Alpha Scaling Revisited

June 21, 2007

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

Motivation

- Portfolio construction
  = transferring investment skill efficiently into positions
    central to the success of an asset management firm

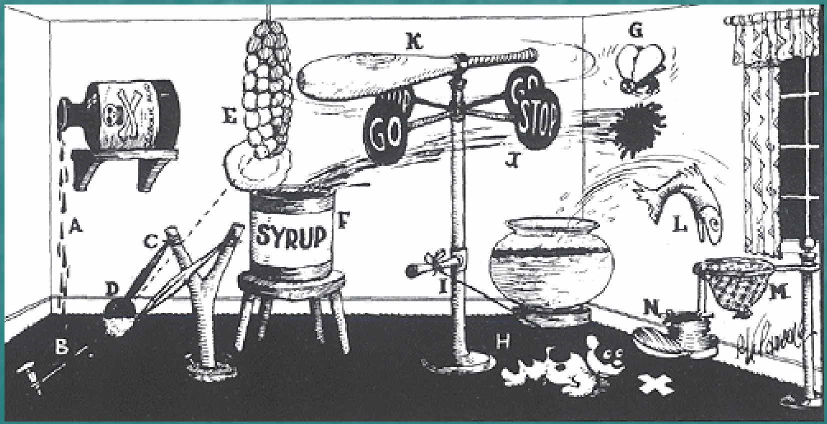
- Traditional portfolio construction incorporates qualitative information
- Quant, particularly optimization, uses information in the form of risk and return

- investment views → return forecasts → positions

Examples of Views

- Tech Analyst – "IBM is a strong buy"
- Strategist – "Financials will mildly outperform over the next year"

- Model – "On a scale from 1-10, Siemens is an 8"



Alpha Scaling/Adjustment

I. Extract all the information contained in the view to formulate a best return forecast

II. Given a set of best forecasts, condition them, so they are suitable for use in an optimizer

I: Extracting Information

- Seek the best prediction of the future given the information
- Suppose the analyst overreacts and is at times wrong
  - The best forecast of the future tempers the analyst's opinion
  - On the other hand, if the analyst is exceedingly cautious, the best forecast should amplify the opinion
- Convert information (e.g. ratings) to returns

II: Conditioning for Optimization

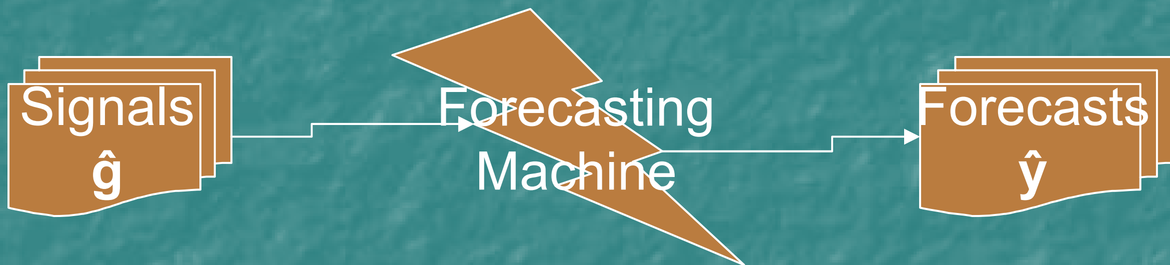
- Optimizers seek extremes (by mandate!)
- Inputs are estimated with error
- Optimized selection introduces bias

- Conditioning deals with optimization under uncertain inputs, a large and separate topic
  - Northfield is building a set of tools to address this

Overview of Alpha Scaling Presentation

- Standard methods of constructing good forecasts spelled out
- Standard method of combining sets of forecasts
- Northfield's upcoming alpha scaling tool

Foundation: Linear Model



- How to make signals (views) into forecasts?
- One approach - fit a linear model

$$\hat{\mathbf{y}}(\hat{\mathbf{g}}) = \mathbf{A}\,\hat{\mathbf{g}} + \mathbf{b}$$

- Minimize expected squared error

$$\hat{\mathbf{y}}(\hat{\mathbf{g}}) = E(\mathbf{y}) + \mathrm{cov}(\mathbf{y}, \mathbf{g})\,\mathrm{cov}(\mathbf{g}, \mathbf{g})^{-1}\,[\hat{\mathbf{g}} - E(\mathbf{g})]$$

Linear Model (cont)

- e.g. g_i = stock i's analyst rating (1-5)
  stock i's earnings surprise
  stock i's percentile rank
  change in 90 day T-bill yield

  e.g. y_k = stock k's return
  stock k's return net of market β and industry

- Important observation: each y_k is built separately
  $\hat{y}_k(\hat{\mathbf{g}}) = E(y_k) + \text{cov}(y_k, \mathbf{g})\ \text{cov}(\mathbf{g}, \mathbf{g})^{-1}\ [\hat{\mathbf{g}} - E(\mathbf{g})]$

One Signal Per Stock – Grinold

- Forecast y_k using only signal g_k
  e.g. forecast IBM's return from only IBM's rating

- $\hat{y}_k(\hat{g}_k) = a\,\hat{g}_k + b$

  choose a and b to minimize expected squared error:
  $\hat{y}_k(\hat{g}) = E(y_k) + \mathrm{cov}(y_k, g_k) / \mathrm{var}(g_k)\,[\hat{g}_k - E(g_k)]$

  $= E(y_k) + [\rho(y_k, g_k)\,\mathrm{std}(y_k)\,\mathrm{std}(g_k)] / \mathrm{var}(g_k)\,[\hat{g}_k - E(g_k)]$

  $= E(y_k) + \underbrace{\rho(y_k, g_k)}_{\text{IC}} \times \underbrace{\mathrm{std}(y_k)}_{\text{volatility}} \times \underbrace{[\hat{g}_k - E(g_k)] / \mathrm{std}(g_k)}_{\text{score}}$

Grinold – No Confusion About Parameters

- IC = correlation (signal, return being forecast)
- Volatility is the volatility of the return being forecast
- Score is the z-score of that instance of the signal
- IC can be estimated over a group of securities (e.g. same cap/industry/volatility) if the model works equally well on them
- Expect lower IC's for volatile securities (harder to predict) than for less volatile ones (easier to predict)
- Using a single IC exaggerates volatile securities' alphas

Grinold Example

- The upcoming period is
  - good for DELL (z-score of 1)
  - better for MSFT (z-score of 2)
  - great for PEP (z-score of 3)

- Stock-specific volatility: $\sigma^{ss}_{DELL} = 27\%$, $\sigma^{ss}_{MSFT} = 25\%$, $\sigma^{ss}_{PEP} = 9\%$

- Skill, corr(signal,return): $IC_{tech} = .10$, $IC_{consumer} = .15$

- Assume E[y] = 0, stock-specific return averages 0 over time

- $\hat{y}_{DELL} = 0 + .10 \times 27\% \times 1 = 2.7\%$
  $\hat{y}_{MSFT} = 0 + .10 \times 25\% \times 2 = 5.0\%$
  $\hat{y}_{PEP} = 0 + .15 \times 9\% \times 3 = 4.0\%$

Grinold Cross-Sectionally

- $\hat{y}_k$ = stock k's return over a benchmark
  $\hat{g}_k$ = the relative attractiveness of stock k

  e.g. forecast IBM's return over the market using IBM's %ile in a stock screen

- $\hat{y}_k(\hat{g}_k) = E(y_k) + IC(y_k, g_k) \times std(y_k) \times score(\hat{g}_k)$

Assume 1) The volatility of what you are predicting is the same across all stocks
2) All stocks are equally likely to have a given level of relative attractiveness
e.g. utility co is as likely to be a strong buy as tech co

Grinold Cross-Sectionally (cont)

- $\hat{y}_k(\hat{g}_k) = E(y_k) + IC(y_k, g_k) \times std(y) \times score(\hat{g}_k)$

  std(y) can be estimated by cross-sectional return vol
  score($\hat{g}_k$) can be estimated by $\hat{g}_k$'s cross-sectional score

  If skill is the same across all securities,
  IC can be estimated by correlation between cross-sectional score and relative return

- $\hat{y}_k(\hat{g}_k) = IC \times$ xc volatility $\times$ xc score

Cross-Sectional Grinold Example

- Relative to other stocks,
  - DELL will outperform (z-score of 2)
  - MSFT will strongly outperform (z-score of 3)
  - PEP will slightly outperform (z-score of 1)

- Cross-sectional volatility of 1 year returns is 15%

- Skill, corr(xc signal, xc return): $IC_{tech\ stocks} = .08$, $IC_{consumer\ stocks} = .12$

- $\hat{y}_{MSFT} = .08 \times 15\% \times 3 = 3.6\%$
  $\hat{y}_{DELL} = .08 \times 15\% \times 2 = 2.4\%$
  $\hat{y}_{PEP} = .12 \times 15\% \times 1 = 1.8\%$

Combining Sets of Good Forecasts: Black Litterman

- Asset managers have different sets of information
  - IBM will return 5%
  - SP500 will beat R2000 by 4%
- Once cleaned up (see previous slides), how can they be fused into 1 forecast per stock?
- Ans: Black-Litterman

Black Litterman

- Motivated by need to stabilize asset allocation optimization
- Bayesian Approach
- Assume a prior distribution on the vector of mean returns
  - Centered at implied alpha that makes market portfolio optimal (stability)
  - Covariance is proportional to covariance of returns

Black-Litterman (cont)

- New information given as portfolio forecasts with error
  - IBM's will return 5% ± 2%
    i.e. return of portfolio holding 100% IBM is 5% ± 2%
  - MSFT will outperform IBM by 3% ± 4%
    i.e. return of portfolio long MSFT short IBM is 3% ± 4%
  - S&P500 will outperform R2000 by 4% ± 2%
- Combined forecast is expected value given prior and information

Black-Litterman (cont)

- prior on mean returns:
  - $\mathbf{m} \sim N(\mathbf{m}_0, \boldsymbol{\Sigma}_0)$
- forecasts impart new information:
  - $\hat{\mathbf{g}} = \mathbf{P}\, E[\mathbf{m} \mid \text{info}] + \boldsymbol{\varepsilon}$
  - $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega})$
- $\hat{\mathbf{y}} = [\boldsymbol{\Sigma}_0^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}\ [\boldsymbol{\Sigma}_0^{-1}\,\mathbf{m}_0 + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\hat{\mathbf{g}}]$

  $= \mathbf{m}_0 + [\boldsymbol{\Sigma}_0^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}\ \mathbf{P}^T\boldsymbol{\Omega}^{-1}\,(\hat{\mathbf{g}} - \mathbf{P}\mathbf{m}_0)$

  $= \mathbf{m}_0 + [\boldsymbol{\Sigma}_0 - (\mathbf{P}\boldsymbol{\Sigma}_0)^T(\boldsymbol{\Omega} + \mathbf{P}\boldsymbol{\Sigma}_0\mathbf{P}^T)^{-1}\,\mathbf{P}\boldsymbol{\Sigma}_0]\ \mathbf{P}^T\boldsymbol{\Omega}^{-1}\,(\hat{\mathbf{g}} - \mathbf{P}\mathbf{m}_0)$
- Because of the prior's covariance, one security tells us about another. e.g. if IBM and DELL are correlated, information about IBM says something about DELL

Black-Litterman Example

- Prior on IBM and DELL of (2%, 5%), with respective variances $4\%^2$, $9\%^2$ and correlation .5
- Predict that IBM will return 5% ± 3%
- $\mathbf{m}_0 = (2\%\ 5\%)^T$, $\boldsymbol{\Sigma}_0 = (4\ 3;\ 3\ 9)\ \%^2$
  $\mathbf{P} = (1\ 0)$, $\hat{g} = 5\%$, $\Omega = 9\%^2$
- Updated forecasts: $\hat{y}_{IBM} = 2.9\%$, $\hat{y}_{DELL} = 5.7\%$

Extending Black Litterman

- Consider as underlying securities all the stock specific returns and all the returns to factors, e.g. $\mathbf{m} = (m^{ss}_{IBM}, m^{ss}_{DELL}, \ldots, m_{E/P}, m_{GROWTH},$ etc.)
- Make forecasts at different levels
  - Net of style and industry, IBM will return 5% ± 4%
  - The dividend yield factor will return 2% ± 3%
  - Inclusive of all effects, DELL will return 9% ± 6%
  - S&P500 will outperform R2000 by 4% ± 2%
- Information gets projected onto all securities. e.g. forecast about S&P500 over R2000 → return on market cap → return on large and small cap stocks which aren't in S&P500 or R2000
- Easy to implement

Signal Decay & Horizon

- Suppose the best forecast is that IBM beats the benchmark by 5% over the next 6 months, and you have no opinion beyond
- What is the forecast alpha if you plan to hold IBM for 6 mos? A year?

- Combined forecast ≈ time-weighted average over reference holding period of each interval's best forecasts

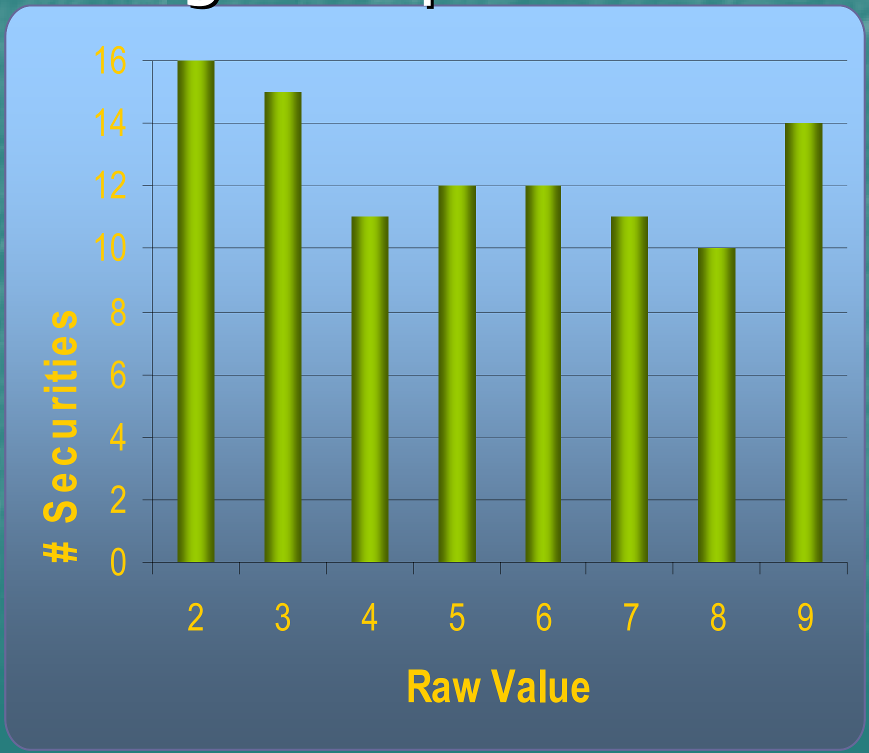
e.g. 2 yr, 8% annualized over 1st 6 mo, 1% over remaining 18 → $\frac{1}{4} \times 8\% + \frac{3}{4} \times 1\% = 2.75\%$

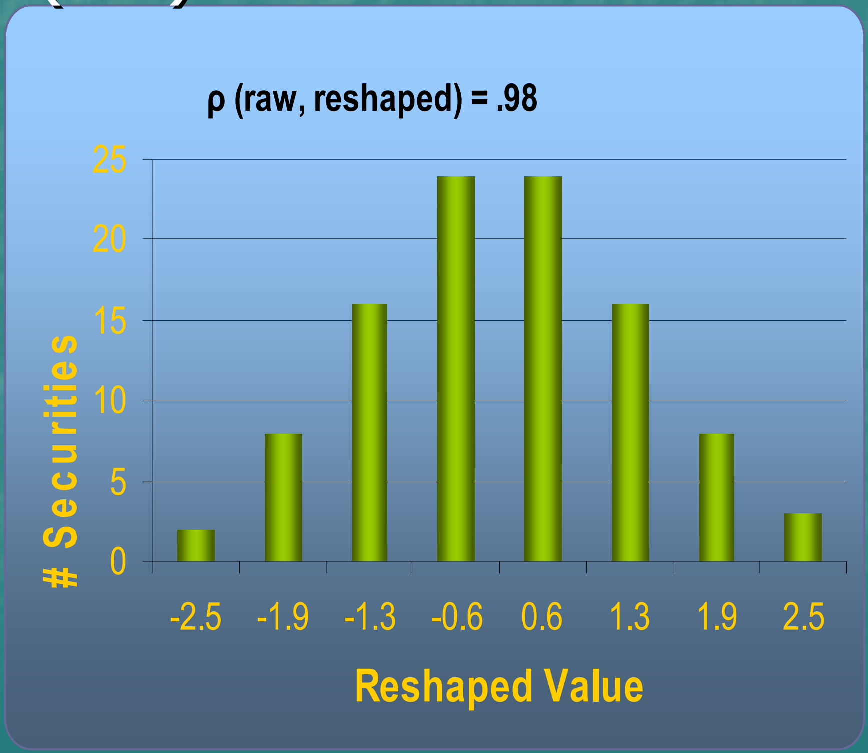
Northfield's Alpha Scaling Tool

- Seek a theoretically sound, information preserving, robust way of refining investment views
- Have client's forecast alpha. Don't know alpha generating process
- Sophisticated methods leverage information
  Better to be simple than falsely precise
- Beginning from alpha forecasts (not individual stock scores) necessitates a cross-sectional framework: Cross-sectional Grinold

Preprocess for Robustness: Rank Rescaling into Scores

- Map raw signals by rank onto standard normal e.g. 25th percentile → $F^{-1}(.25)$





Estimate Cross-Sectional Volatility

- Expected market weighted cross sectional variance
  $= E[\Sigma_s w_s (r_s - r_m)^2]$ where $r_m = \Sigma_s w_s r_s$
  $= E[\Sigma_s w_s (r_s - \mu_s + \mu_s - \mu_m + \mu_m - r_m)^2]$

  $= \Sigma_s w_s \sigma_s^2 - \sigma_m^2 + \Sigma_s w_s (\mu_s - \mu_m)^2$
  $\approx \Sigma_s w_s \sigma_s^2 - \sigma_m^2$
  = avg stock variance – variance of the market

- Numbers come straight from risk model
- If forecasting return net of β, industry, etc., easy to calculate risk net of these effects

Put The Pieces Together

- IC – user parameter
  cross sectional volatility – from risk model
  score – signal after rank mapping to std N

- Forecast of return above market
  = IC × volatility × score

Summary

- Standard practice alpha scaling methods can be arrived at by following your nose. No hidden magic or sophistication
- Being clear about the inputs and what's being forecast is this first step in scaling alphas well
- Adjustments for horizon and signal decay are important, particularly in low-turnover portfolios
- Northfield's upcoming alpha scaling functionality can make your life easier

References

- Grinold, R. "Alpha Is Volatility Times IC Times Score," *Journal of Portfolio Management*, 1994, v20(4,Summer), 9-16.
- Black, F. & Litterman, R. "Global Portfolio Optimization," *Financial Analysts Journal*, 1992, v48(5,Sept/Oct), 22-43.
- Bulsing, Scowcroft, & Sefton. "Understanding Forecasting: A Unified Framework for Combining Both Analyst And Strategy Forecasts," UBS report, 2003.
- Almgren, R. & Chriss, N. "Optimal Portfolios from Ordering Information", *Journal of Risk*, 2006 v9(1,Fall), 1-47.