Mitigating Estimation Error in Optimization

Anish R. Shah, CFA
Northfield Information Services
Anish@northinfo.com

Newport
June 4, 2010

Northfield INFORMATION SERVICES, INC.

Background in 1 Slide

- For the purpose of construction, value is measured by a single number, portfolio utility, which typically is return penalized for risk
- Optimizers find the portfolio weights that maximize this utility
- Alas, inputs into utility are only estimates
- Optimized portfolios perform worse than their forecasted utility and may look funny

Purpose of Talk

- Disabuse the false precision of portfolio utility and optimal portfolio weights
- Describe the fundamental mechanism that causes dissatisfaction with optimization and how to avoid it
- Briefly introduce two recent estimation error adjustment features in Northfield Optimizer

An [Easy-to-Graph] Example

- Recall familiar Markowitz utility function
  Utility = return – λ × variance

- Allocate money across 2 stocks
  - Forecast returns
    - $r_1 = 9\% \pm 1\%$
    - $r_2 = 10\% \pm 1\%$
  - Forecast correlation $\rho = 0.4 \pm 0.2$
  - (For simplicity) Std deviations are known to be 30%
  - User's risk aversion $\lambda = 1\%/200\%^2$

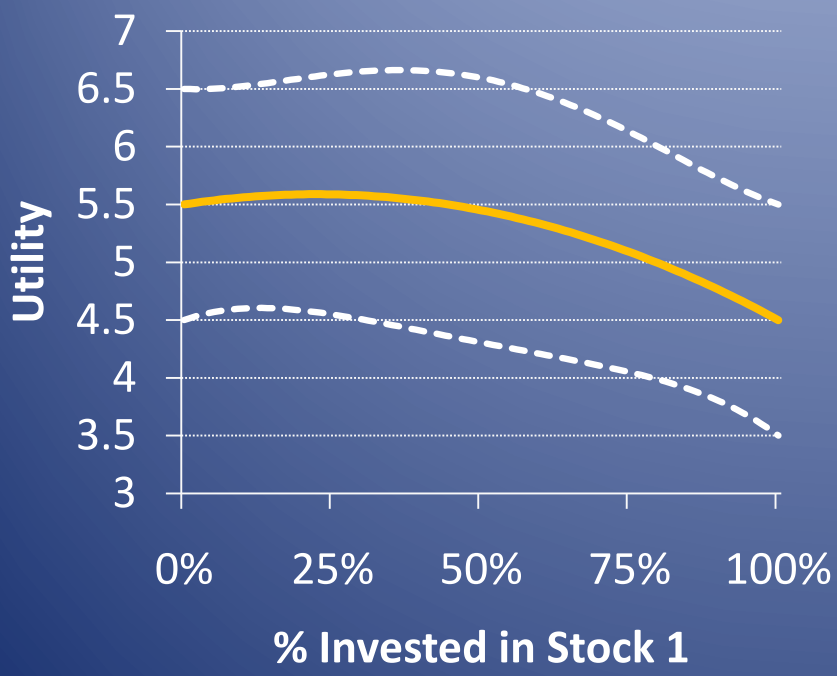
Northfield
INFORMATION SERVICES, INC.

Reference: Propagation of Errors

- Suppose f is a function of variables known with uncertainty
  - $f = f(\mathbf{x}) = f(x^1, \ldots, x^n)$
- Linearly approximate f in the area of interest, around the point $\mathbf{x}_0$
  - $f(\mathbf{x}) \approx f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)\,(\mathbf{x} - \mathbf{x}_0)$
- To approximate the error in f, take the variance
  - $\mathrm{var}[f(\mathbf{x})] \approx \nabla f(\mathbf{x}_0)\,\mathrm{var}(\mathbf{x})\,\nabla^T f(\mathbf{x}_0)$
- An aside: this is the idea behind linear factor risk models

Utility

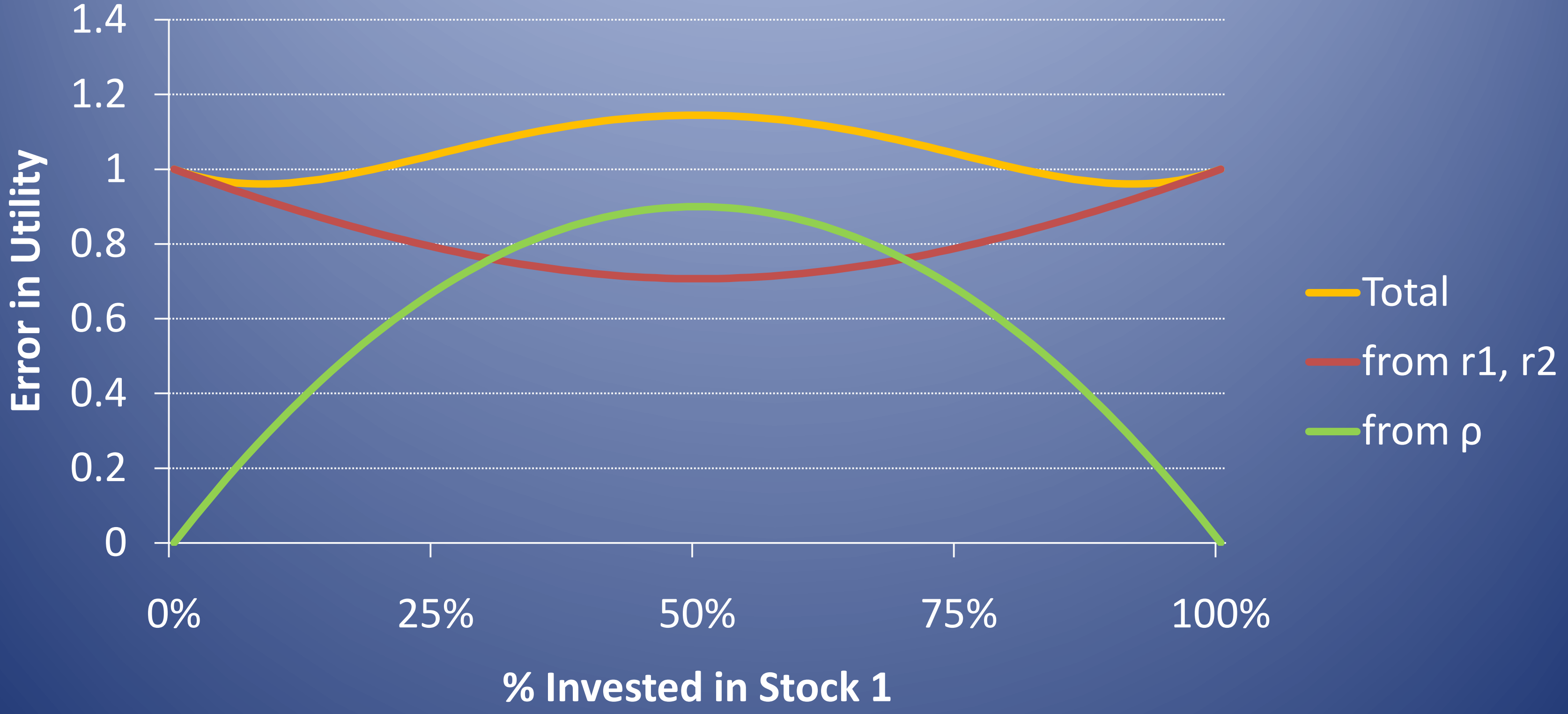
$r_1 = 10 \pm 1$ $r_2 = 9 \pm 1$ $\rho = 0.4 \pm 0.2$ $\sigma = 30$ $\lambda = 1/200$



- Utility ranges from 4.5 – 5.6
- Error is ~ 1 ± 0.1
- Differences in utility are on par with the error
- Many combinations are indistinguishably good
- No single "Optimal Portfolio"

Northfield
INFORMATION SERVICES, INC.

Bonus Slide: Sources of the Error



$r_1 = 10 \pm 1 \quad r_2 = 9 \pm 1 \quad \rho = 0.4 \pm 0.2$

Northfield
INFORMATION SERVICES, INC.

Two Bowls of Baked Beans

Wise monkey says, “They are nearly the same; the second has a bit more onion”

 Confused monkey says, “They’re entirely different; they don’t share a single bean in common”

- Many papers on estimation error in portfolio optimization are written by confused monkeys
  - Mistakenly focus on differences in holdings rather than the difference in portfolio utility

Northfield INFORMATION SERVICES, INC.

No Single Optimal Portfolio
Many, of Vastly Different Composition

A. Markowitz utility function uses only 2 features – return & covariance
B. Securities are similar, particularly in their forecasts
C. Of course the optimal portfolio's composition is sensitive to the inputs, but its utility isn't

After changing a return forecast, confused monkey screams, "Optimization is too sensitive to the inputs!"

- Anthropomorphizes software and calls it an 'evil-doer'
- Kicks computer

Wise monkey understands that the sensitivity reflects reality: many combinations achieve similar utility. Using this to his advantage, he finds a point in the optimal region that minimizes his trading costs

- Optimal utility can be achieved by portfolios of vastly different composition, so it is senseless to fret that 'the optimal portfolio' changes with the inputs

Northfield
INFORMATION SERVICES, INC.

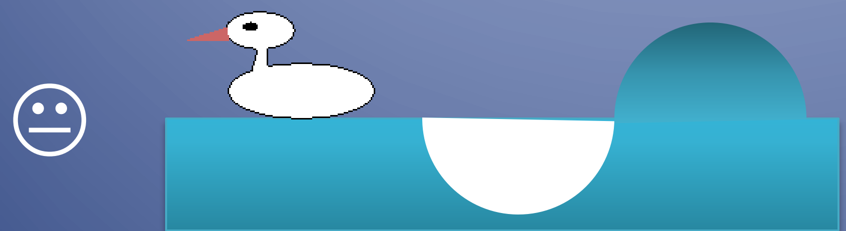
Recall Definition of Unbiased Estimator

- $\hat{g}$ is an unbiased estimator of g if $E[\hat{g}] = g$
  - i.e. if I run the experiment many times, the average of my estimates ($\hat{g}_1$, $\hat{g}_2$, …) will be the true value, g
  - The error of each estimate can be huge
  - Investment managers generally have too few estimates to reach accuracy through averaging
- $\hat{g} = E[g|information]$ is a different statement, which has nothing to do with bias
  - Moreover, talking about E[g] is gibberish until the context becomes Bayesian

Northfield
INFORMATION SERVICES, INC.

Unbiased Estimator:
Height of Duck (Excluding Legs)

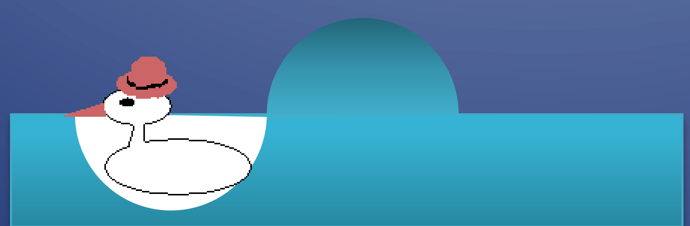
- Duck is sitting on ocean
- We are on land, eyes at sea level



Average of the estimates is .5 m, the duck's true height

Estimated height .5 m 0 m 1 m

Add 2nd duck:

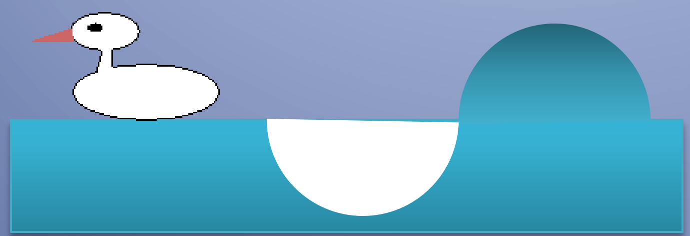


Average of the estimates is .6 m, the duck's true height

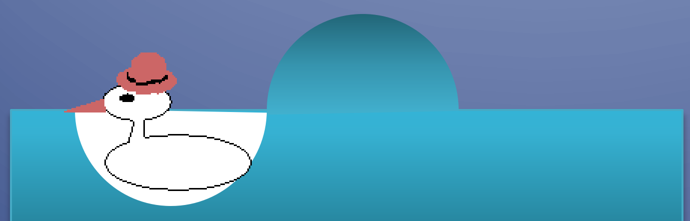
.1 m 1.1 m .6 m

Northfield
INFORMATION SERVICES, INC.

New Question:
Height of Taller Duck?



.5 m 0 m 1 m



.1 m 1.1 m .6 m

Winner:

.5 m 1.1 m 1 m

Average of estimate of taller duck is 0.87 meters!

The taller duck is only .6m

The estimate is biased

Straightforward Implications of *Unbiased* Estimates

- Markowitz utility

  $U(w,r,C) = \text{return} - \lambda \times \text{variance} = w^T r - \lambda\, w^T C w$

  To simplify notation, write as $U(w,G)$

  w = security weights

  G = parameters (returns, covariance) that are estimated

  $\hat{G}$ = an unbiased estimate of G

  $w^*(\hat{G})$ = the value of w that maximizes $U(w,\hat{G})$

1) Good news: For a fixed portfolio, estimated utility is unbiased

   Why?

   – $U(w,G)$ is affine in G

   – $E[U(w,\hat{G})] = U(w, E[\hat{G}]) = U(w,G)$

Northfield
INFORMATION SERVICES, INC.

Implications of Unbiased Estimates (2)

2) Bad news: For an <u>optimized</u> portfolio, estimated utility on average exaggerates true utility

   Why?

   – For any fixed portfolio w,
   – $U(w^*(\hat{G}), \hat{G}) > U(w,\hat{G})$ since $w^*(\hat{G})$ maximizes $U(w,\hat{G})$
   – $E[U(w^*(\hat{G}), \hat{G})] > E[U(w,\hat{G})] = U(w,G)$
   – let $w = w^*(\hat{G})$

3) Worse news: Not only is the optimized portfolio not optimal, but its estimated utility is on average greater than the maximum achievable utility

   Why?

   – Same as above. let $w = w^*(G)$

Northfield
INFORMATION SERVICES, INC.

Illustrating the Point by a False Paradox

- Fix a portfolio P
  - Estimated utility is some number, say 4
  - Estimate is unbiased
- Let O be the (estimated) maximum utility portfolio. Optimize to find it
  - Suppose O turns out to be P
- Is the estimation procedure for P now biased?
  - No
- Is the estimation procedure for O biased?
  - Yes
- But O and P are the same. How is O’s estimate biased and P’s not?
  - The estimated composition of O, the optimal portfolio, depends on the estimate. Although O happens to be P this time around, when estimates change, O changes with them while P doesn’t

Real Life Effect of Optimizing with Unbiased Estimates

- What happens when utility is exaggerated?
  - Markowitz utility has 2 parts, risk and return
    - Return is on average less than estimated
    - and/or Risk is on average greater than estimated
  - Portfolio is more aggressive than preference used in the utility function
    - And (barring a fluke) not optimal at that risk tolerance
- Effect is a mathematical truth
- Easy solution: For optimization, avoid unbiased estimates

Northfield
INFORMATION SERVICES, INC.

Recap of the 3 Major Points

1) Uncertainty in the inputs means there is no single "optimal portfolio"
   - What is optimal is a set of portfolios of indistinguishably different utility

2) Securities are alike, so portfolios comprising no securities in common can have essentially the same utility
   - Measuring distance from the "true optimal portfolio" in terms of difference in positions is senseless
   - Utility is the correct measure

3) Optimization biases estimates
   - Unbiased estimates are unsuitable for use in optimization
   - Quantities should instead be estimated with the goal of minimizing error

Northfield
INFORMATION SERVICES, INC.

Bayesian Inference To Limit Errors

- Classical (also called frequentist) statistics vs. Bayesian
  - Frequentist uses only observations
  - Bayesian has prior beliefs about the likelihood of events. To infer reality, Bayesian combines observations with beliefs
- Oversized hairy biped spotted in the yard
  - Frequentist yells, “Sasquatch!”
  - Bayesian, believing bigfoot sighting nearly impossible, thinks, “Mother-in-law has stopped by”

Tool: Bayes Adjust Alphas

- Idea is described in paper by Black & Litterman (1992) but predates the Kalman Filter (1960)
  - Black, F. & Litterman, R. "Global Portfolio Optimization," Financial Analysts Journal, 1992, v48(5,Sept/Oct), 22-43

- Imagine tracking a collection of sailboats crossing the Atlantic

- The frequentist way of inferring the boats' locations
  - Receive unreliable report – "Boat X is at position Y"
  - Source sometimes overshoots, sometimes undershoots. Take report at face value and estimate boat is at position Y

- A Bayesian way
  - Know the boats' courses, hence where they intend to be today [center of prior]
  - Currents and wind affect all boats [covariance of prior]
  - Receive unreliable report of their locations [observations with error]
  - The best guess of location combines the observations and the prior belief

Bayes Adjust Alphas (2)

- Prior on mean returns
  - $m \sim N(m_0, \Sigma/\tau)$
    $m_0 = 0$ if forecasting benchmark relative returns
    Σ = covariance of benchmark relative returns
    τ = intensity of prior
- Forecasts impart new information but are noisy observations of reality
  - $\hat{g} = m + \varepsilon \quad \varepsilon \sim N(0, \Omega)$
  - With each forecast, user provides standard deviation of the error
  - Users likely don't have opinions about the covariance of the errors, so Northfield assumes the errors are uncorrelated.
- A Bayes forecast combines the prior and the error
  - Most likely returns = $m_0 + \Sigma\,[\tau\Omega + \Sigma]^{-1}\,(\hat{g} - m_0)$

Northfield INFORMATION SERVICES, INC.

Tool: Blend Covariance

- Idea comes from a series of papers by Ledoit & Wolf
  - "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," Journal of Empirical Finance, Dec v10(5):603–621
- To soften extremes, blend original covariance matrix with duller (less differentiated) versions of itself
  - Single Index
  - Constant Correlation
  - Constant Covariance
- Fortunately, each of these can be represented as a one factor model

Northfield
INFORMATION SERVICES, INC.

Blend Covariance (2)

1) Single Index

Reduce the multifactor risk model to CAPM

– $r_s = \beta_s \, r_m + \varepsilon_s$

– $\beta_s = cov(r_m, r_s) / var(r_m)$

– $var(\varepsilon_s) = var(r_s) - \beta_s^2 \, var(r_m)$

– Reference market portfolio contains all stocks in the optimization (excluding cash) weighted by cap

Northfield
INFORMATION SERVICES, INC.

Blend Covariance (3)

2) Constant Correlation

More restrictive than CAPM - all stocks have the same pairwise correlation, ρ, but different variances

– $r_s = (\rho^{½}\ \sigma_s / \sigma_f)\ r_f + \varepsilon_s$ $\quad$ $var(\varepsilon_s) = (1-\rho)\ \sigma_s^{\ 2}$
where r_f is an artificial factor

– ρ is backed out from variance of reference portfolio
$\sigma^2_{\ m} = \sum_i (w_i\sigma_i)^{\ 2} + 2 \sum_{i<k} (w_i\ w_k\ \rho\ \sigma_i\ \sigma_k)$

→ $\rho = [\sigma^2_{\ m} - \sigma^2_{\ m0}] / [\sigma^2_{\ m1} - \sigma^2_{\ m0}]$
where
$\sigma^2_{\ m0} = \sum_i (w_i\sigma_i)^2$ = var of reference if uncorrelated
$\sigma^2_{\ m1} = (\sum_i w_i\ \sigma_i)^2$ = var of reference if perfectly correlated

Northfield
INFORMATION SERVICES, INC.

Blend Covariance (4)

3) Constant Covariance

Least differentiated of the three – all stocks have the same pairwise correlation, ρ, and the same variance, σ^2

– $r_s = (\rho^{1/2}\, \sigma/\sigma_f)\, r_f + \varepsilon_s$ $\quad\quad$ $var(\varepsilon_s) = (1\text{-}\rho)\, \sigma^2$

where r_f is an artificial factor

– σ and ρ are backed out from variance of reference portfolio

$\sigma^2_m = \sum_i (w_i\sigma)^2 + 2 \sum_{i<k} (w_i\, w_k\, \rho\, \sigma^2)$

ρ = as in constant correlation

→ $\sigma^2 = \sigma^2_m \,/\, [\rho + (1 - \rho)\, \Sigma_i w_i^2]$

Summary

- Estimation error is an issue in optimization, but not so nefarious as many believe
- The mechanism, as illustrated by the ducks, is straightforward to understand
- The remedy is to use estimation procedures that limit error in place of procedures that are accurate on average
- The Northfield Portfolio Optimizer has tools to do so
- *For thought, but beyond the scope of this presentation:*
  With uncertain inputs, utility becomes a random variable
  - One can redefine utility as a functional on its distribution
  - For example, people rebalancing once every 5 years would be more concerned with variance of utility than people rebalancing daily