

Financial Engineering & Risk Management

Statistical Biases in Performance Evaluation

M. Haugh G. Iyengar

Department of Industrial Engineering and Operations Research
Columbia University

Performance Evaluation



- Some fund managers claim to have special “skill”
 - at picking securities
 - or **timing** the market.
- This skill is often referred to as **alpha**
 - a term that comes originally from the CAPM

$$\bar{r} - r_f = \alpha + \beta (\bar{r}_m - r_f)$$

- Fund manager with $\alpha > 0$ can justify often substantial management fees.

Question: Is this skill real?

Answer: In general it's very hard to tell

- however can still do some interesting analysis
- we will look at this question from **three** different perspectives.

The first two require the **binomial distribution**.

Recall the Binomial Distribution

We say X has a binomial distribution, or $X \sim \text{Bin}(n, p)$, if

$$P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r}. \quad (1)$$

For example, X might represent the number of heads in n independent coin tosses, where $p = P(\text{head})$. The mean and variance of the binomial distribution satisfy

$$\begin{aligned} E[X] &= np \\ \text{Var}(X) &= np(1 - p). \end{aligned}$$

We see from (1) that

$$P(X \geq r) = \sum_{i=r}^n \binom{n}{i} p^i (1 - p)^{n-i}. \quad (2)$$

Perspective #1: A Single Manager

- Suppose a fund manager has a track record of 10 years and has outperformed the market in 9 of the 10 years.
- He claims to have great skill and that his fees should reflect this.
- How can we assess his claims?
- A first analysis might assume the following:
 1. in a given year he out-performs w.p. p and under-performs w.p. $1 - p$
 2. out-performance or under-performance is independent across years.
- Also assume that we are always referring to risk-adjusted returns.
- If the manager has skill then $p > 1/2$
 - otherwise $p \leq 1/2$ and the manager has no skill.
- The first question that now comes to mind is the following:

Perspective #1: A Single Manager

Question: How likely is such a track record if the fund manager had no skill?

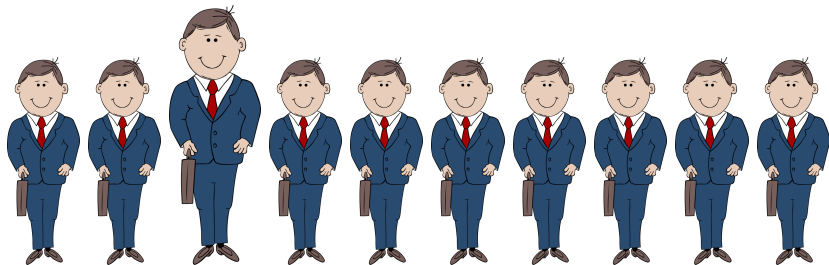
Answer: Let X be the number of outperforming years. If the fund manager has no skill then $X \sim \text{Bin}(n = 10, p = 1/2)$ and

$$\begin{aligned} P(X \geq 9) &= \sum_{r=9}^n \binom{n}{r} p^r (1-p)^{n-r} \\ &= 0.0107 \end{aligned}$$

So if the fund manager has no skill then the probability of having track record as good as his or better is only .0107

– so does it seem fair to conclude the fund manager has skill?

Perspective #2: M Fund Managers



Suppose instead there are M fund managers and that the manager who claims to have skill has the best track record of these managers.

Question: Does this change anything in our analysis? Should it?

To answer this, suppose we start with the hypothesis that **none** of the fund managers have skill and that track records of managers are independent.

Then there are two possible questions:

Perspective #2: M Fund Managers

1. How likely is the 3rd manager to have such a track record if all fund managers have no skill?
2. How likely is the best manager to have such a track record if all fund managers have no skill?

What is the appropriate question here?

– it depends on our **prior hypothesis**.

And what is the answer to the appropriate question?

Perspective #2: M Fund Managers

Assuming none of the managers have skill let:

$Z_i \equiv$ event that i^{th} manager out-performs in $\geq r$ years out of n .

$V \equiv$ event that the **best** manager out-performs in $\geq r$ years out of n .

Then

$$\begin{aligned} P(V) &= 1 - P(\bar{Z}_1, \dots, \bar{Z}_M) \\ &= 1 - P(\bar{Z}_1)^M \\ &= 1 - [1 - P(Z_1)]^M \\ &= 0.1942 \end{aligned}$$

when $M = 20$ and where $P(Z_1) = P(\text{Bin}(n, p) \geq r) = .0107$.

Question: What will happen to $P(V)$ as M gets bigger?

What are our conclusions?

Perspective #3: A Thought Experiment

Another way to make the same point is to consider the following thought experiment:

- Suppose again that all fund managers in the market have no skill.
- At the end of every year:
 - fund-managers who have **out**-performed the market that year survive.
 - fund-managers who have **under**-performed the market that year get fired.

Question: After 1 year of this experiment what will the average track-record of fund managers in the market be?

Question: Is this a fair reflection of the “true” track record?

This is an example of so-called **survivorship bias**

– a very common phenomenon in finance and beyond!

Performance Evaluation: Final Thoughts

- Are we being unfair in **assuming** to begin with that fund managers have no skill?
- After all, in practice there are some managers with skill
 - although surely not many
 - and lots of fund managers **think** they have skill but the reality is different.
- The answer for most people is “no”: it is the responsibility of the fund manager to convince us they have skill.
- In practice it's very hard to find a manager with skill
 - and to then **verify** that he has skill.
- And if we can find such a manager, is the resulting out-performance sufficient to justify the management fees?

Financial Engineering & Risk Management

How Should Average Returns Be Computed?

M. Haugh G. Iyengar

Department of Industrial Engineering and Operations Research
Columbia University

How Should Average Returns Be Computed?

Suppose an investment fund delivers the following performance:

Year	Performance
Year 1	+ 20%
Year 2	-10%

Question: What is the average annual return of the fund?

Answer: Clearly it is 5%.

How Should Average Returns Be Computed?

Suppose an investment fund delivers the following performance:

Year	Performance	\$'s Invested
Year 1	+ 20%	1m
Year 2	-10%	10m

Question: What is the **average** annual return of the fund?

Answer: Now it's not so clear . . . there are two possible answers:

1. 5% as before
2. $-7.27\% = \frac{1 \times 20\% - 10 \times 10\%}{11}$
– the average annual return per \$ invested.

Which answer, if any, is more compelling? Why is this important?

How Should Average Returns Be Computed?

Claim: From investors' perspective, -7.27% is the right answer.

Reason: Investors care about **return on dollars** invested. After all, what would you prefer:

Year	Performance	\$'s Invested
Year 1	+ 20%	1m
Year 2	-10%	10m

or

Year	Performance	\$'s Invested
Year 1	+ 20%	10m
Year 2	-10%	1m

Another Reason to Care About Dollars Invested

In financial markets **expected** returns often **decrease** as dollars invested **increase**.

This is because **liquidity** of a market or “**capacity**” of a trading strategy is not unbounded

- not always obvious to the small investor who only invests in liquid markets
- and therefore does not “**move**” the market.

Large investors often do “move” the market

- and the larger they are, the more they move it
- the more illiquid the market is, the more they move it
- so the cost per security **increases** with the number of securities they **buy**
- and the cost per security **decreases** with the number of securities they **sell**.

This implies that returns **decrease** as dollars invested **increases**.

Just Good Marketing?

- The question of how to compute average returns is important.
- Depending on how you answer it, certain types of investing can seem more or (much) less attractive
 - e.g. the **hedge fund industry**
 - in aggregate, they would prefer to report average returns over time
 - and so they do.
 - but for investors, the average (net) return per dollar invested is surely more meaningful.
- This has caused some controversy and debate
 - there are good financial blogs that cover these topics and others
 - see for example discussion at

<http://blogs.reuters.com/felix-salmon/2012/08/08/why-investors-should-avoid-hedge-funds/>

Another Problem with Averages: Counting Children

These kinds of examples occur frequently and often lead to confusion.

e.g. Suppose I wish to estimate the average # of children per family in the US.

To compute an estimate I do the following:

1. I sample N people randomly
2. For the i^{th} person I determine X_i , the number of siblings in his / her family
3. My estimate, \hat{C} say, is then given by

$$\hat{C} = \frac{\sum_{i=1}^N (X_i + 1)}{N}.$$

Ignore any “minor” problems that you might see with this sampling scheme.

Question: Does the sampling scheme have a fundamental problem?

Average Number of Children in a Family

Question: If so, in what way will \hat{C} be **biased**?

Question: How does this problem compare to the average return problem?

Another Problem with Averages: Waiting Times

Consider the controversy surrounding waiting times to get through immigration at Heathrow airport in London

– a big news story last year.

Here's one way to estimate the average waiting time of a traveler at immigration:

1. sample one person every hour and compute his / her waiting time
2. take the average.

What do you think of this scheme?

Financial Engineering & Risk Management

Survivorship Bias and Data Snooping

M. Haugh G. Iyengar

Department of Industrial Engineering and Operations Research
Columbia University

Survivorship Bias

Consider the following investment: purchase an equi-weighted portfolio of the top 20 stocks in the S&P 500.

- note that the stocks are chosen and fixed **today**.

In order to estimate the performance of this investment you decide to **back-test** it as follows:

1. Get the last 20 years of daily return data for each of the 20 stocks
2. On the first day, i.e. 20 years ago, set up the initial equi-weighted portfolio
 - if the stock, e.g. *Google*, did not yet exist then omit it from the portfolio until it does exist.
3. Every month **rebalance** the portfolio so that it remains equi-weighted
 - and take transactions costs into account.
4. Plot the annual net returns, i.e. r_t against t where r_t is the net return at time t realized over the previous year, of this back-test.

Survivorship Bias

Question: Do you think the plot will be representative of the future performance of the investment?

This is another example of **survivorship bias**

- a problem that arises throughout finance and beyond.

Question: Why is this an important issue?

Question: What other examples are there of survivorship bias in finance?

- needs to be kept in mind by all investors, risk-managers etc.

The Football Game Scam

- On each of 10 consecutive Wednesdays you receive a letter predicting the winner of a big football game the following Sunday.
- Each week the prediction was correct!
- In week 11 a letter arrives but this time it seeks payment of \$10,000 before revealing the prediction for the next game.

Question: What do you do?

Question: What is the scam? You are now the survivor!

Data Snooping

A bank has 4 years worth of daily historical return data on USD-GBP exchange rate. It employs the following mechanism for generating a trading strategy:

1. It first normalizes the **entire** return data so that it has mean 0 and variance 1 (normalizing data is a standard and well justified statistical technique)
2. 75% of the data, i.e. approx 750 returns, is kept for the **training set**
 - used for finding the trading strategy.
3. Remaining 25%, i.e. approx 250 returns, is kept as a hold-out **test set**
 - used to evaluate whatever strategy is yielded by the training data.

The trading strategy appears to be a great success: on any given day it uses the returns of the previous 20 days to forecast the direction of the next day's return.

But the trading strategy performs poorly in practice.

Data Snooping

Question: Why?

(This example was taken from “Learning From Data” by Abu-Mostafa, Magdon-Ismail and Lin)

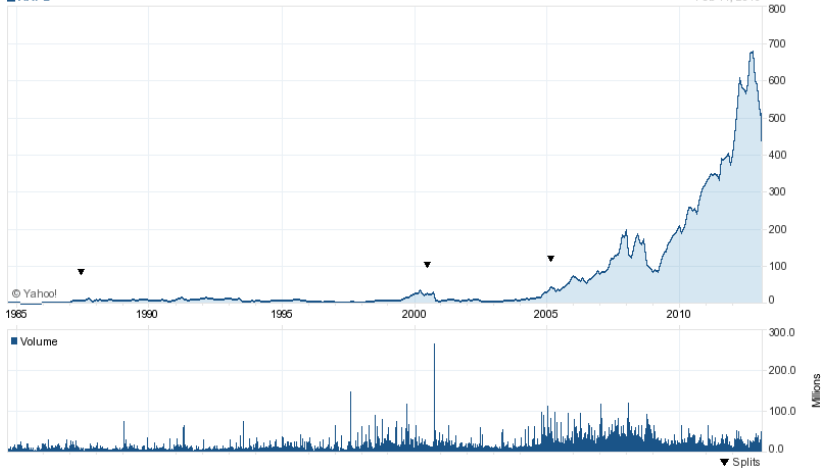
Other Examples of Statistical Biases / Difficulties

- Good risk management always needs to be aware of statistical biases
 - survivorship bias and data snooping are everywhere!
- And how meaningful are statements such as:
“the stock market has never fallen over any 20-year period”?
- Just how likely is a 25 standard deviation move?
 - the size of the move as reported by some participants in August 2007
 - how likely is a 25 standard deviation move several days in a row (!)?
- The market for retail structured products
 - often structured as a note that pays a coupon tied to the performance of another asset
 - often designed to look better than they are, i.e. they tend to
 - they invariably backtest very well
 - e.g. any structured product that is long *Apple* will presumably back-test very well from 2000 onwards
 - investor often exposed to hidden risks, e.g. volatility risk and credit risk
 - often too expensive (too many "middle men" and no price transparency)
 - and no secondary market available in case you need to sell.

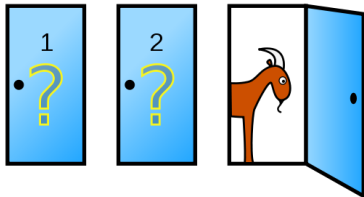
Apple Inc.

■ AAPL

Feb 14, 2013



The Monty Hall Problem



1. There are three closed doors.
2. A goat lies behind two of the doors and \$1m lies behind the other door.
3. You don't know which door has the \$1m and so you have to "guess".
4. Before the door is opened Monty Hall opens a different door
 - this door always has a goat behind it.
5. Monty now gives you the option to change your mind and pick another door.

Question: Should you change your mind?