Advanced Analytics of Finance: Overview

Hui Chen

MIT Sloan

15.457, Spring 2021

Outline

wealthfront

Call us at (888) 227-5498 | FAQ | Contact Us

1 Answer Questions 2 View Your Plan 3 Sign Up

Your Investing Plan

About You

Your Risk Tolerance

– 6.7 +

Change My Answers

Amount to Invest

$50,000

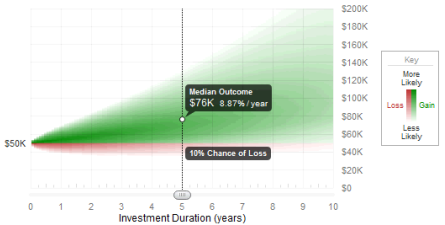
How do I decide?

Your Investment Mix

Why this mix? | *Can I change it?* | *Why Vanguard?*

Asset Class	Investment		Percentage	Amount
US Stocks	Vanguard VTI ETF	why?	36.3%	$18,160
Foreign Stocks	Vanguard VEA ETF	why?	19.8%	$9,920
Emerging Markets	Vanguard VWO ETF	why?	13.7%	$6,860
Real Estate	Vanguard VNQ ETF	why?	8.8%	$4,420
Natural Resources	iPath DJP ETN	why?	5.8%	$2,915
Bonds	Vanguard BND ETF	why?	15.4%	$7,725

Projected Performance | Historical Performance | Your Costs



This projection is net of all fees (our fee, ETF expenses, and commissions) and includes dividends. It is based on each asset class's 25-year historical returns, volatility, and correlations. It's intended to show only an expected range of

Wealthfront Benefits

- We give ***mathematically driven*** advice
- We ***continuously optimize*** your portfolio
- ***Dramatically lower fee*** than traditional advisors

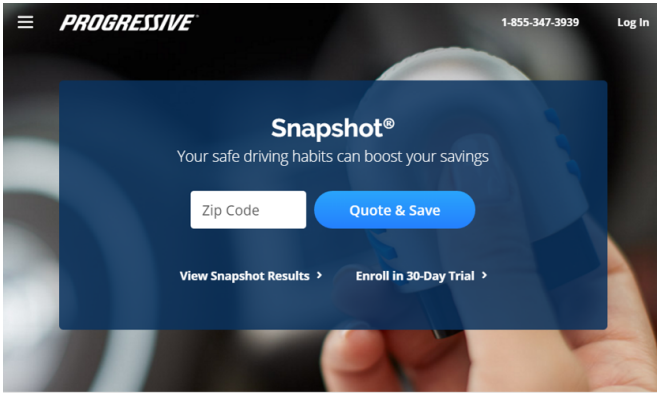
Sign Up & Invest

Save For Later

Common Questions

What about:

- *Rebalancing?*
- *My brokerage account?*
- *Taxes?*
- *My house?*
- *My savings?*
- *"Play money"?*
- *Account Types?*
- See More Questions...



The fair way to pay for car insurance

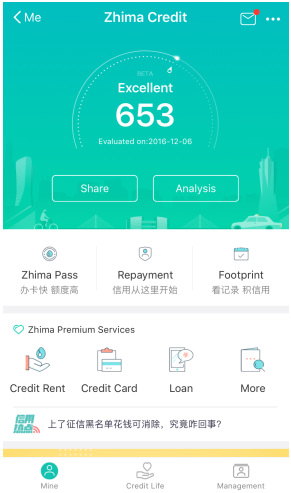
It just makes sense—insurance should be based partly on how you actually drive, rather than just on traditional factors like where you live and what kind of car you have.

That's what Snapshot is all about. Your safe driving habits can help you save on

How good is your credit?

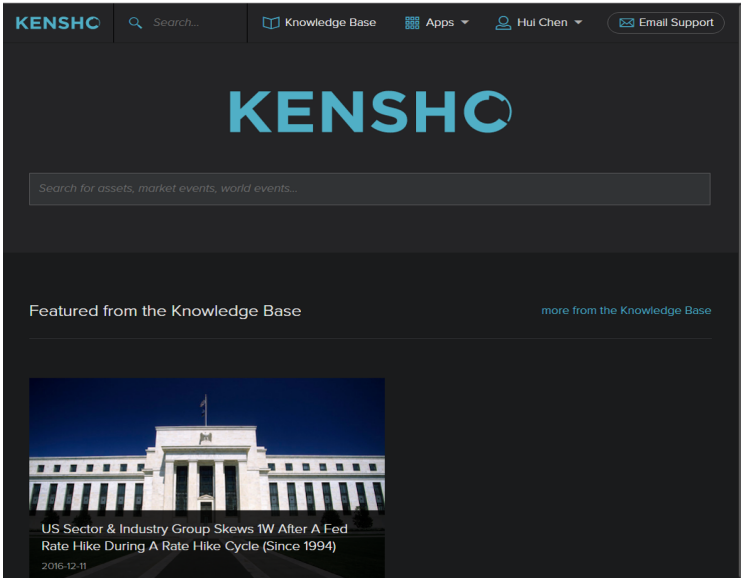
Personal credit monitoring system

- Information from Alibaba's ecosystem
  - ↪ Consumption / sales data from Alibaba (T-Mall, Taobao . . .)
  - ↪ Payment / transaction data from Alipay
  - ↪ Financial data from Ant Financial
  - ↪ Other data from IOT (Wechat, Weibo . . .)
- Pre-lending: credit risk assessment
- During-lending: real-time monitoring
- Post-lending: model updating



Did Target have a good quarter?

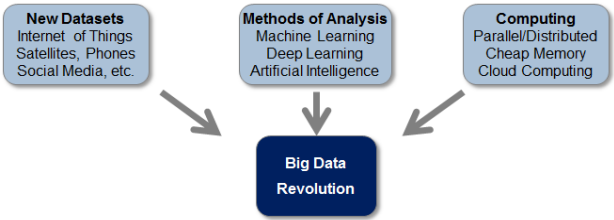
Does the stock market perform better under Rep or Dem presidents?



Finance is going through major changes

- Automated financial reporting
- Robo-advisors
- Big data lending platform
- "Siri" for Wall Street
- Quantitative investing based on alternative data
- Blockchain and disintermediation

What make these innovations feasible?



Source: J.P.Morgan Macro QDS.

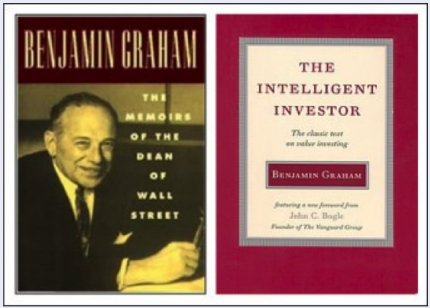
Outline

What is quantitative investing?

- Benjamin Graham: the "father of value investing"
- What does it mean to be a value investor? Let's quantify the strategy:

Quantitative value

1. At least 3 years after IPO
2. Paying dividend for past 3 years or more
3. $\frac{\text{Average EPS (earnings per share) in past 2 years}}{\text{Average EPS for Year 4 and 5 before}} > 1.3$
4. Past 1-year revenue $\geq$ 1 bn
5. Current ratio > 1.5
6. $P/E < 25$
7. $P/E \times P/B < 50$
8. $\frac{\text{Total Debt}}{\text{Current Assets}} < 1.1$



- Express an investment strategy in terms of mathematical rules. Back-test it on (a lot of) historical data to see whether the strategy is reliably profitable.

Types of quantitative investing

Style	Description
Global Macro	Focus on macroeconomic environment, often concentrates on currencies or major interest-rates moves.
Emerging Markets	Invests in the debt or equity (and less frequently, FX) of emerging markets. Markets are typically characterized by their relative lack of transparency and liquidity, in addition to an inability to find viable derivatives contracts for hedging.
Equity Market Neutral	Trades pairs of shares – buying one and selling another – and therefore is typically neutral to market direction (i.e., employs a beta of zero). Also called *statistical arbitrage*.
CTA	Directional strategies using derivatives in a wide range of asset classes, including fixed income, currencies, equity, and commodities. The most common type is systematic trend following.
Convertible Arbitrage	Targets pricing anomalies between convertible bonds and the underlying shares and/or options on shares.
Fixed Income Arbitrage	Exploits anomalies between related bonds, often with high leverage.
Event-Driven	Trades based on anticipated corporate events, such as anticipated merger or take-over activity or bankruptcy filing. Also called *risk arbitrage*.

What is quantitative investing?

- Time series:

$$R_{t+1} = a + bX_t + \varepsilon_{t+1}$$

- Cross section:

$$R_{i,t} = \alpha_i + \beta_i f_t + \varepsilon_{i,t}$$

How much predictability do we need?

Let's play a coin-flip game:

- Cost to play: $1
- Head: Win $100
- Tail: Lose $100
- Opportunity to play: 1 time
- End-of-year balance:

$$100, \quad -100$$

- Would you like to play this game?
- Odds:

$$51/49$$

- Expected end-of-year balance:

How much predictability do we need?

New rules:

- Cost to play: $1
- Head: Win $100/N
- Tail: Lose $100/N
- Opportunity to play: N times
- Odds:

$$51/49$$

- Expected end-of-year balance:

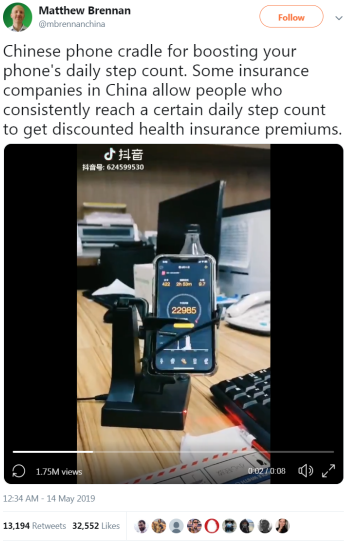
- How much predictability one needs depends on the application:
  - ↪ If there are few opportunities to exploit the predictability, we will need strong signals.
  - ↪ If there are the opportunities are abundant and independent of each other, even weak signals could be highly valuable.

Fundamental Law of Active Investing

$$IR = IC \cdot \sqrt{\text{Breadth}}$$

Outline

Where is your data from?



Challenges for finance analytics

- "Too-big" data?
  - ↪ Engineering and modeling challenges
- Not-so-big data
  - ↪ Low signal-to-noise ratio (1 - 4% for 10s return prediction, < 0.1% daily)
  - ↪ Observations are not independent (1,000 pictures of cats vs. 1,000 small business loans)
  - ↪ Non-stationarity (arbitrages, competition)
  - ↪ Measurement error (accounting data, consumer surveys)
  - ↪ Selection bias (back testing)
  - ↪ Survivorship bias (Apple vs. Blackberry)

> *"The needle comes in an increasingly larger haystack."*
>
> *— Nassim Taleb*

- Big data ⇒ more bias?

Challenges for finance analytics

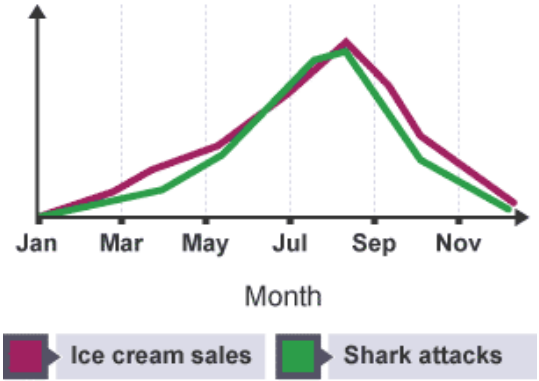
- What if the model is wrong?

> *"All models are wrong, but some are useful."*
>
> — *George E. P. Box*

  - ↪ How to select a good model? **Keep it simple, but not too simple.**
  - ↪ False confidence with a fragile model can be even more dangerous!
  - ↪ How can we still make good decisions when we know the model is possibly wrong?

- Correlation ≠ Causality
  - ↪ And sometimes causality really matters.

Correlation $\not\Rightarrow$ Causality



Isn't correlation enough?

- Prediction is an essential tool for finance:
  - ↪ Valuation
  - ↪ Trading
  - ↪ Risk management
- Isn't correlation enough? In finance applications:
  - ↪ We might care about measuring the magnitude of (marginal) effects.
  - ↪ We might care about model interpretability.
  - ↪ We should care about causality (the foundation of sound quantitative investing).
  - ↪ How to answer the "What if ..." questions?
  - ↪ How to design/improve a policy?
  - ↪ Different from a course in analytics and machine learning.

Example: option pricing

Option pricing model

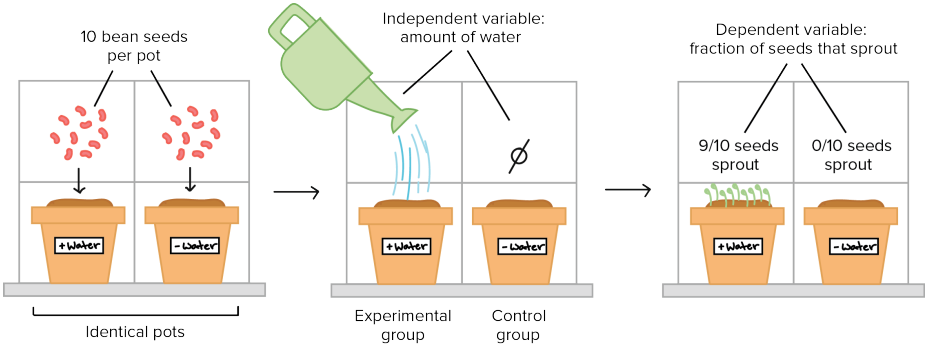
Model I: option price = f(stock price, int rate, past return, volatility, volume, $\cdots$)

Model II: option price = $BSM(S_t, K, \sigma, r, T)$

- Which model is better?
  - ↪ What would be the price of the same option tomorrow?
  - ↪ What would be the price of an option with different maturity?
  - ↪ What if a new option product starts trading today?
  - ↪ What if the exchange implements a short-sale ban on the stocks?

How can we establish causality?

- Does water help the bean seeds sprout?



How effective is a new drug?

Y_i^T = average weight change of individuals in the group receiving treatment

Y_i^C = average weight change of individuals in the group without treatment

$$\text{average treatment effect (ATE)} = E\left[Y_i^T - Y_i^C\right]$$

- Complication: Are those being treated the same as those who are not?

$$\begin{aligned} ATE &= E\left[Y_i^T|T\right] - E\left[Y_i^C|C\right] \\ &= \left(E\left[Y_i^T|T\right] - E\left[Y_i^C|T\right]\right) + \left(E\left[Y_i^C|T\right] - E\left[Y_i^C|C\right]\right) \\ &= \underbrace{E\left[Y_i^T - Y_i^C|T\right]}_{\text{Treatment Effect}} + \underbrace{\left(E\left[Y_i^C|T\right] - E\left[Y_i^C|C\right]\right)}_{\text{Selection Bias}} \end{aligned}$$

Outline

Example: CargoMetrics

Strategies based on alternative data

- Maritime data from the global Automatic Identification System (AIS)
  - ↪ Since 2004, vessels with 300 or more gross tonnage are required to flash AIS positioning signals every few seconds to avoid collisions.
  - ↪ CargoMetrics gains access to such signals through satellite companies.

Summary

What should you expect from this course?

1. Techniques for building financial models and analyzing financial data.
   ↪ Get your hands dirty.
2. Intuition based on a deep level of understanding of the theory.
   ↪ "The Grandma Test"
3. Bridge the gap between theory and practice.
   ↪ Think big!

Readings

- Hernandez et al. (2010): Unleashing the Power of Public Data for Financial Risk Measurement, Regulation, and Governance
- Anderson (2008): The end of theory: The Data Deluge makes the scientific method obsolete
- Harford (2014): Big data: are we making a big mistake?