Class 2: Financial Data and Empirical Estimations

Financial Markets, Fall 2020, SAIF

Jun Pan

Shanghai Advanced Institute of Finance (SAIF)
Shanghai Jiao Tong University

November 23, 2020

Outline

- The thirst for information has made the financial industry an early adopter of data and information technology:
  - Real-time data provider: Bloomberg (since 1981) and Wind (万得).
  - Historical data provider: Datastream (since 1967).
  - Research oriented database: CRSP (since 1960), COMPUSTAT, TAQ, etc.
- Finance is about risk and uncertainty:
  - Theory: modeling random events in financial markets.
  - Data: historical experiences of random events.
  - Empirical estimation: where models meet data.
- Today, we will focus on two examples:
  - Normal distribution and empirical distribution.
  - Estimating the *expected* return $\mu = E(R_t)$.

Where to Get Data

Wharton Research Data Services (WRDS)

Your Subscriptions | Not Subscribed | Your Queries

- Bank Regulatory
- Blockholders
- CBOE Indexes
- COMPUSTAT
- COMPUSTAT Trial
- CRSP
- CUSIP
- DMEF Academic Data
- Dow Jones
- Factset Trial
- Fama French & Liquidity Factors
- Federal Reserve Bank
- GSIOnline
- IBES
- IHS Global Insight
- Markit Trial
- Mergent FISD
- MFLINKS
- Option Metrics
- Option Metrics Trial
- OTC Markets
- Penn World Tables
- PHLX
- Public
- SEC Order Execution
- TAQ
- Thomson Reuters
- TRACE
- WRDS SEC Analytics Suite Trial
- Zacks Trial

Kenneth R. French:

BIOGRAPHY
CURRICULUM VITAE
WORKING PAPERS
DATA LIBRARY
CONSULTING RELATIONSHIPS
FAMA / FRENCH FORUM
CONTACT INFORMATION

Computing Realized Stock Returns

- For a publicly traded firm, we can get
  - its stock price P_t at the end of year t.
  - its cash dividend D_t paid during year t.
- At the end of year t, we calculate the **realized** return on the stock:

$$R_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} = \frac{P_t - P_{t-1}}{P_{t-1}} + \frac{D_t}{P_{t-1}}$$

- Returns = capital gains yield + dividend yield.
- For the US markets, the best place to get reliable and clean holding-period returns is CRSP. I have applied a WRDS account for our class, which gives you access to CRSP.

The Expected Return

- For any financial instrument, the single most important number is its **expected** return.
- Suppose right now we are in year t, let R_{t+1} denote the stock return to be realized next year. Our investment decision relies on the **expectation**:

$$\mu = E\left(R_{t+1}\right) .$$

- Just to emphasize, μ is a number, while R_{t+1} is a random variable, drawn from a distribution with mean μ and standard deviation σ.
- To estimate this number μ with precision is the biggest headache in Finance.

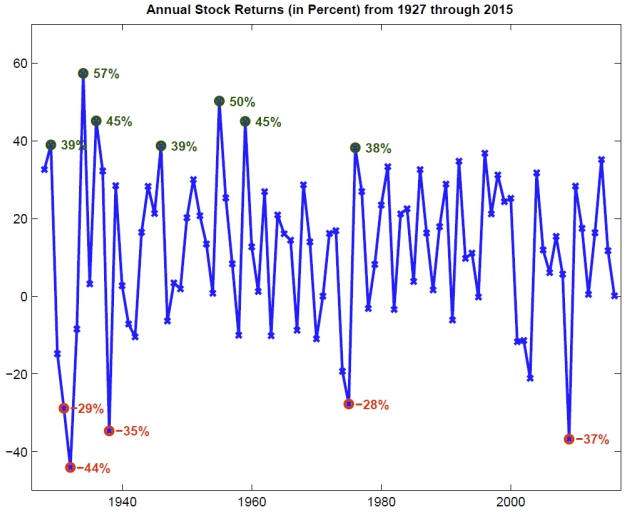
Estimating the Expected Return μ

- We estimate μ by using historical data:

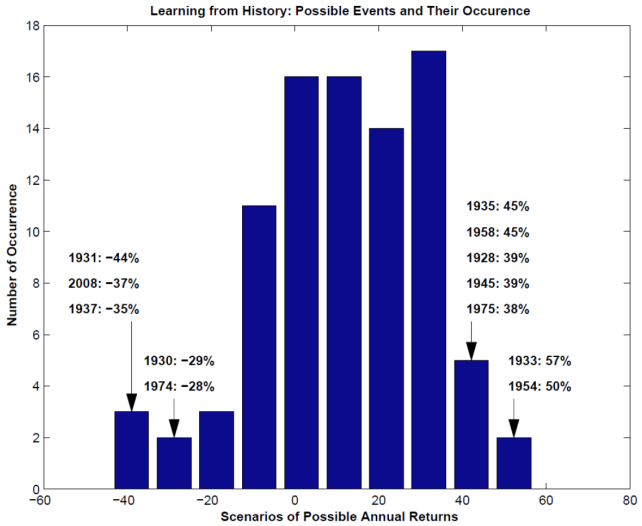
$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^{N} R_t \,.$$

- It is as simple as taking a sample average.
- Why can this sample average of *past* realized returns help us form an expectation of the *future*?
- Because our assumption that history repeats itself. Each R_t in the past was drawn from an identical distribution with mean μ and standard deviation σ.

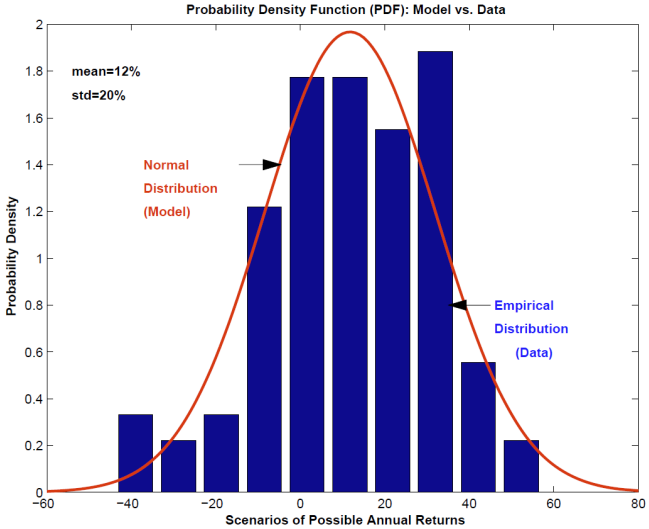
Time Series of Annual Stock Returns



Scenarios and Their Likelihood



Probability Distribution of a Random Event



The Estimator Has Noise

- We use historical returns to estimate the number μ:

$$\hat{\mu} = \frac{1}{N}\sum_{t=1}^{N} R_t$$

- Recall that R_t is a random variable, drawn every year from a distribution with mean μ and standard deviation σ.
- As a result, $\hat{\mu}$ inherits the randomness from R_t. In other word, it is not really a number: var$(\hat{\mu})$ is not zero.
- If this variance var$(\hat{\mu})$ is large, then the estimator is noisy.

The Standard Error of $\hat{\mu}$

- Let's first calculate $\text{var}(\hat{\mu})$:

$$\text{var}\left(\frac{1}{N}\sum_{t=1}^{N} R_t\right) = \frac{1}{N^2}\sum_{t=1}^{N}\text{var}(R_t) = \frac{1}{N^2} \times N \times \sigma^2 = \frac{1}{N}\sigma^2$$

- The **standard error** of $\hat{\mu}$ is the same as $\text{std}(\hat{\mu})$:

$$\text{standard error} = \frac{\text{std}(R_t)}{\sqrt{N}} = \frac{\sigma}{\sqrt{N}}$$

Estimating μ for the US Aggregate Stock Market

- Using annual data from 1927 to 2014, we have 88 data points.
- The sample average is $\text{avg}(R) = 12\%$. The sample standard deviation is $\text{std}(R) = 20\%$.
- The **standard error** of $\hat{\mu}$:

$$\text{s.e.} = \text{std}(R)/\sqrt{N} = 20\%/\sqrt{88} = 2.13\%$$

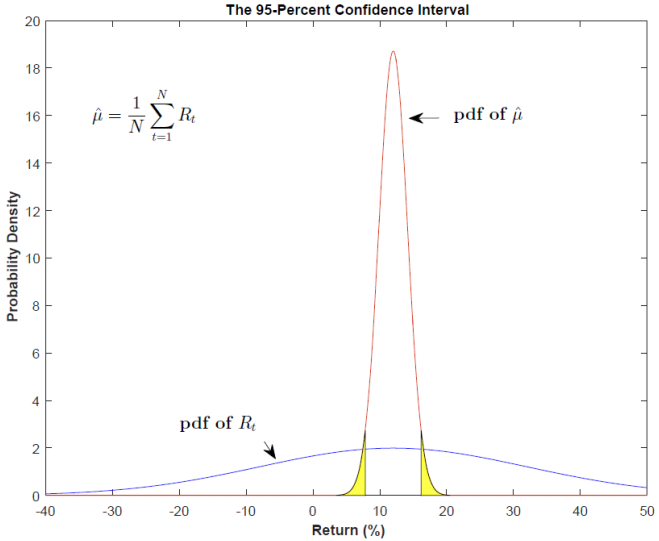
- The 95% confidence interval of our estimator:

$$[12\% - 1.96 \times 2.13\%, 12\% + 1.96 \times 2.13\%] = [7.8\%, 16.2\%]$$

- The **t-stat** of this estimator is (signal-to-noise ratio),

$$\text{t-stat} = \frac{\text{avg}(R)}{\text{std}(R)/\sqrt{N}} = \frac{12\%}{2.13\%} = 5.63\,.$$

The Distributions of R_t and $\hat{\mu}$



How to Improve the Precision?

- Not much, really!
- We got a t-stat of 5.63 for $\hat{\mu}$ using 88 years of data!
- Usually, the time series we are dealing with are much shorter. For example, the average life span of a hedge fund is around 5 years.
- Also, the volatility of individual stocks is much higher than that of the aggregate market. For example, the annual volatility for Apple is 49.16%. For smaller stocks, the number is even higher: around 100%.
- What about designing a derivatives product whose value would depend on μ? (No)
- What about polling investors for their individual assessments of μ and then aggregate the information? (Not very useful)

Estimating μ Using Monthly Returns

- Since the standard error of $\hat{\mu}$ depends on the number of observations, why don't we use monthly returns to improve on our precision?
- Using monthly aggregate stock returns from January 1927 through December 2011, we have 1020 months. So N=1020!
- The mean of the time series is 0.91%, and std is 5.46%.
- So the standard error of $\hat{\mu}$ is:

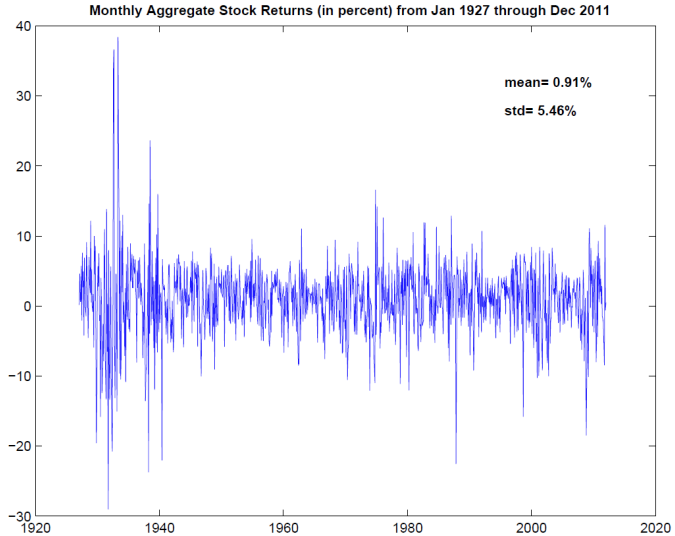
$$\text{s.e.} = 5.46\%/\sqrt{1020} = 0.1718\%$$

- The signal-to-noise ratio:

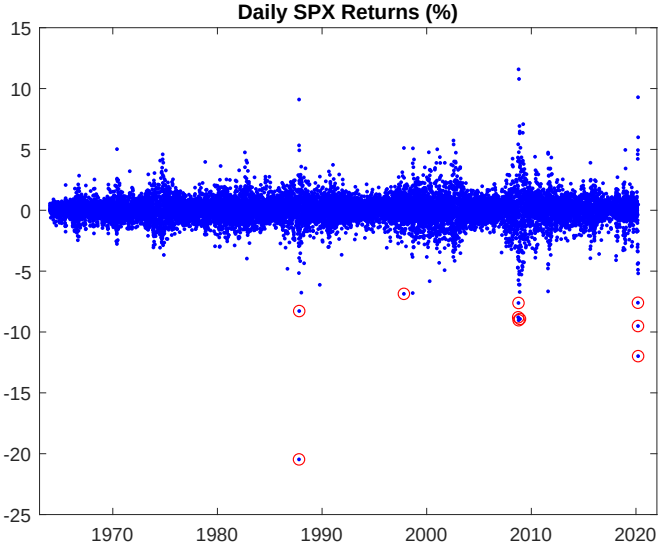
$$\text{t-stat} = \frac{0.91\%}{0.1718\%} = 5.30$$

- We increased N by a factor of 12. Yet, the t-stat remains more or less the same as before. What is going on?

Time Series of Monthly Stock Returns



Time Series of Daily Stock Returns



Chopping the Time Series into Finer Intervals?

- It is actually a very straightforward calculation (give it a try) to show that when it comes to the precision of $\hat{\mu}$, it is the length of the time series that matters. Chopping the time series into finer intervals does not help.
- Professor Merton has written a paper on that. See "On Estimating the Expected Return on the Market," *Journal of Financial Economics*, 1980.
- But when it comes to estimating the volatility of stock returns, this approach of chopping data into finer intervals does help and is widely used. We will come back to this.

The Main Takeaways

- The financial industry has always been data intensive:
  - Data contains information.
  - Data contains noise.
- A good practitioner knows how to extract signal from noise:
  - Knowing how to read tables with standard errors and t-stats is essential.
  - Basic econometrics and statistics will be an important differentiator.
- Questions to be answered by Wednesday's student presentations:
  - What are the means and standard deviations of monthly returns on the US and Chinese equity markets?
  - What is the correlation between the monthly returns?
  - How accurate are these estimates?