Boston University Questrom School of Business

MF793 – Fall 2021
Eric Jacquier

Returns Data Frequency and Precision of sample moments

- Precision of sample moments increases with sample size
- We can obtain more observations by sampling more often
- Are moments more precisely estimated if we sample more often within a given calendar time?

.... the mean, the variance?

1 Sample size and frequency vs calendar span of data

$\mathrm{Var}(\bar{R}) = \sigma^2 / T$

$\mathrm{Var}(\hat{\sigma}) = \sigma^2 / 2T$

- Precision comes from sample size T => Longest possible calendar span
- We may not want to use old data if we think they are not relevant
- Can we increase T by increasing sampling frequency,

Weeks vs Days ?
Days vs Months ?

Log-Normal returns

Low frequency returns: $r_{Mt} \sim N(\mu_M, \sigma_M)$

High frequency returns: $r_{Dt} \sim N(\mu_D, \sigma_D)$

- Goal: Estimate the low frequency parameters (μ_M , σ_M) most precisely.

- Question: Can we achieve this by using higher frequency returns?

- Needed: Aggregation relationships from High to Low frequency
  N high frequency data per low frequency intervals
  E.g., N days per month, N months per year, ….

Link from low to high frequency (log-normal returns) by the aggregation formulas

$$\mu_M = E(r_{Mt}) = E(\sum r_{Dt}) = \sum E(r_{Dt}) = N\,\mu_D$$

$$\sigma^2_M = V(\sum r_{Dt}) = \sum \sigma^2 + \sum_t \sum_{s\neq t} Cov(r_{Dt}\ r_{Ds}) = N\,\sigma_D^2 + \text{STUFF} = N\sigma^2_D$$

assume no returns autocorrelation: STUFF = 0

2 Estimating the mean

Let us estimate the mean via the high or low frequency returns. We will see that

- The t-statistics of the high-frequency and low-frequency sample means are equal!
- The high-frequency and low-frequency data imply exactly the same confidence interval for the low-frequency mean.

2.1 t-test

- Method 1: Get T low frequency returns (monthly).

$\overline{r_M} = \Sigma r_{Mt} / T \qquad E(r_{MT}) = \mu_M \qquad V(\overline{r_M}) = \sigma^2_M / T$

t-test for $\mu_M=0$:

$$\boxed{t_M = \frac{\bar{r}_M - 0}{\sigma_M / \sqrt{T}}}$$

- <u>Method 2</u>: Get N x T high frequency returns (20 T daily returns)

$\mu_D = \mu_M / N$ Similarly: $\overline{r_D} = \sum r_{Dt} / (NT) = \overline{r_M} / N$

$V(\overline{r_D}) = \sigma^2_D / (NT)$

t-test for H_0: $\mu_D = 0$

$$\boxed{t_D = \frac{\overline{r}_D - 0}{\sigma_D / \sqrt{NT}} = \frac{\overline{r}_M / N - 0}{(\sigma_M / \sqrt{N}) / \sqrt{NT}} = \frac{\overline{r}_M / N - 0}{\sigma_M / (N\sqrt{T})}}$$

t_D = t_M !!

2.2 Confidence interval

- Confidence interval from low frequency data:

$$\left[\overline{r_M} - 1.96\frac{\hat{\sigma}_M}{\sqrt{T}}, \overline{r_M} + 1.96\frac{\hat{\sigma}_M}{\sqrt{T}}\right]$$

- Confidence interval from high frequency data:

$$\left[\overline{r_D} - 1.96\frac{\hat{\sigma}_D}{\sqrt{NT}}, \overline{r_D} + 1.96\frac{\hat{\sigma}_D}{\sqrt{NT}}\right]$$

- Second interval is in high frequency (daily) returns: Needs to be converted to low frequency (monthly) returns via the aggregation formula

- *Key: Both values in the interval are values on the distribution of* $\overline{r_D}$

- The low frequency interval implied by the high frequency data:

$$\left[N\,\overline{r_D} - 1.96\frac{\hat{\sigma}_D N}{\sqrt{NT}}, N\,\overline{r_D} + 1.96\frac{\hat{\sigma}_D N}{\sqrt{NT}}\right] = \left[\overline{r_M} - 1.96\frac{\hat{\sigma}_M}{\sqrt{T}}, \overline{r_M} + 1.96\frac{\hat{\sigma}_M}{\sqrt{T}}\right] \qquad !!$$

3 Estimating the variance

3.1 t-test

- T low frequency returns (monthly): $\hat{\sigma}_M$ $\quad V(\hat{\sigma}_M) = \sigma^2_M / 2T$

  t-test for H_0: $\sigma_M = \sigma_{M0}$: $\quad \mathbf{t_M} = (\hat{\sigma}_M - \sigma_{M0}) \,/\, (\sigma_M / \sqrt{2T})$

- NT high frequency returns (20T daily returns): $\hat{\sigma}_D$ $\quad V(\hat{\sigma}_D) = \sigma^2_D / 2NT$

  t-test for H_0: $\sigma_D = \sigma_{D0}$: $\quad \mathbf{t_D} = (\hat{\sigma}_D - \sigma_{D0}) \,/\, (\sigma_D / \sqrt{2NT})$

  Now: $\quad \sigma_M = \sqrt{N}\,\sigma_D \quad and \quad \sigma_{M0} = \sqrt{N}\,\sigma_{D0}$

$$\mathbf{t_D} = (\hat{\sigma}_M/\sqrt{N} - \sigma_{M0}/\sqrt{N}) \,/\, (\sigma_M/(\sqrt{N}\sqrt{2NT})) = \sqrt{N}\ \mathbf{t_M}\,!$$

3.2 Confidence Interval

- Confidence interval from low frequency data:

$$\left[s_M - 1.96\,\frac{s_M}{\sqrt{2T}}\ ,\ s_M + 1.96\,\frac{s_M}{\sqrt{2T}}\right] \qquad [1]$$

- Confidence interval from high frequency data (N points per low frequency interval):

$$\left[s_D - 1.96\,\frac{s_D}{\sqrt{2TN}}\ ,\ s_D + 1.96\,\frac{s_D}{\sqrt{2TN}}\right] \qquad [2]$$

- Now Convert high frequency interval [2] into low frequency to allow for comparison with [1].

  Lower and upper bounds in [2] represent the range of (possible values of) high frequency standard deviations:
  They must each be multiplied by $\sqrt{N}$ for annualization into equivalent low frequency. Then [2] becomes

$$\left[s_M - 1.96\,\frac{s_M}{\sqrt{2T\boldsymbol{N}}}\ ,\ s_M + 1.96\,\frac{s_M}{\sqrt{2T\boldsymbol{N}}}\right] \qquad [3]$$

Precision of sample variance increases by $\sqrt{N}$ when sampling frequency increases by N

4 Summary

- The precision of the estimator of the mean does **not** increase when returns are observed more frequently within the same calendar span.

  Extreme: The Geometric mean, do you care about in-between observations?

- Variance and Standard deviation (and all other second order moments, covariances, betas) are estimated with more precision when the sampling frequency increases.

Black-Scholes-Merton: ``*The agent knows variance*''.

It is not an extra assumption of the Merton model, it follows from the continuous time assumption