

A Statistics Tale

The Zeroth Problem, Building Dams, Black Swans

Chris Comiskey

Open Data Group

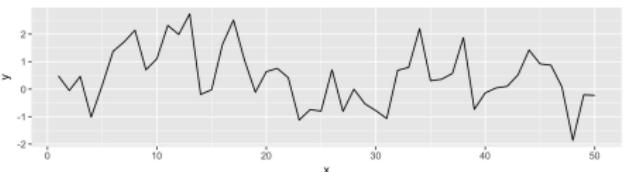
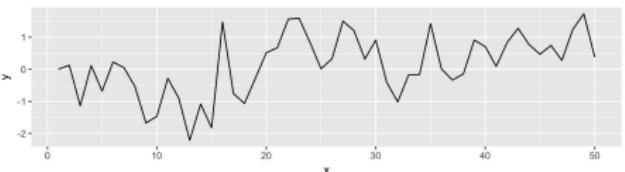
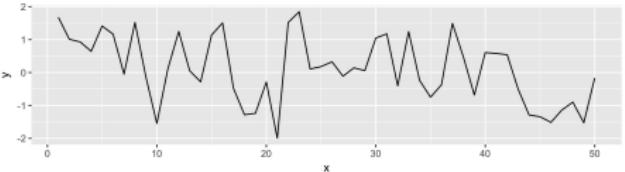
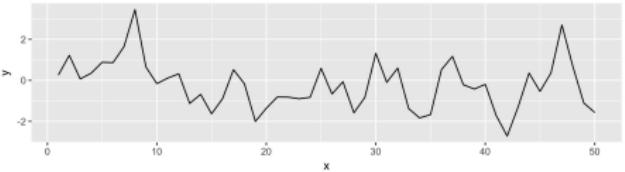
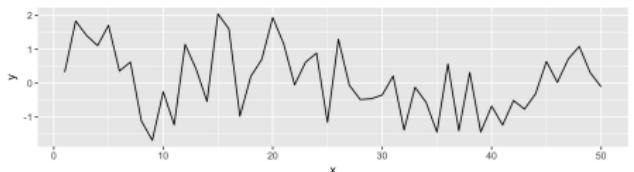
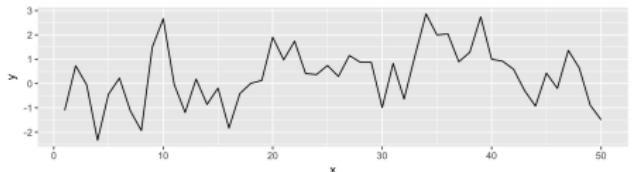
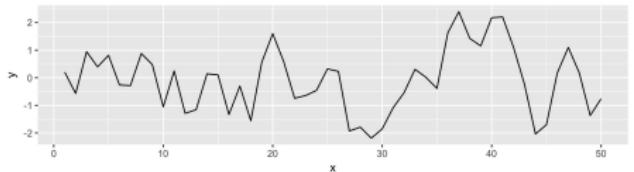
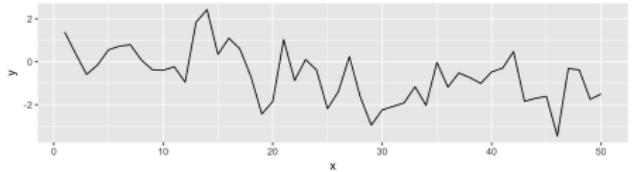
March 21, 2018

“The Zeroth Problem”

Mallows, 1998

- “Formulate questions that can be addressed by statistical methods.”

Which Three Have Trend?

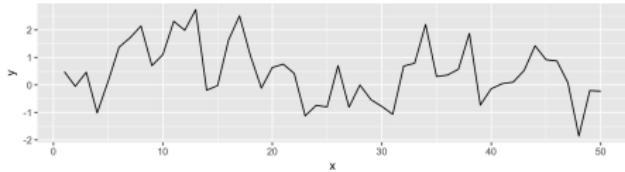
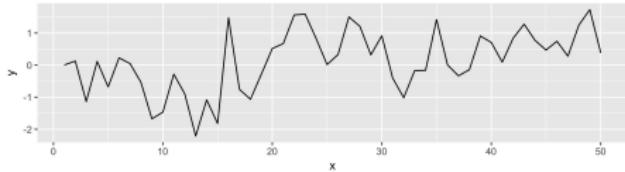
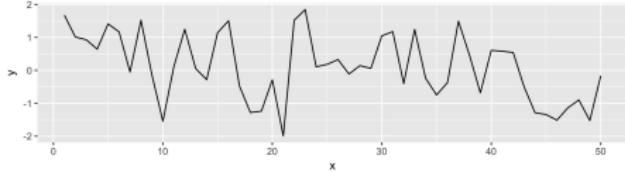
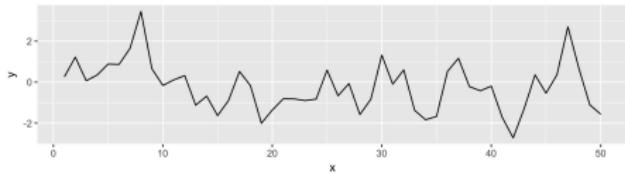
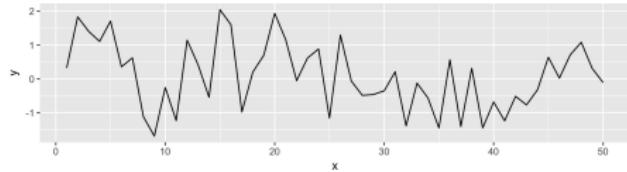
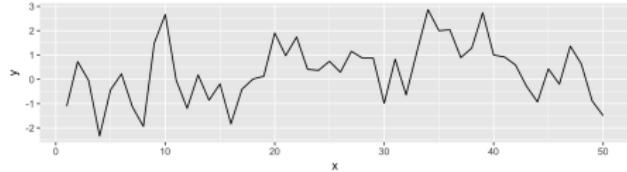
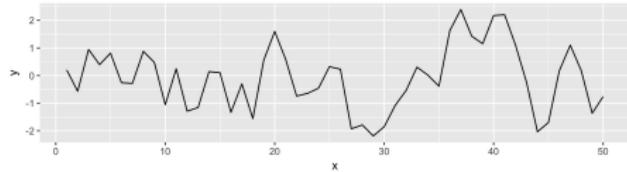
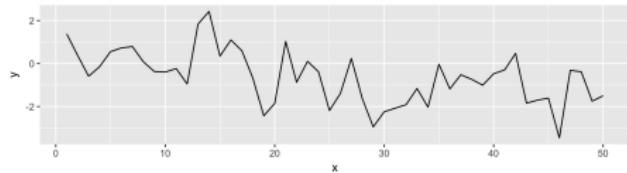


AR(p) Time Series, p = 1

- $X_t = \alpha X_{t-1} + \epsilon_t$
- $|\alpha| < 1 \implies$ stationary
- Typically: $\epsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$
- $\text{Var}(X_t) = \frac{\sigma^2}{1-\alpha^2}$
- $\rho(h) = \alpha^h$, for lag = h

Which Three Have Trend?

None: $\alpha = 0.5$



Nonstationarity versus scaling in hydrology

Demetris Koutsoyiannis

Most common time series approach (hydrologic – Eagleson, 1970) assumes two main components:

- ① A deterministic part → trend (nonstationary)
- ② A non-deterministic, random part (stationary)

Nonstationarity versus scaling* in hydrology

Demetris Koutsoyiannis

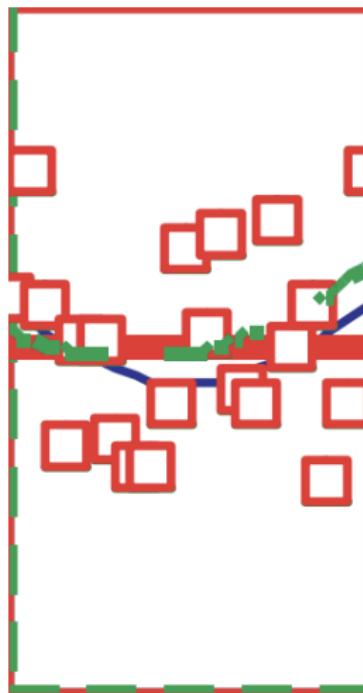
Most common modeling approach:

(business statistics: Freund et al. 1988, p. 583 - 598)

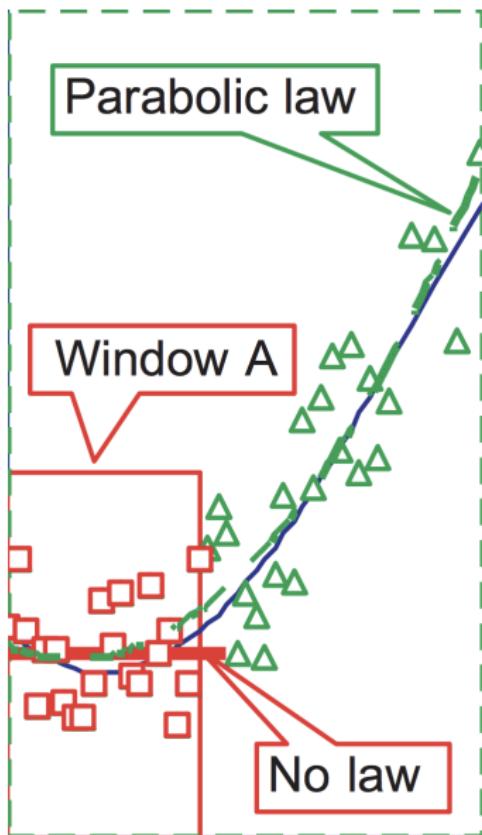
- ① Statistical tests to diagnose: trend vs. no trend
- ② Only continue if trend, otherwise model as random
- ③ Fit function to trend
- ④ Assume modeled trend is deterministic (some law)
- ⑤ ‘Detrend’, model remaining as random process, assess uncertainty therein

* Remind me later if you want to know his scaling hypothesis.

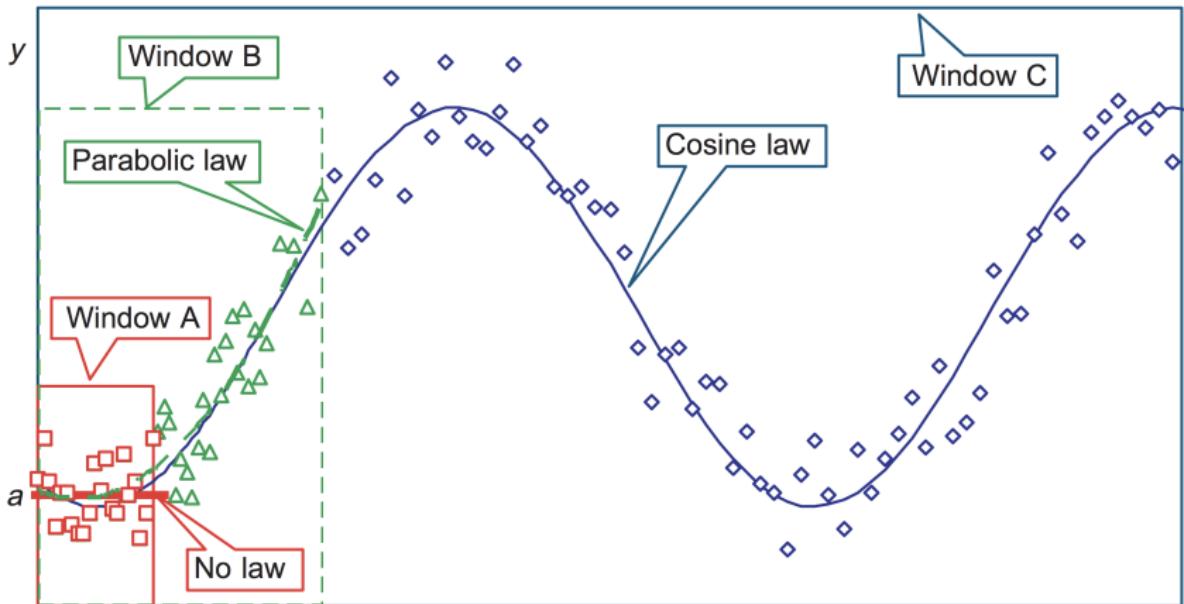
What's the Problem?



What's the Problem?

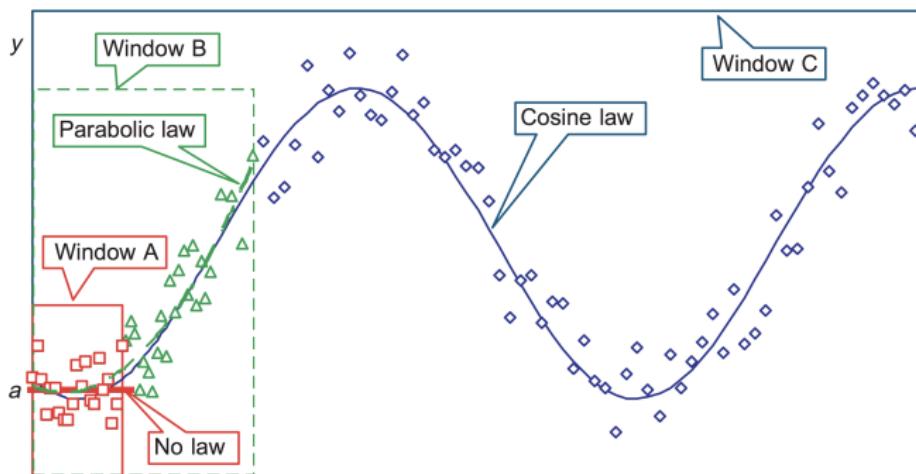


What's the Problem?



Estimating Uncertainty is the Problem.

The deterministic law is unknown, and treating it as such seriously underestimates the volatility of the system.

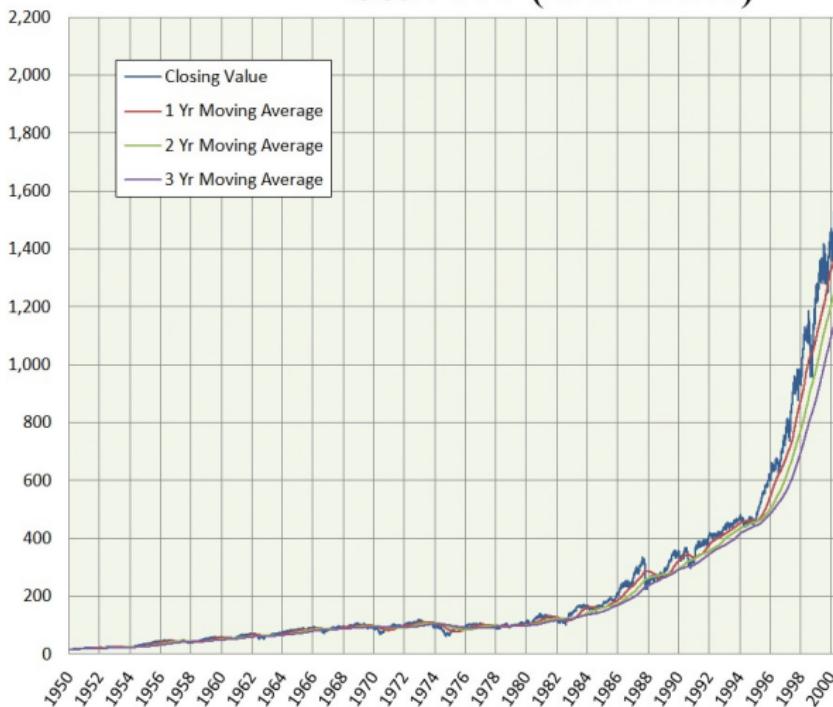


What is going to happen next?

How sure are we?

What law is governing this nonstationary trend?

S&P 500 (1950-2016)

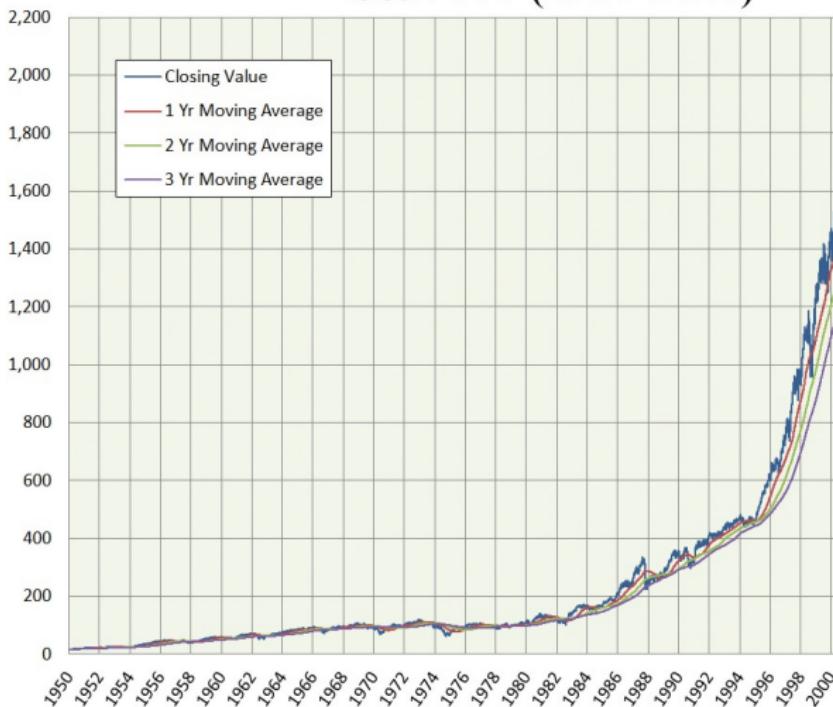


What is going to happen next?

How sure are we?

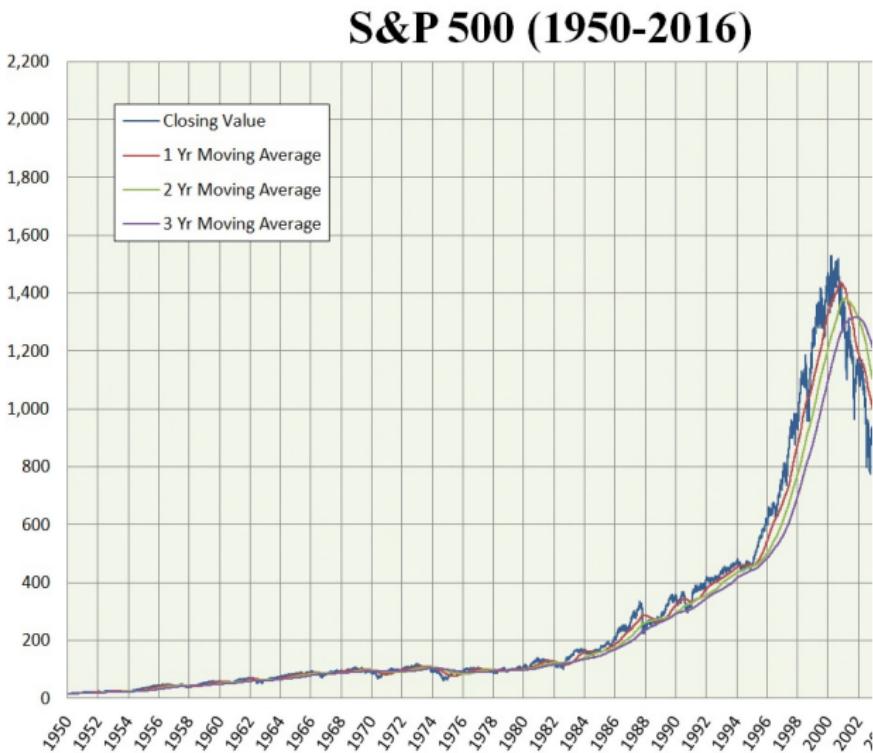
What would a neural net say is going to happen next?

S&P 500 (1950-2016)



What happened?!

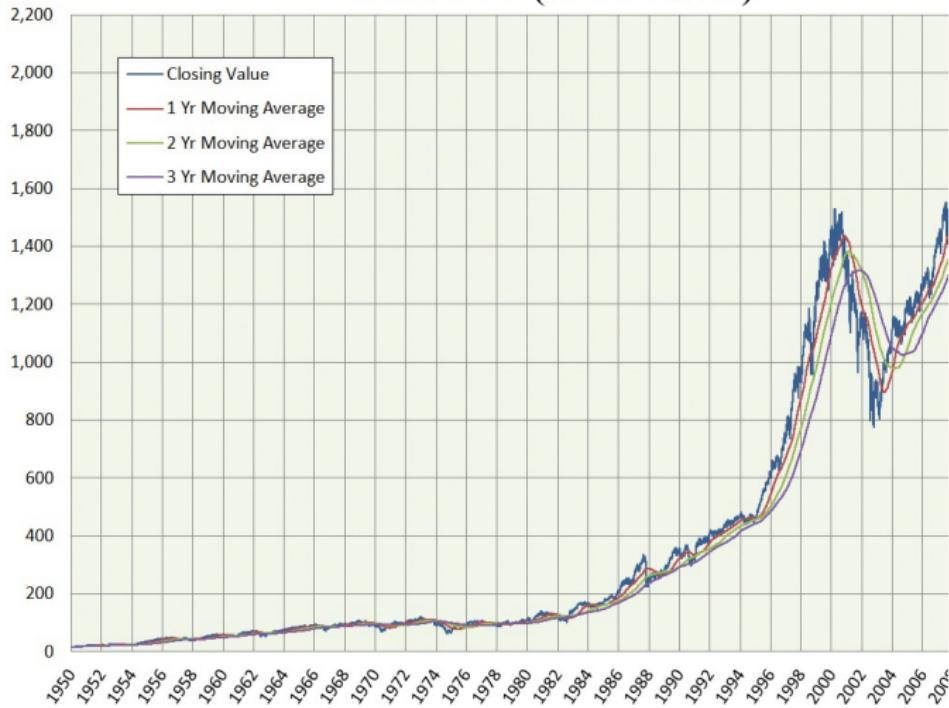
Would a machine learning algorithm have predicted this?



What happened?!

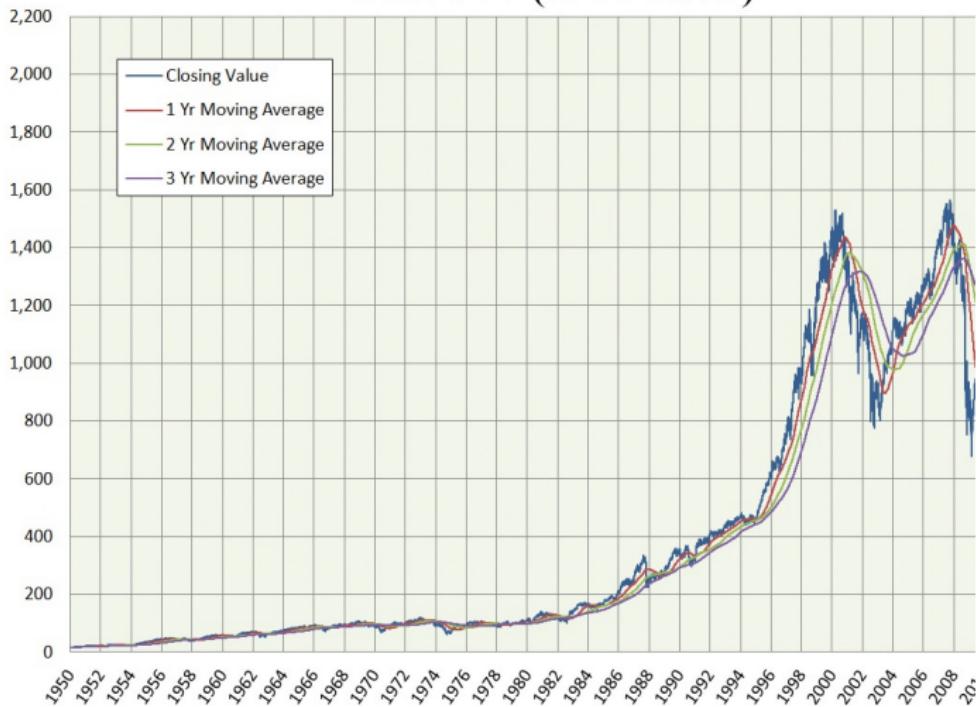
Or this?

S&P 500 (1950-2016)



Roller Coaster Algorithm?

S&P 500 (1950-2016)



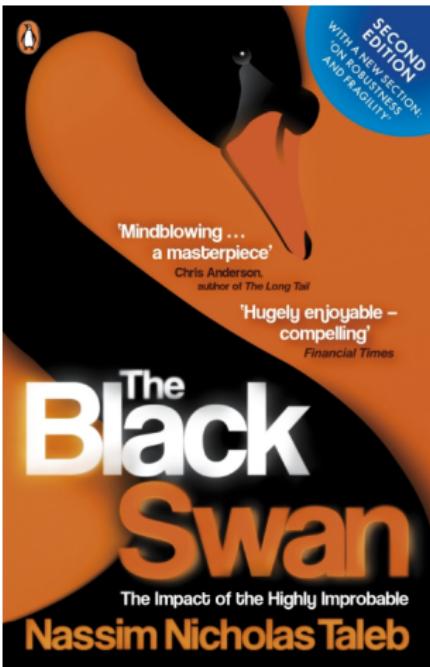
Black Swans

S&P 500 (1950-2016)



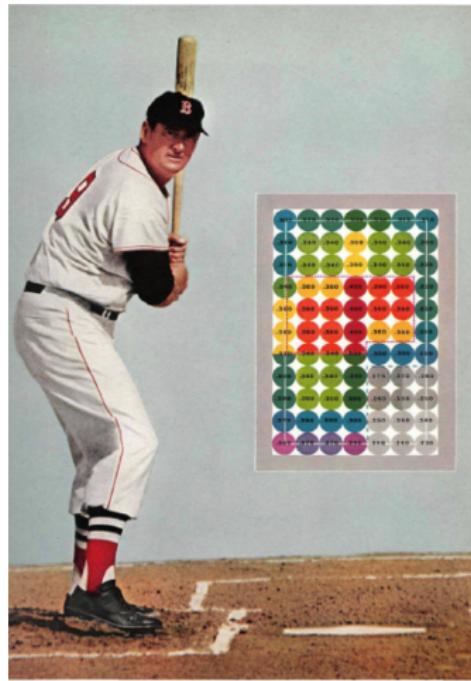
The Black Swan, Nassim Taleb, Empirica Capital

- The Black Swan
 - ▶ The concept
 - ▶ The book
- Nassim Taleb
 - ▶ LSE
 - ▶ Jerk
- Empirica Capital
 - ▶ Malcom Gladwell, New Yorker, “Blowing Up”
 - ▶ Confusing name
 - ▶ Human nature, results bias



Everything Relates Back to Baseball

"Get a good ball to hit." -Rogers Hornsby



- Even Ted Williams had to swing at good pitches.
- New models and technology
→ still need good problem
- Recall Zeroth problem