1. Volatility is the standard deviation of the annual (rate) of return.

   - future volatility is needed for option pricing
   - volatility is not observable in the market
   - it is typically extracted from historic data, e.g., monthly returns.

   Let R_y denote the annual return and $R_1, R_2, \cdots, R_{12}$ denote the monthly returns. The (rate of) return behaves like interest rate.

   $$\frac{S_{12}}{S_0} = (1 + R_y) = (1 + R_1)(1 + R_2) \cdots (1 + R_{12})$$

   This means that

   $$R_y = R_1 + R_2 + \cdots + R_{12}$$

   Assume the monthly return has the same mean $\bar{R}_{\frac{1}{12}}$ and standard deviation $\sigma_{\frac{1}{12}}$ and are uncorrelated. Then

   $$\bar{R}_y = 12\bar{R}_{\frac{1}{12}}$$

   Similarly, using the assumption that $\{R_i\}$ are uncorrelated,

   $$\sigma_y^2 = \mathcal{E}(\left[\sum_{i-1}^{m}(R_i - R_{\frac{1}{12}})\right]^2) = \mathcal{E}(\sum_{i=1}^{12}(R_i - \bar{R}_{\frac{1}{12}})^2) = 12\sigma_{\frac{1}{12}}^2$$

   In other words,

   $$\bar{R}_{\frac{1}{12}} = \frac{1}{12}\bar{R}_y$$

   $$\sigma_{\frac{1}{12}} = \frac{1}{\sqrt{12}}\sigma_y$$

   Let $\delta t > 0$ denote the length of an arbitrarily small sampling period, then the period mean and std are

   $$\bar{R}_{\delta t} = \delta t \bar{R}_y$$

   $$\sigma_{\delta t} = \sqrt{\delta t}\sigma_y$$

   The mean becomes much smaller relative to the std as $\delta t \to 0$!

For example, if $\bar{R}_y = 12\%$ and $\sigma_y = 15\%$. Then $\bar{R}_{\frac{1}{250}} = 0.048\%$ and $\sigma_{\frac{1}{250}} = .95\%$, roughly 20 times the mean. In fact

$$\frac{\sigma_{\delta t}}{\bar{R}_{\delta t}} \to +\infty, \quad \delta t \to 0$$

Assume that $\{S_t\}$ are sampled at interval δt and $\{R_1, R_2, \cdots, R_m\}$ are δt- period returns
The mean of a period δt can be estimated from the historic returns

$$\hat{\bar{R}}_{\delta t} = \frac{1}{m} \sum_{t=1}^{m} R_t.$$

Assuming the sample returns R_t are uncorrelated and have the same mean $\bar{R}_{\delta t}$ and $\sigma_{\delta t}$

$$\mathcal{E}(\hat{\bar{R}}_{\delta t}) = \bar{R}_{\delta t}$$

The likely error is given by the standard deviation of $\hat{\bar{R}}$ which is

$$\text{std}(\hat{\bar{R}}_{\delta t}) = \sqrt{\mathcal{E}((\hat{\bar{R}}_{\delta t} - \bar{R}_{\delta t})^2)} = \frac{1}{\sqrt{m}} \sigma_{\delta t}$$

The likely relative error in mean estimation is

$$\frac{1}{\sqrt{m}} \frac{\sigma_{\delta t}}{\bar{R}_{\delta t}}$$

Consider a typical situation of $\bar{R}_{\frac{1}{12}} = 1\%$ and $\sigma_{\frac{1}{12}} = 4.33\%$, if $m = 12$, $\text{std}(\hat{\bar{R}}_{\delta t}) = \frac{4.33\%}{\sqrt{12}} = 1.25\% > \bar{R}_{\frac{1}{12}}$. To get a good estimate with a relative error of 0.1, $m \approx 156 \times 12$!

Increasing sample frequency does not help because $\frac{\sigma_{\delta t}}{\bar{R}_{\delta t}} \to +\infty, \quad \delta t \to 0$.

The standard deviation can be estimated from

$$\hat{\sigma}_{\delta t}^2 = \frac{1}{(m-1)} \sum_{t=1}^{m} (R_t - \hat{\bar{R}}_{\delta t})^2$$

Use $m - 1$ instead of m because that $\hat{\bar{R}}_{\delta t}$ is used. It can be shown that

$$\mathcal{E}(\hat{\sigma}_{\delta t}^2) = \sigma_{\delta t}^2$$

thus it is an unbiased estimate. The likely error in $\hat{\sigma}_{\delta t}^2$ is

$$\text{std}(\hat{\sigma}_{\delta t}^2) = \frac{\sqrt{2}}{\sqrt{m-1}} \sigma_{\delta t}^2$$

The relative error is reasonable assuming that R_i is uncorrelated and has the same statistics.

Additional Difficulties: serial correlation