in a month). The new times series has now 812 data points.

1.6 Calculate the mean $\langle g' \rangle$ and standard deviation $\sigma_{g'}$ of the monthly log return.

1.7 Plot the monthly log return g'_t (in $\sigma_{g'}$ unit) as a function of the *trading day number t* ({**Fig5**}). Compare qualitatively with a Gaussian random walk and with the daily log return shown in {**Fig2**}.

1.8 Plot, in the same Figure, i) the probability distribution function $P'(g'_t)$ of the log return as a function of $g'_t/\sigma_{g'}$ and ii) the normal distribution with mean $\langle g' \rangle$ and deviation $\sigma_{g'}$ found above. Use first a linear y-axis ({**Fig6**}) and then a logarithmic y-axis ({**Fig7**}).

1.9 Calculate the excess kurtosis κ' of $P'(g'_t)$ and discuss its value.

1.10 What conclusions can you draw by comparing the monthly and the daily changes ?

2 Time Correlations

The time correlation function of two variables X_t and Y_t is defined as

$$C_{XY}(t-t') = \frac{\langle X_t Y_{t'} \rangle - \langle X_t \rangle \langle Y_{t'} \rangle}{C_{XY}(0)} \quad . \tag{1}$$

In the following we consider only the *daily changes* of the S&P500 index.

2.1 Autocorrelation of the S&P500 index

2.1 Calculate and plot the autocorrelation function $C_{gg}(t-t')$ of the daily log return g_t ($X_t = g_t$ and $Y_{t'} = g_{t'}$) for $t - t' < 200$ days ({**Fig8**}). Comment this result.

2.2 To better understand the index change dynamics let's write

$$g_t = \mathrm{sign}(g_t).|g_t| \quad , \tag{2}$$

where $\mathrm{sign}(g_t) = \pm 1$ is the sign of the daily log return g_t and $|g_t|$ its amplitude. Calculate and plot the autocorrelation functions $C_{\mathrm{sign}(g)\mathrm{sign}(g)}(t-t')$ and $C_{|g||g|}(t-t')$ of the sign and of the amplitude respectively of the daily log return g_t for $t - t' < 200$ days on the same Figure ({**Fig9**}). Discuss the behavior of these two functions. Show that $C_{|g||g|}(t-t')$ can be fitted with a power law, $\sim (t-t')^{-\alpha}$, in this time window. Estimate $\alpha > 0$.

2.3 Calculate and plot the autocorrelation function $C_{g^2g^2}(t-t')$ of the daily log return squared g_t^2 for $t - t' < 200$ days ({**Fig10**}). Estimate the exponent $\beta > 0$ of the power law decay of this autocorrelation function. What conclusion can you draw about the volatility of the index ?

2.2 The leverage effect

Let's define the leverage correlation function by

$$\mathcal{L}(t-t') = \frac{\langle (g_t)^2 g_{t'} \rangle}{\langle (g_t)^2 \rangle^2} \quad , \tag{3}$$

where g_t is the daily log return.

2.4 Calculate and plot $\mathcal{L}(t-t')$ for $t - t' < 0$ as a function of $|t - t'| < 200$ ({**Fig11**}). What can you conclude about the correlation between the past volatility at t and the future return at t' ?

2.5 Calculate and plot $\mathcal{L}(t-t')$ for $t - t' > 0$ as a function of $t - t' < 200$ ({**Fig12**}). Show that the curve can be fitted by $-A\mathrm{e}^{-\frac{(t-t')}{\tau}}$. Estimate $A > 0$ and $\tau > 0$ and draw your own conclusions.

3 Open questions

Feel free to use the data by addressing an explicitly well defined question and by providing the result of your work (with up to two figures).