Northfield’s 13th Annual Summer Seminar
Friday, June 8th, 2007
Agenda

Fat Tails, Tall Tales, Puppy Dog Tails
   Dan diBartolomeo, Northfield Information Services

Improving Returns-Based Style Analysis
   Daniel Mostovoy, Northfield Information Services

How Large is the Equity Premium Today?
   Samuel Thompson, Arrow Street Capital

Implied Risk Acceptance Parameters (RAPs) in the Execution of Institutional Equity Trades
   Thorsten Schmidt, CFA, Instinet

Alpha Scaling Revisited
   Anish Shah, Northfield Information Services

Distinguishing Between Being Unlucky and Unskillful
   Dan diBartolomeo, Northfield Information Services
Fat Tails, Tall Tales, 
Puppy Dog Tails

Dan diBartolomeo

Annual Summer Seminar – Newport, RI
June 8, 2007
Goals for this Talk

- Survey and navigate the enormous literature in this area
- Review the debate on assumed distributions for stock returns
- Consider the implications of the various possible conclusions on asset pricing, portfolio construction and risk management
Return Distributions

• While traditional portfolio theory assumes that returns for equity securities and market are normally distributed, there is a vast amount of empirical evidence that the frequency of large magnitude events seems much greater than is predicted by the normal distribution with observed sample variance parameters.

• Three broad schools of thought:
  - Equity returns have stable distributions of infinite variance.
  - Equity returns have specific, identifiable distributions that have significant kurtosis (fat tails) relative to the normal distribution (e.g. a gamma distribution).
  - Distributions of equity returns are normal at each instant of time, but look fat tailed due to time series fluctuations in the variance.
Stable Pareto Distributions

- Mandelbrot (1963) argues that extreme events are far too frequent in financial data series for the normal distribution to hold. He argues for a stable Paretian model, which has the uncomfortable property of infinite variance.
- Mandelbrot (1969) provides a compromise, allowing for "locally Gaussian processes."
- Fama (1965) provides empirical tests of Mandelbrot’s idea on daily US stock returns. Finds fat tails, but also volatility clustering.
- Lau, Lau and Wingender (1990) reject the stable distribution hypothesis.
A Bit on Stable Distributions

• General stable distributions have four parameters
  – Location (replaces mean)
  – Scale (replaces standard deviation)
  – Skew
  – Tail Fatness

• Some moments are infinite

• Except for some special cases (e.g. normal) there are no analytical expressions for the likelihood functions

• Estimation of the parameters is very fragile. Many, many different combinations of the four parameters can fit data equally well

• These distributions do have time scaling (you should be able to scale from daily observations to monthly observations, etc.)
Gulko (1999) argues that an efficient market corresponds to a state where the informational entropy of the system is maximized.

- Finds the risk-neutral probabilities that maximize entropy.
- The entropy maximizing risk neutral probabilities are equivalent to returns having the Gamma distribution.
- Gamma has fat tails but only two parameters and finite moments.
- Has finite lower bound which fits nicely with the lower bound on returns (i.e. -100%).
- Derives an option pricing model of which Black-Scholes is a special case.
The alternative to stable fat-tailed distributions is that returns are normally distributed at each moment in time, but with time varying volatility, giving the illusion of fat tails when a long period is examined.

Rosenberg (1974?)
- Most kurtosis in financial time series can be explained by predictable time series variation in the volatility of a normal distribution.

Engle and Bollerslev: ARCH/GARCH models
- Models that presume that volatility events occur in clusters.
- Huge literature. I stopped counting when I hit 250 papers in referred journals as of 2003.

LeBaron (2006)
- Extensive empirical analysis of stock returns.
- Finds strong support for time varying volatility, but very weak evidence of actual kurtosis.
The Remarkable Rosenberg Paper

- Unpublished paper by Barr Rosenberg (1974?), under US National Science Foundation Grant 3306
- Builds detailed model of time-varying volatility in which long run kurtosis arises from two sources
  - The kurtosis of a population is an accumulation of the kurtosis across each sample sub-period
  - Time varying volatility and serial correlation can induce the appearance of kurtosis when the distribution at any one moment in time is normal
  - Predicts more kurtosis for high frequency data
- An empirical test on 100 years of monthly US stock index returns shows an R-squared of .86
- Very reminiscent of subsequent ARCH/GARCH models
ARCH/GARCH

• Engle (1982) for ARCH, Bollerslev (1986) for GARCH
• Conditional heteroscedasticity models are standard
  operating procedure in most financial market
  applications with high frequency data
• They assume that volatility occurs in clusters, hence
  changes in volatility are predictable
• Andersen, Bollerslev, Diebold and Labys (2000)
  – Exchange rate returns are Gaussian
• Andersen, Bollerslev, Diebold and Ebens (2001)
  – The distribution of stock return variance is right skewed for
    arithmetic returns, normal for log return
  – Stock returns must be Gaussian because the distribution of
    returns/volatility is unit normal
Recent Empirical Research

- Lebaron, Samanta and Cecchetti (2006)
- Exhaustive Monte-Carlo bootstrap tests of various fat tailed distributions to daily Dow Jones Index data using robust estimators
- Propose a novel adjustment for time scaling volatilities to account for kurtosis, in order to use daily data to forecast monthly volatility
- Conclusion: “No compelling evidence that 4th moments exist”
  - If variance is unstable, then it’s difficult to estimate
  - High frequency data is less useful
  - Use robust estimators of volatility
  - Estimation error of expected returns dominates variance in forming optimal portfolios
More Work on Fat Tails

- **Japan Stock Returns**
  - Watanabe (2000)

- **France Stock Returns**
  - Navatte, Christophe Villa (2000)

- **Option implied kurtosis**
  - Corrado and Su (1996, 1997a, 1997b)
  - Brown and Robinson (2002)

- **Sides of the debate**
  - Lee and Wu (1985)
  - Tucker (1992)
  - Ghose and Kroner (1995)
  - Mittnik, Paolella and Rachev (2000)
  - Rockinger and Jondeau (2002)
The Time Scale Issue

• Almost all empirical work shows that fat tails are more prevalent with high frequency (i.e. daily rather than monthly) return observations.

• Lack of fat tails in low frequency data is a problem for proponents of stable distributions, as the tail properties should time scale.
  – Maybe we just don’t have enough observations when we use lower frequency data for apparent kurtosis to be statistically significant.

• Or the observed differences in higher moments could be a mathematical artifact of the way returns are being calculated.
  – Lau and Wingender (1989) call this the “intervaling effect.”
The Curious Compromise of Finanalytica

• The basic concepts of stable fat tailed distributions and time-varying volatility models are clearly mutually exclusive as explanations for the observed empirical data.

• From the Finanalytica website:
  – “uses proprietary generalized multivariate stable (GMstable) distributions as the central foundation of its risk management and portfolio optimization solutions”
  – “Clustering of volatility effects are well known to anyone who has traded securities during periods of changing market volatility. Finanalytica uses advanced volatility clustering models such as stable GARCH…”

• Svetlozar Rachev and Doug Martin are really smart guys so I’m putting this down to pragmatism rather than schizophrenia.
Kurtosis versus Skew

- So far we’ve talked largely about 4\textsuperscript{th} moments
- We haven’t done much in terms of economic arguments about why fat tails exist, and at least appear to be more prevalent with higher frequency data
- Many of the same arguments apply to skew (one fat tail),
  - consistent prevalence of negative skew in financial data series
- Harvey and Siddique (1999) find that skew can be predicted using an autoregressive scheme similar to GARCH
Cross-Sectional Dispersion

- When we think about “fat tails” we are usually thinking about time series observations of returns.
- For active managers, the cross-section of returns may be even more important, as it defines the opportunity set.
- DeSilva, Sapra and Thorley (2001) - if asset specific risk varies across stocks, the cross-section should be expected to have a unimodal, fat-tailed distribution.
- Almgren and Chriss (2004) - provides a substitute for “alpha scaling” that sorts stocks by attractiveness criteria, then maps the sorted values into a fat-tailed multivariate distribution using copula methods.
What’s the Problem with Daily Returns Anyway?

• Financial markets are driven by the arrival of information in the form of “news” (truly unanticipated) and the form of “announcements” that are anticipated with respect to time but not with respect to content.

• The time intervals it takes markets to absorb and adjust to new information ranges from minutes to days. Generally much smaller than a month, but up to and often larger than a day. That’s why US markets were closed for a week at September 11th.
Investor Response to Information

• Several papers have examined the relative market response to “news” and “announcements”
  – Ederington and Lee (1996)
  – Abraham and Taylor (1993)

• Jones, Lamont and Lumsdaine (1998) show a remarkable result for the US bond market
  – Total returns for long bonds and Treasury bills are not different if announcement days are removed from the data set

• Brown, Harlow and Tinic (1988) provide a framework for asymmetrical response to “good” and “bad” news
  – Good news increases projected cash flows, bad news decreases
  – All new information is a “surprise”, decreasing investor confidence and increasing discount rates
  – Upward price movements are muted, while downward movements are accentuated
Implications for Asset Pricing

• If investors price skew and/or kurtosis, there are implications for asset pricing

• Harvey (1989) finds relationship between asset prices and time varying covariances

• Kraus and Litzenberger (1976) and Harvey and Siddique (2000) find that investors are averse to negative skew
  – diBartolomeo (2003) argues that the value/growth relationship in equity returns can be modeled as option payoffs, implying skew in distribution
  – If the value/growth relationship has skew and investors price skew, then an efficient market will show a value premium

• Dittmar (2002) find that non-linear asset pricing models for stocks work if a kurtosis preference is included

• Barro (2005) finds that the large equity risk premium observed in most markets is justified under a “rare disaster” scenario
Portfolio Construction and Risk Management

- Kritzman and Rich (1998) define risk management function when non-survival is possible.

- Satchell (2004)
  - Describes the diversification of skew and kurtosis.
  - Illustrates that plausible utility functions will favor positive skew and dislike kurtosis.

- Wilcox (2000) shows that the importance of higher moments is an increasing function of investor gearing.
Optimization with Higher Moments

• Chamberlin, Cheung and Kwan (1990) derive portfolio optimality for multi-factor models under stable paretian assumptions

• Lai (1991) derives portfolio selection based on skewness

• Davis (1995) derives optimal portfolios under the Gamma distribution assumption

• Hlawitscka and Stern (1995) show the simulated performance of mean variance portfolios is nearly indistinguishable from the utility maximizing portfolio

• Cremers, Kritzman and Paige (2003)
  – Use extensive simulations to measure the loss of utility associated with ignoring higher moments in portfolio construction
  – They find that the loss of utility is negligible except for the special cases of concentrated portfolios or “kinked” utility functions (i.e. when there is risk of non-survival).
Conclusions

• The fat tailed nature of high frequency returns is well established
• The nature of the process is usually described as being a fat tailed stable distribution or a normal distribution with time varying volatility
• The process that creates fat tailed distributions probably has to do with rate at which markets can absorb new information
• The existence of fat tails and skew has important implications for asset pricing
• Fat tails probably have relatively lesser importance for portfolio formation, unless there are special conditions such as gearing that imply non-standard utility functions
References

References

references

References

References


References

• Abraham and Taylor, “Pricing Currency Options with Scheduled and Unscheduled Announcement Effects on Volatility”, Managerial and Decision Science 1993
References

References

Improving Returns-Based Style Analysis

Daniel Mostovoy
Northfield Newport Seminar
June 2007
Main Points For Today

• Over the past 15 years, Returns-Based Style Analysis become a very widely used analytical method

• We’re going to review RBSA and discuss useful several improvements to the basic technique
  – Confidence Intervals
  – Testing for Regime Change
  – Kalman Filter / Exponential Weighting
  – Adjusting for Heteroscedasticity

• Recently, RBSA has gained a new usage in connection with hedge fund replication strategies
Basic RBSA

- First introduced as “Effective Asset Mix Analysis” by Sharpe (1988, 1992). Given a time series of returns of a fund, we try to find the mix of market indices that most closely fits the fund returns.

\[ R_t = \sum_{j=1}^{N} W_j R_{jt} + \varepsilon_t \]

- \( R_t \) is the return on the fund during period t.
- \( W_j \) is the weight of index j.
- \( R_{jt} \) is the return on index j during period t.
- \( N \) is the number of indices.
- \( \varepsilon_t \) is the residual for period t.

Sum of \( W_j = 1, \ 0 < W_j < 1 \)

Basically an OLS multiple regression with constrained coefficients.
Refinement #1
Confidence Intervals

• Like any other estimate we need to know if our style weights results are meaningful
• A style weight estimate of “10% small cap value” isn’t very useful if its really “10% +/- 50%”
• We can only analyze a fund to the extent that the spanning indices are not linear combinations of each other
• Correlation between the spanning indices can frequently cause very large confidence intervals on style weights, as the constraints on the coefficients mask multicolinearity that would be observable in an OLS regression
Confidence Interval Problem Was Solved a While Ago


Oddly, most commercial style analysis software packages still do not incorporate any form of confidence interval on the results.
Refinement #2
Allowing Leverage

• Some portfolios such as hedge fund employ explicit leverage
• Other funds have portfolio properties that are possibly outside the range of the spanning indices
  – An equity portfolio with a beta higher than any available
  – A bond portfolio with maturities longer than any available index
• Solution is to include a cash equivalent among the spanning indices and allow only that index to take on negative values
Refinement #3
Testing for Regime Changes

- Do we want to look at fund results over the last 3 years, 5 years, 32 months, etc.?
- CUSUM is an optimum statistic to determine the change in the mean of a process
  - Was adapted for the purposes of monitoring external asset managers by the IBM pension fund
- Use CUSUM based methods to determine the optimal "look-back" period for the style analysis
  - Mathematically: What is the “look back” date such that the cumulative active return between then and now is least likely to have come about by random chance?
- Forthcoming paper by Bolster, diBartolomeo and Warrick summarized in our February 2005 newsletter
Refinement #4
Capturing Recent Influences

• Traditional style weights represent the fund average behavior over a time sample
  – What we should be worried about is a fund that has changed style recently, rendering “average” past information useless

• One Approach
  – Plot the absolute value of the residual against time during the sample
  – If the slope is positive, the fit is getting worse as we come forward in time. Exponentially weight observations until the slope is not statistically significantly positive

• Another way is to use Kalman filtering
  – Kalman filtering requires use of Markov Chain Monte-Carlo analysis if style weights are constrained to be positive
Refinement #5
Asking the Right Question

• An easy experiment
  – Nine year sample period from 1998 through 2006
  – Make up a monthly return stream for a hypothetical fund whose returns are equal to the S&P 500 for 1/3 of the 9 years, equal to the FTSE Europe for 1/3 of the 9 years, and equal to the Merrill Lynch Global High Yield for 1/3 of the 9 years
  – First intuition suggests style weights should be 1/3 S&P 500, 1/3 FTSE Europe and 1/3 MLGHY

• Generally WRONG. It depends on the order of events
## A Curious Result

<table>
<thead>
<tr>
<th></th>
<th>SE</th>
<th>α</th>
<th>$R^2$</th>
<th>S&amp;P</th>
<th>FTSE</th>
<th>MLGHY</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTMLSP</td>
<td>6.63</td>
<td>2.87</td>
<td>0.67</td>
<td>11.65</td>
<td>28.66</td>
<td>59.69</td>
</tr>
<tr>
<td>FTSPML</td>
<td>5.49</td>
<td>-3.43</td>
<td>0.85</td>
<td>39.9</td>
<td>43.44</td>
<td>16.65</td>
</tr>
<tr>
<td>MLFTSP</td>
<td>6.61</td>
<td>-5.33</td>
<td>0.75</td>
<td>5.54</td>
<td>58.01</td>
<td>36.44</td>
</tr>
<tr>
<td>MLSPFT</td>
<td>6.89</td>
<td>-2.16</td>
<td>0.71</td>
<td>23.85</td>
<td>37.59</td>
<td>38.56</td>
</tr>
<tr>
<td>SPFTML</td>
<td>5.79</td>
<td>-0.74</td>
<td>0.87</td>
<td>64.04</td>
<td>34.39</td>
<td>1.56</td>
</tr>
<tr>
<td>SPMLFT</td>
<td>7.29</td>
<td>8.7</td>
<td>0.71</td>
<td>53.46</td>
<td>0</td>
<td>46.54</td>
</tr>
</tbody>
</table>
What’s Going On

• The results are order dependent
  – The style analysis process, like a regression is minimizing the sum of squared residuals
  – The volatility of markets is different across the three sub-periods, and the more volatile periods are counted more heavily
  – Not only do the weights vary across the different orders but goodness of fit changed a lot too

• Variation in alpha estimates ranged from -5.33 to +8.7
  – This huge difference in alpha arises from the “accidental” market timing arising from the ordering

• Averaged across all six possible orders we get our expected result of 1/3, 1/3, 1/3 for weights
Refinement #5
Asking the Right Question

• The more volatile periods do count for more in the returns experienced by investors
  – If we want to know what market returns influenced the returns of the fund, this is the right answer
  – This corresponds to Sharpe’s original concept of “Effective Asset Mix”

• But if we want to know whether a manager’s style was consistent with a prescribed strategy, we have to filter out the effects of heteroscedasticity within the sample period
  – For each time period calculate the “spanning dispersion”, the average absolute difference in return between all possible pairs within the spanning indices
  – Weight the observations inversely with the square root of the dispersion
Refinement #6
Volatility Based Spanning Indices

• Many hedge fund strategies are predicted on the level of market volatility, rather than expected returns
  – Purported uncorrelated with the direction of markets (e.g. writing option spreads)
• Fung and Hsieh (2002) suggest spanning indices that are volatility related
  – Relative returns between mortgage backed securities and coupon bonds are sensitive to interest rate volatility
  – Bondarenko (2004) constructs a index where the return is based on the difference between implied and realized OEX volatility
Using RBSA to Proxy Hedge Fund Holdings

- A common problem in the hedge fund industry is the need to analyze a hedge fund where the holdings of the fund are not disclosed
  - Create a proxy portfolio for risk management and asset allocation purposes
  - Hold the proxy portfolio as a “synthetic” version of the fund
- We will illustrate a procedure for estimating proxy holdings for a fund where the true underlying holdings are unknown.
  - Using a combination of returns based style analysis and portfolio optimization
- Our proxy portfolio is not meant to be a guess at the true underlying portfolio, but rather an efficient estimate of:
  - The typical style bets of the fund
  - The degree of portfolio concentration
  - The balance between asset specific and factor risks.
Selecting the Spanning Indices

- For each fund we need to select the right set of spanning indices
- Over a universe of funds, some indices will be significant to lots of funds, some indices will be significant to only a few funds
- Use what we know about the fund strategy to manually select a set of “likely suspects”
- Start with a large list of indices. Iteratively run the analysis dropping out the least statistically significant.
  - Easy to get fooled because T stats on indices improve as we drop correlated but less significant indices
- Start with a short list of indices representing major asset classes
  - Run analysis, drop insignificant asset classes. Replace remaining indices with sub-indices. Rerun analysis and again drop out insignificant indices
RBSA Analysis Output

By running the style analysis, we get three pieces of information:

– Observed volatility of the subject hedge fund during chosen sample period

– The "style" exposures of the subject fund (growth, value, short volatility, etc.) expressed as percentages of the different indices that best mimic the fund’s return behavior over time.

– The relative proportion of risk coming from style factors and from fund specific risk.
Now Let Us Start to Form Holdings

• Take the constituents of our spanning indices and form a portfolio of these constituents weighted by results of the style analysis.
• If our style analysis said the fund behaved like 50% the S&P 500 and 50% EAFE
• We would form a portfolio that was 50% the weighted constituents of the S&P 500 and 50% EAFE.
  – At this point, we should have a portfolio that has the right "style" exposures to match our fund
• However, these two indices together have about 1600 stocks.
• The resulting portfolio would be far too diversified to represent a typical hedge fund.
  – It is likely to have far lower risk than a real hedge fund portfolio.
Let’s Refine the Proxy Holdings

Now we’ll consider portfolio volatility
• Load the proxy portfolio into the Optimizer as both the benchmark and the starting portfolio.
  – In our example, our version starting portfolio/benchmark would have 1600 stocks.
• We must reduce the number of positions such that the overall risk of the portfolio approximates the observed risk of the subject fund.
  – We can do this by running an optimization while using the "Maximum number of Assets" parameter.
• With a little trial and error, we can find the portfolio that matches the benchmark (and the subject fund) in style.
  – We reduce the diversification to the point where the expected volatility of the proxy portfolio matches the observed volatility of the subject hedge fund.
Check the Balance Between Factor and Asset Specific Risks

• We now load the refined (reduced number of positions) proxy portfolio into the Optimizer as the portfolio with a cash benchmark.

• By running a risk report, we can determine how much of:
  – the expected risk of the refined proxy portfolio arises from factor bets
  – Arises from asset specific risk.
  – If this is a reasonable match to the subject fund (from the style analysis) we're done.
Changing the Balance Between Factor and Asset Specific Risks

• If we find we don’t have the appropriate balance between factor and asset specific risks
  – Repeat the process of “refining” the proxy portfolio
  – In addition to defining the Max Assets parameter, we can change the Optimizer’s degree of risk tolerance for factor and asset specific risk
  – Again with a little trial and error, we can find risk acceptance parameter values that bring the relative proportions of factor risk and asset specific risk into line with our analysis of the subject fund

• We now have our proxy portfolio to hold or use as a composite asset in other analyses
Conclusions

• The effectiveness of Returns Based Style Analysis can be enhanced in a number of important ways

• These enhancements are particularly important in analyzing funds where substantial shifts in strategy may be expected over time
Empirical Example 100 HF

• Ran 100 HF through a set of 14 spanning indices, retained TVValues
• Reduced number of independent variables by adding them in order of decreasing abs(TValue) & Rerunning
  – For each fund, dropped all independent variables with abs(TValue) < .5
Style Analysis Results

- R2 between .005 and .9, averaging about .36
- Interesting trend: the greater Cash Allocation, the lower the R2:
  - The more “hedged” the fund, the less information there is for the style analysis to pick up on...
%Cash Allocation vs Style R2

R2 = .465
Next Step (Empirical Ex. Cont)

- Combined spanning index constituents according to style weights, used result as benchmark, optimized, max 500 assets.
- Harvested resultant expected standard deviation of returns.
- Calculated historic standard deviation of HF returns.
Results

- Slope = .978
- R2 = .486
- 85% of the modeled portfolio risks were smaller than the respective observed HF risks.
Empirical Conclusions

- The Style Analysis does a good job of modeling a Hedge Fund’s Factor Variance.
- A further adjustment of stock specific is required to beef up the Expected Risks to fit.
  - This can be done by iteratively adjusting:
    - UnsysRAP
    - MaxAssets
How Large is the Equity Premium Today?

June 8, 2007

Samuel Thompson
Partner, Research
Arrowstreet Capital, LP
The equity premium

• The equity premium is the expected excess return on a broad stock index over a safe bond market investment
• To measure it, we need to make choices:
  – What stock index? (I will emphasize the World index but also show results for the US and Canada)
  – What safe investment? (I will use long-term inflation-indexed bonds)
  – What starting point? (I will use current conditions)
  – What investment horizon? (I will discuss forecasting methods suitable for a 5-10 year holding period)
  – What kind of expectation? (I will use a geometric average)
Key questions

• Do we believe that the equity premium has declined since 1950?
  – If so, then the historical mean return over-states future returns

• Do we believe in mean reversion? Price-earnings ratios are at historically high levels. Do we think they will fall?
  – If we believe in mean reversion, then we should be extremely pessimistic about stock market returns over the next five years
  – If we do not believe in mean reversion, then regression-based forecasts of the equity premium are overly pessimistic

• Corporate profits are at historical highs. Do we believe they will remain high?
Alternative ways to estimate the equity premium

Two extreme approaches:
1. Take an average of historically realized excess returns.

I prefer two compromise approaches:
3. Adjust historical averages to reflect the decline in market valuation ratios.
4. Predict the equity premium with valuation ratios. Use logic rather than historical statistics to determine the predictive model.
1. Historical average excess returns

- This gives you a high number
  - Dimson, Marsh, and Staunton (DMS, 2006) report geometric averages of 4.7% for the world, 5.5% for the US, and 4.5% for Canada over the period 1900-2005
  - The numbers are even higher in the late 20th Century

- Problem: you need a long historical sample because stock returns are noisy, but over a long period it is plausible that the equity premium changes
  - With 100 years of data and 15% standard deviation of returns per year, the standard error of the estimate is 1.5%
  - Since stock prices rise when the equity premium falls, a decline in the equity premium leads you to increase your estimate just when the true number is falling
2. Predict the market with yields

• In the US, the dividend-price ratio (dividend yield) is close to a historic low, and the smoothed earnings-price ratio (smoothed earnings yield) is also low relative to its 20th Century average.

• Regression results in US data, 1881-2006:
  
  - Realized Annual Premium = -.05 + 2.52 × Prior Dividend Yield
  - Realized Annual Premium = -.06 + 1.78 × Prior Earnings Yield

• This predicts very low premia!
  - World dividend yield currently 1.8% → premium is -.5%
  - World earnings yield currently 4% → premium is 1.1%
2. Predict the market with yields

- Extrapolating from the historical relationship between yields and subsequent returns gives a very gloomy view.
- Why are the regression coefficients greater than 1?
  - Realized Annual Premium = -.05 + 2.52 \times \text{Prior Dividend Yield}
  - When the dividend yield falls by 1%, the equity premium falls by 2.52%. Why is the effect so large?
- Historically, low dividend yields hurt you two ways
  - You earned low dividends
  - *Mean reversion*: dividend yields tended to rise back to historical norms through price declines
- Suppose that there has been a permanent shift in valuations, so we never return to historical norms
  - Then we get low dividends, but do not expect price declines
  - If we do not expect mean reversion, then the future is okay
Low 1973 D/P followed by price declines

1.7% in 2006

Historical average
Two flawed approaches

- The equity premium may have fallen
- When the equity premium falls, the historical mean becomes unreasonably bullish
- When the equity premium falls, forecasts based on the historical relationship between returns and yields become unreasonably bearish
- What I advocate: use yields to forecast the equity premium, but do not assume mean reversion. Low dividend yields mean low dividends, but do not mean that prices will collapse
3. Adjusting the historical average

- DMS and Fama-French (2002) propose the following:
- Historical average returns:
  \[ \text{Avg\{stock returns\}} = \text{Avg\{dividend yield\}} + \text{Avg\{price growth\}} \]
- Adjusted estimate:
  \[ \text{Avg\{stock returns\}} = \text{Avg\{dividend yield\}} + \text{Avg\{earnings growth\}} \]
- What’s the idea?
  - If the equity premium falls, historical price growth will be higher than in the future. Historical earnings growth will not be similarly overstated
  - Suppose that the price-earnings ratio is expected to be stable (so no mean reversion). Then going forward, average price growth equals average earnings growth
  - We estimate price growth going forward by averaging over historical earnings growth
3. Adjusting the historical average

- Adjustment to the 1900-2005 average returns give us a geometric equity premium of 4.0% for the world, 4.8% for the US, 3.5% for Canada
  - This adjustment lowers the historical average by about 0.7% in the US and globally, and about 1% in Canada

- We can further adjust the estimate using today’s dividend yield:
  - \( \text{Avg\{stock returns\}} = \{\text{Today’s dividend yield}\} + \text{Avg\{earnings growth\}} \)
  - This adjustment leads to a geometric equity premium of 2.5% for the world, 3.3% for the US, 2% for Canada

- The adjustments lead to lower but still sizeable equity premiums
4. Steady-state valuation models

• The simplest steady-state model is the Gordon growth model: \( R = \frac{D}{P} + G \)

• That is, returns come from income and capital gains, which in steady state must equal dividend growth

• Use current \( \frac{D}{P} \) and an estimate of \( G \) to infer \( R \)

• The problem with this is that US firms have shifted from dividends to share repurchases, which has altered \( G \) in a way that is hard to estimate

• Campbell and Thompson (2006) find that an earnings-based approach works better
4. Steady-state valuation models

• Use two facts:
  – $D/P = (D/E)(E/P)$
  – $G = (1-D/E) \text{ ROE}$, where ROE is accounting return on equity
• Get an earnings-based formula:
  – $R = (D/E)(E/P) + (1-D/E) \text{ ROE}$
• The rate of return is a weighted average of the earnings yield and profitability, where the payout ratio is the weight on the earnings yield
• In practice, you need to smooth earnings, ROE, and payout ratio to eliminate short-run cyclical noise
• Finally, to get an equity premium number you must subtract an estimate of the real interest rate
4. Steady-state valuation models

- Steady-state approach vs. regression
  - Assume that ROE = E/P. The steady-state prediction is
    \[ R = \frac{E}{P} \]
  - Recall the regression results:
    \[ R = -0.06 + 1.78 \times \frac{E}{P} \]
  - The steady state approach over-rules the regression coefficients of -0.06 and 1.78 with 0 and 1.

- The steady-state approach uses logic rather than historical statistics to determine the relationship between valuation and future stock returns
- The steady-state approach assumes no mean reversion
- Campbell and Thompson (2006) find that in historical data the steady-state approach leads to more accurate stock forecasts than regression-based approaches
Earnings yield

3-Year Smoothed Earnings / Current Price

- World
- US
- Canada
Profitability

Smoothed Real ROE
(3-Year Smoothed Earnings / Current Book Value) - Inflation

- World
- US
- Canada
Payout ratio

Smoothed Dividend Payout Ratio
Current Dividend / 3-Year Smoothed Earnings

- World
- US
- Canada

Real interest rate

Inflation-Indexed Government Bond Yields

UK
US
The equity premium today

Implied equity premium, 03/30/2007

- Method 1: Assume constant real ROE of 6%, dividend payout ratio of 50%
- Method 2: Use 3-year smoothed ROE and dividend payout ratio
- Weighted average: 75% weight on Method 1, 25% weight on Method 2

<table>
<thead>
<tr>
<th></th>
<th>Method 1</th>
<th>Method 2</th>
<th>Weighted Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>World</td>
<td>3.20%</td>
<td>5.51%</td>
<td>3.77%</td>
</tr>
<tr>
<td>US</td>
<td>3.05%</td>
<td>6.75%</td>
<td>3.97%</td>
</tr>
<tr>
<td>Canada</td>
<td>3.02%</td>
<td>4.89%</td>
<td>3.49%</td>
</tr>
</tbody>
</table>
Implied world equity premium

Equity Premium -- World

Assume Constant Real ROE 6%, Dividend Payout Ratio 50%
Use 3-Year Smoothed ROE and Dividend Payout Ratio
Assume Constant Real ROE 6%, Dividend Payout Ratio 50%
Use 3-Year Smoothed ROE and Dividend Payout Ratio
Implied Canada equity premium

Equity Premium -- Canada

- Assume Constant Real ROE 6%, Dividend Payout Ratio 50%
- Use 3-Year Smoothed ROE and Dividend Payout Ratio
The world equity premium today

- The steady-state approach gives results that are highly sensitive to one’s beliefs about corporate profitability.
- If recent profitability is sustainable, with a high reinvestment rate, then the world equity premium is 5.51%.
- If profitability and reinvestment rates return to their late 20th Century averages, then the world equity premium is only 3.20%.
- A reasonable compromise number is 3.8%.
- This is almost one percentage point lower than the 1900-2005 historical average reported by DMS.
- Note that the equity premium is this high only because long-term real interest rates are low.
The equity premium in the US and Canada

- The US numbers are even more sensitive to the assumption about profitability. In Canada the recent profit boom is smaller, so profit sustainability is less important.
- In the US, the compromise number of 4% is 1.5% below the 1900-2005 historical average.
- In Canada, the compromise number of 3.5% is 1% below the 1900-2005 historical average.
- Thus in the US and Canada, we should not expect the future to be as good as the past.
- Reality check: Graham and Harvey (2007) survey CFO’s of US corporations and report a premium of 3.4%.
Conclusion

• Sensible methods for estimating the equity premium give
  – Positive, significant numbers
  – World forecasts are 3.8% today versus 4.7% historically
  – If corporate profitability reverts to long run averages, the world premium falls to 3.2%.
  – Absolute returns will be lower still: real interest rates are about 2% today versus 3.5% in the 1990s

• If we believe in mean reversion, we become very pessimistic. In the past, rising earnings yields have come from falling prices

• If equities have been permanently revalued, then we are much less pessimistic
Northfield Summer Seminar
Newport, RI
June 8, 2007

Implied Risk Acceptance Parameters in the Execution of Institutional Equity Trades

Thorsten Schmidt, CFA
Objective

- Assess the risk-aversion implicit in the execution of institutional trades

Methodology

- Generate implied RAPs from Instinet’s database of institutional trades

Sample

- 21,959 institutional orders between 10/1/06 and 4/20/07
Sample Characteristics

• Algorithmic trade executions
• Orders without limits (avoid selection biases)
• Fully completed orders

Grouping

• We utilize 3 different strategies which are chosen by the trader as a proxy for the level of risk aversion
• These strategies are distinct in terms of the amount of market risk associated with them
Results Measurement

- We use ‘implementation shortfall,’ i.e. difference b/w volume-weighted execution price and last market print at order arrival

Order Strategy Groupings

- **Volume Participation**: Most aggressive trading style, least market risk exposure
- **Implementation Shortfall (“Wizard”)**: Medium aggressive trading style, medium market risk exposure
- **VWAP**: Most passive trading style, highest market risk exposure

* Avg Volume Participation Rate is 26%
<table>
<thead>
<tr>
<th>Strategy</th>
<th># Orders</th>
<th>Average % Avg Daily Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>VWAP</td>
<td>15,847</td>
<td>0.92%</td>
</tr>
<tr>
<td>Impl. Shortfall</td>
<td>4,316</td>
<td>1.95%</td>
</tr>
<tr>
<td>Volume Part.</td>
<td>1,796</td>
<td>2.3%</td>
</tr>
</tbody>
</table>

**Efficient Frontier - All Orders**

- Standard Deviation Bps (Risk)
- Avg Slippage in Bps vs Arrival

- VWAP: 82, -3
- Impl Shortfall: 67, -12
- Volume Part.: 62, -22
Utility Evaluation

Utility (Northfield Definition)

\[ U = \alpha - \left[\frac{STE^2}{SYSRAP} + \frac{UTE^2}{URAP}\right] - \text{Cost} - \text{Penalties} \]

Utility (adapted for this analysis)

\[ U = (0 - \frac{\text{StdDev}^2}{SYSRAP/100} - \text{abs(Slippage)^1}) \times 100 \]

Note

• Because institutional trade executions lose money on average, the utility will always be negative
• We are trying to minimize disutility

\(^1\) Expressed in decimal format
Impl Shortfall (Medium Risk) starts to have more utility than VWAP (High Risk) below RAP of 3.
Volume Participation, as typically utilized* (Low Risk) is the strategy with the most utility only below RAP of 0.7! Is anyone this risk averse?

* Avg Volume Participation Rate is 26%
Among orders ADV <2%, Impl Shortfall does not show a strong comparative advantage.
For ADV < 2%, Volume Participation becomes a strategy with most utility at RAP 3 and below because of lower potential to 'do damage' with smaller orders.
### Strategy Performance (2% to <4% ADV)

<table>
<thead>
<tr>
<th>Strategy</th>
<th># Orders</th>
<th>Average % Avg Daily Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>VWAP</td>
<td>914</td>
<td>2.8%</td>
</tr>
<tr>
<td>Impl. Shortfall</td>
<td>736</td>
<td>2.8%</td>
</tr>
<tr>
<td>Volume Part.</td>
<td>253</td>
<td>2.8%</td>
</tr>
</tbody>
</table>

**Efficient Frontier - 2% to <4% ADV**

- **VWAP**: 
  - Standard Deviation Bps (Risk)
  - Avg Slippage in Bps vs Arrival

Among ADVs 2% - 4%, Impl Shortfall is the most efficient strategy.
VWAP and Impl Shortfall intersect b/w RAP 7-8, Volume Participation has lowest utility through the entire RAP spectrum down to 1.
Among ADVs 4% - 6%, Impl Shortfall is the most efficient strategy.
VWAP and Impl Shortfall intersect around RAP 5, Volume Participation intersects with VWAP at RAP 1.
Summary

• A lot of trade executions performed at implied RAPs <10, and even at RAP <1
• Most balanced institutional portfolios have RAPs of 50-75 (rule of thumb: desired tracking error * 6)

Why are many institutional trades this risk averse?
• Short-term alpha
• Workflow reasons (get done and move on)
• Psychology of trader-PM relationship
Short-term alpha test

- Assume 50 RAP
- VWAP strategy has most utility
- Spread b/w VWAP and Volume Participation is 19bps (-3bps VWAP, -22 Volume Participation)
- 19bps intraday alpha decay * 252 trading days = 47.88%
- Do you have 47.88% annualized alpha?
Conclusions

• The reasons for low implied RAPs are structural and behavioral
• Rigorous analysis of trading performance can shed light on implied RAPs of executions
• It can help align the RAPs of your trading with those of your portfolio
• Talk to your broker (e.g. Instinet) to help you do this
Thorsten Schmidt, CFA
FVP Algorithmic Trading
thorsten.schmidt@instinet.com

Tel 212.310.4177
This presentation is for information purposes only and does not constitute an offer, solicitation, or recommendation with respect to the purchase or sale of any security. System response times may vary for a number of reasons, including market conditions, trading volumes and system performance.

© 2007 Instinet, LLC. All rights reserved. INSTINET is a registered mark. Instinet, LLC, member NASD/SIPC.
Alpha Scaling Revisited

June 8, 2007

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

- A procedure that forecasts variable(s) $y$ from signals $g$

- One way - fit a linear model
  \[ \hat{y}(\hat{g}) = A \hat{g} + b \]

Minimizing expected squared error,
\[ \hat{y}(\hat{g}) = E(y) + \text{cov}(y, g) \text{cov}(g, g)^{-1} [\hat{g} - E(g)] \]

- End of presentation
The Real Goal

- Build forecasts that are suitable for optimization

- Additional considerations
  - optimizers seek extremes (by mandate!)
  - all forecasts have error
  - optimized selection introduces bias

- Most forecasting methods, e.g. Grinold formula, address the 1st half. This talk focuses on the 1st half

- Shrinking all forecasts towards the average forecast across securities, for example, is one way to deal with the 2nd half

- The 1st is about forecasting, the 2nd about optimizing with errors
More About The Difference

- Forecasting returns for many stocks from a set of signals
- In basic linear model, each forecast is constructed separately
  \[
  \hat{y}(\hat{g}) = E(y) + \text{cov}(y, g) \text{cov}(g, g)^{-1} [\hat{g} - E(g)]
  \]

  \[
  \hat{y}(\hat{g}) = [\hat{y}_1(\hat{g}), \hat{y}_2(\hat{g}), \ldots]^T
  \]
  \[
  \hat{y}_k(\hat{g}) = E(y_k) + \text{cov}(y_k, g) \text{cov}(g, g)^{-1} [\hat{g} - E(g)]
  \]

- So, stock 1’s forecast is concerned with signals that affect stock 2 to the extent that those signals also affect stock 1. It is not at all concerned with the final forecast for stock 2
- In contrast, portfolio optimization compares securities against each other and cares only about relative value
Articulate So Formulas Make Sense

- Say what the signals are
  e.g. $g_i = \text{stock i’s raw alpha forecast}$
  \( \text{stock i’s most recent earnings surprise} \)
  \( \text{stock i’s cross-sectional rank} \)
  \( \text{change in 90 day T-bill yield} \)

- Say what you are forecasting
  e.g. $y_k = \text{stock k’s return}$
  \( \text{stock k’s return – market return} \)
  \( \text{stock k’s return net of market } \beta \text{ and industry} \)
  \( \text{stock k’s stock specific return} \)
I: One Signal Per Stock – Grinold

- Recall basic linear model
  \[ \hat{y}_k(\hat{g}) = E(y_k) + \text{cov}(y_k, g) \cdot \text{cov}(g, g)^{-1} [\hat{g} - E(g)] \]

- Forecast \( y_k \) using only signal \( g_k \)
  \[ \hat{y}_k(\hat{g}) = E(y_k) + \text{cov}(y_k, g_k) \cdot \text{var}(g_k)^{-1} [\hat{g}_k - E(g_k)] \]
  \[ = E(y_k) + \rho(y_k, g_k) \times \text{std}(y_k) / \text{var}(g_k) \cdot [\hat{g}_k - E(g_k)] \]
  \[ = E(y_k) + \rho(y_k, g_k) \times \text{std}(y_k) / \text{var}(g_k) \times [\hat{g}_k - E(g_k)] / \text{std}(g_k) \]
  
  IC volatile score
I: 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
- Likewise, you would expect lower IC’s for volatile securities (harder to predict) than for less volatile ones (easier to predict.) A single IC exaggerates volatile securities’ alphas
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: $\sigma_{DELL}^{ss} = 27\%$, $\sigma_{MSFT}^{ss} = 25\%$, $\sigma_{PEP}^{ss} = 9\%$

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

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

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

- let \( y_k \) = stock \( k \)'s return above the market
  \( g_k \) = stock \( k \)'s relative attractiveness, e.g. xc percentile rank

- \( \hat{y}_{IBM} = E(y_{IBM}) + \rho(y_{IBM}, g_{IBM}) \times \text{std}(y_{IBM}) \times \left[ \hat{g}_{IBM} - E(g_{IBM}) \right] / \text{std}(g_{IBM}) \)

- Assume that returns and signals are alike across stocks
  \[ E(g_k) = E(g) \quad \text{std}(g_k) = \text{std}(g) \]
  \[ E(y_k) = E(y) = 0 \quad \text{std}(y_k) = \text{std}(y) \]
  \[ \rho(y_k, g_k) = \rho(y, g) \]

- \( \hat{y}_k(\hat{g}) = \rho(y, g) \times \text{std}(y) \times \left[ \hat{g}_k - E(g) \right] / \text{std}(g) \)
II: Cross-Sectional Grinold (cont)

- \( \hat{y}_k(\hat{g}) = \rho(y, g) \times \text{std}(y) \times \frac{[\hat{g}_k - E(g)]}{\text{std}(g)} \)
  - \( [\hat{g}_k - E(g)] / \text{std}(g) = \text{xc z-score} \)
  - \( \text{std}(y) = \text{xc volatility of returns} \)
  - \( \rho(y, g) = \text{xc correlation}(y, g) \)

- Ignored Realities
  - High/low \( \beta \) securities are less likely to be near 0
  - Specific risk increases the chance that a security is in the tails
  - Volatile securities are harder to predict

- Solution: vary IC and vol by group
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): \( \text{IC}_{\text{tech stocks}} = 0.08 \), \( \text{IC}_{\text{consumer stocks}} = 0.12 \)

\[
\hat{y}_{\text{MSFT}} = 0.08 \times 0.15 \times 3 = 0.36 \\
\hat{y}_{\text{DELL}} = 0.08 \times 0.15 \times 2 = 0.24 \\
\hat{y}_{\text{PEP}} = 0.12 \times 0.15 \times 1 = 0.18
\]
Motivated by need to stabilize asset allocation optimization

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

New information given as portfolio forecasts with error
- IBM’s return is 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%

Combined forecast is expected value given prior and information
prior on mean returns:
- \( m \sim N(m_0, \Sigma_0) \)

forecasts impart new information:
- \( \hat{g} = P \, E[m \mid \text{info}] + \varepsilon \)
- \( \varepsilon \sim N(0, \Omega) \)

\[
\hat{y} = \left[ \Sigma_0^{-1} + P^T \Omega^{-1} P \right]^{-1} \left[ \Sigma_0^{-1} m_0 + P^T \Omega^{-1} \hat{g} \right]
= m_0 + \left[ \Sigma_0^{-1} + P^T \Omega^{-1} P \right]^{-1} P^T \Omega^{-1} (\hat{g} - Pm_0)
= m_0 + \left[ \Sigma_0 - (P \Sigma_0 P^T)^{-1} P \Sigma_0 \right] P^T \Omega^{-1} (\hat{g} - Pm_0)
\]

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.

Note: Holding the benchmark in benchmark relative optimization implies an alpha of 0 for all securities.
III: 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%

- \( \mathbf{m}_0 = (2\% \ 5\%)^\text{T}, \ \Sigma_0 = (4 \ 3; \ 3 \ 9) \ %^2 \)

- \( \mathbf{P} = (1 \ 0), \ \hat{\mathbf{g}} = 5\%, \ \Omega = 9\%^2 \)

- Updated forecasts: \( \hat{y}_{\text{IBM}} = 2.9\% \), \( \hat{y}_{\text{DELL}} = 5.7\% \)
IV: Extending Black Litterman

- Consider as underlying securities all the stock specific returns and all the returns to factors, e.g. \( \mathbf{m} = (m^{ss}_{IBM}, m^{ss}_{DELL}, \ldots, m^{E/P}, m^{GROWTH}, \ldots) \)

- 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
V: Bulsing, Scowcroft, Sefton

- General framework for integrating stock-specific, portfolio, and factor forecasts

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

built out with places for the different pieces

- Encompasses Grinold and Black-Litterman as specific cases

- More flexible and precise, but required information about error covariances makes implementation hard
Signal Decay & Turnover

- 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 $\approx$ time-weighted average over reference holding period of each interval’s best forecasts.

\[
e.g. 2 \text{ yr, } 8\% \text{ annualized over } 1^{st} 6 \text{ mo, } 1\% \text{ over remaining } 18 \rightarrow \frac{1}{4} \times 8\% + \frac{3}{4} \times 1\% = 2.75\%
\]
How Long Is A Stock Held?

- Turnover is 25% / yr → ~ 4 yrs.

- Turnover is 25% / yr, but 90% of the securities are in common with the benchmark
  → (10 %) / (25 % / yr) = 0.4 yrs.
Northfield’s Alpha Scaling Tool

- Seek a theoretically sound, information preserving, robust way of refining alpha forecasts
- Don’t know client’s alpha generating process
- Only have client’s forecast alpha
- Sophisticated methods leverage information. Fully reasonable yet wrong guesses at what we don’t know erase any benefits over simpler approaches
- Beginning from alpha forecasts (not individual stock z-scores) necessitates a cross-sectional framework
Preprocessing for Robustness: Rank Rescaling

- Map raw signals by rank onto standard normal
e.g. 25th percentile $\rightarrow F^{-1}(0.25)$
Estimating Cross-Sectional Volatility

- Expected market weighted cross sectional variance
  \[
  E[\Sigma_s w_s (r_s - r_m)^2] \quad \text{where } r_m = \Sigma_s w_s r_s
  \]
  \[
  = E[\Sigma_s w_s (r_s - \mu_s + \mu_s - \mu_m + \mu_m - r_m)^2]
  \]
  \[
  = \Sigma_s w_s \sigma_s^2 - \sigma_m^2 + \Sigma_s w_s (\mu_s - \mu_m)^2
  \]
  \[
  \approx \Sigma_s w_s \sigma_s^2 - \sigma_m^2
  \]
  \[
  = \text{avg stock variance} - \text{variance of the market}
  \]

- Numbers come straight from risk model
  - If forecasting return net of \(\beta\), 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 prevents confusion.
- It is important to be aware of horizon, particularly in low-turnover portfolios.
- For many practitioners, Northfield’s upcoming alpha scaling functionality can make your life easier by doing the work for you.
References


Distinguishing Between Being Unlucky and Unskillful

Dan diBartolomeo
Northfield Information Services
June 2007
What We’re After

• There are numerous performance metrics used as proxies for manager skill such as alpha, and information ratio
  – Most of these rarely have statistically significant values because you need a long time series of data, over which time conditions are presumed but not guaranteed to be stable
  – We would like a measure that uses more information so we can get statistically meaningful results over a shorter window

• Manager’s occasionally experience very bad return outcomes for a period of time
  – We need a means to discriminate the manager being bad from a truly random event
What We Probably Don’t Care About

• There is an enormous literature in finance regarding whether asset managers collectively exhibit skill
  – Obvious implications for concepts of market efficiency
  – Most of this work is based on the concept of “performance persistence”: those that perform consistently well must be skillful

• But we want to evaluate only one manager
Usual Methods

• There is lots of literature on using traditional return performance metrics such as alpha and information ratio as proxies for manager skill:

• You need very long time series of return observations to have enough data to get anything statistically significant by which time conditions may change

• Just going to daily data doesn’t help
More on Skill Detection

• Some research has been done on CUSUM methods

• Tries to isolate what portion of a manager’s history is likely to be relevant to current activities
  – Throw away data from before the most likely date of a regime shift
The Breakdown Problem

• Consider a real manager who maintains a below 3% ex-ante tracking error and has a cumulative return of -6.3% over a one year period
  – Is this a 2.4 standard deviation event? If so the manager was very unlucky
  – Was the risk model wrong? Maybe the risk model was underestimating the risk so it’s not such a rare event
  – Expected tracking error averaged 2.74%, realized was 2.80%

• Ex-ante tracking error estimate is the expectation of the standard deviation of the active return, which is measured around the mean
  – Mean active return was -.54% per month

IR as Skill


• IR = IC * Breadth\(^5\)
  - IR = alpha / tracking error
    IC = correlation of your return forecasts and outcomes
    Breadth = number of independent “bets” taken per unit time

• If we know how good we are at forecasting and how many bets we act on, we know how good our performance should be for any given risk level
The Fundamental Law Makes Big Assumptions

• There are no constraints at all on portfolio construction
  – Positions can be long or short and of any size
• We measure only “independent” bets
  – Buying 20 different stocks for 20 different reasons is 20 different bets
  – Buying 20 stocks because they all have a low PE is one bet, not 20!
• Transaction costs are zero, so bets in one time period are independent of bets in other periods
  – This is the property that casinos depend on. Once we have the odds in our favor, we want to make lots of bets
• Research resources are limitless so our forecasting effectiveness (IC) is constant as we increase the number of eligible assets
Enter the Transfer Coefficient


- \[ IR = IC \times TC \times \text{Breadth}^5 \]
  - \[ IR = \frac{\text{alpha}}{\text{tracking error}} \]
  - \[ IC = \text{correlation of your return forecasts and outcomes} \]
  - \[ TC = \text{the efficiency of your portfolio construction (TC < 1)} \]
  - \[ \text{Breadth} = \text{number of independent “bets” taken per unit time} \]
What Drives the Transfer Coefficient?

• Imagine a manager with a diverse team of analysts that are great at forecasting monthly stock returns on a large universe of stocks, but whose portfolio is allowed to have only 1% per year turnover
  – Good monthly forecasts, diverse reasons and a large universe imply high IC and high breadth
  – But if we can never act on the forecasts because of the turnover constraint TC can be zero or even negative

• If we can’t short a stock that we correctly believe is going down, or take a big position in a stock that we correctly believe is going up, TC declines
  – The more binding constraints we have on our portfolio construction, the more return we fail to capture when our forecasts are good
  – For bad forecasters, a low TC is good. You hurt yourself less when you constrain your level of activity
Limitations of IR

• Managers often talk about IR, but it really doesn’t correspond to investor utility except in extreme cases
  – Consider a manager with an alpha of 1 basis point and a tracking error of zero
  – IR is infinite but value added for the investor is very, very small

• The statistical significance of a ratio is hard to calculate
Our Solution is to Incorporate Cross-Sectional Information

• Successful active management involves forecasting what returns different assets will earn in the future, and forming portfolios that will efficiently use the valid information contained in the forecast
  – We usually have a large universe of assets to work with, so we get statistical significance quickly
  – In mathematical terms, this means that the position sizes within our portfolios balance the marginal returns, risks and costs
  – If we know how good we are at forecasting future asset returns, we can forecast how well our portfolios should perform if they are efficiently constructed. If we do less well than we should, our portfolio construction is at fault
A Quant Way to Think About It

• Every portfolio manager must believe that the portfolio they hold is optimal for their investors
  – If they didn’t they would hold a different portfolio

• If we describe investor goals as maximizing risk adjusted returns, we know that the marginal risks associated with every active position must be exactly offset by the expected active returns
  – Guaranteed by the Kuhn Tucker conditions for finding the maximum of a polynomial function
  – For every portfolio, there exists a set of alpha (active return) expectations that would make the portfolio optimal. We call these the implied alphas
The Effective Information Coefficient

• We define the EIC as the skill measure
• EIC is the pooled average rank correlation of the implied alphas and the realized returns at the security level
  – If our forecasting skill is good (high IC) and our portfolio construction skill is good (high TC) then EIC will be high
  – If either IC or TC is low, EIC will be low
• As this measurement involves every active position during each time period, the sample is large and statistical significance is obtained quickly
Using EIC to Dissect Performance

• If we have EIC values for a given period (e.g. month), we can estimate the expected magnitude of alpha
• The expected alpha is just the EIC times the cross-sectional dispersion of the asset returns
• So we can look at returns as:

\[ P_t - B_t = EIC_t * D_t + \varepsilon_t \]

- \( P_t \) = portfolio return during period \( t \)
- \( B_t \) = benchmark return during period \( t \)
- \( EIC_t \) = skill for period \( t \)
- \( D_t \) = cross sectional dispersion of asset returns
- \( \varepsilon_t \) = residual returns due to luck

• You can now look at the time series standard deviation of the \( \varepsilon_t \) to see if the risk model is predicting risk accurately
An Alternative View

- Active managers can add value in two ways:
  - Being right more often than they are wrong about which securities will outperform the market. Sort of like a batting average in baseball
  - Getting bigger magnitude returns on gainers than on losers. You can have a batting average below 50% and still make money if you hit a decent number of “home runs”

- Peter Lynch used to refer to “ten baggers”
  - Stocks that go up ten fold in value while you hold them
  - Just a couple can have a huge effect on portfolio returns

- Batting average concept first formalized in:
Attribution to Batting Average and Active Return Skew

• A formalization was proposed by hedge fund manager Andrei Pokrovsky (formerly Northfield staff) in 2006

• We can easily measure batting average
  – It is the percentage of cases in which active returns and active weights are of the same sign
  – High numbers are good

• Take the vector product of active weights and active returns. Measure the skew statistic of the distribution
  – Positive skew in active returns is a measure of portfolio construction efficiency
Style Dependency of Batting Average and Active Skew

- Value managers will tend to have high batting average and low skew
- Growth/momentum managers will tend to have lower batting average but higher skew
- Trend following behavior creates the skew
  - diBartolomeo (2003)
Conclusions

• It is often difficult to assess whether a period of extraordinary performance (good or bad) is the result of luck or skill

• We propose the Effective Information Coefficient as a measure of skill
  – It is estimated both over time and across assets so sample sizes get large quickly
  – It incorporates both key aspects of investment skill, forecasting returns and forming efficient portfolios

• We present an alternative representation of skill as “batting average” and “payoff skew”