Lecture 03: Signal vs. Noise

Statistical Arbitrage

MGMT MFE 431-3

Professor Olivier Ledoit

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

Fall 2019

Review of Last Lecture

Plan

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

Marching Orders: 1st Stage

PORTFOLIO SELECTION*

HARRY MARKOWITZ
The Rand Corporation

THE PROCESS OF SELECTING a portfolio may be divided into two stages. The first stage starts with observation and experience and ends with beliefs about the future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with the choice of portfolio. This paper is concerned with the second stage. We first consider the rule that the investor does (or should)

Optimization Problem

$$\min_x \; w'\Sigma w$$

<u>subject to: </u>$\alpha' w = [\text{target}]$

If Σ is proportional to identity then

$$w = [\text{scalar}] \cdot \alpha$$

Caring about risk induces diversification

Investment Rule

- Capital invested should be proportional to expected return
- Is it the same as proportional to past return?

Merton (1980)

A more likely explanation is simply that estimating expected returns from time series of realized stock return data is very difficult. As it is shown in appendix A, the estimates of variances or covariances from the available time series will be much more accurate than the corresponding expected return estimates. Indeed, even if the expected return on the market were known to be a constant for all time, it would take a very long history of returns to obtain an accurate estimate. And, of course, if this expected return is believed to be changing through time, then estimating these changes is still more difficult. Further, by the Efficient Market Hypothesis, the unanticipated part of the market return (i.e., the difference between the realized and expected return) should not be forecastable by any predetermined variables. Hence, unless a significant portion of the variance of the market returns is caused by changes in the expected return on the market, it will be difficult to use the time series of realized market returns to distinguish among different models

Big Picture

- Markowitz 2nd stage: taking risk into account, do not put all your capital in the asset with highest expected return (diversification).

- 1st stage: taking estimation error into account, do not allocate your capital in proportion to past returns (shrinkage).

Plan

1. Positioning
2. Intuition
3. Formula
4. Estimate the formula
5. 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:

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

where $\mathrm{Perf}_{p,t}$ 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?

Matlab

- Shrinkage demo

Plan

1. Positioning
2. Intuition
3. Formula
4. Estimate the formula
5. 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 $\mu_i - \bar{\mu}$

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

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

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

Interpretation

Shrinkage Slope

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

Dispersion of Expected Returns

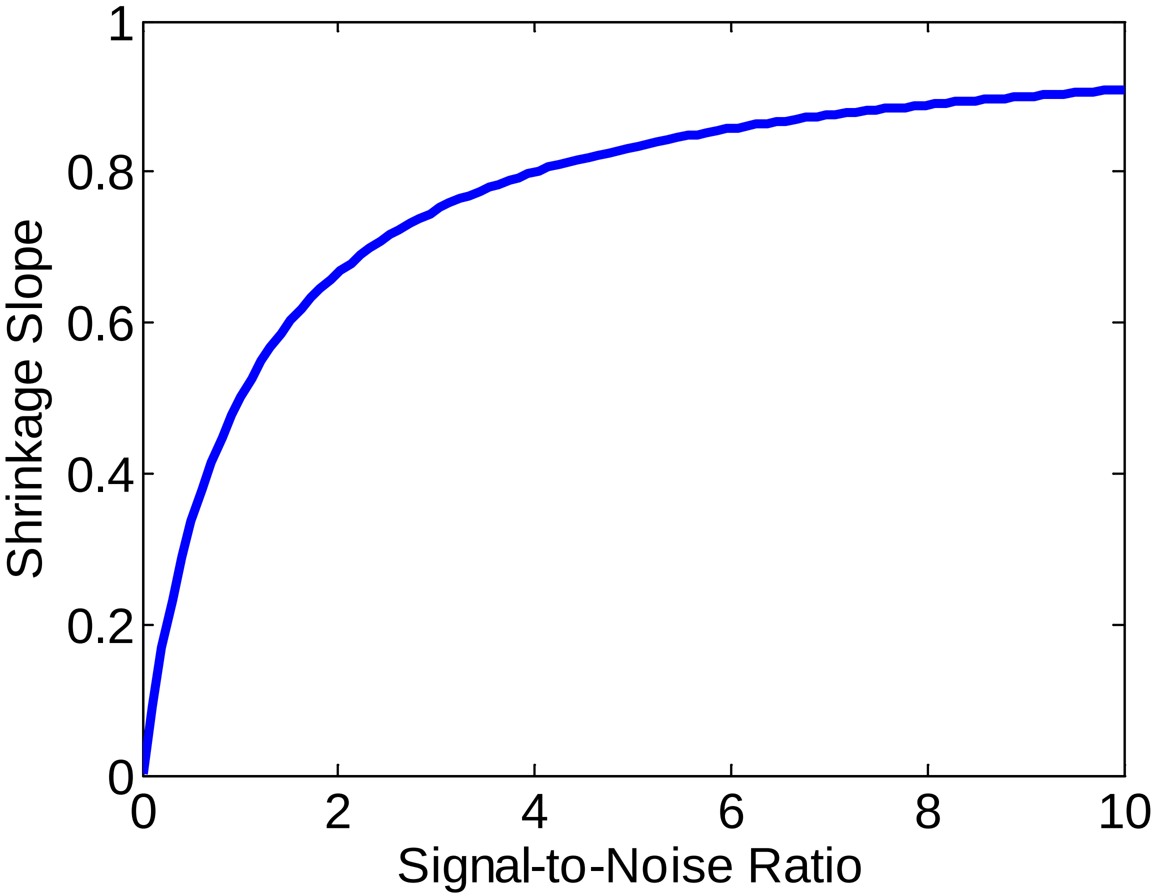
Estimation Error

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 $\mu_1, \mu_2, \ldots \mu_n$
- 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. Positioning
2. Intuition
3. Formula
4. Estimate the formula
5. 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: $\hat{\sigma}_i^2 = \dfrac{s_i^2}{T}$
- 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:

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

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

3b) Estimate δ^2

δ^2 = cross-sectional dispersion of expected returns

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

$E[(m_i - \overline{\mu})^2]$ 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

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

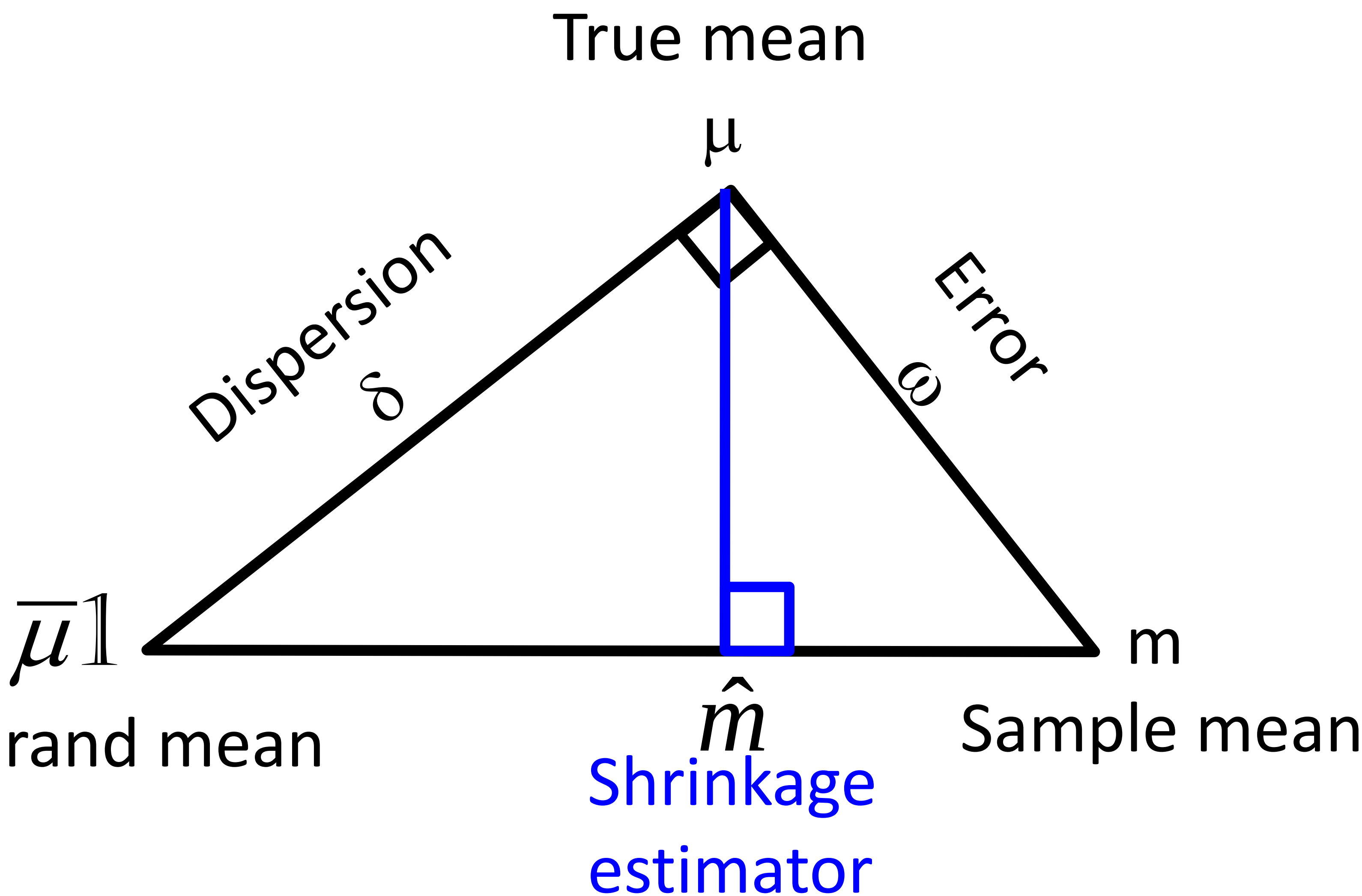
Cross-section

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

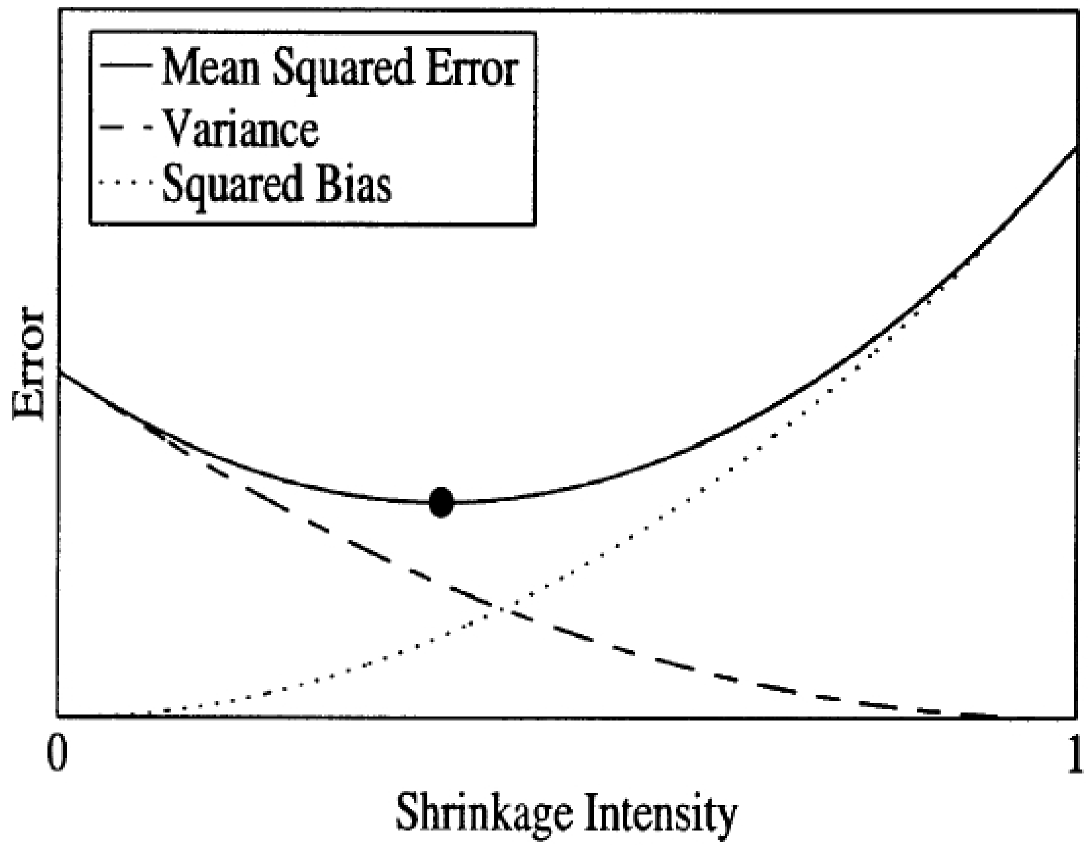
Excel Spreadsheet

- 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. Positioning
2. Intuition
3. Formula
4. Estimate the formula
5. 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: $w' \times \mathbb{1} = 1$

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

$\Rightarrow$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