Lecture 03: Signal vs. Noise

Statistical Arbitrage

Mgmt 237M, Section 02

Professor Olivier Ledoit

University of California Los Angeles
Anderson School of Management
Master of Financial Engineering
Fall 2016

Wall Street Job Interview Question

- Russian Roulette

How to Estimate Transaction Costs?

Keim and Madhavan (JFE 1997)

- 62,000 stock orders
- 21 institutional traders
- Volume = $83 billion
- 1991 – 1993

Regression Model

Variable	Buyer-initiated orders		Seller-initiated orders		All orders	
	Estimate	Standard error	Estimate	Standard error	Estimate	Standard error
Intercept	0.767	0.325	0.505	0.449	0.687	0.269
D^{NASDAQ}	0.336	0.052	0.058	0.085	0.239	0.045
Trsize	0.092	0.016	0.214	0.030	0.165	0.005
Logmcap	− 0.084	0.019	− 0.059	0.027	− 0.076	0.016
$1/P_i$	13.807	1.356	6.537	1.482	9.924	1.029
D^{TECH}	0.492	0.050	0.718	0.049	0.607	0.035
D^{INDEX}	0.305	0.049	0.432	0.074	0.451	0.040
Adjusted R^2	0.046		0.086		0.060	

Simple Transaction Cost Model

- Commission + 1bp + median bid-ask spread / 2

- Good approximation to get started (small book)
- Ignores price impact
- In reality, t-cost depends on trade size
- As you grow book, build your own t-cost model

- This is why you must increase book size slowly!

Equity Market Impact

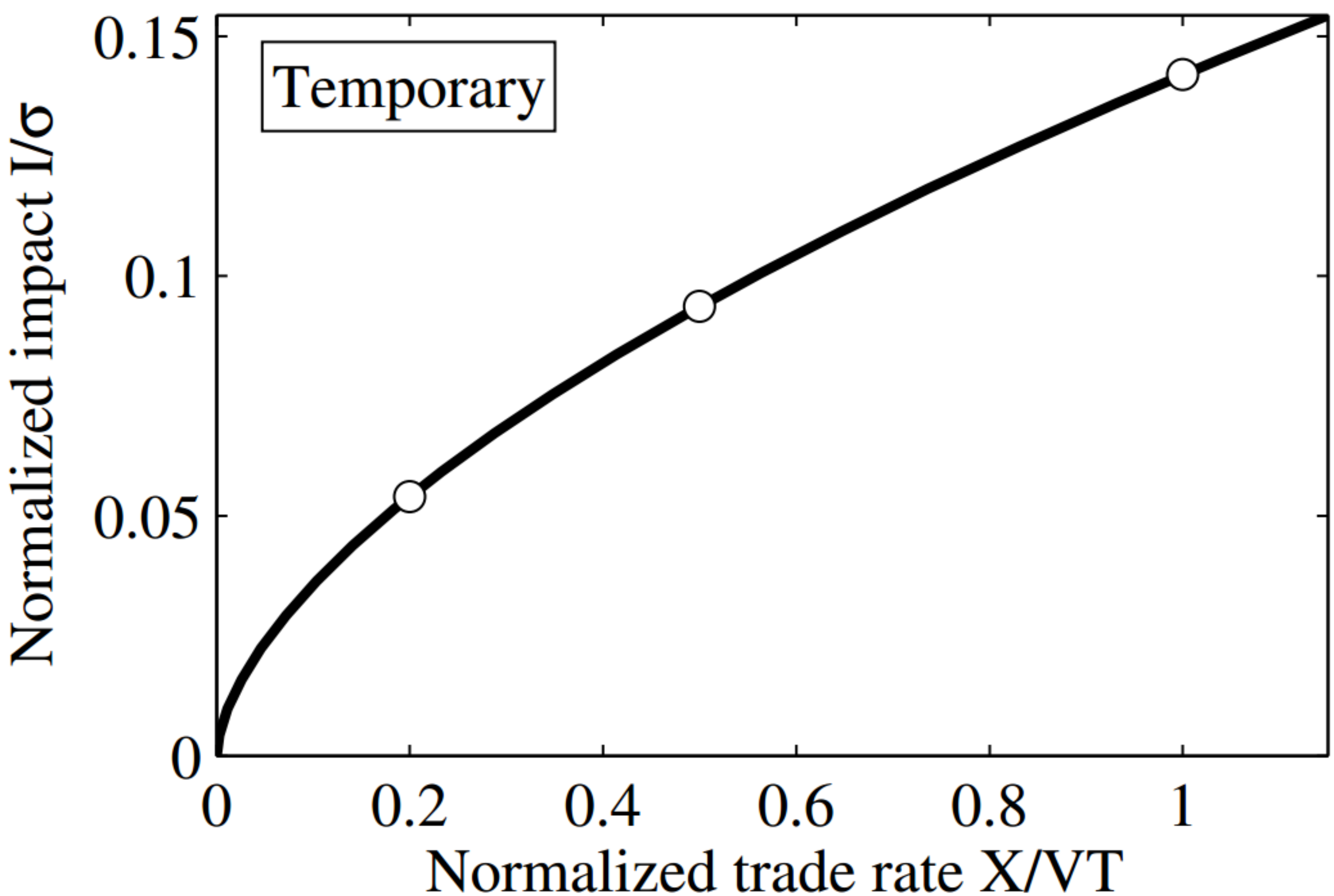
- Robert Almgren + 3 Citigroup Quants
- 700,000 US orders executed by Citigroup Equity Trading desks 12/2001 $\rightarrow$ 6/2003
- I = temporary price impact
- s = daily volatility
- X = trade size
- V = average daily volume
- T = trade duration (in days)

Market Impact Model

- I = temporary price impact
- σ = daily volatility
- X = trade size
- V = average daily volume
- T = trade duration (in days)

$$I\,/\,\sigma = \text{constant} \cdot \text{sign}(X) \cdot |X/VT|^{\beta} + \text{noise}$$

Price Impact in Power 3/5

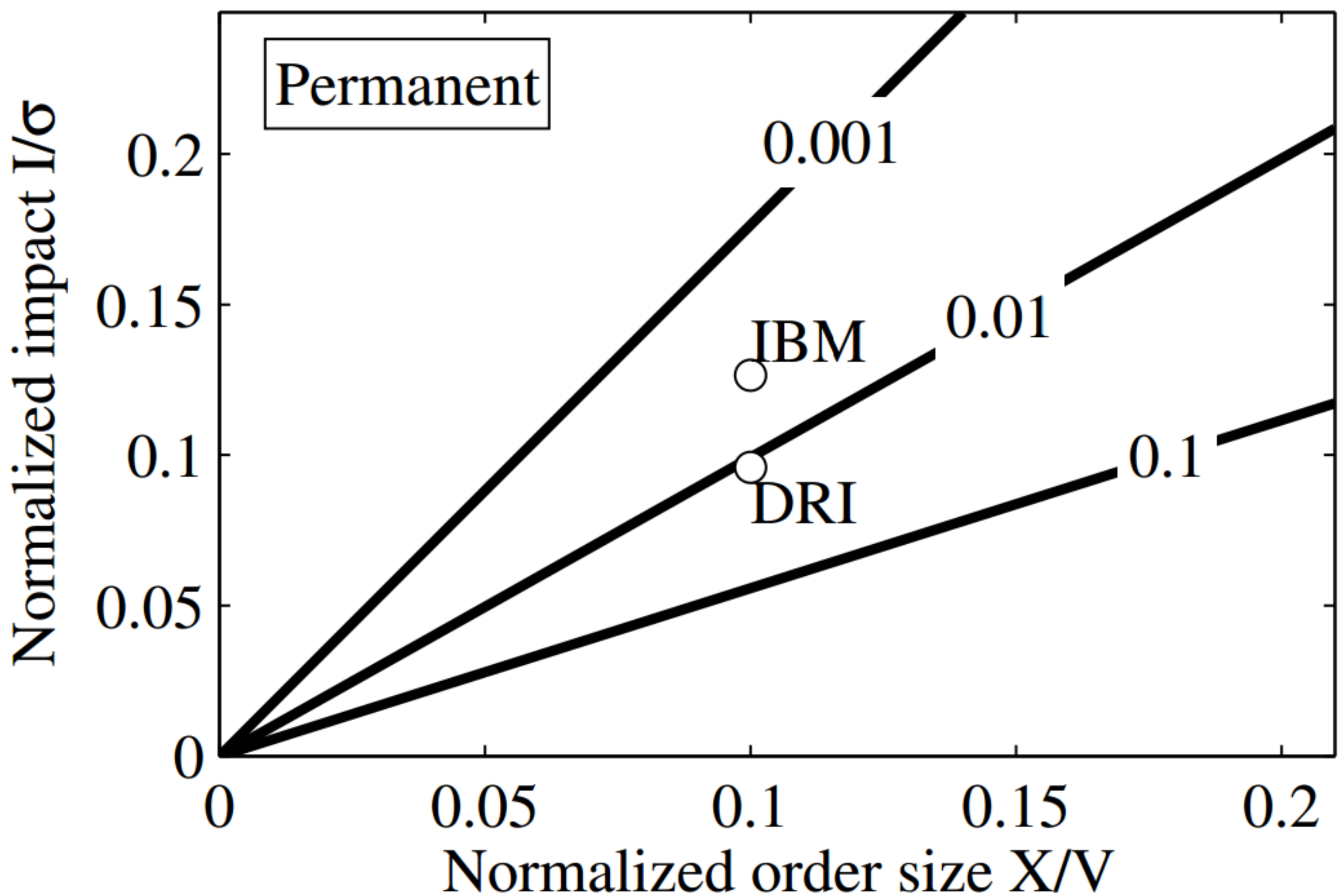


Permanent Price Impact

- I = permanent price impact
- Θ = shares outstanding
- X = trade size
- V = average daily volume

$$I/\sigma = \text{constant} \cdot (X/V) \cdot (\Theta/V)^{\delta} + \text{noise}$$

Permanent Impact in Power 1/4



Question of the Day

- How to make sense of quant trader or strategy track record?
- Distinguish skill from luck
- Quantitatively

- Learn statistical technique also useful to estimate the covariance matrix

Plan

1. Intuition
2. Formula
3. Estimate the formula
4. Applications

Goals per Game 2009-2010

		2009	2010	Change
W. Rooney	Man Utd	0.70		
F. Torres	Liverpool	0.60		
Agbonlahor	Aston Villa	0.40		
R. van Persie	Arsenal	0.37		
F. Lampard	Chelsea	0.30		
N. Anelka	Chelsea	0.25		
S. Gerrard	Liverpool	0.25		
Dirk Kuyt	Liverpool	0.25		
John Carew	Aston Villa	0.20		
Kevin Davies	Bolton	0.17		

Goals per Game 2009-2010

		2009	2010	Change
W. Rooney	Man Utd	0.70	0.67	↘
F. Torres	Liverpool	0.60	0.33	↘
Agbonlahor	Aston Villa	0.40	0.28	↘
R. van Persie	Arsenal	0.37	0.11	↘
F. Lampard	Chelsea	0.30	0.89	↗
N. Anelka	Chelsea	0.25	0.33	↗
S. Gerrard	Liverpool	0.25	0.22	↘
Dirk Kuyt	Liverpool	0.25	0.22	↘
John Carew	Aston Villa	0.20	0.33	↗
Kevin Davies	Bolton	0.17	0.20	↗

Past Performance Is Not...

Bollen and Busse (2004)

tion and market timing ability, Table 6 shows the results of the following cross-sectional regression of performance on its lagged value:

$$\text{Perf}_{p,t} = a + b\text{Perf}_{p,t-1} + \varepsilon_{p,t}, \tag{6}$$

where $\text{Perf}_{p,t}$ is either raw return or the contribution of active management to fund returns as defined above. A positive slope coefficient would indicate that past performance predicts the following period's perform-

Table 6
Cross-sectional regression tests of performance persistence

	Returns, R_p (%)	Stock selection, α_p (%)	Market timing (%)		Mixed (%)	
			TM	HM	TM	HM
A	0.044	−0.002	−0.003	−0.003	−0.002	−0.002
p-value	.006	.213	.160	.126	.164	.181
B	0.036	0.122	0.118	0.117	0.118	0.117
p-value	.502	.000	.000	.000	.000	.000
R^2	0.101	0.038	0.034	0.032	0.034	0.036

INADMISSIBILITY OF THE USUAL ESTIMATOR FOR THE MEAN OF A MULTIVARIATE NORMAL DISTRIBUTION

CHARLES STEIN
STANFORD UNIVERSITY

1. Introduction

If one observes the real random variables $X_1, \cdots, X_n$ independently normally distributed with unknown means $\xi_1, \cdots, \xi_n$ and variance 1, it is customary to estimate ξ_i by X_i. If the loss is the sum of squares of the errors, this estimator is admissible for $n \leqq 2$, but inadmissible for $n \geqq 3$. Since the usual estimator is best among those which transform correctly under translation, any admissible estimator for $n \geqq 3$ involves an arbitrary choice. While the results of this paper are not in a form suitable for immediate practical application, the possible improvement over the usual estimator seems to be large enough to be of practical importance if n is large.

Shrinkage

- Compute the grand mean (cross-sectionally)
- Shrink every estimator towards the grand mean
- Grand mean = shrinkage target

- What is the shrinkage slope?

Plan

1. Intuition
2. Formula
3. Estimate the formula
4. Applications

The Model

- Stage 1: God draws skill according to $N\left(\overline{\mu}, \delta^2\right)$
  Fund *i* has expected return $\mu_i \sim N\left(\overline{\mu}, \delta^2\right)$

- Stage 2: **Independently,** Lady Luck draws *T* observations around expected value μ_i with random error: $x_{ti} \sim N(\mu_i, T\omega^2)$

Shrinkage Target

- From the *T* observations $x_{1i}, \ldots, x_{Ti}$ we compute the sample mean: $$m_i = \frac{x_{1i} + \cdots + x_{Ti}}{T}$$
- From the *n* sample means we compute the **grand mean**: $$\overline{m} = \frac{m_1 + \cdots + m_n}{n}$$
- Shrink every sample mean towards grand mean:

$$\hat{m}_i = (1 - \beta)\underbrace{\overline{m}}_{\text{Shrinkage Target}} + \beta m_i$$

Shrinkage Slope

- Need optimal shrinkage slope β

$$\hat{m}_i = (1 - \beta)\overline{m} + \beta m_i$$

- $\beta = 1$: no shrinkage: use sample means
- $\beta = 0$: full shrinkage: all means are equal (Global Minimum Variance Portfolio)
- Optimum: *somewhere* between 0 and 1

Excel Spreadsheet

- Regress truth on observables

Linear Regression

- Regress true μ_i onto observed m_i
- Grand mean $\overline{m}$ is close enough to $\overline{\mu}$

$$\mu_i - \overline{\mu} = \beta\left(m_i - \overline{m}\right)$$

$$\beta = \frac{Cov\left(m_i - \overline{m}, \mu_i - \overline{\mu}\right)}{Var\left(m_i - \overline{m}\right)}$$

$$\beta = \frac{Cov\left[\left(m_i - \mu_i\right) + \left(\mu_i - \overline{\mu}\right), \mu_i - \overline{\mu}\right]}{Var\left[\left(m_i - \mu_i\right) + \left(\mu_i - \overline{\mu}\right)\right]}$$

Independence

- Lady Luck is independent from God (skill)
- $m_i - \mu_i$ is independent from $\mu_i - \overline{\mu}$

$$\beta = \frac{Cov\left[(m_i - \mu_i) + (\mu_i - \overline{\mu}), \mu_i - \overline{\mu}\right]}{Var\left[(m_i - \mu_i) + (\mu_i - \overline{\mu})\right]}$$

$$\beta = \frac{Var\left[\mu_i - \overline{\mu}\right]}{Var\left[m_i - \mu_i\right] + Var\left[\mu_i - \overline{\mu}\right]}$$

$$\beta = \frac{\delta^2}{\omega^2 + \delta^2}$$

Interpretation

Shrinkage Slope

$$\beta = \frac{\delta^2}{\omega^2 + \delta^2}$$

Dispersion of Expected Returns

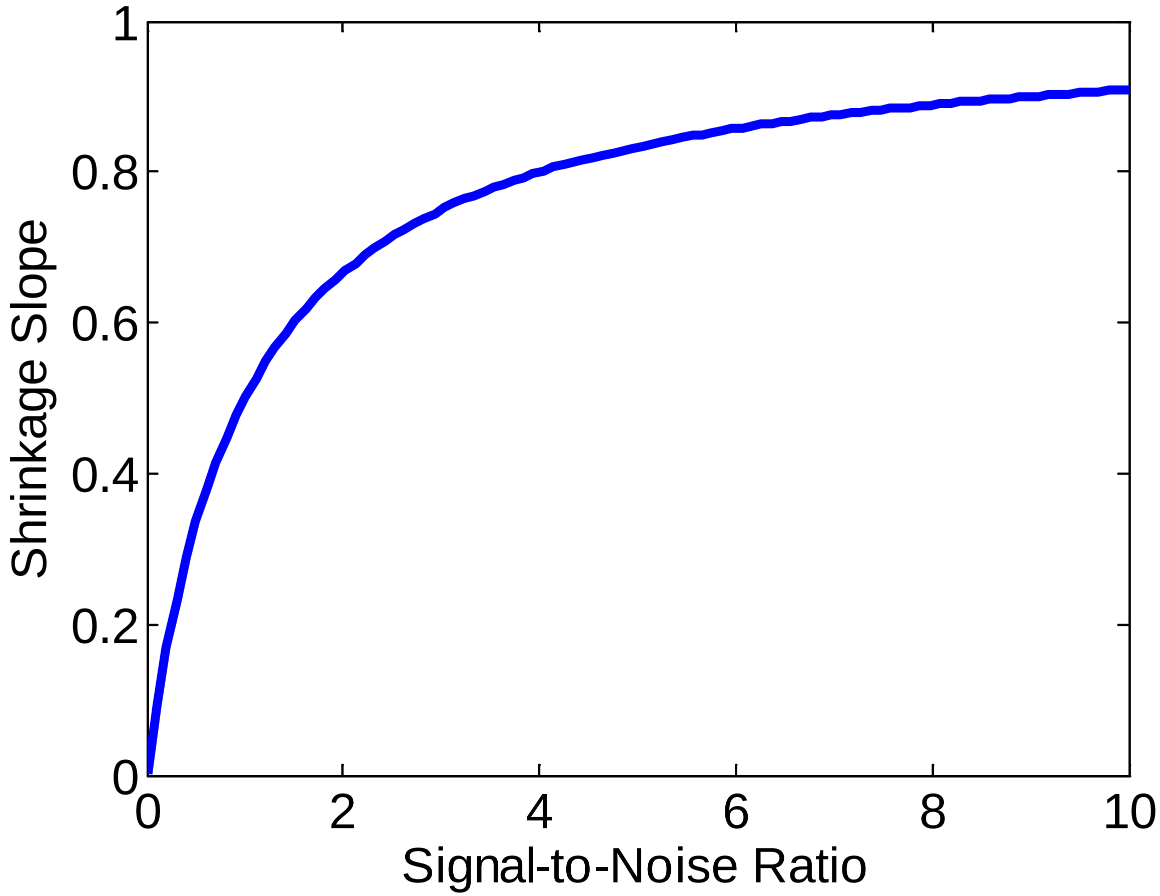
Estimation Error

Shrinkage Target

$$\hat{m}_i = (1-\beta)\overline{m} + \beta m_i$$

What happens to β when δ or ω go to 0 or infinity?

β as Function of δ^2/ω^2



How Does It Help?

- Problem was to know *n* true means μ_1, μ_2, ...μ_n
- That was not possible... 
- Boiled it down to just 2 parameters: δ^2 and ω^2
- This is possible! 

- This is *not* about getting a crystal ball...
- Playing the cards we have the best we can
- Using time-series & cross-section information

Plan

1. Intuition
2. Formula
3. Estimate the formula
4. Applications

3a) Estimate ω^2

- ω^2 = estimation error on sample mean m_i
- T observations: $m_i = (x_{1i} + x_{2i} + ... + x_{Ti}) / T$
- Sample variance of the T observations:

$s_i^2 = [(x_{1i} - m_i)^2 + (x_{2i} - m_i)^2 + ... + (x_{Ti}-m_i)^2]/ (T-1)$

- Variance of estimation error on m_i is: $\hat{\sigma}_i^2 = \frac{s_i^2}{T}$
- This is the usual way to construct a confidence interval: $m_i \pm 2\hat{\sigma}_i$

Estimator of ω^2

- Average across all variables:

$$\hat{\omega}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}_i^2 = \frac{1}{nT(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{ti}-m_i\right)^2$$

- Intuition: Dispersion in the time-series contains information about the amount of noise
- How far away from its own average is each observation?

3b) Estimate δ²

δ^2 = cross-sectional dispersion of expected returns

$$E[(m_i - \overline{\mu})^2] = Var[(m_i - \mu_i) + (\mu_i - \overline{\mu})]$$
$$= Var[m_i - \mu_i] + Var[\mu_i - \overline{\mu}]$$
$$= \omega^2 + \delta^2$$

$E[(m_i - \overline{\mu})^2]$ can be estimated by $\frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^2$

- Therefore: $\hat{\delta}^2 = \frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^2 - \hat{\omega}^2$

Estimated Shrinkage Slope

$$\hat{\beta} = \frac{\hat{\delta}^2}{\hat{\omega}^2 + \hat{\delta}^2}$$

Time-series

$$= 1 - \frac{1}{T(T-1)} \cdot \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \left(x_{ti} - m_i \right)^2}{\sum_{i=1}^{n} \left(m_i - \overline{m} \right)^2}$$

Cross-section

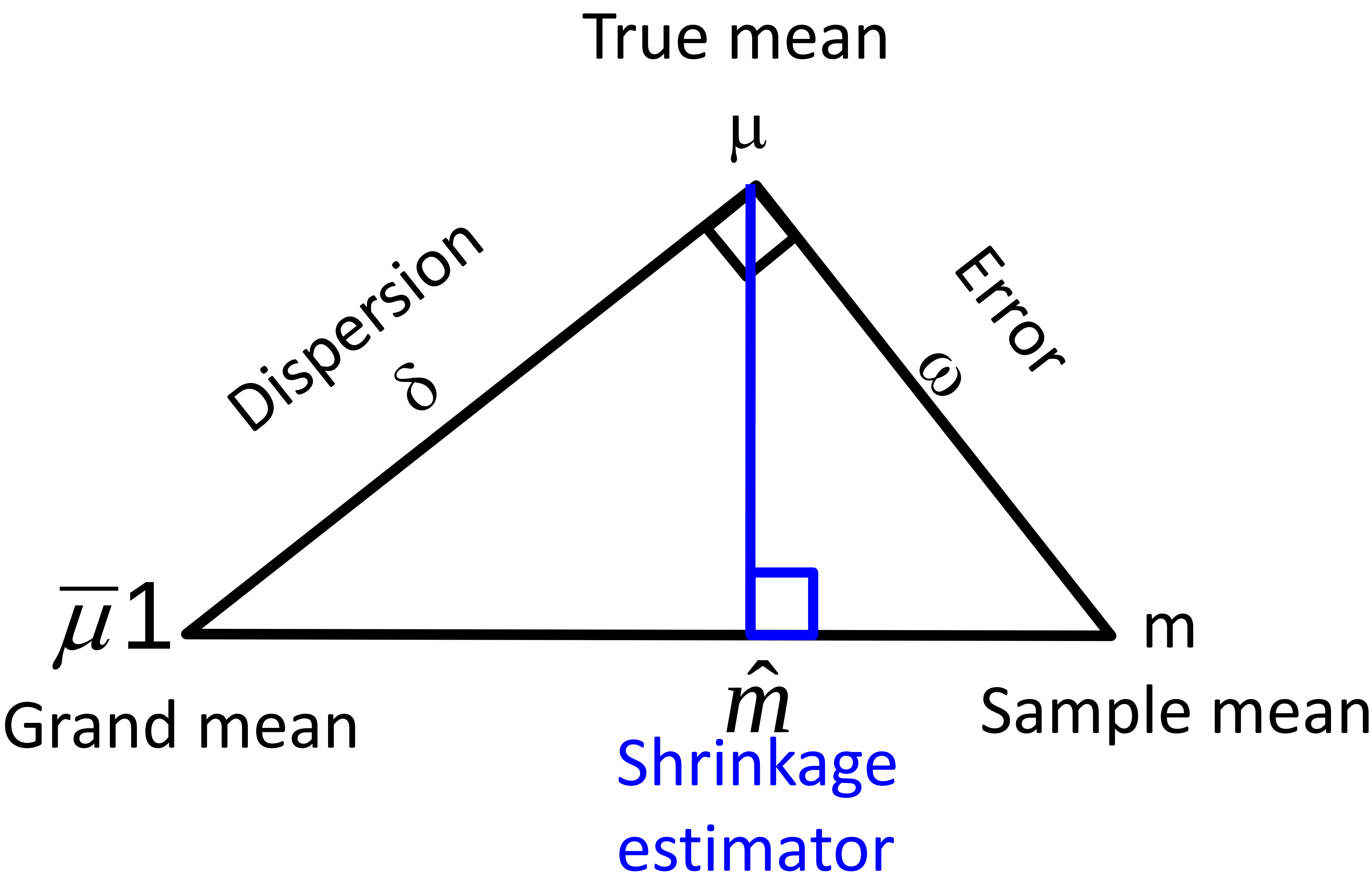
$$\hat{m}_i = (1 - \hat{\beta})\overline{m} + \hat{\beta} m_i$$

Excel Spreadsheet

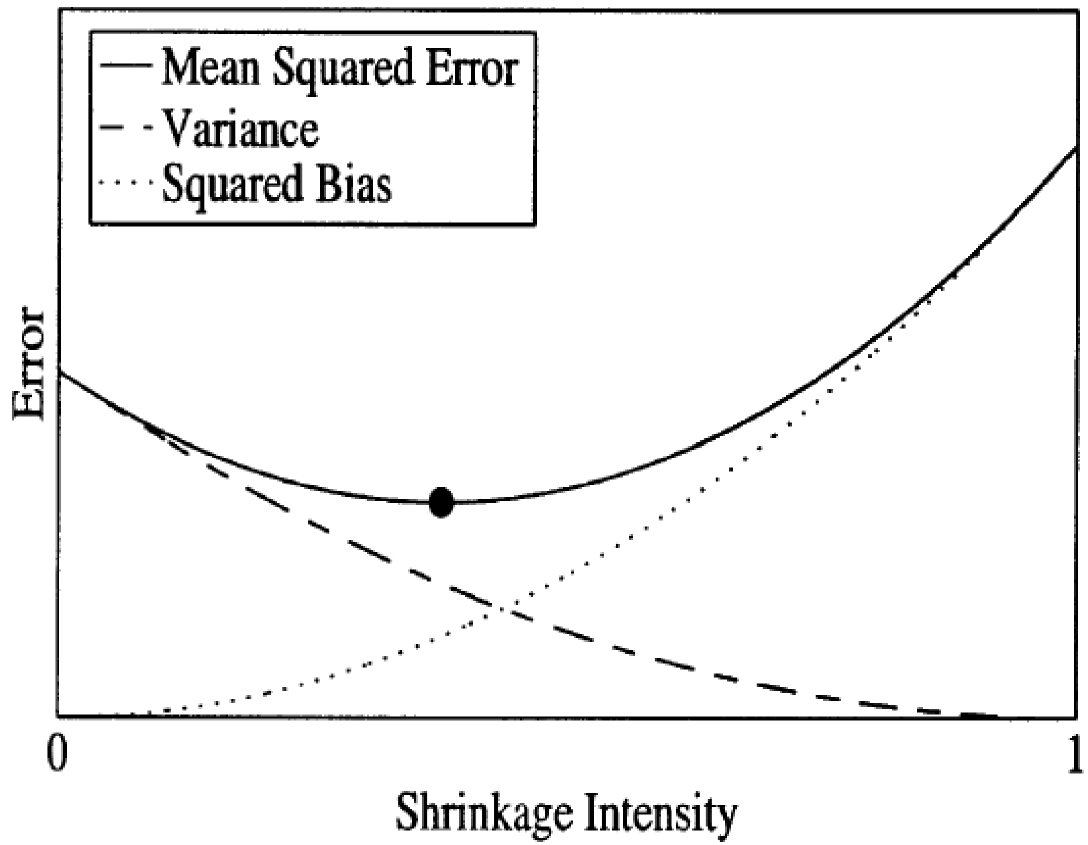
- Take it for a spin

- Works better the higher the number of variables, and the higher the number of observations

Geometric Interpretation



Trade-Off



Mean Squared Error
Variance
Squared Bias
Error
0
1
Shrinkage Intensity

Plan

1. Intuition
2. Formula
3. Estimate the formula
4. Applications

Can You Use It to Estimate Expected Returns?

- Maximize quadratic utility function:

$$w' \times m - \lambda \cdot w' \times \Sigma \times w$$

- Subject to: $w' \times 1 = 1$

- Replace m by: $\hat{m} = (1 - \hat{\beta})\overline{m}1 + \hat{\beta} m$

$\Rightarrow$Mean-variance efficient frontier *unchanged*

$\Rightarrow$Same as changing the risk-aversion coefficient

Applications

- Anything where selection $\rightarrow$ investment
- Data mining
- Hedge Fund Rankings
- Marriage
- Socialism
- Winning back-to-back championships
- Getting fired
- Estimating the Covariance Matrix!

Required Reading for Next Lecture

- The Markowitz Optimization Enigma: Is 'Optimized' Optimal? by Richard Michaud
- A well-conditioned estimator for large-dimensional covariance matrices, by Olivier Ledoit and Michael Wolf