

Estimating Expected Returns and Variance

Estimating Expected Returns and Variance

Expected Value and Variance for Log Returns of Different Periods

Consider a log-return over the period t to T and it's relation to a log return of an interval of length Δt , with $(T - t)/\Delta t = m$.

$$\begin{aligned}\ln\left(\frac{S(T)}{S(t)}\right) &= \ln\left(\frac{S(t+m\Delta t)}{S(t)}\right) \\ &= \ln\left(\frac{S(t+m\Delta t)}{S(t+(m-1)\Delta t)} \cdot \frac{S(t+(m-1)\Delta t)}{S(t+(m-2)\Delta t)} \cdots \frac{S(t+2\Delta t)}{S(t+\Delta t)} \frac{S(t+\Delta t)}{S(t)}\right) \\ &= \sum_{i=1}^m \ln\left(\frac{S(t+i\Delta t)}{S(t+(i-1)\Delta t)}\right)\end{aligned}$$

Let $X_T = \ln\left(\frac{S(T)}{S(t)}\right)$, be a random variable with mean \bar{X}_T and variance σ_T^2 .

Let $X_{i\Delta t} = \ln\left(\frac{S(t+i\Delta t)}{S(t+(i-1)\Delta t)}\right)$, be a collection of random variables (indexed by i), that are uncorrelated, and identically distributed, each with mean $\bar{X}_{\Delta t}$ and variance $\sigma_{\Delta t}^2$.

Since

$$X_T = \sum_{i=1}^m X_{i\Delta t},$$

we obtain

$$\text{Var}(X_T) = \text{Var}\left(\sum_{i=1}^m X_{i\Delta t}\right) = \sum_{i=1}^m \text{Var}(X_{i\Delta t}) = m \cdot \sigma_{\Delta t}^2$$

and

$$\text{E}(X_T) = \text{E}\left(\sum_{i=1}^m X_{i\Delta t}\right) = \sum_{i=1}^m \text{E}(X_{i\Delta t}) = m \cdot \bar{X}_{\Delta t}.$$

Such that

$$\sigma_T^2 = m \cdot \sigma_{\Delta t}^2 \quad \text{and} \quad \bar{X}_T = m \cdot \bar{X}_{\Delta t}.$$

Note:

If $T = 1$ and $t = 0$, such that X_1 is the annual return, and $\Delta t = \frac{1}{255}$ is a single trading day, we obtain the classical result that

$$\sigma_{\frac{1}{255}} = \frac{\sigma_1}{\sqrt{255}}.$$

Estimating Expected Returns

Consider that we have N uncorrelated samples of X_T , denoted $X_T^{(i)}$ for $i = 1, \dots, N$. The typical estimator for the sample mean is given by

$$\hat{X}_T = \frac{1}{N} \sum_{i=1}^N X_T^{(i)}.$$

We want to investigate the properties of this *estimator*, for example, how much data would we need to have a "good" estimate?

First, we consider the mean of the estimator

$$\mathbb{E}(\hat{X}_T) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}(X_T^{(i)}) = \frac{1}{N} \cdot N \cdot \bar{X}_T = \bar{X}_T.$$

Thus, this estimator is *unbiased*: its expected value is the true value of the parameter being estimated.

Next, we consider the variance of the estimator

$$\text{Var}(\hat{X}_T) = \text{Var}\left(\frac{1}{N} \sum_{i=1}^N X_T^{(i)}\right) = \frac{1}{N^2} \text{Var}\left(\sum_{i=1}^N X_T^{(i)}\right) = \frac{N}{N^2} \sigma_T^2 = \frac{\sigma_T^2}{N}.$$

On the surface, this result is not surprising: as we increase the number of samples in our data set, the variance of our mean estimator increases, going to zero as our number of samples goes to infinity. Unfortunately, this is not great given the typical values for financial data.

Consider the case where $\bar{X}_1 = 10\%$ and $\sigma_1 = 20\%$, i.e., the expected annual return is 10% and the annual volatility is 20%. The table below shows the standard deviation of the estimator as a percentage of the true value we are trying to estimate.

Expected Annual Return	10%		
Annual Volatility	20%		
Number of Samples	10	100	200
Stdev of Estimator	0.06325	0.02	0.01414
As % of Parameter	63.2456	20	14.1421

We would need *at least* 100 years of independent annual returns to estimate the expected return such that our estimate has a standard deviation of only 20% of the true value.

Perhaps we can improve our estimator for X_T by consider returns over smaller intervals, $X_{\Delta t}$?

Note:

The Bank of New York Mellon Corporation (BNY Mellon) is the longest listed company on the NYSE. It was listed in 1792. That's the equivalent of 229 samples of annual returns (X_1).

Period-Length Effects for Expected Returns

We can attempt to exploit the relationship between X_T and $X_{i\Delta t}$ to improve \hat{X}_T . For example, if we have $N = 100$ samples of annual returns ($X_1^{(i)}$), we may have access to $N \cdot m = 100 \cdot 255$ samples of daily returns ($X_{\frac{1}{255}}^{(i)}$).

We consider a new estimator for the sample mean of X_T , denoted \hat{X}_T^* , using our samples of $X_{i\Delta t}$,

$$\hat{X}_T^* := m \cdot \hat{X}_{\Delta t} = m \frac{1}{Nm} \sum_{i=1}^{Nm} X_{i\Delta t}^{(i)} = \frac{1}{N} \sum_{i=1}^{Nm} X_{i\Delta t}^{(i)}.$$

This estimator is unbiased,

$$\mathbb{E}(\hat{X}_T^*) = \frac{1}{N} \sum_{i=1}^{Nm} \mathbb{E}(X_{i\Delta t}^{(i)}) = \frac{Nm}{N} \bar{X}_{\Delta t} = \bar{X}_T.$$

Unfortunately, it has the same variance as the previous estimator,

$$\text{Var}(\hat{X}_T^*) = \text{Var}\left(\frac{1}{N} \sum_{i=1}^{Nm} X_{i\Delta t}^{(i)}\right) = \frac{1}{N^2} \text{Var}\left(\sum_{i=1}^{Nm} X_{i\Delta t}^{(i)}\right) = \frac{Nm}{N^2} \sigma_{\Delta t}^2 = \frac{\sigma_T^2}{N}.$$

Thus, **our estimate for expected return cannot be improved by subdividing our interval**. This is a direct result of the *additive or compound nature of returns*.

Estimating Variance

Consider the unbiased sample estimator for σ_T^2 ,

$$\hat{\sigma}_T^2 = \frac{1}{N-1} \sum_{i=1}^N (X_T^{(i)} - \hat{X})^2.$$

IF we assume that $X_T \sim \mathcal{N}(\bar{X}_T, \sigma_T^2)$, then the [variance of this estimator](#) is given by

$$\text{Var}(\hat{\sigma}_T^2) = \frac{2\sigma_T^4}{N-1}.$$

This is already an enormous improvement on our estimate for the mean. Consider again the case where $\bar{X}_1 = 10\%$ and $\sigma_1 = 20\%$, i.e., the expected annual return is 10% and the annual volatility is 20%. The table below shows the standard deviation of the estimator as a percentage of the *true value we are trying to estimate*.

Annual Variance	4%		
Annual Volatility	20%		
Number of Samples	10	100	200
Stdev of Estimator	0.00075	0.00023	0.00016
As % of Target Parameter	1.88562	0.56854	0.401

Even with only 10 samples, the standard deviation of our estimate is less than 2% of the true, target value of 4%.

With an additional assumption, we can also further improve this estimate by subdividing our interval.

Period-Length Effects for Variance

We assume that the length Δt log returns are normally distributed as well, that is, $X_{i\Delta t} \sim \mathcal{N}(\bar{X}_{\Delta t}, \sigma_{\Delta t}^2)$. We consider a new estimator for the sample variance of X_T , denoted $\hat{\sigma}_T^{2,*}$, using our samples of $X_{i\Delta t}^{(i)}$,

$$\hat{\sigma}_T^{2,*} := m \cdot \hat{\sigma}_{\Delta t}^2$$

where $\hat{\sigma}_{\Delta t}^2$ is the standard unbiased sample estimator for the variance of $X_{i\Delta t}$. It should immediately be clear that this estimator is unbiased. However, the variance of this new estimator becomes

$$\text{Var}(\hat{\sigma}_T^{2,*}) = m^2 \frac{2\sigma_{\Delta t}^4}{Nm-1} = \frac{2\sigma_T^2}{Nm-1}.$$

Now, the variance of the new estimator is less than the variance of the old estimator, as long as $m > 1$, which occurs whenever $\Delta t < T - t$.

Thus, **our estimate for variance can be improved by subdividing our interval**. In fact,

the more frequently we can sample our return (X_T), the better our estimate for variance.