# How is math used outside academia?

Cathy O'Neil
mathbabe.org

# Answers:

- Climate modeling
- Financial modeling
- Medical research
- Biology modeling
- Sports modeling
- Healthcare costs modeling

# Answers:

- Cryptography
- Advertising (purchase/click prediction)
- Recommendation engines
- Fraud detection of all kinds
- Geophysics (predicting oil accumulation)
- Defense modeling

# Answers:

- Image processing
- Handwriting recognition
- Language recognition
- Materials simulation
- Teacher evaluations
- The h-score for published researchers

# But for this talk...

- I’m focusing on predictive models
- 

# But for this talk...

- I'm focusing on predictive models
- Is this math?
- 

# But for this talk...

- I’m focusing on predictive models
- Is this math?
- What mathematicians in industry do
-

# But for this talk...

- I’m focusing on predictive models
- Is this math?
- What mathematicians in industry do
- Public face of math (besides calculus)

# What is a model?

- Something that takes data in

-

# What is a model?

- Something that takes data in
- And a toy model of how things are related
-

# What is a model?

- Something that takes data in
- And a toy model of how things are related
- Gives out prediction

- 

# What is a model?

- Something that takes data in
- And a toy model of how things are related
- Gives out prediction
- *Should come with* an evaluation method
- 

# What is a model?

- Something that takes data in
- And a toy model of how things are related
- Gives out prediction
- *Should come with* an evaluation method
- Incredibly sensitive to manipulation

# Why should you care?

- Models are powerful
-

# Why should you care?

- Models are powerful
- But they are not oracles

- 

# Why should you care?

- Models are powerful
- But they are not oracles
- They rely on trust people have of math
- 

# Why should you care?

- Models are powerful
- But they are not oracles
- They rely on trust people have of math
- The authority of the inscrutable
-

# Why should you care?

- Models are powerful
- But they are not oracles
- They rely on trust people have of math
- The authority of the inscrutable
- The mathematician as super human

# Why you should care

- Conflict of interest or disinterest?

-

# Why you should care

- Conflict of interest or disinterest?
- Largely used to manipulate politics
-

# Why you should care

- Conflict of interest or disinterest?
- Largely used to manipulate politics
- Mathematicians are generally moral
-

# Why you should care

- Conflict of interest or disinterest?
- Largely used to manipulate politics
- Mathematicians are generally moral
- We shouldn't let this happen

# Salient properties

- Name
-

# Salient properties

- Name
- Underlying model
-

# Salient properties

- Name
- Underlying model
- Underlying assumptions

-

# Salient properties

- Name
- Underlying model
- Underlying assumptions
- Input/output
- 

# Salient properties

- Name
- Underlying model
- Underlying assumptions
- Input/output
- Purported/political goal

# Salient properties

- Name
- Underlying model
- Underlying assumptions
- Input/output
- Purported/political goal
- Evaluation method
-

# Salient properties

- Name
- Underlying model
- Underlying assumptions
- Input/output
- Purported/political goal
- Evaluation method
- Gaming potential
-

# Salient properties

- Name
- Underlying model
- Underlying assumptions
- Input/output
- Purported/political goal
- Evaluation method
- Gaming potential
- Reach

# Ex 1: VaR

- Name: Value-at-Risk
-

# Ex 1: VaR

- Name: Value-at-Risk
- Underlying model: Monte Carlo engine, ranking by P&L loss, percentile
-

# Calculating VaR

- Build a covariance matrix M of log returns of N instruments in your portfolio
-

# Calculating VaR

- Build a covariance matrix M of log returns of N instruments in your portfolio
- Typically with a decay of a few days or weeks - EWMA
- e

# Example: EWMA for variance depends on "decay factor" s

Next, assume we have the current variance estimate as

$$V_{old} = (1-s) \cdot \sum_i r_i^2 s^i$$

and we have a new return $r_0$ to add to the series. Then it's not hard to show we just want

$$V_{new} = s \cdot V_{old} + (1-s) \cdot r_0^2.$$

# Calculating VaR

- Build a covariance matrix M of log returns of N instruments in your portfolio
- Typically with a decay of a few days or weeks - EWMA
- Starting with N independent normal draws, make them correlated using M
-

# Calculating VaR

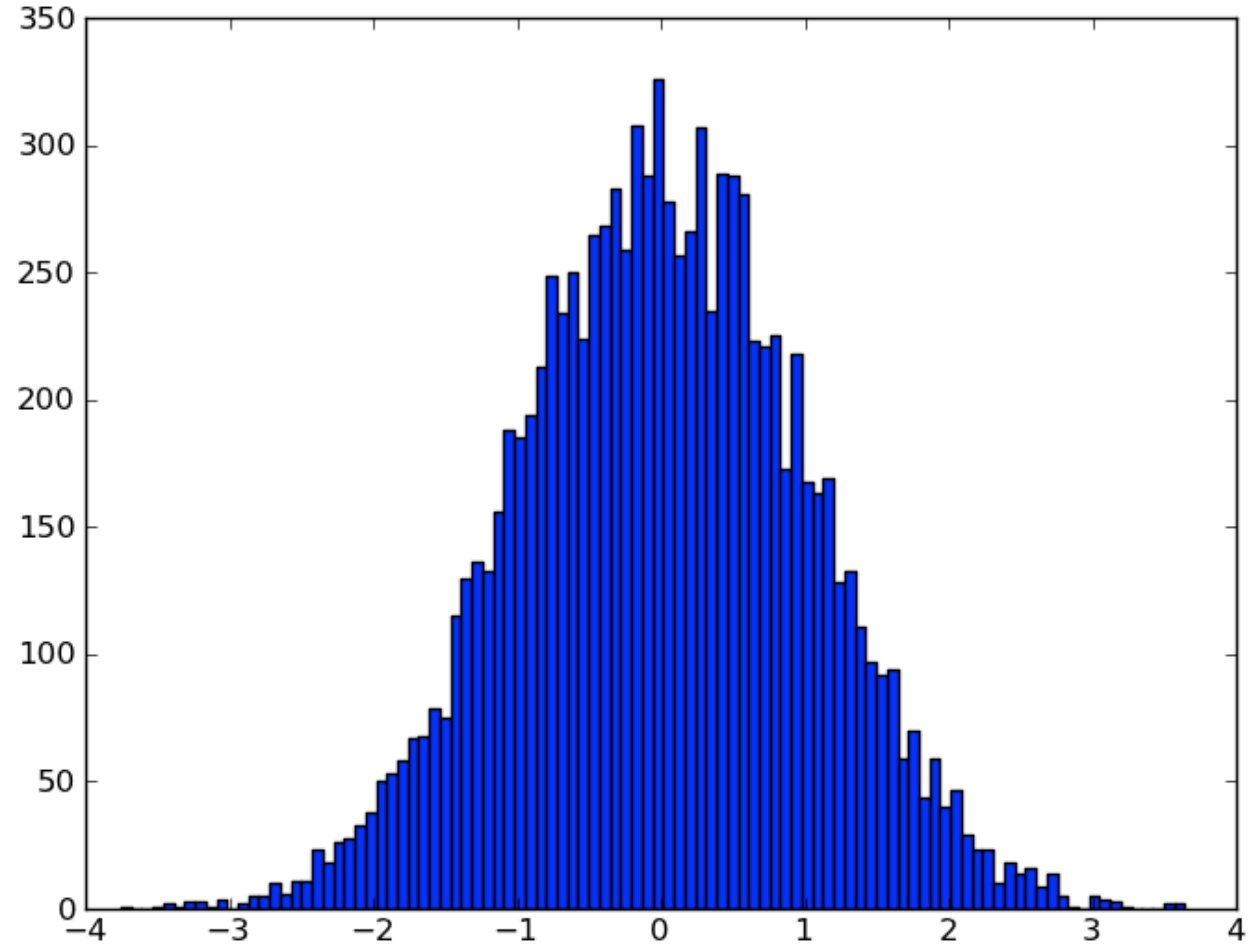
- Build a covariance matrix M of log returns of N instruments in your portfolio
- Typically with a decay of a few days or weeks - EWMA
- Starting with N independent normal draws, make them correlated using M
- Calculate PnL based on those log returns
-

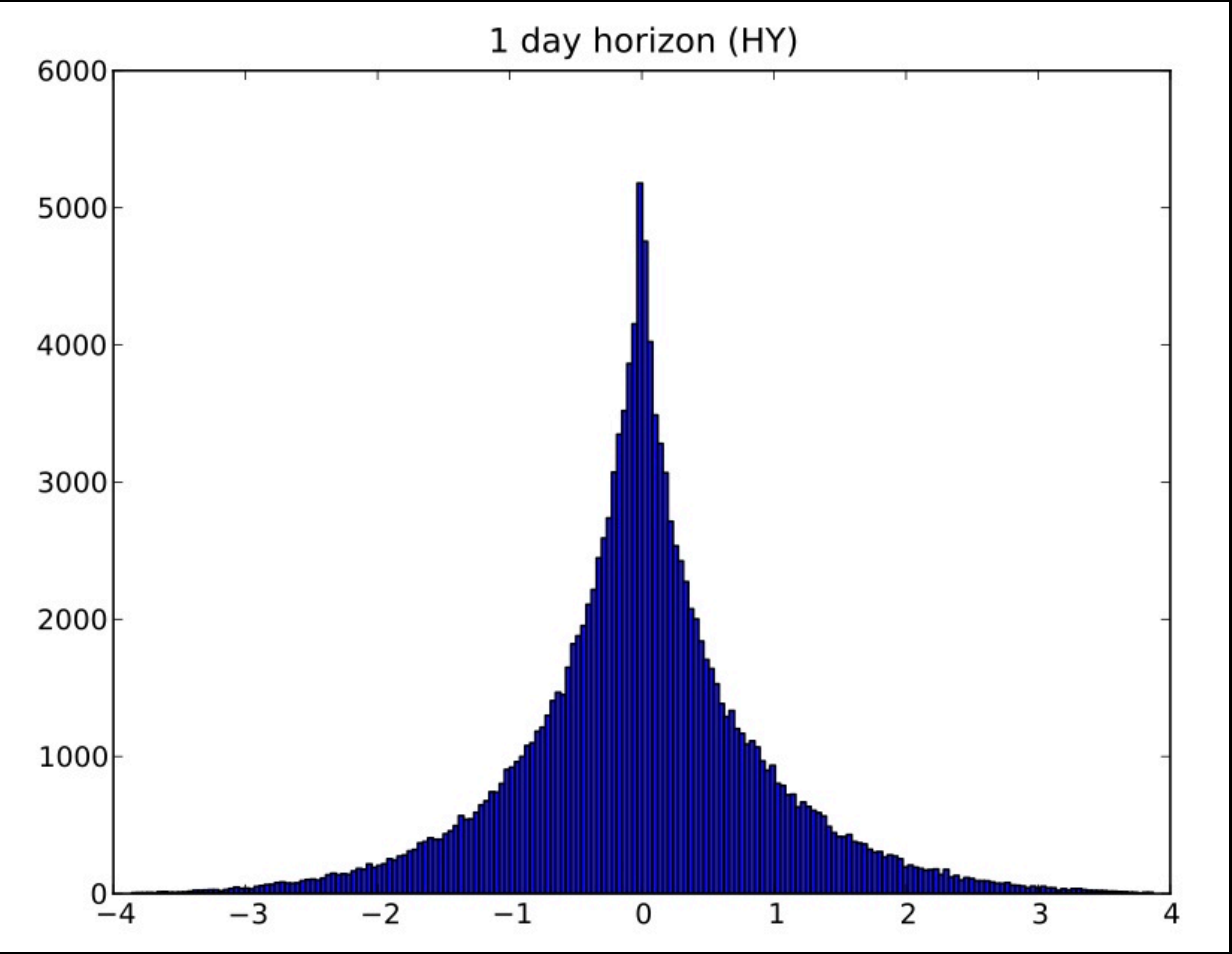
# Calculating VaR

- Build a covariance matrix M of log returns of N instruments in your portfolio
- Typically with a decay of a few days or weeks - EWMA
- Starting with N independent normal draws, make them correlated using M
- Calculate PnL based on those log returns
- Do this 1000 times, find 95th or 99th percentile

# Ex 1: VaR

- Name: Value-at-Risk
- Underlying model: Monte Carlo engine, ranking by P&L loss, percentile
- Underlying assumptions: Normal distributions, consistent correlations
-


350
300
250
200
150
100
50
0
−4
−3
−2
−1
0
1
2
3
4


1 day horizon (HY)
6000
5000
4000
3000
2000
1000
0
−4
−3
−2
−1
0
1
2
3
4

# Ex 1: VaR

- Name: Value-at-Risk
- Underlying model: Monte Carlo engine, ranking by P&L loss, percentile
- Underlying assumptions: Normal distributions, consistent correlations
- Input/output: Daily market data, misaligned time zones, proxies. Single #.
-

# Ex 1: VaR

- Name: Value-at-Risk
- Underlying model: Monte Carlo engine, ranking by P&L loss, percentile
- Underlying assumptions: Normal distributions, consistent correlations
- Input/output: Daily market data, misaligned time zones, proxies. Single #.
- Purported/political goal: Risk measurement/ CEO & regulators image
-

# Ex 1: VaR

- Name: Value-at-Risk
- Underlying model: Monte Carlo engine, ranking by P&L loss, percentile
- Underlying assumptions: Normal distributions, consistent correlations
- Input/output: Daily market data, misaligned time zones, proxies. Single #.
- Purported/political goal: Risk measurement/ CEO & regulators image
- Evaluation method: Not hard - extravagant failure in 2008
-

# Ex 1: VaR

- Name: Value-at-Risk
- Underlying model: Monte Carlo engine, ranking by P&L loss, percentile
- Underlying assumptions: Normal distributions, consistent correlations
- Input/output: Daily market data, misaligned time zones, proxies. Single #.
- Purported/political goal: Risk measurement/ CEO & regulators image
- Evaluation method: Not hard - extravagant failure in 2008
- Gaming potential: See Whale, London, among others.
-

# Ex 1: VaR

- Name: Value-at-Risk
- Underlying model: Monte Carlo engine, ranking by P&L loss, percentile
- Underlying assumptions: Normal distributions, consistent correlations
- Input/output: Daily market data, misaligned time zones, proxies. Single #.
- Purported/political goal: Risk measurement/ CEO & regulators image
- Evaluation method: Not hard - extravagant failure in 2008
- Gaming potential: See Whale, London, among others.
- Reach: The entire financial system

# Ex 2: VAM

- Name: Value-Added Teacher model
- 

# Ex 2: VAM

- Name: Value-Added Teacher model
- Underlying model: How much teacher raised scores vs. expectation
- 

# This is called "Counterfactual"

- In other words, the underlying model tries to predict what the score of a given student would be in a "random" class
- Takes into account student-level, classroom-level, and teacher-level attributes
- Hard to know how accurate this is!

# Ex 2: VAM

- Name: Value-Added Teacher model
- Underlying model: How much teacher raised scores vs. expectation
- Underlying assumptions: Account for externalities, small errorbars
- 

# Short list of sources of errors in VAM

- Need to score test
- Some problems harder than others?
- Some years smarter than others?
- Some tests harder than others?
- Normalized differently for different years
- Correlation of errors
- Model error
- Bayesian “shrinkage”

# Accounting for externalities in VAM

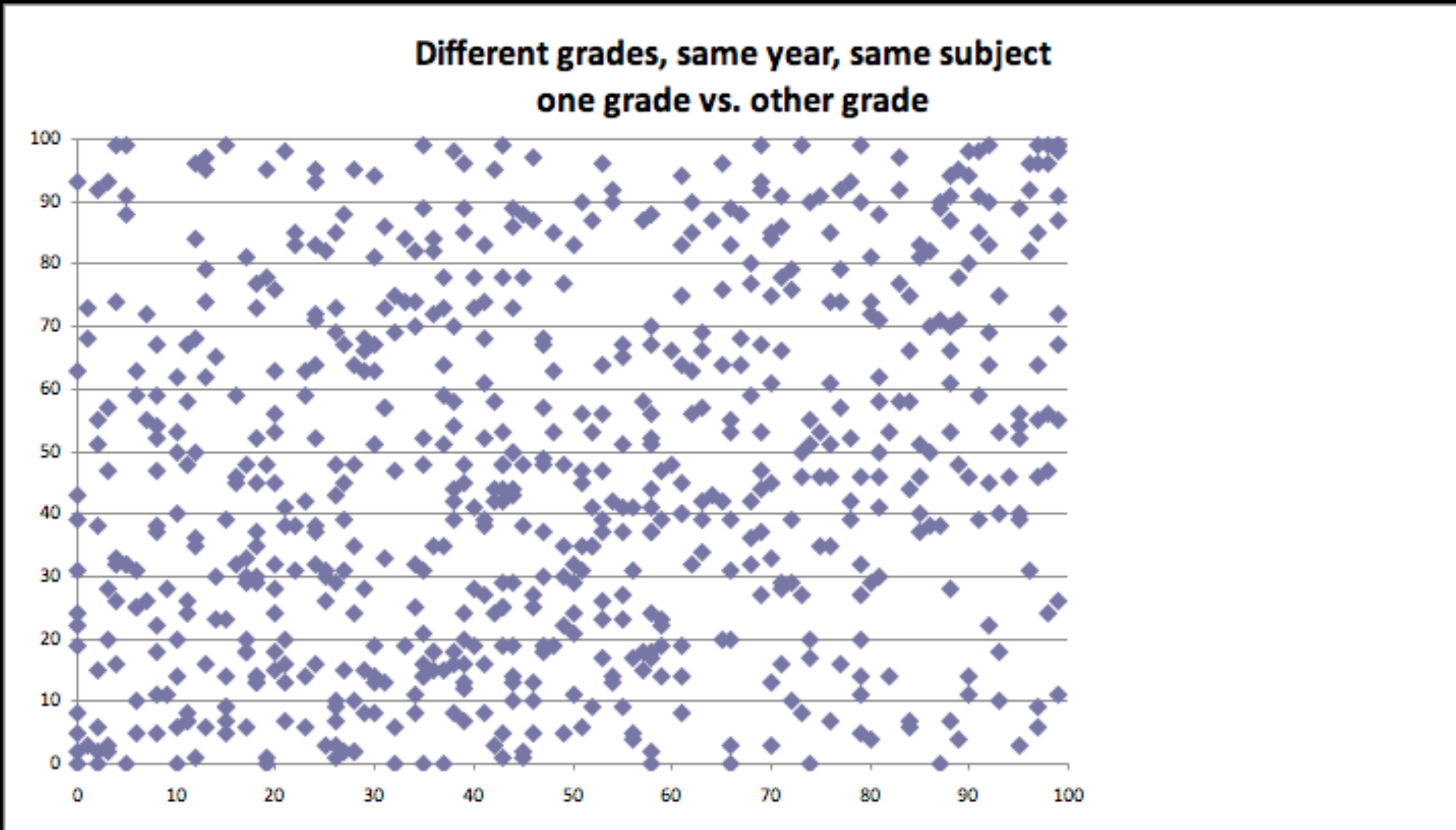
- Account for what is “under control”
- Tests better at testing middle than ends
- % of free school lunches very fat tailed
- Summer vacation loss
- “no child left behind” mindset
- Punishes teachers at tough schools

# Ex 2: VAM

- Name: Value-Added Teacher model
- Underlying model: How much teacher raised scores vs. expectation
- Underlying assumptions: Account for externalities, small errorbars
- Input/output: Student standardized test scores, attributes. Single #
- 

# The underlying model

- Linear regression with multiple sub-models
- Opaque correction terms and techniques
- Small samples (by grade, subject, year)
- Lots of missing data
- 14% correlation on NYC teachers


Different grades, same year, same subject
one grade vs. other grade
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100

# Ex 2: VAM

- Name: Value-Added Teacher model
- Underlying model: How much teacher raised scores vs. expectation
- Underlying assumptions: Account for externalities, small errorbars
- Input/output: Student standardized test scores, attributes. Single #
- Purported/political goal: Better teacher/ power, privatization
-

# Ex 2: VAM

- Name: Value-Added Teacher model
- Underlying model: How much teacher raised scores vs. expectation
- Underlying assumptions: Account for externalities, small errorbars
- Input/output: Student standardized test scores, attributes. Single #
- Purported/political goal: Better teacher/ power, privatization
- Evaluation method: **None!**
-
-

# Ex 2: VAM

- Name: Value-Added Teacher model
- Underlying model: How much teacher raised scores vs. expectation
- Underlying assumptions: Account for externalities, small errorbars
- Input/output: Student standardized test scores, attributes. Single #
- Purported/political goal: Better teacher/ power, privatization
- Evaluation method: **None!**
- Gaming potential: Cheating, etc. - mostly gamed by administrators
-

# Ex 2: VAM

- Name: Value-Added Teacher model
- Underlying model: How much teacher raised scores vs. expectation
- Underlying assumptions: Account for externalities, small errorbars
- Input/output: Student standardized test scores, attributes. Single #
- Purported/political goal: Better teacher/ power, privatization
- Evaluation method: **None!**
- Gaming potential: Cheating, etc. - mostly gamed by administrators
- Reach: LA, NY, Chicago public school systems...

# Ex 3: Credit Scores

- Name: Credit Score
-

# Ex 3: Credit Scores

- Name: Credit Score
- Underlying model: Unknown, but takes into account past paid bills etc.
-

# Ex 3: Credit Scores

- Name: Credit Score
- Underlying model: Unknown, but takes into account past paid bills etc.
- Underlying assumptions: Behavior consistent over time
- 

# Ex 3: Credit Scores

- Name: Credit Score
- Underlying model: Unknown, but takes into account past paid bills etc.
- Underlying assumptions: Behavior consistent over time
- Input/output: Regulated, open to consumers for free 1x per year. Single #
-

# Ex 3: Credit Scores

- Name: Credit Score
- Underlying model: Unknown, but takes into account past paid bills etc.
- Underlying assumptions: Behavior consistent over time
- Input/output: Regulated, open to consumers for free 1x per year. Single #
- Purported/political goal: Measurement/ buy-in, fear of default, quantification
-

# Ex 3: Credit Scores

- Name: Credit Score
- Underlying model: Unknown, but takes into account past paid bills etc.
- Underlying assumptions: Behavior consistent over time
- Input/output: Regulated, open to consumers for free 1x per year. Single #
- Purported/political goal: Measurement/ buy-in, fear of default, quantification
- Evaluation method: Constant but not public. Example of death spiral.
-

# Ex 3: Credit Scores

- Name: Credit Score
- Underlying model: Unknown, but takes into account past paid bills etc.
- Underlying assumptions: Behavior consistent over time
- Input/output: Regulated, open to consumers for free 1x per year. Single #
- Purported/political goal: Measurement/ buy-in, fear of default, quantification
- Evaluation method: Constant but not public. Example of death spiral.
- Gaming potential: Not high
-

# Ex 3: Credit Scores

- Name: Credit Score
- Underlying model: Unknown, but takes into account past paid bills etc.
- Underlying assumptions: Behavior consistent over time
- Input/output: Regulated, open to consumers for free 1x per year. Single #
- Purported/political goal: Measurement/ buy-in, fear of default, quantification
- Evaluation method: Constant but not public. Example of death spiral.
- Gaming potential: Not high
- Reach: National, possibly international

# Aside: the death spiral of modeling

- Insurance: pooled risk
-

# Aside: the death spiral of modeling

- Insurance: pooled risk
- Add segmentation/ good health modeling

-

# Aside: the death spiral of modeling

- Insurance: pooled risk
- Add segmentation/ good health modeling
- Lose original goal

-

# Aside: the death spiral of modeling

- Insurance: pooled risk
- Add segmentation/ good health modeling
- Lose original goal
- People who benefit don’t see the problem directly
-

# Aside: the death spiral of modeling

- Insurance: pooled risk
- Add segmentation/ good health modeling
- Lose original goal
- People who benefit don't see the problem directly
- Same can be said for credit card offers via credit scoring
- 

# Aside: the death spiral of modeling

- Insurance: pooled risk
- Add segmentation/ good health modeling
- Lose original goal
- People who benefit don't see the problem directly
- Same can be said for credit card offers via credit scoring
- In general if someone benefits someone loses
-

# Aside: the death spiral of modeling

- Insurance: pooled risk
- Add segmentation/ good health modeling
- Lose original goal
- People who benefit don't see the problem directly
- Same can be said for credit card offers via credit scoring
- In general if someone benefits someone loses
- Systematized racism etc.
-

# Aside: the death spiral of modeling

- Insurance: pooled risk
- Add segmentation/ good health modeling
- Lose original goal
- People who benefit don't see the problem directly
- Same can be said for credit card offers via credit scoring
- In general if someone benefits someone loses
- Systematized racism etc.
- Philosophically, what do we want our culture to be?

# Ex 4: E-Score

- Name: E-Score (or “buying power” score)

-

# Ex 4: E-Score

- Name: E-Score (or “buying power” score)
- Underlying model: Unknown; takes into account past google searches etc.
-

# Ex 4: E-Score

- Name: E-Score (or “buying power” score)
- Underlying model: Unknown; takes into account past google searches etc.
- Underlying assumptions: Behavior consistent over time, correct ID
- 

# Ex 4: E-Score

- Name: E-Score (or "buying power" score)
- Underlying model: Unknown; takes into account past google searches etc.
- Underlying assumptions: Behavior consistent over time, correct ID
- Input/output: Unregulated, could use race, age, whatever
-

# Ex 4: E-Score

- Name: E-Score (or “buying power” score)
- Underlying model: Unknown; takes into account past google searches etc.
- Underlying assumptions: Behavior consistent over time, correct ID
- Input/output: Unregulated, could use race, age, whatever
- Purported/political goal: Measurement/ quantification, skimming $
-

# Ex 4: E-Score

- Name: E-Score (or “buying power” score)
- Underlying model: Unknown; takes into account past google searches etc.
- Underlying assumptions: Behavior consistent over time, correct ID
- Input/output: Unregulated, could use race, age, whatever
- Purported/political goal: Measurement/ quantification, skimming $
- Evaluation method: Death spiral, this time not regulated.
-

# Ex 4: E-Score

- Name: E-Score (or "buying power" score)
- Underlying model: Unknown; takes into account past google searches etc.
- Underlying assumptions: Behavior consistent over time, correct ID
- Input/output: Unregulated, could use race, age, whatever
- Purported/political goal: Measurement/ quantification, skimming $
- Evaluation method: Death spiral, this time not regulated.
- Gaming potential: Not high
- 

# Ex 4: E-Score

- Name: E-Score (or “buying power” score)
- Underlying model: Unknown; takes into account past google searches etc.
- Underlying assumptions: Behavior consistent over time, correct ID
- Input/output: Unregulated, could use race, age, whatever
- Purported/political goal: Measurement/ quantification, skimming $
- Evaluation method: Death spiral, this time not regulated.
- Gaming potential: Not high
- Reach: International

# Ex 4: h-index

- Name: E-Score (or "buying power" score)

-

# Ex 4: h-index

- Name: E-Score (or "buying power" score)
- Underlying model: max N where there are N papers with N citations
-

# Ex 4: h-index

- Name: E-Score (or "buying power" score)
- Underlying model: max N where there are N papers with N citations
- Underlying assumptions: papers and citations, and quantity, meaningful
- 

# Ex 4: h-index

- Name: E-Score (or “buying power” score)
- Underlying model: max N where there are N papers with N citations
- Underlying assumptions: papers and citations, and quantity, meaningful
- Input/output: Academic publishing records, single number
-

# Ex 4: h-index

- Name: E-Score (or “buying power” score)
- Underlying model: max N where there are N papers with N citations
- Underlying assumptions: papers and citations, and quantity, meaningful
- Input/output: Academic publishing records, single number
- Purported/political goal: Measurement, self-advancement
-

# Ex 4: h-index

- Name: E-Score (or “buying power” score)
- Underlying model: max N where there are N papers with N citations
- Underlying assumptions: papers and citations, and quantity, meaningful
- Input/output: Academic publishing records, single number
- Purported/political goal: Measurement, self-advancement
- Evaluation method: Fields vs. not?
-

# Ex 4: h-index

- Name: E-Score (or "buying power" score)
- Underlying model: max N where there are N papers with N citations
- Underlying assumptions: papers and citations, and quantity, meaningful
- Input/output: Academic publishing records, single number
- Purported/political goal: Measurement, self-advancement
- Evaluation method: Fields vs. not?
- Gaming potential: Highly vulnerable

# Ex 4: h-index

- Name: E-Score (or "buying power" score)
- Underlying model: max N where there are N papers with N citations
- Underlying assumptions: papers and citations, and quantity, meaningful
- Input/output: Academic publishing records, single number
- Purported/political goal: Measurement, self-advancement
- Evaluation method: Fields vs. not?
- Gaming potential: Highly vulnerable
- Reach: As far as h-score reaches

# Others

- Education - who will graduate
- Debt collectors - who will pay
- Political ads - uberpersonal targeting
- Health and DNA models

# Modeling physics vs. people

- There's a feedback loop for modeling
-

# Modeling physics vs. people

- There's a feedback loop for modeling
- Sometimes indicates the model is bad
-

# Modeling physics vs. people

- There's a feedback loop for modeling
- Sometimes indicates the model is bad
- "People models" ≠ "statistical models"

# Keep in mind

- You can’t manage what you don’t measure
-

# Keep in mind

- You can’t manage what you don’t measure
- What are we *not* quantifying for each ex?
-

# Keep in mind

- You can’t manage what you don’t measure
- What are we *not* quantifying for each ex?
- Should we be?

# Where do we go now?

- Defend math
-

# Where do we go now?

- Defend math
- First step: educate ourselves
-

# Where do we go now?

- Defend math
- First step: educate ourselves
- Anticipate gaming
- 

# Where do we go now?

- Defend math
- First step: educate ourselves
- Anticipate gaming
- Require transparent evaluation methods
- 

# Where do we go now?

- Defend math
- First step: educate ourselves
- Anticipate gaming
- Require transparent evaluation methods
- Let’s not become economists though

# Suggestions

- Referee process for public models
-

# Suggestions

- Referee process for public models
- Effort to make models “simple”

-

# Suggestions

- Referee process for public models
- Effort to make models “simple”
- Effort to educate public
-

# Suggestions

- Referee process for public models
- Effort to make models "simple"
- Effort to educate public
- Panels of mathematicians (& others)
-

# Suggestions

- Referee process for public models
- Effort to make models “simple”
- Effort to educate public
- Panels of mathematicians (& others)
- Don’t take money from industry for this