

---

# Portfolio Construction and Analytics 读书笔记

目录 .....	2
Contents .....	1
Contents .....	1
1 资产管理的介绍 .....	1
2 随机变量、概率分布和重要的统计概念 .....	1
3 常见的分布函数介绍 .....	1
4 统计学模型 .....	1
5 模型模拟 .....	1
6 模型优化 .....	1
7 非确定优化 .....	1
8 资产多样化 .....	1
9 因子模型 .....	1
9.1 金融经济学文献中的因子模型 .....	1
9.2 因子模型的均值方差优化 .....	2
9.3 实践中的因子选择 .....	4
9.4 阿尔法构建中的因子模型 .....	5
9.5 因子模型用于风险评估 .....	5
9.5.1 宏观经济因子模型 .....	5
9.5.2 基本面因子模型 .....	5
9.5.3 统计因子模型 .....	6
9.5.4 混合因子模型 .....	6
9.5.5 选择“适合的”因子模型 .....	7
9.6 数据管理与质量问题 .....	7
9.6.1 数据调整 .....	7
9.6.2 幸存者偏差 .....	8
9.6.3 前视偏差 .....	8
9.6.4 数据迁就 .....	8

9.7 风险分解、风险归因和业绩归因 .....	9
9.8 因子投资 .....	9

- 1 资产管理的介绍
- 2 随机变量、概率分布和重要的统计概念
- 3 常见的分布函数介绍
- 4 统计学模型
- 5 模型模拟
- 6 模型优化
- 7 非确定优化
- 8 资产多样化
- 9 因子模型

## 9.1 金融经济学文献中的因子模型

线性因子模型代表了应用于金融中的绝大多数因子模型：

$$r_i = \alpha_i + f_1\beta_{i1} + \cdots + f_p\beta_{ip} + \varepsilon_i \quad i = 1, \dots, n,$$

$r_i$ :第*i*份资产收益率

$f_k$ : 影响因子

$\beta_{ik}$ :回归系数（灵敏度）

$\alpha$ :常数

$\varepsilon_i$ 随机扰动

最简单的线性因子模型—单一指数模型,是Markowitz(1959)为简化均值一方差投资组合分析的目标而提出的。虽然马科维茨在其著作的脚注中提到了这个方法,但Shape(1963)进一步研究了这一方法。这个模型包含一个称为指数的单一因子。理论上,该指数是经济中所有资产的价值加权投资组合,其中“价值加权”意味着投资组

合中每种资产回报的权重代表了该资产的市值(发行股数乘以每股价格)在经济中所有资产的总市值的份额。在实践中，几个宽基指数(如标普500指数)都得到使用。假定因子是市场指数(类如标普500指数)回报的模型称为单一指数市场模型。

在单一指数市场模型中，相对于市场变化的证券的敏感度(证券的贝塔值)是利用简单线性回归从一组观察到的证券回报和市场回报中估计得到的：

$$r_i = \alpha_i + \beta_i r_M + \varepsilon_i$$

$r_i$ :第*i*份资产收益率

$r_M$ : 市场组合的回报

$\beta_{ik}$ :回归系数（灵敏度）

$\alpha_i$ :证券*i*特有的回报

$\varepsilon_i$ 随机扰动

单一指数模型有三个假设：

- (1) 模型中的所有随机变量都有有限的均值和方差
- (2) 市场指数的随机回报与证券的随机误差项之间的协方差为零。
- (3) 任何两个证券*i*, *j*回报的特有部分的协方差是零

我们可以得到,对 $\forall i \neq j$ :

$$Cov(r_i, r_j) = Cov(\beta_i r_M, \beta_j r_M) = \beta_i \beta_j \sigma_M^2$$

上面的式子表明如果我们的假设是正确的，那么投资组合的风险仅仅由贝塔系数决定！

本章的后续讨论了Capm模型，但是笔者认为假设条件过于强，所以没有特定写出。

## 9.2 因子模型的均值方差优化

考虑一个具有*N*个资产的投资组合，其期望回报向量 $\mu = (\mu_1, \dots, \mu_N)$ ，回报协方差矩阵为 $\Sigma$ 。如果资产的权重是 $W = (w_1, w_2, \dots, w_N)$ 那么回忆第8章，投资组合期望回报和方差可以分别写成 $\mu^T w$ 和 $w^T \Sigma w$ 。

为了在不使用因子模型的情况下，计算出必要的输入信息以估计一个具有*N*个证券的投资组合的风险(定义为投资组合方差)，那么人们需要发现*N*个证券的协方差矩阵。这将需要估计*N*个方差(*N*个证券中的每一个证券都有一个方差)和*N*(*N* - 1)/2个协方差。人们还需要估计*N*个期望回报。总计需要2*N* + *N*(*N* - 1)/2 个估计。

假设投资组合经理正在考虑构建一个投资组合，并且投资组合有1000个候选证券（即 $N = 1\,000$ ）。为了计算投资组合，投资组合经理需要估计1000个方差和499 500 协方差。加上1000个期望回报的估计，总共需要对501500个值进行估计，这是一个非常艰难的任务。除了准备用于估计的数据是困难的之外，估计也将是不稳定的。在时间上，一个证券对特定因子的敏感性可能比该证券对投资组合中其他证券的敏感性更加稳定。如果使用基于证券之间直接关系的协方差估计，那么来自投资组合优化流程的输出值将是不可靠且不稳定的。输入参数的一个微小变化可能导致最优投资组合配置的大幅波动。

现在假设投资组合经理希望从 $N$ 个证券中构建一个投资组合，并且 $N$ 个证券回报的向量可以通过因子模型表示：

$$r = \alpha^T w + Bf + \varepsilon$$

$r$ : 预期收益率

$f$ :  $K$ 维因子向量

$B$ :  $N \times K$ 维因子载荷矩阵

$\varepsilon$ :  $N$ 维残差

投资组合收益率的期望：

$$E(r) = \alpha^T w + BE(f)$$

投资组的方差：

$$w^T (B \Sigma_f B^T + \Sigma_\varepsilon) w$$

其中 $\Sigma_f$ 是因子的协方差矩阵。 $\Sigma_\varepsilon$ 是对角残差矩阵。

假设投资组合使用单一指数模型来估计输入信息。那么，一个含有 $N$ 个证券的投资组合将需要估计以下参数以便进行均值一方差优化：

$N$ 个阿尔法来计算期望回报。

$N$ 个贝塔来计算期望方差和协方差。

1个指数方差 $\sigma_M^2$ 的估计来计算方差和协方差。

$N$ 个残差的方差

总共是 $3N+1$ 个估计。对于一个具有1000个候选证券的投资组合，这转换成3001个估计，比不使用因子模型的501500个估计更加易于操作。假设投资组合经理使用具有 $K$ 因子的模型，逻辑是类似的，只不过需要估计：

N个矩阵中每一个证券的K个因子载荷

$K(K-1)/2$ 个因子协方差

K个因子方差

N个残差的方差

### 9.3 实践中的因子选择

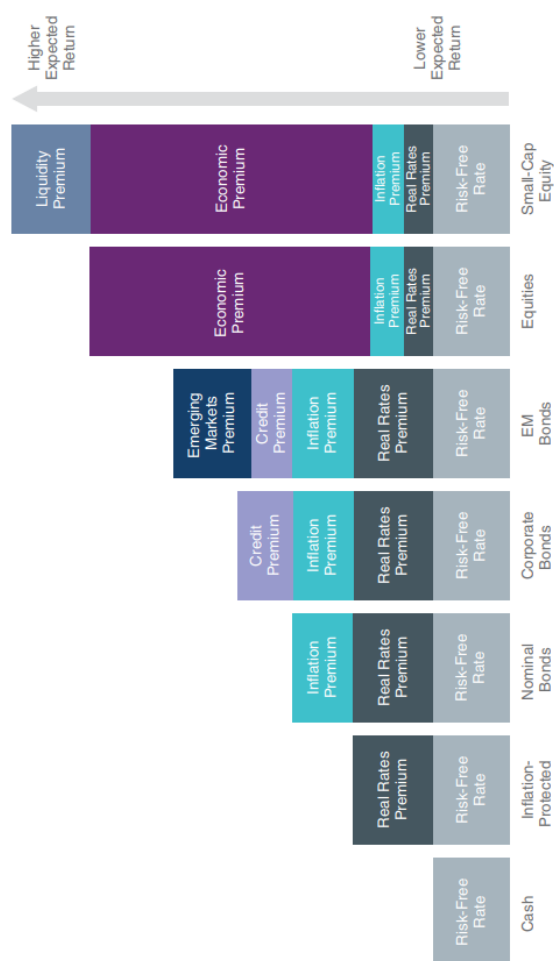
一般地，因子模型中使用的经验因子分为以下三类：

(1) 外部宏观经济因子，如国内生产总值(GDP)、消费者价格指数(CPI)、失业率、债券信用息差，以及收益率曲线的陡峭度。

(2) 基本面因子（公司特征），如市盈率，股息支付率盈利增长预测和财务杠杆

(3) 提取的统计因子，例如市场投资组合的回报率特定行业(公用事业、交通、航空航天)股票回报率的平均值等。

下图展示了将资产类型分解成为因子的案例



**Exhibit 9.1** An example of breaking down asset classes into factors.  
Source: Hogan, Hodges, Potts, and Ransenberg (2015). Reprinted with permission.

## 9.4 阿尔法构建中的因子模型

几十年来，构建阿尔法一直是主动量化投资管理的核心。主要思路是识别带来出色绩效的回报（因子）的经验动因。

在实践中，构建阿尔法的过程开始于观察到的现象或信号，例如“小盘股优于大盘股”“低波动股票优于高波动股票”“每股收益高的公司优于每股收益低的公司”。在每一个这些观察到的例子背后是一个因子（规模、波动性、每股收益），投资组合经理试图用因子模拟投资组合来描绘这个因子。

我们来抽象化这一问题，令 $B$ 为因子载荷矩阵， $\alpha$ 是包含阿尔法信号的向量，那么优化问题可以写成：

$$\begin{aligned} \min_w w^T \Sigma w \\ s.t. B^T w = 0 \\ \alpha^T w = 1 \end{aligned}$$

## 9.5 因子模型用于风险评估

### 9.5.1 宏观经济因子模型

逻辑上，资产回报率应该收到宏观经济的影响。例如，如果美联储削减利率，债券和股票市场都会做出反应。原因是当利率降低时，投资者需要较低的投资回报率。较低的必要回报率一般带来证券价值的升值。

宏观经济模型中的每个因子的每个贝塔系数通常根据时间序列回归方程来估计：

$$r_{it} = \beta_{i1}f_{1t} + \cdots + \beta_{iK}f_{Kt} + \varepsilon$$

### 9.5.2 基本面因子模型

基本面因子模型依赖于基本面研究——梳理行业和个体公司数据以找到具有市场价值的公司特征。这样的公司特征可能与公司估值或投资者期望有关，并且通常称为风格因子。包括公司规模、财务比率（如市盈率）、股息率、动量、增长、财务杠杆、收益变动和波动性。行业因子（例如，证券的发行人属于哪个行业类别）是组成基本面因子模型部分的另一组因子。通常，出于建模目的，为了处理多重共线性效应并提高模型的解释力，常常使用风险指数。这些风险指数是基本面因子的汇集，而不是单一财务比率。

基本面因子模型中的每个因子的每个贝塔系数通常根据时间序列回归方程来估计：

$$r_{it} = \beta_{i1}f_{1t} + \cdots + \beta_{iK}f_{Kt} + \varepsilon$$

### 9.5.3统计因子模型

在数学上，证券收益率的协方差矩阵的主成分通过分解协方差矩阵 $\Sigma$ 来提取：

$$\Sigma = PAP^T$$

其中 $P$ 是正交矩阵。

主成分分析的结果可以被认为是不相关的 $N$ 个投资组合，此外，主成分分析可以帮助降低维度。如果最终选取了 $K$ 个因子，那么模型解释的方差占比：

$$\frac{\sum_{k=1}^K \lambda_k}{\sum_{k=1}^N \lambda_k}$$

一个确定隐含的统计因子的替代技术是因子分析。当基本因子模型的理论基础存在时，因子分析比主成分分析更为合适。因子分析通常使用回归来估计因子载荷矩阵，因此证券回报可以表示为：

$$r = \alpha^T w + Bf + \varepsilon$$

其中 $\alpha$ 是回报的 $N$ 维向量， $f$ 是因子的 $K$ 维向量， $B$ 是因子载荷的 $K \times N$ 矩阵， $\varepsilon$ 是残差的 $N$ 维向量。需要估计的协方差矩阵可以假设为：

$$\Sigma = BB^T + \Sigma_\varepsilon$$

估计过程分为两个步骤。第一步中，通过线性回归或最大似然估计法来估计协方差矩阵和因子载荷。第二步中，使用协方差矩阵和因子载荷来估计因子。在每个时间点 $t$ 的因子估计为：

$$f_t = (B^T \Sigma_\varepsilon^{-1} B)^{-1} B^T \Sigma_\varepsilon^{-1} (r_t - \alpha)$$

### 9.5.4混合因子模型

如今几个因子模型供应商都提供混合因子模型。这些模型使用传统的因子组(货币、行业、国家等)来尽可能多地解释回报的不定性，然后使用统计因子(通过主成分分析获得)吸收因子模型中的剩余风险，即因子模型误差项中剩余的不定性。



一方面，使用基本面宏观经济因子模型为风险和投资组合管理提供了背景，另一方面使用统计因子模型使得捕获残差回报中的任何剩余隐藏因子得以可能。因此，混合因子模型试图做到两者最好：它对模型强加了一些结构设置，但也从数据中挖掘那些在理论模型中无法得到的附加信息。

### 9.5.5 选择“适合的”因子模型

如前所述，许多资产管理公司使用软件供应商提供的因子模型。“基本面”因子模型仅使用公司的基本面信息。“全球”因子模型纳入了关于世界货币和利差的信息。“混合”因子模型将基本面和/或宏观经济因子模型与统计因子模型相结合。

那么哪个因子模型是“最好的”呢？这没有单一的答案，但投资组合经理的投资策略细节可以缩小选择。

首先，投资经理应选择一个能涵盖其投资领域的因子模型：单一国家、地区或全球。这是因为全球因子模型考虑了区域投资时不那么重要的货币和其他因子。

第二，投资经理应该考虑投资组合的投资时限。供应商因子模型通常具有目标时限——短期、中期或长期，其中投资时限是投资组合的平均周转时间。不同的因子在短期内和长期内是不同的。短期因子模型更适食期限为一天至两个月的投资者，例如对冲基金。长期时限是指长于一年的投资期，中期时限是指介于短期和长期之间的任何时限。

## 9.6 数据管理与质量问题

### 9.6.1 数据调整

数据通常存储在多个数据库中需要合并才可以进行分析。合并来自不同数据库的数据可能是非常具有挑战性的。例如，有时不同数据库为同一公司使用不同的标识符，使得难以对应记录。即使使用诸如CUSIP或股票代码之类的公共标识符，后者也可能随时间而改变，并且有时会因为公司不再存在于数据库中而被重新使用，因此可能难以跨数据库正确地连接所有公司。

数据调整中的一个更微妙的问题是，因为数据的差异，但基于不同变量的数据记录，用不同的方法来计算模型输入信息，可能会导致不一致的估计。Faboi、Focardi和Kolm展示了两个替代但等价的方法来计算一个基本面因子(EBITDA/EV)，数据来源是Capital IQ Compustat Point-In-Time数据库中的记录两种方法得出非常不同的EBITDA/EV估计。这两种方法的不同在于用于计算分母(EBITDA)的数据不同：

方法一： $EBITDA = \text{销货} - \text{销货成本} - \text{销货和一般管理费用}$

方法二： $EBITDA = \text{未许折旧前的经营收入}$

---

二者的关系：

未计折旧前的经营=销货-销货成本-销货和一般管理费用

然而，从数据库中提取数据不是一样的Fabozzi列出了这种数据不一致的多种原因。首先，可能存在四舍五入和轻微误差。第二，记录中可能有错误。第三，有时不同的数据项被结合在一起。有时旧和摊销费用不是损益表上的单独项目；而是包括在销售商品的成本中。第四，数据项可能在不同的公司、部门或行业中报告不一致当财务数据提供者不正确地将财务指标从公司报告映射到特定数据库项目时，也会发生这种情况。

### 9.6.2幸存者偏差

如果当公司不再存在而从数据库中删除，并且只有还运行的公司的数据被用于分析时，幸存者偏差(survival bias)便会发生。（公司可能因不同的原因而不再存在，例如破产或合并。）幸存者偏差导致错误的分析。只关注幸存者和忽略那些不再存在的公司，都可能会在分析中忽略重要的信息。通过下面的例子可以理解这个问题。假设一个人将在比较2005年3 000家最大(最高市值)的公司与2015，3 000家最大的公司后发现，3 000家最大公司的价值上升了40%。有人可能会认为，包括这些公司在内的投资组合的回报率是40%。然而，如果一些在2005年存在的公司不再存在(或不在2015年3000家最大公司的名单上)，那么实际上是无法构建包含最大公司的投资组合的！！

### 9.6.3前视偏差

在预测模型中使用在预测发生时还没有的数据，就会发生前视偏差。例如，假设将年终收益用作1月份回报预测模型中的一个因子这些间报在年度结束后至少一个月后才会公布。因此，它们不能用于预测1月份的收益。它们不可能在1月初获得。如果使用回填和重述，前视偏差可能变得更糟。回填(backfilling)是将先前丢失的数据在收到后输入数据库，这可能发生在比必须使用该信息以用于预测的时间晚很多。重述(restatement)发生于(例如)当公司修改了其最初发布的财报时。许多数据库公司(例如，Capital IQ)会覆盖最初记录的数字。然而，如果将信息用于预测模型，则只有原始数字可获有。使用更新的信息来构建预测模型引入了偏差，因为它不能正确地反映在需要做出投资决定时可用的信息！

### 9.6.4数据迁就

数据迁就是在数据中发现“模式”，即使这些模式实际上并不存在。当人们不能进行控制实验而必须依靠记录的数据和统计推断来寻找模式时，这个问题尤其突出。

这正是金融数据的情况。

拥有足够的时间和足够的数据，人们可以找到想找到任何数据模式。因此，确保投资组合经理使用的任何分析模型都建立在合理的经济理论和常识基础上，而不仅仅是经验证据，是非常重要的。

## 9.7 风险分解、风险归因和业绩归因

风险分解就是将投资组合方差根据是否持有或风险类型分解为个体贡献在使用因子模型进行风险分解的情况下，人们确定投资组合对因子模型中不同因子的风险敞口。

## 9.8 因子投资

因子模型自20世纪70年代以来就一直存在，并用于阿尔法构建和风险管理环境

因子投资策略——有时称为聪明贝塔、科学贝塔、基于因子的投资或基本面指数化有巨大的多样性，但它们所共有的是对历史表现良好或投资经理认为可能继续增加风险调整后回报的因子的持续敞口。因子投资已经改变了特定资产类别中的传统资产分配和投资组合选择。

在传统的资产配置策略中，投资组合在不同资产类别之间是多元化的(例如，60/40投资组合将包含投资于股票的资金为60%和债券为40%)。然而，这样的投资组合不一定能够很好地多元化——资产类别之间的相关性往往很高，并且实际上市场困难时会爆发。例如，在“正常”的市场，如2009—2014年期间，美国大盘股与标普500指数完全相关;美国中盘股与标300指数的相关系数为0.96;美国小盘股与标普500指数的相关系数为0.93。根据晨星机构的数据，在2008 2009年金融危机后，这些相关性分别为1.00、0.97和0.96。同样高的相关和相同的现象也可见于发达和新兴的股票市场。

更缜密的投资方法可能需要将资产类别分解成可以解释其相关性和回报/风险特征的共同因子。例如，贝莱德(BlackRock)——一个聪明贝塔产品的供应商，识别出6个跨越不同资产类别的因子:经济(与经济增长不确定性相关的风险信用(违约风险)、实际利率(暴露于实际利率变化的风险)、通货膨胀(暴露于价格变化的风险)、新兴市场(政治动荡的风险)和流动性(在特定时间无法卖出投资的风险通过将投资配置到这些因子而不是资产类别，投资可以避免因资产类别的因子敞口重叠而导致的风险集中。因子本身显示较低的相关性，投资组合更好的多元化。下图显示贝莱德所确定的因子之间的相关性。

	ECONOMIC	CREDIT	REAL RATES	INFLATION	EMERGING MARKETS	LIQUIDITY
ECONOMIC	1					
CREDIT	0.58	1				
REAL RATES	0.16	-0.09	1			
INFLATION	-0.37	-0.34	-0.14	1		
EMERGING MARKETS	0.52	0.51	0.08	-0.28	1	
LIQUIDITY	0.44	0.44	0.09	-0.31	0.38	1

**Exhibit 8.2** Correlations of major factors that explain asset class returns and co-movements. Correlations measured over the period August 1988 through September 2014.  
*Source:* Hogan, Hodges, Potts, and Ransenberg (2015). Reprinted with permission.