Quantitative Equity Portfolio Management

[Book Link](#)
Author(s): Ludwig and Deahwan
Publish Year: 2006
Tags: #Portfolio_Management #Factor_Investing #idea_factor_modelling

Part One: An Overview of QEPM

- "If return are somewhat predictable over the long run, then stable quantitative models should work more reliably than picking individual stocks intermittently on qualitative information"

Chapter 1: The Power of QEPM

- "Factor tilting can be a very effective way of managing an equity portfolio and is one of the many powerful tools of QEPM"
- "The first step in building a model of stock returns is to choose a mix of explanatory variables for the model"

Chapter 2: The Fundamentals of QEPM

- "Given the returns of the reference instrument, it is possible to [...] decompose the portfolio's returns into two parts—one that is related to the reference instrument and one that is not. The related part is typically called *expected returns* or *consensus returns*, and the second part is known as *residual return* or *the returns not explained by the model*"
- Benchmark α
    - "Given the portfolio returns r_P and the benchmark return r_B, we can estimate:"

$$r_P = \alpha + \beta r_B + \epsilon \tag{2.1}$$

    - $\alpha + \epsilon$ is the residual return. "Benchmark α is the expected value of the residual return, and [...] ϵ is the deviation of the residual return from its mean. Equation (2.1) is constructed so that ϵ averages zero"
- CAPM α
    - "Given the portfolio returns r_P and the market return r_M, we can estimate:"

$$r_P = \alpha + \beta r_M + \epsilon \tag{2.2}$$

    - "The CAPM α, or α^{CAPM}, is the risk adjusted excess return over the market"
- Multifactor α
    - "Given the portfolio returns r_P and the a series of factor returns $f_1, f_2, \ldots, f_K$, we can estimate:"

$$r_P = \alpha + \beta_1 f_1 + \cdots + \beta_K f_K + \epsilon \tag{2.3}$$

- "If all the other factors in the APT model are uncorrelated with the first factor (the benchmark), then the benchmark α is given by

$$\alpha^B = E(\alpha + \beta_1 f_1 + \cdots + \beta_K f_K) \tag{2.5}$$

The benchmark α includes the effect of all the other factors. [...] In general, if we allow the benchmark α to be correlated with all other factors, the benchmark α is given by

$$\alpha^B = E(\alpha + \beta_1 f_1 + \cdots + \beta_K f_K) - (\beta_2 \gamma_2 + \cdots + \beta_K \gamma_K)E(f_1) \tag{2.6}$$

where γ_j is the coefficient in the regression of f_1 on f_j ",i.e.:

$$\frac{Cov(f_1, f_j)}{Var(f_1)}$$

"when measuring the performance of a manager with benchmark α, benchmark α can be positive even if multi-factor α equals zero".

- "The *ex-ante* α is the expected α, whereas the *ex-post* α is the realised one. The *ex-ante* α is of interest to the quantitative portfolio manager when he or she is constructing his or her portfolio"
- "The ex-ante information ratio *(IR)* is given by:

$$IR = \frac{\alpha^B}{\omega} \tag{2.7}$$

  where α^B is the expected or predicted α of the portfolio manager, and ω is the predicted standard deviation of the residual (i.e., S(ϵ))"
- **Seven Tenets of QEPM**:
    1. *Markets are mostly efficient.*
    2. *Pure arbitrage opportunities do not exist.*
    3. *Quantitative analysis creates statistical arbitrage opportunities.*
    4. *Quantitative analysis combines all the available information in an efficient way.*
    5. *Quantitative models should be based on sound economic theories.*
    6. *Quantitative models should reflect persistent and stable patterns.*
    7. *Deviations of a portfolio from the benchmarks are justified only if the uncertainty is small enough.*
- "the fundamental law [Advances in Active Portfolio Management](#) is for *understanding* QEPM, not for *doing* QEPM. A quantitative portfolio manager does not use the law to actually create portfolios. Rather, the law provides the manager with a useful conceptual framework for analysing the QEPM process"
- "Given the definition of the information ratio in Eq. (2.7):"

$$\frac{\alpha^B}{\omega} = IC \cdot \sqrt{BR} \tag{2.9}$$

- "One of the essential tasks of QEPM is to predict future stock returns by estimating a model that specifies a relationship between the stock return and a list of explanatory variables (a.k.a *factors*). Suppose that the model specifies that the return of stock i at time t, r_{it}, is a linear function of the value of K factor premiums at time t, that is $f_{1t}, f_{2t}, \ldots, f_{Kt}$:

$$r_{it} = \alpha_i + \beta_{i1}f_{1t} + \cdots + \beta_{iK}f_{Kt} + \epsilon_{it} \tag{2.10}$$

  where $\alpha_i, \beta_{i1}, \ldots, \beta_{iK}$ are parameters to be estimated, and ϵ_{it} is the random-error term (i.e., the deviation of the stock from its expected value)"
- "The fundamental law assesses how well Eq. (2.10) explains the stock-return process, and it expresses the equation's goodness of fit as the product of *number of explanatory variables* and each variable's *average contribution*."
- "When constructing the portfolio, a manager should use all the information gathered from the estimation of Eq. (2.10)"
- **Lemma 1 (THE INFORMATION CRITERION)** "*The fundamental law of active portfolio management as expressed in Eq. (2.9) is valid only if the portfolio manager combines all the information in the most efficient way*"
    - Example: portfolio manager following analyst-revision strategy: "Construct an equal-weight portfolio of stocks whose analyst rations improved in the last month. We can consider 'inclusion in the portfolio' a variable. That is, we may define a variable β_{it} that has a value of 1 if stock i's rating improved at a time t and a value of 0 otherwise. The manager estimates:

$$r_{i,t+1} = \alpha + \beta_{it}f + \epsilon_{i.t+1} \tag{2.11}$$

  where $r_{i,t+1}$ is the return to stock i at time $t + 1$, α and f are the parameters to be estimated, and $epsilon_{i,t+1}$ is the error (the part of the return that is not explained by the model). If the value of f is positive, the equation suggests that those stocks with $\beta_{it} = 1$ will have higher future returns. Thus the portfolio manager constructs an equal-weighted portfolio of stocks for which $\beta_{it} = 1$, indicating that their ratings improved in the last month". But the portfolio manager is not using all the available information. "Stocks with higher volatility should smaller weights in the eventual portfolio. Since the portfolio manager did not use this piece of information, the information criterion is not satisfied."
- Data Mining
    - "*Data mining* is the practice of running regressions of historical stock returns on so many combinations of factors that one is virtually guaranteed of finding a model or handful of factors that seem significantly correlated with stock returns but are in fact not particularly meaningful."
    - "We cannot completely avoid data mining or data snooping when we test a factor unless we use data that have never before been used to test that particular factor"

- "careful empirical study is not enough to avoid the data-mining problem. The best way to generate meaningful statistics is to follow tenet 5 and make sure that the model is based on sound economic theory and common sense"
- Parameter Stability
  - "For example, if we believe that β is generally not constant for more than five years, we should not estimate it from a sample that includes data covering more than five years."
- Parameter Uncertainty
  - "Even if the active portfolio does not look too risky on its surface, it could have a high degree of parameter uncertainty, which is a component of risk"
  - "The standard error, which measures the error of the estimation, depends on two things: the variance of the error in the model and the variance of the explanatory variables"
  - "the portfolio manager should consider not only the estimated value but also the standard error of the estimates. If the standard error is large, then that should be considered part of the portfolio's risk. This is one of the fundamental implications of Bayesian econometrics for portfolio selection". Discussed in Chapter 14.

Chapter 3: Basic QEPM Models

	Fundamental Factor Model	Economic Factor Model
Data type	P/E ratio, Mcap, etc.	GDP, macro variables; other variables
Factor Exposure (β)	directly observed	must be estimated using time series regression
Factor Premium (f)	must be estimated using cross-sectional regression	Directly observed (may be determined by zero-investment portfolio or PCA)
Expected Return	factor exposure x factor premium	factor exposure x factor premium
Risk	factor exposure x risk of factor premium + specific risk	factor exposure x risk of factor premium + specific risk
Criterion for selecting		
Theoretical basis		Better
Factor accommodation		Better
Ease of implementation	Better	
Data requirement	Better	
Intuitive appeal	Depends	Depends
Assumption	stable factor premium	stable factor exposure

- **Basic QEPM Models and Portfolio Construction Procedures**

- "*Factors* are explanatory variables that represent different types of risk. A factor model shows that the average stock return is proportional to the stock's exposure to the risk that factor represents (the *factor exposure*) and to the payoff for each unit of exposure to risk (the *factor premium*)"
  1. **Fundamental factor model**
     - "uses fundamental factors, which are stock characteristics such as P/E ratio and market cap"
     - Factor exposure directly observable using financial statements and other data sources.
     - "factor premium must be estimated from the historical relationship between stock returns and the factor exposures"
  2. **Economic factor model**
     - originally only macroeconomic variables "but is general enough to handle other types of factors as well"

- "stock's exposure to economic (or other) factors is not observable and instead must be estimated from the historical relationship between stock returns and factor premiums"
- "factor premium be determined up to a proportionality without a statistical estimation in certain cases. In other cases, it is determined by constructing *zero-investment portfolios* or through a mathematical method called *principal-component analysis*"
- "*fundamental factor model and the economic factor model employ different techniques for modelling stock returns.*"
- "The specific steps involved in constructing a portfolio depend partly on which model is used"
- Factor choice
  - "Good factors are really the secret sauce of QEPM"
  - "The economic factor model is more flexible, allowing for both economic factors and all fundamental factors"
- Data decision
  - cross-sectional dimension: universe selected; "attributes of the data set will affect the attributes of the ultimate portfolio and number of stocks will affect the ease of estimation"
  - time-series dimension: periodicity and time period of the entire data set; "may affect parameter stability and parameter uncertainty"
  - "decide whether to use the gross return and whether to use the return in excess of a risk-free rate"
- Factor exposure (*factor loading*)
  - Straight-forward in fundamental factor model; "if the factor is the P/B ratio, the factor exposure to the stock i is the latest observed value of the P/B ratio for stock i"
  - economic model: "the factor premiums must be determined first. Then the factor exposure can be estimated from the relationship between returns and factor premiums"
  - given factor premium $f_1, \ldots, f_K$, return of stock i, r_i, can be estimated:

  $$r_i = \alpha_i + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + \epsilon \qquad (3.1)$$

  - "estimated by a time-series regression using observations made at various time periods. [...]. For example, the factor premium f_j might be the real GDP growth. The factor exposure β_{ij} found by estimating the regression shows how sensitive stock returns are to real GDP growth"
- Factor premium
  - fundamental factor model: "factor premium is estimated from the historical relationship between the stock return and factor exposure. [...] by means of either cross-sectional regression, which involves using observations for various stocks at a single point in time, or panel regression, which entails using observations for various stocks at many points in time"
  - Macroeconomic factors: "the values of the variable themselves are taken as factor premiums. For example, if the inflation rate is 3% this month, then the premium for the inflation factor is 3%"
  - "We can determine the factor premium by calculating the return on a zero-investment portfolio of investments that exhibit the factor in question"
    - example for size factor: "A zero-investment portfolio could be created by going long a small-cap sub-portfolio and shorting a equivalent amount of a large-cap sub-portfolio. The factor premium on size would be the difference between the return of the small-cap sub-portfolio and the return of the large-cap sub-portfolio"
    - `This is the approach I am currently thinking of taking, i.e. calculate factor premiums first and then calculate the factor loadings. This is the approach Sanskar was talking about 'being more statistical' in nature. Sanskar is using the Fundamental Factor Model, as opposed to the approach I have in mind: Economic Factor Model`
- Expected return
  - Estimation of the model "essentially allows us to determine the expected returns of stocks based on their factor exposures and the factor premiums"
- Risk
  - "Non-diversifiable risk is represented by $r_i = \alpha_i + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K$ in Eq. (3.1)"
  - "Diversifiable risk, on the other hand, can be removed from the portfolio by diversifying holdings. It is captured by ϵ_i. often called *stock-specific risk*"
  - Stock's total risk can be estimated by "calculating the variance or the standard deviation of each component of risk"
- Forecasting

- "In all likelihood, the factor exposure will not change in the immediate future, but the factor premium *is* likely to change in a very short amount of time. Thus it is usually necessary to forecast the factor premium. We explain how to forecast in Chapter 8"
    - Security Weighting
        - "With the estimates of stock returns and risks based on the new forecasts, the portfolio manager can use optimisation techniques to select the stocks for the portfolio and assign them their respective weights"
- **The Equivalence of The Basic Models** #idea_factor_modelling
    - **Lemma 5 (THE EQUIVALENCE OF FACTOR MODELS)** "*The fundamental factor model and the economic factor model are equivalent to each other if the expected stock return is a linear function of the fundamental factor exposure*"
    - Proof: "First, we assume that expected stock returns are a linear function of fundamental factor premiums, as described by the fundamental factor mode. Then, under this assumption, we show that the expected stock return can be expressed as a linear function of the factor exposure, as described by the economic factor model"
    - "For practitioners, the equivalence of the factor models means that to the extent that the fundamental factor model is a correct model, it does not really matter whether one uses the fundamental factor model or the economic factor model. The two models produce essentially identical results. The only catch is that if the fundamental factor is not correct, then it is best to use the economic factor model because it has a stronger theoretical basis than the fundamental factor model does"
- The Screening and Ranking of Stocks with the z-score
    - "a Z-score is stock's standardised exposure to a fundamental factor"
- Hybrids of the Models and the Information Criterion
    - "Practitioners often try to add extra inputs to the basic QEPM models by combining them with ad hoc models. [...] The most critical problem with hybrids is that they frequently violate the information criterion."
    - Example of the danger of combining a fundamental factor model with a z-score analysis, i.e. "calculates the expected return based on the z-score method and adds this expected return to the fundamental factor model as a constant"
    - Two stocks, and their returns determined by "P/E ratios of the previous period":

$$r_{A,t+1} = 0.1\left(\frac{P}{E}\right)_{A,t} + \epsilon_{A,t+1} \quad \epsilon_{A,t+1} \sim N(0,20) \tag{3.8}$$

$$r_{B,t+1} = 0.1\left(\frac{P}{E}\right)_{B,t} + \epsilon_{B,t+1} \quad \epsilon_{B,t+1} \sim N(0,10) \tag{3.9}$$

    "We also assume that the covariance between the two random variables ϵ_A and ϵ_B is zero." Suppose P/E for A = 20 and B = 10. "Suppose further that the average and the variance [of P/E] are constant over time and across stocks"
    - **The Z-Score Model**
        - Table 3.2 shows the assumed P/E ratios and the subsequent z-score.

TABLE 3.2

Price-to-Earnings Ratio and the Z-Score

	Stock *A*	Stock *B*	Mean	Variance	SD
$\left(\frac{P}{E}\right)_T$	20	10	15	25	5
z_T	1	−1	0	1	1

        - Using the z-score, prediction of the expected stock return:

$$r_{i,t+1} = a_i + bz_{it} + v_{i,t+1} \quad i = A, B; t = 1, \cdots, T-1 \tag{3.10}$$

        where z_{it} is the z-score of the stock i at time t, a_i is a constant, and v_{t+1} is the error. Using Table 3.2 estimates of a_A, a_B, and b ("by demeaning i.e., subtracting the mean from both sides, of the equation and applying the standard ordinary least squares (OLS) formula"):

$$b = \frac{1}{2}\frac{C(r_A, z_A)}{V(z_A)} + \frac{1}{2}\frac{C(r_B, z_B)}{V(z_B)}$$

$$= \frac{1}{2}\sqrt{V\left(\frac{P}{E}\right)_A}\frac{C\left(r_A, \left(\frac{P}{E}\right)_A\right)}{V\left(\frac{P}{E}\right)_A} + \frac{1}{2}\sqrt{V\left(\frac{P}{E}\right)_B}\frac{C\left(r_B, \left(\frac{P}{E}\right)_B\right)}{V\left(\frac{P}{E}\right)_B} \tag{3.11}$$

$$= \frac{1}{2}\cdot 5 \cdot 0.1 + \frac{1}{2}\cdot 5 \cdot 0.1 = 0.5$$

$$a_i = E(r_i) - bE(z_i) \quad i = A, B$$
$$= 1.5 - 0.5 \cdot 0 = 1.5 \tag{3.12}$$

- Covariance(x,y)/Variance(x) gives the linear association between x and y, i.e. if y = mx + b, then C(x,y)/V(x) = m, and C(x,y)/V(y) = 1/m
- "Given the time T value of the Z-score, the preceding estimates imply the following expected returns for stock A and stock B:"

$$E(r_{A,T+1}) = a_A + bz_{A,T} = 1.5 + 0.5 \cdot 1 = 2 \tag{3.14}$$

$$E(r_{B,T+1}) = a_B + bz_{B,T} = 1.5 + 0.5 \cdot (-1) = 1 \tag{3.15}$$

"The Z-score model itself does not create any problem with efficient use of information. The problem arises when the Z-score model is combined incorrectly with other models"

- **A Hybrid of the Z-Score Model and a Fundamental Factor Model**
    - "Adam first estimated the fundamental factor model, i.e., Eq. (3.1). Then he did the Z-score analysis. Now he create his hybrid model by setting the term α in the fundamental factor model according to the expected return from the Z-score model. [...] For time T + 1, he incorrectly modifies Eq. (3.1) into:"

$$r_{A,t+1} = 1 + 0.1\left(\frac{P}{E}\right)_{A,t} + \epsilon_{A,t+1} \quad \epsilon_{A,t+1} \sim N(0, 20) \tag{3.16}$$

$$r_{B,t+1} = 1 + 0.1\left(\frac{P}{E}\right)_{B,t} + \epsilon_{B,t+1} \quad \epsilon_{B,t+1} \sim N(0, 20) \tag{3.17}$$

   - "To construct his optimal tangent portfolio, Adam finds" the weight of stock A to be 100% and stock B to be 0%.
- Information Loss
    - "he is actually not achieving the maximum Sharpe ratio because his estimates of the expected return are not correct"
    - "Given the actual Sharpe ratio and the maximum Sharpe ratio, the information loss (IL) is the difference between the two" which is 0.1005 in this case.
- "The moral of the story is clear. The best thing to do is to choose one model and stick with it. Combining basic models is hard to justify [...]. It does no good to tack an ad hoc model onto the one that the software generates, so designing an entire custom model may be the best course of action."
- Choosing the right model
    - "The economic factor model upholds tenet 5 better than the fundamental factor does. [...] An economic factor model can measure the sensitivity of stocks to various economic risk factors."
        - fundamental factor typically assumes "the factor premium is stable for some time. Similarly, in implementing the economic factor model, the assumption is often that the factor exposure is stable for some time".
    - "If a portfolio manager wants to use a mix of all types of factors, the economic factor model is the necessary choice."
    - "For fundamental factor model, only factor premiums need to be estimated. For the economic factor models, factor exposures of individual stocks need to be estimated, and forecasts of factor premiums are usually needed as well."
    - Fundamental factor models do not require a large amount of historical data other than to "estimate the variance of individual stock returns, which is ultimately required to build optimised portfolios." "On the other hand, to estimate the economic factor model, portfolio managers have to gather a relatively large period of historical returns because the estimation of factor exposures requires a time-series analysis of returns."

Part Two: Portfolio Construction and Maintenance

Chapter 4: Factors and Factor Choice

	Univariate Regression Tests	Multiple Regression Tests	Unidimensional Zero-Investment Portfolio	Multidimensional Zero-Investment Portfolio
Used in	Fundamental Factor Model	Fundamental Factor Model	Economic Fac… Model	Economic Factor Model
Equation	$r_{i,t} = \alpha + f\beta_{i,t} + \epsilon_{i,t}$	$r_{i,t} = \alpha + f_1\beta_{i1,t} + \cdots + f_K\beta_{iK,t} + \epsilon_{i,t}$		
Pros	simple to perform; to get an early idea of what might be relevant	considers relationship among factors		
Cons	not helpful if the factors are correlated	Data mining is a big risk		
Remarks		ideal for <15 factors		

- "Factors come in many varieties: fundamental, technical, macroeconomic, and alternative".
- "Good factors therefore exhibit relationships with stock returns that not only are stable and persistent but also can be explained by economic theory".
- Fundamental Factors
    - "We group the fundamental factors into six subcategories:"
    1. Valuation factors
    2. Solvency factors
        - "companies with high solvency factor values are more liquid or solvent and less likely to be forced into bankruptcy by payments owed on debt and other liabilities"
    3. Operational efficiency factors
        - "attempt to describe how efficiently a firm operates over both the short and the long term"
        - Inventory turnover ratio, total asset turnover, etc.
    4. Operational profitability factors
        - "One or more of these operating profitability factors (or transformations of these factors) appear in most QEPM models"
    5. Financial risk factors
        - high debt-to-equity ratios are not always bad; "businesses that generate large, stable cash flows can handle relatively large amounts of debt because they can finance the interest payments on the debt"
        - In 2003, IBM's D/E ratio of 0.61 was well above the industry average of 0.35, "but its interest coverage of 63.11 times also was greater than the industry average. Its earning power covered its debt burden."
        - "Low financial risk does not just avert disaster; it is a big plus for expansion and new investment"
    6. Liquidity risk factors
        - "gauge whether a stock's liquidity matches trading demand"
        - Trading Turnover "is calculated as the average daily trading volume (ADV) divided by market capitalisation, measures the percentage of outstanding shares traded in a given period."
- Technical Factors
    - "Portfolio managers typically use technical factors to capture very short-term changes in the relative value of stocks."
- Economic Factors
    - "For macroeconomic models of stock returns, the portfolio manager should choose factors that affect all stock returns to some degree"
- Alternative Factors
    - Analyst factors
        - "Wall Street analysts' earnings forecasts, buy-sell recommendations, and related information can be valuable in forecasting stock returns."
        - "many quants look for good neglected (not widely covered by analysts) stocks using a multi-factor model that examines fundamental characteristics and assigns a bonus to neglected status"
    - Corporate finance policy factors
        - stock buybacks and insider purchases.
    - social responsibility factors

- "Probably the most useful factor in this arena is the employee relations factor"
- Factor Choice `#idea_factor_modelling`
  - "To choose factors for the fundamental factor model, we suggest using the **univariate** and **multiple regression** techniques. To choose factors for the economic factor model, we suggest using the **unidimensional** and **multidimensional zero-investment portfolio** techniques"
  - **Univariate regression Tests**
    - "Many portfolio managers [...] performing a series of simple regressions of factors. [...] run panel regressions on each factor versus the stock returns in the universe. Thus"

$$r_{i,t} = \alpha + f\beta_{i,t} + \epsilon_{i,t} \tag{4.1}$$

    "estimate f from this panel regression [...]. If a factor has a significant value of f, then the factors is useful in explaining stock returns".
    - This is not really helpful if the factors chosen are highly correlated with each other.
    - Helpful to provide first level cut if there are too many factors
  - **Multiple regression Tests**
    - "if the portfolio manager includes too many factors in this regression, there is a chance of *misspecification bias*"
    - If there are fewer than 10 factors, "then a multivariate panel regression can be estimated:"

$$r_{i,t} = \alpha + f_1\beta_{i1,t} + \cdots + f_K\beta_{iK,t} + \epsilon_{i,t} \tag{4.2}$$

    "Once this regression is estimated, all insignificant factors are dropped. The significant factors go into the factor model."
    - "One of the biggest dangers here is data mining"
    - "Data mining can be a serious problem when one uses the *stepwise regression* technique. This method involves starting with a collection of factors and then running a regression on every possible combination of the factors and picking the set of factors with the largest R^2."
  - **Unidimensional Zero-Investment Portfolio**
    - "Typically, a manager will split the universe of stocks conditional on a particular factor exposure into thirds, quintiles, or deciles. [...] a portfolio is created from the first division, and another portfolio is created from the last division. The returns of the top division are subtracted from those of the last division"
    - Steps for P/B ratio factor:
      1. "rank the universe of stocks by their P/B ratio exposure".
      2. "create an equal-weighted portfolio stocks in the first quintile and an equal-weighted portfolio of stocks in the fifth quintile".
      3. "compute the returns of the two portfolios for each month"
      4. "calculate statistics on the historical returns of the first-quintile portfolio minus the fifth-quintile portfolio"
      - "repeated for every factor"
    - One can do a statistical test of whether the average return of the zero-investment portfolio "is significantly different from zero". Calculate portfolio returns for T periods and obtained $r_1, \ldots, r_T$. Then the t statistic:

$$\hat{t} = \frac{\bar{r}}{\hat{S}(\bar{r})} = \frac{\bar{r}\sqrt{T}}{\sqrt{\frac{1}{T-1}\sum_{s=1}^{T}\left(r_s - \bar{r}\right)^2}} \tag{4.3}$$

    where $\bar{r}$ is the average return and $\hat{S}(\bar{r})$ is the standard error of the mean.
    - If t statistic > 2, "the average portfolio return is significantly different from zero, and the factor is statistically significant."
  - **Multidimensional Zero-Investment Portfolio**
    - "Zero-investment portfolio also can be created by considering many factor simultaneously".
    - If there are two factors, each stock is assigned a ranking based on each factor. "Based on these rankings, the portfolio manager can group stocks into a number of joint quintiles or deciles". Say the portfolio manager wants to create 10 groups out of each factor, "then there will be 100 groups eventually [...] By taking intersections of these portfolios"
    - "The method to create the zero-investment portfolio depends on what the portfolio manager is interested in. If he or she is interested in whether small size and low P/B ratio together influence stock returns, he or she

could create the zero-investment portfolio by taking a long position on a small-low portfolio a short position on a large-high portfolio"

- Techniques to Reduce the Number of Factors #idea_factor_modelling
    1. *within-correlation coefficient*
        - "correlation coefficient adjusted for the panel structure of the data":

$$\rho(X, Y) = \frac{\sum_t C_t(X, Y)}{\sum_s \sqrt{V_s(X)V_s(Y)}} \tag{4.4}$$

    2. Kendall's τ
        - "shows the rank correlation (i.e., whether the ranking by one variable is related to the ranking by another variable)"
        - "We calculate the numerator (the number of pairs for which the rankings by two variable are the same minus the number of pairs for which the rankings by two variables are the opposite of each other) and the denominator (the total number of pairs) for each period and then aggregate over time"

- **Appendix 4A link to the tables**

Chapter 5: Stock Screening and Ranking

- "There are two kinds of stock screening, sequential and simultaneous"
    - Sequential: "eliminate stocks from the investment universe fist based on the most important criterion, then based on the second most important criterion, and so on"
    - Simultaneous: "The more advanced version of stock screening [...]. the manager applies all criteria to the investment universe at once." *Aggregate Z-score approach* "involves choosing factors that explain stock returns and aggregating the factors into one score with which to rank the stocks"
- "The stock screening and ranking methods in this chapter are useful starting points for people just beginning to build quantitative portfolios"
- **Sequential Stock Screening**
    - example: if you want to invest in 60 stocks with low P/B ratio, you rank the universe according to the P/B ratio and buy the lowest P/B ratio stocks.
    - Good stock screen: easy to automate, and easy to replicate.
    - "The portfolio manager's success lies in his or her ability to identify what it is that leads to superior stock returns."
    - `Authors go on to discuss different screening strategies used by famous investors`
- **Simultaneous Screening and the Aggregate Z-Score**
    - "The entire list of factors is taken as a single set of criteria with which to evaluate all stocks in the investment universe."
    - "there is no chance of eliminating an otherwise good stock simply because it does not measure up during an early round of screening."
    - "make sure that the Z-scores match the underlying factor values. That is, if high Z-scores are going to represent good stocks and low Z-scores bad stocks, then the factors themselves must be expressed such that high values are good and the low values are bad."
    - Each stock's aggregate Z-score is a linear combination of the Z-scores of all the factors of that stock.
    - To deal with outliers, "round all Z-scores above 3 down to 3 and similarly round up all Z-scores below -3 to -3."
    - "Some portfolio managers weight the factor Z-scores according to the factors' information ratios, obtained by creating decile of quintile portfolios of each factor and computing the factor's historical information ratio."
    - Optimal weighting of Z-scores: "The portfolio manager takes a series of monthly returns on all the stocks in the universe, combined with the factor Z-score values for each stock at the beginning of the month, and runs a set of cross-sectional regressions over the sample period to find the optimal Z-score weights." #idea_factor_modelling
        - "Suppose that we have K factor Z-scores for each of the N stocks in the universe and that we have T periods of return information.[...] One can estimate the following regression with a sample of historical data:"

$$r_{i,t} = \gamma_i + \delta_1 z_{i,1,t-1} + \delta_2 z_{i,2,t-1} + \cdots + \delta_K z_{i,K,t-1} + \epsilon_{i,t} \tag{5.3}$$

        where γ_i represents a constant term, δ_K is the coefficient estimate for the contribution of the factor Z-score k to the stock returns. "The δ estimates are the optimal combination of the factor Z-scores."
    - Dividing the factors into groups (for example: P/B, P/E, and P/S in valuation group) offers several benefits.

- **Aggregate Z-Score and Expected Return**
  - "Expected stock returns can be found from aggregate Z-scores by performing a regression of actual stock returns against the aggregate Z-scores. This can be done using a panel regression of stock returns on the Z-score of the prior period"

$$r_{i,t} = \gamma_i + \delta z_{i,t-1} + \epsilon_{i,t} \tag{5.8}$$

    Given the estimate of γ_i and δ, expected return of stock i for time $T+1$:

$$E(r_{i,t+1}) = \gamma_i + \delta z_{i,T} \tag{5.9}$$

    Problems with this approach: "Z-scores may not change from one period to another, but factor premiums may change substantially. [...] there may be a low correlation between the Z-score and the subsequent returns. [...] this procedure adds complexity to the process, which is a weakness given that the biggest strength of the aggregate Z-score model is its simplicity"
  - **The equivalence between the Z-score model and the fundamental factor model**
    - "**the Z-score model of Eq. (5.8) produces an expected return identical to that of the fundamental factor under certain conditions**"
- **Aggregate Z-Score and The Multifactor α**
  - "Typically, many portfolio managers will use some commercial software to perform the optimisations versus the benchmark". Common ones: Barra, Birr, Bitaplus, FactSet and Northfield.
  - "Many portfolio managers transform the aggregate Z-score into a multifactor α". Popular methods:
    1. One-to-one transformation that uses the actual Z-score as α's. This is helpful only if the portfolio manager has trimmed extreme Z-scores. "One therefore, could add the aggregate Z-score to the expected return of each stock predicated by a risk model. [...] One then performs an optimisation to form the portfolio".
    2. "use the expected excess return from Eq. (5.10) as α"
    3. "obtain all stocks' factor exposures and factor premiums and use them to calculate historical α".

$$r_{i,t} - \boldsymbol{\beta}_i' \boldsymbol{f_t} = \gamma_i + \delta z_{i,t-1} + \epsilon_{i,t} \tag{5.14}$$

    "Then the portfolio manager can use the estimates from this equation to forecast the α's of all the stocks for the next period ($t+1$), given their aggregate Z-scores for time t"
- "Stock screening can serve as the basis for a model of stock returns, but in QEPM it is usually just a preliminary step in creating the portfolio."

Chapter 6: Fundamental Factor Models

TABLE 6.1

Creating a Fundamental Factor Model

Preliminary work:

1. Choose the factors for the model.
2. Determine the treatment of the risk-free rate.
3. Define the investment universe.
4. Decide on the time interval and time period of the data.

Main steps for modeling total stock return (for nonbenchmark managers):

1. Stock returns are estimated with the following equation: $r_i = \alpha_i + \beta_{i1}f_1 + \cdots + \beta_{iK}f_K + \epsilon_i$.
2. Collect stock-return and factor-exposure data for the time period at each time interval.
3. Estimate the factor premium from a panel regression of the stock return on the factor exposure using either the OLS, MAD, or the GLS method.
4. Check the robustness of the factor premium estimation by splitting the data into subsets and comparing the estimates for each subset. If the estimates are similar across subsets, the estimation is robust.
5. If the estimation is not robust, try a different estimation method.
6. Calculate each stock's average return as the product of the factor premium and the vector containing the stock's factor exposures.
7. Decompose the risk of the stock return into its diversifiable and nondiversifiable components.
8. Calculate the correlation between the returns of the stocks in the investment universe.

Main steps for modeling residual stock return (for benchmark managers):

1. Stock returns are estimated with the following equation: $r_i = \tilde{\alpha}_i + \tilde{\beta}_i r_B + \tilde{\epsilon}_i$.
2. Collect stock return and benchmark return data for the given time period at the given time intervals.
3. After estimating $\tilde{\alpha}$ and $\tilde{\beta}$, calculate the residual return as $\tilde{r}_i \equiv \tilde{\alpha}_i + \tilde{\epsilon}_i$.
4. Estimate residual return with the following equation: $\tilde{r}_i = \alpha_i + \beta_{i1}f_1 + \cdots + \beta_{iK}f_K + \epsilon_i$.
5. Follow steps 2 through 8 of the main steps for modeling total stock return, substituting the residual return for the total stock return.

- In the fundamental factor model, factor exposure "it some observable fundamental characteristic of the stock. [...] The factor premium, on the other hand, is not known. It is the proportionality between the average stock return and the factor exposure, and it must be estimated empirically."
- "*The average stock return is the payoff for taking nondiversifiable risk*"

$$Nondiversifiable\ risk = factor\ exposure^2 \cdot factor\ risk \tag{6.9}$$

  where factor risk is the variance of "one unit of exposure".
- Total risk of a stock:

$$V(r_i) = V(\alpha_i + \beta_{i1}f_1 + \cdots + \beta_{iK}f_K) + V(\epsilon_i) \tag{6.10}$$

  "The diversifiable component, [...] is the term $V(\epsilon_i)$"
- **Preliminary work**
  - "The primary decision if the choice of factors"
  - "In practice, financial economists tend to prefer a monthly interval and, to a lesser degree, a weekly interval. Daily and annual intervals are rather uncommon"

- "Financial economists tend to include between 36 and 60 time intervals in their analyses when using a monthly interval [...]. When using a weekly interval, the number of time intervals tends to be greater than 60"

- **Benchmark and α**
  - "If the portfolio manager aims to outperform a benchmark while minimising the portfolio's tracking error, then the benchmark must play a role in the model. One approach to incorporating the benchmark into the fundamental factor model is to use the model to predict only the residual return rather than the entire stock return". Estimate:

$$r_i = \tilde{\alpha}_i + \tilde{\beta}_i r_B + \tilde{\epsilon}_i \tag{6.14}$$

  using a time-series regression. "After estimating $\tilde{\alpha}$ and $\tilde{\beta}$, we define the residual return $\tilde{r}_i$ as"

$$\tilde{r}_i \equiv \tilde{\alpha}_i + \tilde{\epsilon}_i \tag{6.15}$$

  "simply substitute the residual return for the stock return given in the fundamental factor model. [...] An alternative way to account for the benchmark, rather than predicting the residual stock return, is to add the benchmark to the model as one of the factors."

- **Factor Exposure**
  - "once the factors are chosen, determining the factor exposure is rarely a challenging task"
  - "After all the factor exposures are determined, we collect the numbers and save them as a set of vectors $\{(\beta_{11}, \ldots, \beta_{N1}), \ldots, (\beta_{1T}, \ldots, \beta_{NT})\}$ where β_{it} is the factor exposure of stock i for time t"

- **The Factor Premium**
  - "the factor premium is estimated from the pooled cross-sectional regression (i.e., panel regression) of the stock return on the factor exposure"
  - "While the OLS estimators is simple to obtain, it may not be the most reliable estimator. We suggest that portfolio managers do a number of robustness checks on it"

  - **OLS estimator of the factor premium**

    - "Given the returns of N stocks over T time periods, $\{(r_{11}, \ldots, r_{N1}), \ldots, (r_{1T}, \ldots, r_{NT})\}$, and the factor exposures of N stocks over T time periods, $\{(\beta_{11}, \ldots, \beta_{N1}), \ldots, (\beta_{1T}, \ldots, \beta_{NT})\}$, we can estimate the following:"

$$r_{it} = \beta'_{it} f + \epsilon_{it} \tag{6.17}$$

    - "OLS estimator of the factor premium f is given as:"

$$\hat{f} = \sum_{t=1}^{T} \sum_{i=1}^{N} (\beta_{it} - \bar{\beta})(r_{it} - \bar{r}) \left[\sum_{t=1}^{T} \sum_{i=1}^{N} (\beta_{it} - \bar{\beta})(\beta_{it} - \bar{\beta})' \right]^{-1} \tag{6.18}$$

    where

$$\bar{\beta} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \beta_{it} \tag{6.19}$$

$$\bar{r} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} r_{it} \tag{6.20}$$

    - "The standard error for the factor premium is the square root of (the diagonal elements of) the following variance:"

$$\hat{V}(\hat{f}) = \hat{\sigma}^2 \left(\sum_{t=1}^{T} \sum_{i=1}^{N} (\beta_{it} - \bar{\beta})(\beta_{it} - \bar{\beta})' \right)^{-1} \tag{6.21}$$

    where $\hat{\sigma}^2$ is the estimated variance of ϵ_{it}:

$$\hat{\sigma}^2 = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(r_{it} - \beta'_{it} \hat{f} \right)^2 \tag{6.22}$$

  - "Table 6.4 present the factor premium estimates and their standard errors. We made the estimations for Standard & Poor's (S&P) 500 stocks in the period from January 2000 to December 2003"

TABLE 6.4

The Factor Premium

P/E	P/B	D/E	Size	M12M
−0.0006	−0.0334	0.0341	−0.0129	−0.0814
(0.0005)	(0.0075)	(0.0125)	(0.0021)	(0.0229)

Note: Estimated for the period from January 2000 to December 2003. Standard errors are inside parentheses. Factors are as in Table 6.2.

- "To check the robustness of the estimation, split the data set into a few subsets and see whether the estimates are very different across the subsets"
  - Subsets can be created along the time dimension or along the cross-sectional dimension.
- "The weakness of the OLS method is that in trying to minimise the sum of the squared residuals, it is highly sensitive to outliers."
  - Minimum Absolute Deviation (a.k.a. *median estimation*): "minimises the sum of the absolute value of residuals rather than the squared residuals."
- "The OLS is the best estimation method only if the variance of ϵ_{it} is constant across different stocks"
  - "In econometrics, the problem of non-constant variance is called *heteroscedasticity*. When the variance of ϵ_{it} is not constant across stocks, the best estimation procedure is the generalised least square (GLS) method. The <u>GLS estimator</u> of f:"

$$\hat{\hat{f}} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} \left(\frac{\beta_{it}}{\hat{\sigma}_i} - \tilde{\beta} \right) \left(\frac{r_{it}}{\hat{\sigma}_i} - \tilde{r} \right)}{\sum_{t=1}^{T} \sum_{i=1}^{N} \left(\frac{r_{it}}{\hat{\sigma}_i} - \tilde{r} \right)^2} \tag{6.23}$$

where

$$\tilde{\beta} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{\beta_{it}}{\hat{\sigma}_i} \tag{6.24}$$

$$\tilde{r} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{r_{it}}{\hat{\sigma}_i} \tag{6.25}$$

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^{T} \left(r_{it} - \beta_{it}' \hat{f} \right)^2 \tag{6.26}$$

- **Decomposition of Risk**
  - "In the fundamental factor model, the risk of the stock return has two components: non- diversifiable risk captured by $\beta_i' \hat{f}$ and diversifiable risk captured by ϵ_i. The total risk of the stock return is simply the sum of these two risks."
  - Variance of the factor premium:

$$\hat{V}(\hat{f}) = \left(\sum_{t=1}^{T} \sum_{i=1}^{N} \beta_{it} \beta_{it}' \right)^{-1} \left(\sum_{t=1}^{T} \sum_{i=1}^{N} \hat{\sigma}_i^2 \beta_{it} \beta_{it}' \right) \left(\sum_{t=1}^{T} \sum_{i=1}^{N} \beta_{it} \beta_{it}' \right)^{-1} \tag{6.28}$$

where $\hat{\sigma}_i^2$ is an estimate (Eq. 6.26).

  - Non-diversifiable risk of stock i is a quadratic equation of the factor exposure

$$\beta_{iT}' \hat{V}(\hat{f}) \beta_{iT} \tag{6.29}$$

"Note that we used the last value of the factor exposure to obtain the non-diversifiable risk of stock i"

  - Diversifiable risk: "part of the variation in the stock return that the variation in the model's factors cannot explain. This part of the stock's variation appears as the error term ϵ_i, that is, σ_i^2". Eq (6.26) shows the estimate for this variation.
  - Total risk of stock i:

$$\beta_{iT}' \hat{V}(\hat{f}) \beta_{iT} + \hat{\sigma}_i^2 \tag{6.31}$$

- To build optimal portfolios, "we also need to know the correlations between stock returns. [...] The correlation between the returns on a pari of stock consists of the correlation between their non-diversifiable components and the correlation between their diversifiable components. [...] the covariance between r_{it} and r_{jt}:"

$$C(r_{it}, r_{jt}) = \beta'_{it} V(f) \beta_{jt} + C(\epsilon_{it}, \epsilon_{jt}) \tag{6.32}$$

"In practice, we estimate only the correlation between the non-diversifiable components of two stocks' returns. [...] the convention is to assume that diversifiable risk $C(\epsilon_{it}, \epsilon_{jt})$ equals zero"

Chapter 7: Economic Factor Models

TABLE 7.1

Creating an Economic Factor Model

Preliminary work:

1. Choose the factors for the model.
2. Determine the treatment of the risk-free rate.
3. Define the investment universe.
4. Decide on the time interval and time period of the data.

Main steps for modeling total stock return (for nonbenchmark managers):

1. Stock returns are estimated with the following equation: $r_i = \alpha_i + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + \epsilon_i$.
2. Collect stock-return and factor-premium data for the time period at each time interval. Factor-premium data are readily available for economic/behavioral/market factors. To calculate factor premiums for fundamental/technical/analyst factors, use the zero-investment portfolio method. To calculate factor premiums for statistical factors, use principal-component analysis.
3. If there are sufficient factor-premium data, estimate the factor exposure from a time-series regression of the stock return on the factor premium.
4. If there are insufficient factor-premium data to perform a time-series regression, use characteristic matching to identify similar firms that can be used as proxies for the firms lacking data. Collect stock-return and factor-premium data for the proxies. Using the proxies' data, estimate the factor exposure from a time-series regression of the stock return on the factor premium.
5. Calculate each stock's average return as the product of the factor exposure and the vector containing the stock's factor premiums.
6. Decompose the risk of the stock return into its diversifiable and nondiversifiable components.
7. Calculate the correlation between the returns of the stocks in the investment universe.

Main steps for modeling residual stock return (for benchmark managers):

1. Stock returns are estimated with the following equation: $r_i = \tilde{\alpha}_i + \tilde{\beta}_i r_B + \tilde{\epsilon}_i$.
2. Collect stock-return and benchmark-return data for the given time period at the given time intervals.
3. After estimating $\tilde{\alpha}$ and $\tilde{\beta}$, calculate the residual return as $\tilde{r}_i \equiv \tilde{\alpha}_i + \tilde{\epsilon}_i$.
4. Estimate residual return with the following equation: $\tilde{r}_i = \alpha_i + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + \epsilon_i$.
5. Follow steps 2 through 7 of the main steps for modeling total stock return, substituting the residual return for the total stock return.

- "In the economic factor model, the factor premium is the known value (or, at least, the value that can be calculated from given data), whereas the factor exposure must be estimated by a regression of stock returns on factor premiums"
- "The economic factor model does not assume that the amount of reward demanded by investors is the actual inflation rate. What it assumes is that the the amount of reward is a linear function of the actual inflation rate. [...] This assumption removes the necessity to distinguish between the true factor premium and the observed inflation rate."
- "the economic factor model defines the return to stock i:"

$$r_i = \alpha_i + \beta_{i1}f_1 + \cdots + \beta_{iK}f_K + \epsilon_i \tag{7.1}$$

  where $f_1, \ldots, f_K$ are factor premiums (which do not vary across stocks), and $\beta_{i1}, \ldots, \beta_{iK}$ are factors exposures (which do vary across stocks).
- "As with the fundamental factors, we again define $(K+1)$-dimensional column vector $\boldsymbol{f}$ and $\boldsymbol{\beta_i}$ as follows:"

$$\boldsymbol{f} = (1, f_1, \ldots, f_K)' \tag{7.2}$$

$$\boldsymbol{\beta_i} = (\alpha_i, \beta_{i1}, \ldots, \beta_{iK})' \tag{7.3}$$

  Using this notation,

$$r_i = \boldsymbol{\beta_i'f} + \boldsymbol{\epsilon}_i \tag{7.4}$$

- "We use lowercase letter t to indicate one time interval and capital letter T to indicate the number of time intervals. Thus the return to stock i at time t:"

$$r_{it} = \boldsymbol{\beta_i'f_t} + \epsilon_{it}, \quad t = 1, \ldots, T \tag{7.7}$$

- "the academic literature fails to distinguish between economic and fundamental factor models, using instead the term *multifactor pricing model* to refer generally to the framework of the economic factor model"
- Three categories of factors:
  1. *Economic, behavioural, market factors*: GDP, inflation, unemployment, interest rates; consumer sentiment index, other survey-based indexes; return on broad market, returns on other market group / industry indexes.
  2. *Fundamental/technical/analyst factors*: Size, P/E ratio; momentum, trading volume; analyst rating changes, earnings revisions. "Some studies suggest that thoroughly vetted fundamental/technical/analyst factors predict stock returns more effectively than other factor groups do".
  3. *Statistical factors*: Factors obtained from principal-component analysis applied to historical returns.
- **Benchmark and α**
  - "As a measure of portfolio performance, it can be useful to create a model that reflects, or adjusts for, the effects of the benchmark".
    - One way is to only use the residual return of stocks to construct the model by first regressing the stock returns on the benchmark returns and calculate the residual, "the part of the stock returns that is not correlated with the benchmark return"
    - "Alternatively, we can include the benchmark in the economic factor model explicitly".
  - If we add benchmark as the Kth factor in the economic factor model:

$$r_i = \alpha_i + \beta_{i,1}f_i + \cdots + \beta_{i,K-1}f_{K-1} + \beta_{i,K}r_B + \epsilon_i \tag{7.8}$$

  benchmark alpha :

$$\alpha^B = \alpha_i + \beta_{i,1}f_1 + \cdots + \beta_{i,K-1}f_{K-1} \tag{7.9}$$

- **The Factor Premium**
  - "one cannot always observe a factor premium directly, though. For economic/behavioural/market factors, the computation is rather trivial. For fundamental/technical/analyst factors, though computation is somewhat more demanding, and for statistical factors, it poses quite a bit of a challenge."
  - **Economic/behavioural/market factors**:
    - "does not involve any computation. All we need to do is simple 'copy and paste' the values". People sometimes use deviation from mean to reflect the idea of 'surprise'.
    - examples: "For consumer sentiment, we take the consumer sentiment index compiled by the University of Michigan, which shows the level of consumer sentiment for each month. The factor premium is simply the growth rate of this index."
    - "After we have all the factor premiums for a certain time interval, we may store them in a set of vectors $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T\}$, where $\boldsymbol{f}_t$ is the factor premium for time t"

- **Fundamental/technical/analyst factors**:
    - "involves constructing zero-investment portfolios and calculating their returns"
    - "In principle, we want to create a zero-investment portfolio for each time interval, but in practice, the data may not permit us to do so. It would not make sense to construct a zero-investment portfolio at intervals shorter than a quarter because the P/B ratio would not contain updated book values. Any fluctuations in the ratio would only reflect changes in the stock price"
    - **Steps**:
        1. Rank all stocks in terms of the factor.
        2. Create high-exposure (low-exposure) portfolio by equally weighting top (bottom) 33% of the list.
        3. Calculate return difference between high-exposure and low-exposure portfolios; this is the factor premium.
- **Statistical factors**:
    - "involves a rather intensive computation. [...] known as *principal-component analysis*"
    - Principal component analysis:
        1. Estimate variance-covariance matrix of stock returns. "N stock returns at time t as an N-dimensional column vector; that is, $r_t = (r_{1t}, \ldots, r_{Nt})'$. We have total of T such vectors $\{r_1, \ldots, r_T\}$." Variance-covariance matrix of returns Σ is estimated:

        $$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} r_t r_t' - \bar{r}\bar{r}' \tag{7.11}$$

        where $\bar{r}$ is the average return vector:

        $$\bar{r} = \frac{1}{T} \sum_{t=1}^{T} r_t \tag{7.12}$$

        2. "diagonalise" the variance-covariance matrix by finding an orthogonal matrix Q (that is, $Q^{-1} = Q'$) such that

        $$Q'\hat{\Sigma}Q = D \tag{7.13}$$

        "where D is a diagonal matrix whose diagonal elements are eigenvalues (i.e., characteristic values) of $\hat{\Sigma}$. [...] each column of Q is an orthonormal eigenvector corresponding to eigenvalues of $\hat{\Sigma}$"
        "let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of $\hat{\Sigma}$ such that $\lambda_1 \geq \ldots \geq \lambda_N \geq 0$"

        $$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix} \tag{7.14}$$

        3. "If we want to find K factors, then we obtain K factor premiums by weighting individual stock returns using the first K columns of Q". K-th factor premium: $f_{K,t} = q_K' r_t$
- **Factor Exposure**
    - "factor exposure typically are determined from the time-series regression of stock returns on factor premiums". a.k.a *factor sensitivities* or *factor loadings*
    - returns of stock i and the factor premium for T periods of time, $\{r_{i1}, \ldots, r_{iT}\}$ and $\{f_1, \ldots, f_T\}$, we can estimate:

    $$r_{it} = \beta_i' f_t + \epsilon_{it}, \qquad t = 1, \ldots, T \tag{7.17}$$

    - The ordinary least squares (OLS) estimator of β_i is given by:

    $$\hat{\beta}_i = \left(\sum_{t=1}^{T} f_t f_t' \right)^{-1} \left(\sum_{t=1}^{T} f_t r_{it} \right) \tag{7.18}$$

    - "The standard error of the factor exposure is the square root of (the diagonal elements of) the following variance:

    $$\hat{V}(\hat{\beta}_i) = \hat{\sigma}_i^2 \left(\sum_{t=1}^{T} f_t f_t' \right)^{-1} \tag{7.19}$$

    where $\hat{\sigma}_i^2$ is the estimated variance of ϵ_{it}:

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^{T} \left(r_{it} - \hat{\boldsymbol{\beta}}_i' \boldsymbol{f}_t \right)^2 \tag{7.20}$$

- "For a stock with insufficient data, we may infer the factor exposures by weighting groups of similar stocks and using the weighted factor exposures as proxies for the original stock's exposures"
- **Decomposition of Risk**
  - total risk of stock i:

$$V(r_i) = \boldsymbol{\beta}_i' V(\boldsymbol{f}) \boldsymbol{\beta}_i + V(\boldsymbol{\epsilon}_i) \tag{7.24}$$

  - "we already have the estimate for $\boldsymbol{\beta}_i$. Given the factor-premium data $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T\}$, finding the estimate for $V(\boldsymbol{f}_t)$":

$$\hat{V}(\boldsymbol{f}_t) = \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{f}_t - \bar{\boldsymbol{f}})(\boldsymbol{f}_t - \bar{\boldsymbol{f}})' \tag{7.25}$$

  - "The estimate for $V(\boldsymbol{\epsilon}_i)$ is obtained naturally from estimation of the factor exposure if the factor exposure was estimated by" Eq. (7.18). Look at Eq. (7.20) for $\hat{V}(\boldsymbol{\epsilon}_i)$.
  - "the correlation between the nondiversifiable components and the correlation between the diversifiable components. The covariance between two stock returns r_i and r_j is"

$$C(r_i ,\ r_j) = \boldsymbol{\beta}_i' V(\boldsymbol{f}) \boldsymbol{\beta}_j + C(\boldsymbol{\epsilon}_i ,\ \boldsymbol{\epsilon}_j) \tag{7.27}$$

Estimate for $C(\boldsymbol{\epsilon}_i ,\ \boldsymbol{\epsilon}_j)$ from the estimation of the factor exposure:

$$\hat{C}(\boldsymbol{\epsilon}_i ,\ \boldsymbol{\epsilon}_j) = \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{r}_{it} - \hat{\boldsymbol{\beta}}_i' \boldsymbol{f}_t)(\boldsymbol{r}_{jt} - \hat{\boldsymbol{\beta}}_j' \boldsymbol{f}_t) \tag{7.28}$$

"In practice, however, unless T is quite large, it is conventional to assume that $C(r_i ,\ r_j)$ is zero. [...] If the model is good, it does not create too much distortion to assume that the covariance is zero"

Chapter 8: Forecasting Factor Premiums and Exposures

- "the return on stock i at time $T+1$, $r_{i,T+1}$, is

$$r_{i,T+1} = \boldsymbol{\beta}_{i,T+1}' \boldsymbol{f}_{T+1} + \boldsymbol{\epsilon}_{i,T+1} \tag{8.1}$$

where $\boldsymbol{\beta}_{i,T+1}$ is the $(K+1)$-dimensional vector of the factor exposure for stock i at time $T+1$, $\boldsymbol{f}_{T+1}$ is the $(K+1)$-dimensional vector of the factor premium at time $T+1$, and $\boldsymbol{\epsilon}_{i,T+1}$ is the deviation of the stock i's return from its average"
- Fundamental factor model: "only the factor exposure changes over time, and given stock returns and factor exposures up to time T, we estimated a fixed factor premium through time T, namely, $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T\}$"
- Economic factor model: "only the factor premium changes over time, and given stock returns factor premiums up to time T, we estimated a fixed factor exposure through T, namely, $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$"
- When is forecasting necessary?
  - "if we are using the fundamental factor model from Chapter 6, we do not need to forecast anything. [...] The explanatory variable (factor exposure β_{it}) is measured at the beginning of time t, whereas the dependent variable (stock return r_{it}) is measured during time t"
  - "Forecasting becomes a necessary step when we are using the economic factor model. [...] the factor premium in the economic factor model"
- `skipped 4 sections on factor premia forecasting`
- Forecasting the stock return
  - From the relationship

$$r_{i,T+1} = \boldsymbol{\beta}'_{i,T+1} \boldsymbol{f}_{i,T+1} + \boldsymbol{\epsilon}_{i,T+1} \tag{8.9}$$

the mean and the variance of the stock return are

$$E(r_{i,T+1} = \boldsymbol{\beta}'_{i,T+1} E(\boldsymbol{f}_{i,T+1}) \tag{8.10}$$

$$V(r_{i,T+1}) = \boldsymbol{\beta}'_{i,T+1} V(\boldsymbol{f}_{i,T+1}) \boldsymbol{\beta}_{i,T+1} + V(\boldsymbol{\epsilon}_{i,T+1}) \tag{8.11}$$

Chapter 9: Portfolio Weights

- **Ad Hoc Methods (Non-Benchmark)**
  - "Equal weighting makes sense only when the portfolio manager has very poor information about the expected return and the risk of stocks selected."
  - "*Value weighting* is assigning weights proportional to the stocks' market capitalisations."
- **Standard Mean-Variance Optimisation (Non-Benchmark)**
  - "quadratic programming technique to find the portfolio that the *minimum* risk"
  - "The portfolio manager must do his or her best to account for estimation error."
  - Common constraints:
    - restricting short-sale;
    - restricting any particular stock from having more than a certain threshold weighting;
    - restricting the composite weight of any group of stocks in a certain sector from being beyond a level.
  - No constraints:
    - using expected stock returns and risk, "include all the relevant information about individual stock returns in a vector μ and a matrix Σ."

$$\boldsymbol{\mu} = \begin{pmatrix} E(r_1) \\ \vdots \\ E(r_N) \end{pmatrix} \tag{9.3}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} Var(r_1) & Cov(r_1, r_2) & \cdots & Cov(r_1, r_N) \\ Cov(r_2, r_1) & Var(r_2) & \cdots & Cov(r_2, r_N) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(r_N, r_1) & Cov(r_N, r_2) & \cdots & Var(r_N) \end{pmatrix} \tag{9.4}$$

    - "portfolio is specified by a weight vector $\boldsymbol{w}$"

$$\boldsymbol{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix} \tag{9.5}$$

    such that sum of all weights should be one, i.e., $\boldsymbol{w}'\boldsymbol{l} = 1$ where $\boldsymbol{l} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

    - The expected return of the portfolio is $\boldsymbol{w}'\mu$, whereas the risk of the portfolio is $\boldsymbol{w}'\Sigma\boldsymbol{w}$. The portfolio that has the minimum risk with expected return μ_P:

$$\min_{w} \boldsymbol{w}'\Sigma\boldsymbol{w} \; s.t. \quad \boldsymbol{w}'\boldsymbol{\mu} = \mu_P \quad and \quad \boldsymbol{w}'\boldsymbol{l} = 1 \tag{9.7}$$

    "mathematicians would prefer to rewrite the constraints in the following way:"

$$\boldsymbol{Aw} = \boldsymbol{b} \tag{9.8}$$

    where

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \mu_1 & \mu_2 & \cdots & \mu_N \end{pmatrix} \tag{9.9}$$

$$\boldsymbol{b} = \begin{pmatrix} 1 \\ \mu_P \end{pmatrix} \tag{9.10}$$

    "this particular form of the quadratic minimisation problem with equality constraints has a closed-form solution"

$$\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}')^{-1}\boldsymbol{b} \tag{9.11}$$

  - Shorting and diversification constraints
    - Shorting contraint: $\boldsymbol{w} \geq 0$
    - Diversification contraint: $\underline{\boldsymbol{w}} \leq \boldsymbol{w} \leq \overline{\boldsymbol{w}}$; where $\underline{\boldsymbol{w}}$ and $\overline{\boldsymbol{w}}$ are N-dimensional vectors of maximum and minimum allowed weights.
  - Sector/Industry constraints
    - $\underline{w}_j \leq w_j \leq \overline{w}_j$ where w_j represents the weight of sector j in the portfolio.
  - Trading volume constraint

- $w \leq cx$ where x is a vector of average daily trading volume in dollar terms, and c is a constant indicating the threshold.
  - Mean variance optimisation for expected return maximisation:

$$\max_{w} w'u \quad s.t. \quad w'\Sigma w = \sigma_P^2 \tag{9.16}$$

"more useful if the portfolio manager has a specific target risk level σ_P^2 whereas the risk minimisation may be more useful if the portfolio manager has a target expected return" `#idea_factor_modelling`
  - "the mean-variance optimisation can be expressed in terms of *risk-adjusted* expected return. We can adjust the expected return for the risk by subtracting some multiple of the risk, that is, $\mu_P - A\sigma_P^2$. The multiplier A is called the *risk-aversion parameter* [...]. If the value of A is 2, it means that the portfolio manager equates a 1% increase in the variance with a 2% decrease in the expected return"

$$\max_{w} w'\mu - Aw'\Sigma w \tag{9.17}$$

"This formulation has a theoretical appeal as well"
- "Minimising tracking errors when the benchmark is inefficient may lead to inefficient portfolios in a variety of cases"
- **Ad Hoc Methods (Benchmark)**
  - Select the stocks from the benchmark and "tilt weights slightly toward his or her preferred stocks"
  - Assuming the portfolio manager ranked stocks according to a normalised measure such that the sum of ranks is equal to 0; let's say he/she uses Z-score.
    - $\eta \to$ maximum percentage deviation (away from the benchmark) in weight allowed.
    - $z^{max} \to$ maximum of absolute value of all the Z-scores among all the stocks.
    - $m = \eta/z^{max} \to Z$-score multiplier
    - Tilted weights w_i^* are calculated by $w_i^* = w_i + mz_i$
- **Stratification**
  - proportional random sampling.
  - Steps:
    - assuming the portfolio manager already has predicted the excess returns of all the stocks in his or her universe.
    - "stratify the universe of stocks by dividing it into J non-overlapping groups"
    - "select representative stocks from each stratum. [...] select a proportion of stocks from each stratum that is representative of the universe of stocks. [...] before making selections of stocks from each stratum, rank the stocks in each stratum according to the aggregate Z-score, expected returns, excess expected return, or α and then choose the stocks with the highest values of the chosen criteria"
  - "Stratification is an easy way to choose one's preferred stocks while also controlling risk via broad diversification."
- **Factor-Exposure Targeting**
  - "we may prefer to stipulate a range for each of the portfolio's factor exposures. This is sometimes referred to as *factor tilting* because we are tilting the portfolio toward greater exposure to certain factors and less exposure to others, according to our view on the factors"
  - "The factor exposure of a portfolio is the weighted average of the factor exposures of individual stocks. Let B be an $N \times K$ matrix of factor exposures"

$$B = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1K} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2K} \\ \vdots & & & \\ \beta_{N1} & \beta_{N2} & \cdots & \beta_{NK} \end{pmatrix} \tag{9.21}$$

where K is the number of relevant factors, and β_{ik} is the exposures of stock i to factor k.
  - Factor exposure of a portfolio with weight w is simply $B'w$; factor exposure constraint:

$$\underline{\beta} \leq B'w \leq \bar{\beta} \tag{9.22}$$

"The portfolio manager can exercise his or her particular management style through this sort of constraint. Assigning a minimum exposure (say, 0.9) to the growth factor orients the portfolio toward growth investments. If the first factor is the growth factor, then the manager would set the first element of $\underline{\beta}$ to 0.9"
- **Tracking-error Minimisation**
  - Minimise the tracking error for a given expected excess return over the benchmark

- *Tracking error* (TE) defined by most portfolio managers as standard deviation of portfolio returns minus benchmark returns:

$$TE = S(r_P - r_B) = \sqrt{V(r_P - r_B)} \qquad (9.23)$$

- Components of the variance term:

$$V(r_P - r_B) = V(r_P) - 2C(r_P, r_B) + V(r_B) \qquad (9.24)$$

$V(r_B)$ is out of our control, so we minimise $V(r_P) - 2C(r_P, r_B)$.

- Covariance between portfolio and benchmark can be computed from covariances between individual stock returns and the benchmark return. $\boldsymbol{\gamma}$ is an N-dimensional vector of covariances b/w individual stock and benchmark

$$\boldsymbol{\gamma} = \begin{pmatrix} C(r_1, r_B) \\ \vdots \\ C(r_N, r_B) \end{pmatrix} \qquad (9.25)$$

- Covariance between portfolio and benchmark equals $\boldsymbol{w}'\boldsymbol{\gamma}$, where $\boldsymbol{w}$ is the portfolio weights. To find portfolio that minimises tracking error, we solve:

$$\min_{w} \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - 2\boldsymbol{w}'\boldsymbol{\gamma} \quad \text{s.t.} \quad \boldsymbol{w}'\boldsymbol{\mu} = \mu_P \qquad (9.26)$$

"Typically, the chosen portfolio mean μ_P will be some excess over the benchmark"

- Tracking by Factor exposure:
  - variance of stock i:

$$V(r_i) = \boldsymbol{\beta_i}'V(\boldsymbol{f})\boldsymbol{\beta}_i + V(\epsilon_i) \qquad (9.27)$$

var-covar matrix of all stock returns: $\Sigma = \boldsymbol{B}V(\boldsymbol{f})\boldsymbol{B}' + V(\epsilon)$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \beta_{1,1} & \cdots & \beta_{1,K} \\ \vdots & \vdots & \vdots \\ \beta_{N,1} & \cdots & \beta_{N,K} \end{pmatrix} \begin{pmatrix} V(f_1) & \cdots & C(f_1, f_K) \\ \vdots & \vdots & \vdots \\ C(f_K, f_1) & \cdots & V(f_K) \end{pmatrix} \begin{pmatrix} \beta_{1,1} & \cdots & \beta_{1,K} \\ \vdots & \vdots & \vdots \\ \beta_{N,1} & \cdots & \beta_{N,K} \end{pmatrix} + \begin{pmatrix} V(\epsilon_1) & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & V(\epsilon_N) \end{pmatrix} \text{ (9}$$

- <u>Maximise the expected excess return over the benchmark without exceeding a maximum tracking-error constraint</u>
  - "The objective function will be the expected return of the portfolio rather than the tracking error, and the constraint will include the tracking error"

$$\max_{w} \boldsymbol{w}'\boldsymbol{\mu} \quad \text{s.t.} \quad V(r_P - r_B) = \sigma_x^2 \qquad (9.34)$$

  - If there is neither a target tracking error nor a target mean, "we can adjust the expected return for the tracking risk by subtracting some multiple of the squared tracking error $\mu_P - A\sigma_x^2$. The multiplier A is called the *tracking-error aversion parameter*"

$$\max_{w} \boldsymbol{w}'\boldsymbol{\mu} - AV(r_P - r_B) \qquad (9.35)$$

Chapter 10: Rebalancing and Transaction Costs

- "Changes in the underlying parameters of the stock-return model do not automatically require adjustments in the portfolio. The portfolio manager needs to decide whether a parameter change is big enough to warrant incurring transactions costs"
- **Rebalancing and Model Periodicity**
  - "'update' the estimation of the models at least once a month if the model uses monthly returns. Updating the estimation does not necessarily lead to rebalancing. The portfolio manager simply should make sure that his or her information is up to date"
- **Understanding Transactions Costs**
  - Broker commission, *bid-ask spread*, and *price impact*
  - "impossible to anticipate the exact cost of the transaction in advance of its execution"
- **Modelling Transactions Costs**

- "it is useful to write down the transactions cost (TC) as a linear function"

$$TC = V_t \sum_{i=1}^{N} c_i(w_i^a - w_i^b) \tag{10.4}$$

  where

$$c_i = \begin{cases} c & \text{if } w_i^a > w_i^b \\ -c & \text{if } w_i^a < w_i^b \end{cases} \tag{10.5}$$

  where "$V_t w_i^b$ is the current holding of stock i in dollar terms, and $V_t w_i^a$ is the prospective holding of a stock in dollar terms".

- $\boldsymbol{w}^b$ is a vector of current weights and $\boldsymbol{w}^a$ a vector of prospective weights. The vector of transactions cost as $\boldsymbol{c}$. Then:

$$TC = V_t(\boldsymbol{w}^a - \boldsymbol{w}^b)'\boldsymbol{c} \tag{10.6}$$

-