Outline

- Where to get financial data?
- Modeling random events in financial markets.
- Test financial models using financial data.
- Estimating the expected return.
### Where to Get Data

- **Bloomberg**
- **Datastream**
- **WRDS:** →

**CRSP:**
- Stock
- Treasury Bonds
- Mutual Funds

**Prof. Ken French’s Website.**

<table>
<thead>
<tr>
<th>Your Subscriptions</th>
<th>Not Subscribed</th>
<th>Your Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank Regulatory</td>
<td>Fama French &amp; Liquidity Factors</td>
<td>OTC Markets</td>
</tr>
<tr>
<td>Blockholders</td>
<td>Federal Reserve Bank</td>
<td>Penn World Tables</td>
</tr>
<tr>
<td>CBOE Indexes</td>
<td>GSOnline</td>
<td>PHLX</td>
</tr>
<tr>
<td>COMPSTAT</td>
<td>IBES</td>
<td>Public</td>
</tr>
<tr>
<td>COMPSTAT Trial</td>
<td>IHS Global Insight</td>
<td>SEC Order Execution</td>
</tr>
<tr>
<td>CRSP</td>
<td>Markit Trial</td>
<td>TAQ</td>
</tr>
<tr>
<td>CUSIP</td>
<td>Mergent FISD</td>
<td>Thomson Reuters</td>
</tr>
<tr>
<td>DMEF Academic Data</td>
<td>MFLINKS</td>
<td>TRACE</td>
</tr>
<tr>
<td>Dow Jones</td>
<td>Option Metrics</td>
<td>WRDS SEC Analytics Suite Trial</td>
</tr>
<tr>
<td>Factset Trial</td>
<td>Option Metrics Trial</td>
<td>Zacks Trial</td>
</tr>
</tbody>
</table>
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’ve applied a CRSP class account for us.
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**:
  \[ \mu = E(R_{t+1}). \]
- Just to emphasize, $\mu$ is a number, while $R_{t+1}$ is a random variable, drawn from a distribution with mean $\mu$ and standard deviation $\sigma$.
- To estimate this number $\mu$ with precision is the biggest headache in Finance.
We estimate \( \mu \) 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 \( \mu \) and standard deviation \( \sigma \).
Time Series of Annual Stock Returns

Annual Stock Returns (in Percent) from 1927 through 2015

- 57%
- 45%
- 45%
- 39%
- 39%
- 38%
- 28%
- 29%
- 35%
- 44%
- 37%
Scenarios and Their Likelihood

Learning from History: Possible Events and Their Occurrence

- 1931: -44%
- 2008: -37%
- 1937: -35%
- 1930: -29%
- 1974: -28%
- 1933: 57%
- 1954: 50%

Number of Occurrence

Scenarios of Possible Annual Returns

- 1935: 45%
- 1958: 45%
- 1928: 39%
- 1945: 39%
- 1975: 38%
Probability Distribution of a Random Event

Probability Density Function (PDF): Model vs. Data

- Mean = 12%
- Std = 20%

- Normal Distribution (Model)
- Empirical Distribution (Data)

Scenarios of Possible Annual Returns

Probability Density
We use historical returns to estimate the number $\mu$:

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

Recall that $R_t$ is a random variable, drawn every year from a distribution with mean $\mu$ and standard deviation $\sigma$.

As a result, $\hat{\mu}$ inherits the randomness from $R_t$. In other word, it is not really a number: $\text{var}(\hat{\mu})$ is not zero.

If this variance $\text{var}(\hat{\mu})$ 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 $\text{std}(\hat{\mu})$:

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

\[
\text{s.e.} = \frac{\text{std}(R)}{\sqrt{N}} = \frac{20\%}{\sqrt{88}} = 2.13\%
\]

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),

\[
t-\text{stat} = \frac{\text{avg}(R)}{\frac{\text{std}(R)}{\sqrt{N}}} = \frac{12\%}{2.13\%} = 5.63.
\]
The Distributions of $R_t$ and $\hat{\mu}$

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

pdf of $R_t$

pdf of $\hat{\mu}$
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 $\mu$? (No)
- What about polling investors for their individual assessments of $\mu$ and then aggregate the information? (Not very useful)
Estimating $\mu$ 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:

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

- The signal-to-noise ratio:

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

- 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

Monthly Aggregate Stock Returns (in percent) from Jan 1927 through Dec 2011

- mean = 0.91%
- std = 5.46%
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.


- But when it comes to estimating the volatility of stock returns, this approach of chopping does help and is widely used. We will come back to this.