1 Style & Factors

1.1 Introduction to factor investing

- Traditional investment choices **active and passive**.
  - Passive is put your money in the index S&P 500,and not try to outperform the index.
  - Active management is to try to outperform index.(that's why active managers are expensive and inconsistent).
- **Indexation** of a portion of active management. Indexation refers to the trend of creating new indices that capture the portion of active management that is rules based and systematic, and in the long run should outperform the cap-weighted benchmark. It is both active(not just cap weighted index) and passive(no subjective views).
- Factor: a variable which influences the returns of assets. It represents commonality in the returns something outside of the individual asset.

  Factor premium: a risk premium associated with the factor. The excess return will be gained if one is willing to expose him/herself to the factor.
- Three types of factors.
  - Macro factors: such as growth and inflation - not easy to trade
  - Statistical/implicit factors: something extracted from the data that may or may not be identifiable. The commonality is intrinsic or implicit in the data. - not easy to trade
  - Intrinsic/style factors(most common): such as value versus growth, momentum and low volatility.

1.2 Factor models and CAPM

- Factor model: decompose R into sum of premia

$$R_i = \beta_1 f_1 + \beta_2 f_2 + \beta_3 f_3 + ... + \alpha + \epsilon$$

- CAPM: a factor model which only has only factor - market

$$E(r_i) - r_f = \frac{cov(r_i, r_m)}{var(r_m)}(E(r_m) - r_f)$$
$$E(r_i) - r_f = \beta_i(E(r_m) - r_f)$$

- Various anomalies: have explanations that arise from a **rational risk-based story** or a **behavioral story**
  - **Rational**: High expected returns compensate for negative returns during certain periods. The key is defining what those bad times are
  - **Behavioral**: High expected returns result from agents' under- or over-reaction to news and/or the inefficient updating of beliefs. Behavioral biases persist because there are barriers of the entry of capital

- For some anomalies, the explanations are largely rational for others. Mostly behavioral.
  - value/growth: rational and behavioral
  - momentum, low-vol: behavioral, mostly
- Value stocks usually outperform growth stocks. Low-vols outperform high-vols.Low beta stocks outperform high beta stocks.
  - Value stocks:
    Def: Shares of a company that appears to trade at a lower price relative to its fundamentals, such as dividends, earnings, or sales, making it appealing to value investors.

    Characteristics: High dividend yield, low P/B ratio, and a low P/E ratio.

    Typically has a bargain-price as investors see the company as unfavorable in the marketplace.
  - Growth stocks:
    Share in a company that is anticipated to grow at a rate significantly above the average growth for the market.
    These stocks generally do not pay dividends. This is because the issuers of growth stocks are usually companies that want to reinvest any earnings they accrue in order to accelerate growth in the short term.
    When investors invest in growth stocks, they anticipate that they will earn money through capital gains when they eventually sell their shares in the future.
    Growth stocks often look expensive, trading at a high P/E ratio, but such valuations could actually be cheap if the company continues to grow rapidly which will drive the share price up
- Supplements
  1. P/B ratio(price-to-book):
     compare a firm's market capitalization to its book value

     $$\frac{\text{stock price per share}}{\text{book value per share (BVPS)}}$$

  2. P/E ratio(price-to-earnings):
     the ratio for valuing a company that measures its current share price relative to its per-share earnings (EPS).A high P/E ratio could mean that a company's stock is overvalued, or else that investors are expecting high growth rates in the future.

     $$\frac{\text{current share price}}{\text{per-share earnings (EPS)}}$$

1.3 Multi-Factor and Fama-French

- **Size factor**: book/price(measure of 'value').The size effect is a well-observed effect that seems to suggest that small cap stocks outperform large cap stocks.

- Fama-French model uses three factors:
  - value versus growth(HML): High$\frac{\text{Book}}{\text{Price}}$(value) minus Low $\frac{\text{Book}}{\text{Price}}$(Growth)
  - size factor(SMB): small versus big - Small minus big stocks
    Provides a portfolio which long in small stocks and short in big stocks. The difference becomes the return of the factor.
  - market risk factor(MKT): beta
  -

$$E[r_i] = r_f + \beta_{i,MKT}E[r_m - r_f] + \beta_{i,SMB}E[SMB] + \beta_{i,HML}E[HML]$$

- Carhart model uses four factors:
  - HML
  - SMB
  - MKT
  - momentum factor: winner versus losers
- Fama and French interpret the small stock effect and the value effect as being systematic factors.
- SMB and HML are zero-cost portfolios, so $\beta_{i,SMB}$ and $\beta_{i,HML}$ are centred around zero.(?)
- Factors today: Low-vol(low beats high), value, momentum and quality(high beats low). Size in practice is treated as a universe choice.

1.4 Factor benchmarks and Style analysis

- The simple CAPM is re-interpretable as a factor benchmark. Consider a portfolio with a beta of 1.3.

$$\begin{aligned} E(r_i) - r_f &= \alpha + 1.3(E(r_m) - r_f) \\ E[r_i] &= \alpha + [-0.3r_f + 1.3E(r_m)] \end{aligned}$$

  Factor benchmark is a short position of \$0.30 in cash(t-bills) and a leveraged position of \$1.30 in the market portfolio.

  α shows the value added by the manager.

- Style analysis is introduced by Sharpe(1992), and allows the benchmark to change over time. It decomposes the returns into some explanatory returns. By looking at the coefficients of these explanatory returns,one can tell the style of the manager.

  Factor/style loadings are re-estimated every period using data up until the current period.

  Sliding window allows for style drift of the manager.

- Model:

$$\begin{aligned} R_m = W_1R_{i1} + W_2R_{i2} + W_3R_{i3} + ... + \alpha + \epsilon \\ \text{S.T. } W_i > 0 \text{ and sum}(W_i) = 1 \end{aligned}$$

Solved through quadratic programming repeated for a sliding window of 1-3 years over time.

Quality of fit:

$$\text{PSEUDO } R^2 = \frac{(\text{VAR(Rm)-VAR})(\epsilon))}{\text{VAR(Rm)}}$$
$$\alpha = \text{manager value added}$$

- Sharpe's paper(1992) - Fidelity Magellan fund

1.5 Shortcomings of cap-weighted indices

- Default approach to investment management is to use a cap-weighted index as a benchmark.

  It has two benefits:

  1. It is simple and very easy to understand, their construction process is very simply.
  2. It is consistent with the aggregate holding of the stocks within the context of the market portfolio by the aggregate investor.

  However, it is inefficient. According to Haugen and Baker(1991), CW portfolios occupy positions inside the efficient set.(also look at Schwartz(2000), and Platen and Rendek(2010))

  Why? It is due to lack of diversification. Cap-weighted indices tend to be heavily concentrated poorly diversified portfolios.

- Other benchmark choices:
  - Equally-weighted benchmarks
  - Minimum variance benchmarks
  - Risk-parity benchmarks

1.6 From cap-weighted benchmarks to smart-weighted benchmarks

- Shortcomings of CW
  1. CW indices provide an inefficient diversification of unrewarded and specific risks

     Solution: smarter weighting schemes

  2. CW indices provide an inefficient exposure to rewarded systematic risks. The portfolio will by construction have a bias towards large cap stocks as opposed to mid or small cap stocks, and a bias in favour of growth stocks as opposed to value stocks.

     Just the opposite to the work of Fama and French.

     Solution:
     Step 1: Select your desired factor exposure
     Step 2: Select your preferred weighting schemes

2 Robust estimates for the covariance matrix

2.1 The Curse of Dimensionality

- To obtain efficient frontier of n securities, one must estimate
  - n expected returns
  - n volatility parameters
  - $[n(n-1)/2]$ correlations

- We expect to run into a serious as we need lots of data to estimate that many parameters with any degree of accuracy. Hence, we reduce the number of parameters to estimate.
  - increase sample size - sample period/frequency
  - decrease number of parameters - number of assets n / number of parameters for a fixed n

- Quiz 1:
  What is the number of parameters required for mean-variance optimization based on the S&P 500 universe, which contains 500 stocks?

  Answer: We need 500 expected return estimates and $\frac{500*499}{2} = 124,750$ covariance parameter estimates, for a total of 125,250 parameter estimates.

- Extreme example 1:
  No model risk - high sample risk

  The simplest estimate is given by the sample covariance estimate:

$$\hat{S}_{ij} = \frac{1}{T}\, \Sigma_{t=1}^{T} \left(R_{it} - \bar{R}_i\right)\left(R_j t - \bar{R}_j\right)$$
$$\text{with } \bar{R}_i = \frac{1}{T}\, \Sigma_{t=1}^{T} \left(R_{it}\right)$$
$$\bar{R}_j = \frac{1}{T}\, \Sigma_{t=1}^{T} \left(R_{jt}\right)$$

- Extreme example 2:
  High model risk - low sample risk

  Constant Correlation Model: assume identical ρ for all ρ_{ij}

$$\hat{\sigma}_{ij}^{CC} = \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}$$

  Cut the number $N\frac{n(n-1)}{2}$ of correlation parameters down to 1

  The optimal estimator of this constant correlation is the 'global' average

$$\hat{\rho} = \frac{1}{n(n-1)}\, \Sigma_{i,j=1,i\neq j}^{n} \hat{\rho}_{ij}$$

- Quiz 2:
  What is the number of parameter estimates required for mean-variance optimization based on the S&P 500 universe, when using the constant correlation covariance matrix estimate?

  Answer: We need 500 expected return estimates, 500 volatility parameter estimates, and also one correlation parameter estimate.

- 1973 by Elton and Gruber

  The **out-of-sample estimate** for the minimum variance portfolio constructed using the **constant correlation estimate** was better than the one using the sample estimate. So indeed the trade-off was worth it reducing sample risk, even though it came at the cost of huge amount of **model risk** that eventually kind of paid off.

- Wrap-up
  - In the presence of large portfolios the number of parameters is often larger than the sample size.
  - Increasing frequency is necessary but not always sufficient in terms of increasing sample size.
  - Introducing structure helps dealing with sample risk. But this comes at the cost of model risk.

2.2 Estimating the Covariance Matrix with a Factor Model

- Factor-based covariance estimate

  Assume stock returns are driven by a limited set of factors

  $$R_it = \mu_i + \beta_{i1}F_{1t} + ... + \beta_{ik}F_{kt} + ... + \beta_{iK}F_{Kt} + \epsilon_{it}$$

  where β_{ik} is the sensitivity of asset i with respect to factor k(k = 1,...K)

- An example with two factors
  - Variance

    $$\sigma_i^2 = \beta_{i1}^2\sigma_{F_1}^2 + \beta_{i2}^2\sigma_{F_2}^2 + 2\beta_{i1}\beta_{i2}Cov(F_1, F_2) + \sigma_{\epsilon_i}^2$$

  - Covariance

    $$\sigma_{ij} = \beta_{i1}\beta_{j1}\sigma_{F_1}^2 + \beta_{i2}\beta_{j2}\sigma_{F_2}^2 + (\beta_{i1}\beta_{j2} + \beta_{i2}\beta_{j1})Cov(F_1, F_2) + Cov(\epsilon_{it}, \epsilon_{jt})$$

  Introduce structure by imposing that residuals are uncorrelated. If the factor model is well specified, assuming that the specific components are uncorrelated is not too bad an assumption

  $$Cov(\epsilon_{it}, \epsilon_{jt}) = 0$$

- Case with K uncorrelated factors

- General decomposition of returns

$$cov\left(R_i(t), R_j(t)\right) = \Sigma_{k=1}^{K}\beta_{ik}\beta_{jk}\sigma^2_{F_k} + cov\left(\epsilon_i(t), \epsilon_j(t)\right)$$

- Assume uncorrelated residuals

$$\sigma_{ij} = cov\left(R_i(t), R_j(t)\right) = \Sigma_{k=1}^{K}\beta_{ik}\beta_{jk}\sigma^2_{F_k}\text{for } i \neq j$$
$$\sigma_{ii} = cov\left(R_i(t), R_i(t)\right) = \Sigma_{k=1}^{K}\beta^2_{ik}\sigma^2_{F_k} + \sigma^2_{\epsilon_i}\text{for } i = j$$

- Quiz 1
  How many parameters do you need to estimate when using a 2-factor models for estimating the covariance matrix of a universe of 500 stocks?

  Answer:
  We first need 500 volatility estimates for individual stock returns, plus 500 estimates of betas of stocks with respect to factor 1, 500 estimates of betas of stocks with respect to factor 2, and finally 2 volatility estimates for factor returns, which gives a total of 500+500+500+2=1,502, which compares favorably to $\frac{500*499}{2} = 124,750$ when using the sample covariance matrix estimate.

- Choice of the factor model
  - Sharpe's single-factor market model
  - Multi-factor model
    * Explicit factor model - Macro factors
    * Explicit factor model - Micro factors
    * Implicit factor model - Statistical factors

- Wrap-up
  - Using a factor model is a convenient way to reduce the number of risk parameters, and to estimate while introducing a reasonable amount of model risk.
  - An implicit factor model is often preferred since it lets the data tell us what the relevant factors are, thus alleviating model risk.

2.3 Honey I Shrunk the Covariance Matrix!

- Trade-off between sample risk and model risk

  Sample risk:
  too many parameters to estimate

  model risk:
  constant correlation methodology or the factor base methodology

  a shrinkage approach to estimating the covariance matrix will mix sample risk and model risk

- Statistical shrinkage

$$\hat{S}_{shrink} = \hat{\delta}^* \hat{F} + \left(1 - \hat{\delta}^*\right) \hat{S}$$

- Weight constraints versus statistical shrinkage
  - Academic research shows that the mixture works better than either approach.
  - (Jagannathan and Ma 2003):
    Imposing constraints on weights is equivalent to perform statistical shrinkage

- Quiz:
  Consider two stocks with sample volatility estimates at 20% and 30%, respectively, and sample correlation at 0.75. Further assume that the average of the sample correlation estimates of all stocks in the universe is 0.5. What is for these two stocks the sample-based covariance estimate, the constant correlation covariance estimate and the covariance estimate based on statistical shrinkage with a shrinkage factor of 50%?

  Answer:
  The sample-based estimate is $20\% * 30\% * 0.75 = 0.045$.

  The constant correlation estimate is$20\% * 30\% * 0.5 = 0.03$.

  The shrinkage estimate is $\frac{0.045+0.03}{2} = 0.0375$.

- Wrap-up
  - Statistical shrinkage allows one to find the optimal trade-off between sample risk and model risk
  - It is based on an average of two covariance matrix estimates. One with high sample risk and one with high model risk.
  - performing statistical shrinkage is formally equivalent to introducing min/max weight constraints

2.4 Portfolio Construction with Time-Varying Risk Parameters

- Curse of non-stationarity - risk parameters might be time-varying

- Estimating volatility

  Denote σ_T as the volatility per day between day T and day T+1 as estimated at end of day T

$$\sigma_T^2 = \frac{1}{T} \Sigma_{t=1}^T (R_t - \bar{R})^2$$
$$\bar{R} = \frac{1}{T} \Sigma_{t=1}^T R_t$$

Assume mean value of R_t is zero

$$\sigma_T^2 = \frac{1}{T}\Sigma_{t=1}^{T}(R_t)^2$$

- Quiz 1:
  Consider the following stream of returns: +1%, -2%, -1%, +2%. What are the corresponding (arithmetic) average return and volatility estimates?

  Answer:
  Average return is $\frac{1\%-2\%-1\%+2\%}{4} = 0\%$. Variance of returns is $\frac{1\%^2-2\%^2-1\%^2+2\%^2}{4} = 0.025\%$. Volatility is square-toot of variance: $\sqrt{0.025\%} = 1.58\%$.

- Increasing **frequency** is better than increasing **sample period** in case of non-stationary return distributions.

- Trade-off between **expanding window analysis** and **rolling window analysis**

  Expanding window analysis:
  - A situation whereby as time goes by and you come up with new estimate for volatility, you're expanding the window as the name says, and you're increasing sample size.
  - Suitable for the case where asset returns are stationary and volatility is constant, because that's going to give a bigger sample.

  Rolling window analysis:
  - Keeps the constant,the size of the sample and at every point in time you're moving forward the window while keeping a given constant size.
  - Suitable for the case where asset returns are not stationary and volatility moves around, so as to only look at the most recent data.

- Quiz 2:
  What type of data would give you the best estimation power for covariance matrix parameters, assuming constant parameters?
  Answer:
  This gives you $52 * 5 = 260$ data points. Of course, if risk parameters are time-varying, it may not be such a good idea to use data extending over such a long time period.

- Wrap-up

  We have reasons to believe that volatility changes over time.
  In this context using rolling windows is better than using expanding windows.

In all cases, historical volatility estimates are backward looking in nature. They give an estimate for the average volatility over the sample period.

2.5 Exponentially weighted average

- Weighting scheme

  Instead of assigning equal weights to the observations we can set

$$\sigma_T^2 = \Sigma_{t=1}^{T} \alpha_t R_t^2$$
$$\text{where } \Sigma_{t=1}^{T} \alpha_t = 1$$

- EWMA model

  In an exponentially weighted moving average model, the weights decline exponentially as we move back through time, which leads to:

$$\alpha_t = \frac{\lambda^{T-t}}{\Sigma_{t=1}^{T} \lambda^{T-t}}$$

  Covariance parameter estimate:

$$cov(R_i, R_j) = \Sigma_{t=1}^{T} \alpha_t \left(R_{i,t} - \bar{R}_i\right)\left(R_{j,t} - \bar{R}_j\right)$$

  The lowest the Lambda and the fastest the decrease of the importance of the oldest data point in the sample when estimating covariance parameters.

  Value for lambda around 0.9 has been found to be a reasonable value.

  Ok to use expanding windows.

- Quiz:

  If youre using 10 years of daily returns and an exponentially weighted moving average estimator for the covariance matrix, should you be using a rolling window analysis or an expanding window analysis?

  Answer:

  Since we are using an exponentially weighted moving average estimator, the importance of past data gradually fades away as time goes by, and there is therefore no need to use rolling window analysis.

- Rolling window analysis makes sense if you're concerned about time-varying risk parameters but the **problem with rolling window analysis** is as long as the data point is within the rolling window, it matters, and whenever it is out of the rolling window, it doesn't matter at all. So the day before it gets out, it's as important as the most recent observation, the next day it's out of the rolling window, it doesn't matter.

2.6 ARCH and GARCH Models

- Arch model
  In an ARCH(T)model, we also assign some weight to the **long-run variance** V_L:

$$\sigma_T^2 = \gamma V_L + \Sigma_{t=1}^{T} \alpha_t R_T^2$$
$$\text{where } \gamma + \Sigma_{t=1}^{T} \alpha_t = 1$$

  ARCH(1) model:

$$\sigma_T^2 = \gamma V_L + \alpha R_T^2$$
$$\text{where } \gamma + \alpha = 1$$

- GARCH model
  In GARCH(1,1), we additionally assign some weight to the previous variance estimate to capture **volatility clustering**

$$\sigma_T^2 = \gamma V_L + \alpha R_T^2 + \beta \sigma_{T-1}^2$$
$$\text{with } \gamma + \alpha + \beta = 1$$

- Quiz 1:
  Suppose the estimation of a GARCH(1,1) model on daily data gives:

$$\sigma_T^2 = 0.000002 + 0.13 R_T^2 + 0.86 \sigma_{T-1}^2$$

  and also suppose the last daily estimate of the volatility is 1.6% per day and the most recent percentage change in the market variable is 1%. What is the new daily volatility estimate?

  Answer:
  The new variance estimate is: $0.000002 + 0.13 * 0.0001 + 0.86 * 0.000256 = 0.00023516$. Taking the square root, we obtain the new volatility estimate as 1.53% per day.

- Variations on GARCH
  GARCH(p,q):

$$\sigma_T^2 = \omega + \Sigma_{i=1}^{p} \alpha_i R_{T-i}^2 + \Sigma_{j=1}^{q} \beta_j \sigma_{T-j}^2$$
$$\omega = \gamma V_L$$

- Factor GARCH
  GARCH models are very convenient in terms of time-varying parameters, but they increase the curse of dimensionality.

  Orthogonal (0)GARCH model:

  $$\sigma_{ij}^{OGARCH} = \hat{\sigma}_{ij}(t) = \Sigma_{k=1}^{K} \hat{\beta}_{ik} \hat{\beta}_{jk} \hat{\sigma}_{F_k}^2(t)$$

- Quiz 2:
  How many parameters do you need to estimate when using a 2-factor models with GARCH(1,1) model for the volatility of each one of the two factors?

  Answer:
  We first need 500 volatility estimates for individual stock returns, plus 500 estimates of betas of stocks with respect to factor 1, 500 estimates of betas of stocks with respect to factor 2, and finally 3 GARCH parameter estimates for each factor, which gives a total of $500 + 500 + 500 + 2 * 3 = 1,506$, which is not much more than if we had assumed constant volatility parameters.

- Wrap-up
  GARCH models explicitly account for the time-varying nature of volatility, but they require additional parameter estimates.

  A parsimonious approach consists of mixing them with a factor model:
  Allowing the variance of the (uncorrelated factors to follow a GARCH model)