

因子投资：实证分析empirical analysis(OPP:theoretical analysis)

1)样本sample:实证分析中样本选取全面(eg行研数据、私募量化所选取的数据需要中国所有股票的分析)

选择偏差，survival bias，addition/deletion在剔除和加入时收益变化

2)缩尾winsorization，截尾truncation：若不对异常现象进行剔除，可能产生偶然性

3)描述性统计：Xi，t (min，max,mean均值,skew,kurt，median中位数，P0.5，P25，P75，P95在对应百分数的数值，N观测值个数，std)

4)相关性(Ai,t Bi,t):自相关(今天的交易与昨天的交易有关)(diff=At+1-At,一阶差分可以剔除季节效应若否做二阶差分)，持续性，MDH：mixture distribution hypothesis：price-volume量价关系 本质上是信息和价值的关系，volume proxy of information成交量推动价格变化 corr(volumei,t，reti,t)==>MDH:positive&&significant 交易量和收益率具有正且显著的相关性则证明了MDH假说

SIAH/SAIH sequential information arrival hypothesis顺序信息到达假说 volumei,t--reti,t具有领先滞后关系lead-lag-->Granger Causality因果检验(仅代表数据上具有因果关系不代表现实中具有因果关系) Granger不成立也不能推翻SIAH 考虑nonlinear情况

5)单变量分析univariate--双变量分析bivariate independent-sort analysis

第1章

因子投资基础

因子投资涵盖的内容包罗万象。有人用因子来获取收益，也有人用因子来控制风险；有人用因子进行被动的资产配置，也有人用因子博取主动的套利收益；有人用因子研究预期收益率，也有人用因子研究收益率的波动.....这些差异性的使用方式足以把刚接触因子投资的人搞得晕头转向。而如果将“什么是因子投资”这个问题抛给从事因子投资的人，一千个人也可能给出一千种答案。因子投资涉及的内容虽然看似纷繁，但它们其实非常协调地构成了一个有机的整体。本书将围绕一个公式，站在一个统一的视角下探讨因子投资的方方面面。

1.1 统一视角下的因子投资

1.1.1 一个公式

20世纪60年代，资本资产定价模型（**Capital Asset Pricing Model, CAPM**^[1]）问世。在CAPM被提出之前，人们对于风险如何影响一个公司的资本成本（cost of capital），进而如何影响预期收益率并没有清晰的认识。1958年，Modigliani and Miller（1958）提出了著名的资本结构无关原理（也被称为MM定理）。该定理认为在不考虑税收、破产成本、信息不对称，且市场有效的假设下，企业价值不会因为企业融资方式改变而改变。然而，MM定理使得如何刻画风险和预期收益率之间的关系变得更加扑朔迷离。直到CAPM的诞生，才首次清晰地描绘出风险和收益率之间的关系。

根据CAPM理论，资产的预期超额收益由下面这个一元线性模型决定：

$$E[R_i] - R_f = \beta_i (E[R_M] - R_f) + 1 * SML (\text{规模因素}) + 2 * HML (\text{价值因素}) \quad (1.1)$$

SML、HML构造因子，构造方法多

eg : T-bili, shibor国债利率, 银行同业拆借利率

其中 $E[\cdot]$ 是期望符号， R_i 为某资产 i 的收益率， R_f 为无风险收益率， R_M 为市场组合的预期收益率。式中 $\beta_i = \text{cov}(R_i, R_M) / \text{var}(R_M)$ 刻画了该资产收益对市场收益的敏感程度，它也被称为资产 i 对市场风险的暴露程度。CAPM正是最简单的线性因子模型，它指出资产的预期超额收益率由市场组合的预期超额收益率和资产对市场风险的暴露大小决定，而市场组合也被称为市场因子。

式（1.1）这个简单的关系为后续大量线性多因子定价模型的研究拉开了序幕。人们发现不同资产的收益率并非由单一的市场因子决定，而是同时受到其他因子的影响。以此为契机，Ross（1976）提出了著名的套利定价理论（Arbitrage Pricing Theory, APT），在CAPM的基础上做了进一步延伸，构建了线性多因子定价模型（简称多因子模型）。多因子模型假设资产 i 的预期超额收益由以下多元线性模型决定：

$$E[R_i^e] = \beta_i' \lambda \quad (1.2)$$

其中 $E[R_i^e]$ 表示资产 i 的预期超额收益， β_i 是资产 i 的因子暴露（factor exposure）或称因子载荷（factor loading）， λ 是因子预期收益（factor expected return），也常被称为因子溢价（factor risk premium）^[2]。式（1.2）中引入符号 $E[R_i^e]$ 代表预期超额收益而舍弃了 $E[R_i] - R_f$ ，原因如下。对于个股或者由若干支股票组成的纯多头投资组合这类资产，它的预期超额收益是 $E[R_i]$ 和无风险收益率 R_f 之差，即 $E[R_i] - R_f$ 。在实证

资产定价中，另一类常见的资产是通过多空对冲构建的资金中性投资组合，例如做多一定金额的股票，同时做空同样金额的股票。依照定义，这类资产的预期超额收益率就是多头和空头预期收益之差，无须额外减去无风险收益率。因此，使用 $E[R_i^e]$ 代表资产的预期超额收益更具一般性。此外，在资产定价研究中，由于学术界默认的研究对象始终是资产的预期超额收益，因此约定俗成在称呼上舍去“超额”二字而仅用预期收益率。为了表达的便利性，本书将遵循这种惯例，后续统一称 $E[R_i^e]$ 为预期收益率。

和CAPM类似，多因子模型假设资产的预期收益率 $E[R_i^e]$ 由公式（1.2）右侧的一系列因子的预期收益率和资产在这些因子上的暴露决定。它表明多因子模型研究的是不同资产预期收益率之间的差异，而非单一资产收益率在时间序列（又叫时序，书中为了不同对应关系，两种叫法会混用）上的变化。在多因子模型研究的术语中，不同资产预期收益率的差异通常被称为（横）截面（cross-sectional）差异，给定因子的预期收益率，资产预期收益率的高低由它在这些因子上的暴露 β_i 大小决定。多因子模型在数学上简单、优雅，它允许人们在式（1.2）的右侧加上任何可能影响资产预期收益率的因子。然而，实际的金融市场十分复杂，影响资产收益率的原因层出不穷又千变万化。因此，虽然式（1.2）给出了市场均衡状态下资产的预期收益率，但在实际中式（1.2）左、右两侧往往并不相等，而是存在一个定价误差（pricing error）：

$$E[R_i^e] = \alpha_i + \beta_i' \lambda \quad (1.3)$$

上式中 α_i 就是资产 i 的实际预期收益率和多因子模型隐含的预期收益率之间的定价误差。该误差产生的原因可以从以下两方面来考虑：

（1）模型设定偏误，即式（1.2）右侧遗漏了重要的因子。当被遗漏的因子被加入后，即可消除误差。

（2）模型本身没有问题，但由于资产收益率的实际数据仅仅是总体的一个样本，因此误差总是存在的。这时需要通过统计方法检验误差 α_i 是否显著不为零：

- 如果 α_i 并非显著的偏离零，那么可以认为它的出现仅仅是因为运气的原因；

- 如果 α_i 显著偏离零，它则代表了某个可以通过套利而获得超额收益的机会^[3]；它也同时说明由于某些原因，市场对该资产出现错误定价

（mispricing），从而导致其实际预期收益率和多因子模型下的预期收益率出现了偏离。

如何选择因子来构建多因子模型，如何计算资产在因子上的暴露以及因子的收益率，如何使用统计学的方法对定价误差 α_i 进行正式检验就成了使用多因子模型研究资产定价时必须回答的问题。而多因子模型一经提出，用来回答上述问题的相关理论也得到了迅速的发展。如今多因子模型早已成为实证资产定价

（empirical asset pricing）研究的最主要方法，也由此在投资实务中催生了一个全新的类别——因子投资（factor investing）。

作为当今一个重要的投资类别，因子投资中涉及的内容非常丰富，但所有的方面都可以围绕式（1.3）来展开和讨论，形成一个关于因子投资的统一视角。为了更好地解释因子投资包含的内容，以下先对式（1.3）的三个成分加以说明，它们是因子、多因子模型以及异象。

1.1.2 因子、多因子模型和异象

图1.1展示了式（1.3）中的三个成分。在多因子模型中， $\beta_i'\lambda$ 中包含了多个影响式（1.3）左侧资产收益率的解释变量，每个解释变量代表了一个因子。多个因子的预期收益率以及资产在它们上的暴露放在一起构成的 $\beta_i'\lambda$ 就是多因子模型。对于某个给定的资产*i*，如果它的实际预期收益率和多因子模型隐含的预期收益率之间的误差 α_i 显著不为零，则称这个资产为一个异象（anomaly）。

因子投资中最重要的自然是因子。那么，因子到底是什么？如何选择因子？由式（1.3）可知，它左侧不同资产的收益率均可以归结到有限个因子的收益率上，而不同资产预期收益率的高低由它们对因子的暴露大小决定。由此可以得出因子的定义：

一个因子描述了众多资产共同暴露的某种系统性风险，该风险是资产收益率背后的驱动力；因子收益率正是这种系统性风险的风险溢价或风险补偿，它是这些资产的共性收益。

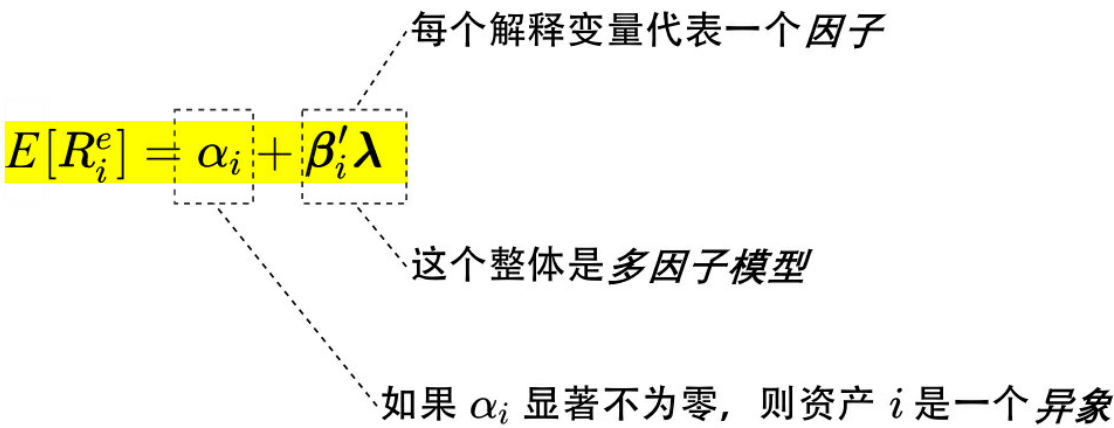


图1.1 式（1.3）的三个成分

下面通过一个来自Ang（2014）的类比加深对上述定义的理解。Ang（2014）将因子之于资产类比于营养之于食物。人们的日常生活离不开摄入各种食物，比如米饭、谷物、肉类和乳制品等。这些食物之所以对人体重要是因为它们所包含的营养成分，例如水、碳水化合物、蛋白质、纤维和脂肪等。这些营养成分就是食物的“因子”。每种食物对人体的价值体现在两方面：（1）它包含哪些营养成分以及包含了多少（类比多因子模型中的 β_i ）；（2）该食品包含的每种营养成分对人体健康的重要性（类比多因子模型中的 λ ）。因此，虽然食物种类五花八门，但归根结底它们都可以归结为不同营养成分的组合；同理，虽然股票资产的组合形

式成百上千，但归根结底它们都可以归结为不同因子的组合。这就是因子的作用和多因子模型的价值。

从上述定义可以总结出因子需要满足的两个必要条件：（1）因子驱动了资产收益率的共同运动（co-movement），因此因子一定和资产收益率的协方差矩阵有关；（2）从长期来看因子是可以获得正收益的，这意味着因子必须是被定价的。明确了因子的定义，下面再来看看什么是多因子模型。

把若干个因子放在一起使用就构成了一个多因子模型，它们对应着式（1.3）中的 $\beta_{i\lambda}$ 部分。通常来说，人们总可以找到非常多的能够解释资产预期收益率截面差异的因子，但无论是学术理论还是投资实践都表明并不应该把所有的因子都放到一个多因子模型里。如何确定一个多因子模型中到底有几个以及哪些因子，可以从以下两点考虑。首先，很多因子之间都存在一定的相关性，高度相关的因子代表的是资产收益率所暴露的同一类风险。在选择模型包括的因子时，必须要考虑相关性的影响，人们希望加入模型的因子是相互独立的、每个因子都能对解释资产预期收益率截面差异有显著的增量贡献。其次要考虑简约法则（The Law of Parsimony）。从定义出发，每个因子代表的是资产暴露的某种共性风险。因此从常识来说，多因子模型中因子的个数一定是有限的^[4]。学术界的主流多因子模型通常包括3~5个因子，本书的第4章将对这些模型进行系统的阐述。

最后来看异象。在选定了多因子模型后，如果按照某个特征构建的多空组合的预期收益率中存在一部分无法被多因子模型解释，且这部分显著大于零，那么该资产就是一个异象。举个例子，在股票市场中，人们经常根据某个公司财务指标或者量价指标来把股票排序，并按照排序高低构建一个多、空对冲投资组合。将该投资组合作为一个资产放在多因子模型中，如果其 α_i 显著大于零，那么就称该投资组合为一个异象，而构建该组合的指标就被称为异象变量（anomaly variable）。

1.1.3 再论异象和因子

1.1.2节从式（1.3）出发探讨了因子投资中的三个部分：因子、多因子模型和异象。下文将围绕这三部分，从学术界和业界各自的立场来讲述因子投资的内容。不过在那之前，本节想花一点笔墨说明一下学术界和业界对异象和因子看法上的差异，这将有助于下文介绍因子投资的内容。

在因子投资的术语中，无论是式（1.3）中 α_i 代表的异象还是 $\beta_{i\lambda}$ 代表的因子往往都被称为因子。为了区分它们，只需要记住 $\beta_{i\lambda}$ 是一个定价模型，因此它包含的因子又被称为定价因子（pricing factors）；而 α_i 代表的是按某种方法构建出的能够获得多因子模型无法解释的超额收益的资产，因此它又被称为异象因子（anomaly factors）。在本书中，如无特殊说明，当谈及异象时，指的就是 α_i 代表的部分；当谈及因子时，指的则是定价因子 $\beta_{i\lambda}$ 的部分。

在因子投资中，学术界往往更加严谨地区分异象因子和定价因子。由因子的定义可知，它必须能够解释资产预期收益率截面上的差异。经过几十年的发展，学术界已经从上市公司财务指标和股票交易中的量价指标中找到了上百个满足上述要求的因子。然而，从1.1.2节的论述可知，一个多因子定价模型应满足简约法

则，这意味着定价因子个数是有限的。为此，学术界往往从经济学和金融学的原理出发，从大量相关的因子中找到有限个相对独立的因子构成多因子模型，而把其他没有被加入模型中的、能获得模型无法解释的超额收益的因子都视为异象。反观业界，无论是管理人还是投资者，他们往往不严格区分异象因子和定价因子，而把它们统称为因子。业界往往更关心因子能否获取定价模型无法解释的超额收益，因为这代表着构建因子的变量可以区分高收益股票和低收益股票，因此人们可通过买入高收益股票而“战胜市场”。理解了学术界和业界对定价因子和异象因子看法的差异后，接下来就来看看因子投资包含哪些内容。

1.1.4 因子投资包含的内容

因子投资的内容可以因研究的对象是式（1.3）右侧的 $\beta_i\lambda$ 或 α_i 而分成两大类。每一大类下又可以因站在学术界和业界的不同立场再进一步划分。按照重要性的先后顺序，首先来看 $\beta_i\lambda$ 这部分。

1. 关注于 $\beta_i\lambda$ 的研究

对于 $\beta_i\lambda$ ，学术界关注的是找到“最好”的多因子模型“最好”通常意味着在该模型下资产的 α_i 尽可能接近零，即该模型无法解释的异象越少越好。以此为目标，学术界在过去30年针对股票市场推出了很多经典的多因子模型。在提出每个多因子模型的同时，学术界也格外关注每个因子背后的原因。即便到了今天，因子背后的原因也仍然是研究的热点之一。经过几十年的发展，学术界也发明出很多计量经济学的方法，它们可以计算因子暴露、因子收益率并对多因子模型进行假设检验。除此之外，学术界研究 $\beta_i\lambda$ 的另一个重点是对主动基金管理人的业绩进行归因。大量的实证数据研究表明，在选择了适当的多因子模型后，绝大多数主动基金管理人并不能获得超额收益。

与学术界不同，业界进行因子投资最重要的目标是使用因子来获取超过基准的收益^[5]。因此业界从资产配置的角度聚焦于找到从长期来看有风险溢价的因子（即 λ 大），并以尽可能高的暴露（即 β_i 高）配置在这些因子上。仍然以Ang（2014）中食物和营养的类比做解释。人们每天摄入不同的食品，其本质是为了吸收食品背后的营养成分，这些食品只是获得营养的渠道。人们想要保持健康则需要这些营养成分均衡的搭配。回到因子投资中，业界关心如何以资产为媒介配置在不同的因子上，并在分散化的前提下最大化在不同因子上的暴露，从而获得稳健的（超额）收益。除此之外，随着因子投资的发展，很多Smart Beta ETF金融产品相继问世，极大地丰富了资产配置的工具。

2. 关注于 α_i 的研究

再来看 α_i 部分。 α_i 代表了可以通过套利获得的超额收益。出于截然不同的原因，学术界和业界都对它非常感兴趣。先说学术界，从有效市场假说（Efficient-Market Hypothesis，出自Fama 1970）的观点出发，如果多因子模型不存在模型设定偏误，那么市场中不应该存在太多异象。因此，一旦一个异象被发现，就会被拿来评判市场并非有效，或用来描述市场均衡状态的因子模型有误，这是学术界研究异象的主要动机。另外，解释异象的能力是评价多因子模型优劣的重要标准之一。比如两个多因子模型A和B，如果使用模型A时异象的个数少于使用模型B时

异象的个数，就会认为模型A比模型B更好。模型之间的对比也从侧面推动了对异象的研究。

相比之下，业界对异象的研究动机则“单纯”得多。正如1.1.3节所描述的那样，业界并不严格区分定价因子和异象因子。在业界看来，一个因子是否被纳入某个定价模型（从而从异象因子转变为定价因子）并不重要，而更关心在考虑了交易成本之后，使用该因子是否仍然能够获得超额收益。

3. 截面角度vs时序角度

目前，在关于因子投资内容的介绍中，无论研究的关注点是 $\beta_i\lambda$ 还是 α_i ，它们都是在给定的因子模型下研究资产的预期收益和 β_i 之间的关系，即资产预期收益率在截面上的差异，因此多因子模型是关于均值的模型（model of the mean）。从数学定义可知，预期收益率是收益率在时间序列上的平均，因此均值的模型仅关心不同资产的收益率均值为什么会有差异，而非每个资产的收益率如何随时间变化。把式（1.3）沿时间轴展开，资产超额收益和因子收益率在时序上满足如下多元线性回归模型：

$$R_{it}^e = \alpha_i + \beta_i' \lambda_t + \varepsilon_{it} \tag{1.4}$$

其中 R_{it}^e 表示 t 时刻资产 i 的超额收益， λ_t 表示 t 时刻因子收益率， ε_{it} 表示 t 时刻的随机扰动。这就是研究多因子模型的时间角度。图1.2展示了截面角度和时序角度之间的切换。

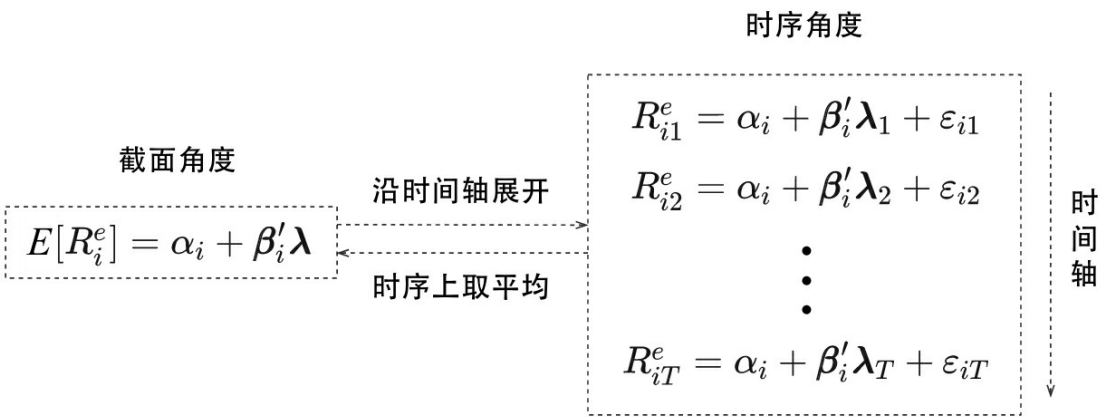


图1.2 截面角度vs时序角度

在时序角度下，因子投资中的两个重要问题是方差模型（model of the variance）和因子择时。为了解释前者，把 N 个资产的时序多元回归模型放在一起表达：

$$R_t^e = \alpha + \beta \lambda_t + \varepsilon_t \tag{1.5}$$

其中 $R_t^e = [R_{1t}^e, R_{2t}^e, \dots, R_{Nt}^e]'$ 是 N 维超额收益向量； $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]'$ 是 N 维定价误差向量； $\beta = [\beta_1, \beta_2, \dots, \beta_N]'$ 是 $N \times K$ 因子暴露矩阵； $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt}]'$ 是 N 维随机扰动向量，满足 $E[\varepsilon_t] = 0$ 以及 $\text{cov}(\lambda_t, \varepsilon_t) = 0$ 。对式（1.5）两侧求协方差矩阵并利用 $\text{cov}(\lambda_t, \varepsilon_t) = 0$ 可

得：

$$\Sigma = \beta \Sigma_k \beta' + \Sigma_\varepsilon \quad (1.6)$$

其中 Σ (N 阶矩阵)、 Σ_k (K 阶矩阵) 以及 Σ_ε (N 阶矩阵) 分别为 N 个资产的协方差矩阵、 K 个因子的协方差矩阵以及 N 个随机扰动的协方差矩阵。由于 ε_{it} 相互独立，因此 Σ_ε 是一个对角阵。式 (1.6) 就是关于方差的模型。

从学术界的立场来看，式 (1.6) 意味着因子必须和资产的协方差矩阵有关，而这也正是1.1.2节中关于因子需满足的两个必要条件中第一个条件的出处。值得一提的是，当下学术界流行的所有主流多因子模型（本书第4章会详细介绍）要么来源于金融学或经济学原理，要么来源于自行为金融学的启发，但它们都没有直接从式 (1.6) 传达的因子协方差矩阵和资产协方差矩阵的关系入手。一个多因子模型代表着资产超额收益的某种结构，因此一个好的模型应该尽可能解释不同资产收益率间的共性运动。这反映到式 (1.6) 上就是模型中因子的选择应使得 $\beta \Sigma_k \beta'$ 尽可能接近资产协方差矩阵 Σ 的非对角线元素。但无论是经典的Fama–French三因子模型（Fama and French 1993）还是当下热门的Hou–Xue–Zhang四因子模型（Hou et al. 2015）等，均没有对此给予充分的考量。在最新的研究中，Pukthuanthong et al. (2019) 指出了这种关联的重要性，并以此为依据提出了一个甄别真正因子的框架。可以预期，从式 (1.6) 出发找到真正的因子、构建多因子模型将会是学术界未来的一个研究重点。

再来看看业界。业界对于式 (1.6) 的关注和学术界有所不同，而是基于风险控制这个非常现实的考量。在投资中，一个很重要的风险指标是投资组合的波动率。投资组合的波动率由其包含的资产在组合中的权重、资产收益率的方差以及资产收益率之间的协方差决定。因此，计算投资组合的波动率就必须知道资产之间的协方差矩阵。从数学上说，如果使用历史收益率序列计算样本协方差矩阵，那么历史数据的期数 T 需满足 $T \geq N$ ，否则用历史数据估算出的样本协方差矩阵就是不可逆的^[4]。当资产个数 N 很大时，要求 $T \geq N$ 是不切实际的。然而，如果把资产的收益率通过多因子模型转化为因子的收益率，则可以大大简化上述问题。

在多因子模型中，因子的个数 K 往往远小于资产的个数，因此使用多因子模型分析资产收益率相当于一种降维处理。式 (1.6) 通过多因子模型把求解 N 阶的协方差矩阵巧妙地转化为求解 K 阶因子收益率的协方差矩阵。有了因子协方差矩阵，利用资产和因子收益率之间的线性关系就可以反推出资产收益率的协方差矩阵并进行风险控制，这种做法在投资实务中有巨大的价值。在海外的大机构中，使用多因子模型计算资产协方差矩阵并进行风险控制的有很多，但其中最著名的大概要数Barra多因子模型^[5]。以Barra最新一代的中国股票多因子模型CNE6为例，它使用1个国家因子、9个风格因子以及一系列行业因子来刻画A股市场上超过3000支股票的收益率，极大简化了股票协方差矩阵的求解。在过去好几代多因子模型中，Barra开发了一系列统计学方法，逐步提高了协方差矩阵估计的准确性。

时序角度的另一个研究重点是因子择时。与截面角度只关心因子的预期收益率不同，在时序角度，人们关心的是因子收益率如何随时间波动。由因子的定义可知，因子代表了资产共同暴露的某种系统性风险，而长期大于零的因子溢价正是对该系统性风险的补偿。这说明每个因子的收益率在时间序列上是有波动的，有时收益高，有时则面临亏损；且不同因子的收益率在不同宏观经济周期对应的

市场环境中表现也是不同的。在这方面，学术界通常关心的是不同因子和宏观经济以及投资者情绪之间的关系。而反观业界，由于业界使用因子的目的是进行资产配置和获取超额收益，因此自然会对因子收益率在时序上的变化产生极大的兴趣。这是因为如果能够通过择时来挑出某个因子“好使”的时候配置、在它“不好使”的时候规避，那么无疑会提高使用因子的风险收益水平。正是在这个动机的趋势下，因子择时始终是业界的研究热点之一。

4. 统一视角

以上从式（1.3）出发，从 $\beta_i'\lambda$ 、 α_i 、 α_i 以及“截面角度vs时序角度”三个维度梳理了因子投资中所涉及的内容。这也正是本书希望传递出来的统一视角下的因子投资。图1.3对这三个维度进行了总结，它可以被视作全书的思维导图。

1.1.5 实证资产定价与因子投资

严格地说，在1.1.4节涉及的内容中，学术界视角下因子投资的内容属于实证资产定价研究的范畴；而一般当人们谈论因子投资的时候，通常指的是业界的视角。然而，出于以下两个原因，本书提倡把学术界和业界关注的不同内容放入统一的视角，学习和实践因子投资。

首先，投资学大概是学术界和业界结合的最紧密的学科之一，而因子投资则是这种紧密结合的代表。经过了几十年的努力，学术界在实证资产定价方面提出了丰富的金融理论模型和计量经济学统计手段，并发现了诸如价值、盈利、动量、低波动等如今已被广泛认可和接受的因子。这些发现自问世以来就持续地指导着因子投资实践。毫不夸张地说，因子投资的兴起深深地植根于学术界关于资产定价的研究中。另外，业界的因子投资活动也为已有理论的反复检验和新理论的提出提供了充足的数据，促使学术界不断提出新的研究成果。

$E[R_i^e] = \alpha_i + \beta_i'\lambda$			学术界	业界
因子投资	截面角度	$\beta_i'\lambda$	<ul style="list-style-type: none">“最好的”模型管理人业绩归因	<ul style="list-style-type: none">配置因子、获取超过基准的收益
		α_i	<ul style="list-style-type: none">市场有效性检验模型比较	<ul style="list-style-type: none">通过异象获得收益
	时序角度		<ul style="list-style-type: none">因子和资产协方差矩阵有关因子和宏观经济以及投资者情绪的关系	<ul style="list-style-type: none">风险控制因子择时

图1.3 统一视角下的因子投资

其次，作为一个理论指导实践的细分学科，虽然历经了几十年的发展，但因子投资依然在不断地完善和创新。最初，多因子模型仅是在人们拒绝了CAPM之后的代替之选，其后它逐渐发展成为股票投资的有效手段之一。时至今日，多因子模型早已渗透到不同大类资产的投资中，人们在股票、债券、商品期货甚至是加密货币领域都可以看到它大展身手。在当下最新的趋势中，人们更是从因子的角

度对不同类资产收益率的底层驱动因素和逻辑进行分析，使用因子进行跨类别的大类资产配置。对因子投资的实践者来说，只有了解已有的各种研究方法，才能更好地迎接和拥抱因子投资未来的发展和变革。

基于上述两个原因，本书作者认为掌握学术界的研究成果对于培养正确的因子投资科学观至关重要。而将学术界的方法论和业界的实践相结合，依照图1.3勾勒出的统一视角系统地解读因子投资则是本书最重要的特色。在后面的第2章到第7章将会对这个框架下的不同部分进行探讨。在那之前，接下来的1.2节和1.3节先从学术界和业界的角度梳理一下因子投资的起源和发展，1.4节介绍本书的结构和写作目标。

最后需要指出的是，虽然因子投资如今已被应用于各类资产，但相比于其他类别的资产，它在股票市场的应用无疑是压倒性的。当人们谈到因子投资时，也往往默认是股票市场。这背后的原因是股票是最重要的一类资产，而因子投资的发展也源自学术界对股票市场的研究。为此，本书将完全针对股票市场阐述因子投资的方方面面。感兴趣的读者也可以尝试将本书介绍的方法论应用于其他资产。

[1]CAPM由Treyner（1961，1962），Sharpe（1964），Lintner（1965a，1965b）以及Mossin（1966）分别独立提出。公众号“川总写量化”的文章《CAPM的一小段历史》对其进行了介绍。

[2]预期收益可用来描述各类资产提供给投资者的期望回报，包括但不限于股票、债券、大宗商品、房地产等。而对于本书关注的因子，学术界往往更习惯将因子的预期收益称为因子溢价。本书中会交替使用“因子预期收益”和“因子溢价”这两种称谓来指代因子的预期收益。

[3]金融市场中通常使用 α 代表资产相对于某基准的超额收益。

[4]如果一共有 N 个资产，那么总可以使用这 N 个资产作为“因子”来完美地解释它们预期收益率的差异。但这么做毫无意义。

[5]以股票市场为例，市场指数就是基准，而因子投资的目标是使用不同的因子获得比市场收益率更高的收益。

[6]协方差矩阵应是可逆的。在投资组合优化中，协方差的逆矩阵经常出现在资产最优权重的表达式中。因此，当使用样本协方差矩阵作为协方差矩阵的估计时，如果因 $T < N$ 造成样本协方差不可逆，就会使得投资组合优化不可解。除此之外，即便 $T \geq N$ 成立，仍需考虑使用样本协方差矩阵作为协方差矩阵的估计的准确性的问题。Ledoit and Wolf（2003）指出，通常 T 需要比 N 高一个数量级才能保证估计的准确性。在实际的资产配置中， T 比 N 高一个数量级几乎是不可能的。

[7]Barra如今已经被MSCI收购，但依照惯例业界仍然使用Barra多因子模型这个叫法。

1.2 因子投资的学术起源

学术界对于因子的研究可以追溯到20世纪30年代。Graham and Dodd (1934) 提出了价值溢价，而这本*Security Analysis* (证券分析) 也早已成为业界的圣经。之后的60年代和70年代，CAPM和APT相继被提出，它们为研究因子提供了定量的分析工具。自20世纪70年代以来，学者们逐渐发现按照某种风格“打包”的股票能够战胜市场。这其中最值得一提的是Basu (1977) 发现的便宜股效应和Banz (1981) 发现的小市值效应。由于它们和当时主流的有效市场假说相违背，因此被称为异象。也正是从那个时期开始，学术界开始了轰轰烈烈的因子研究大潮。

1.2.1 实证资产定价

对于学术界而言，研究因子的最主要目标是提出更好的资产定价模型。在这方面的研究中，2013年的诺贝尔经济学奖获得者Eugene Fama、Lars Peter Hansen以及Robert Shiller三位教授功不可没。

对于Eugene Fama，各位读者可能并不陌生，他被誉为有效市场假说之父。他对市场有效性的最大贡献出自Fama (1970) 这篇文章。该文提出了联合假说 (joint hypothesis) 问题，即要想检验市场的有效性就必须先有一个合理的资产定价模型。只有知道了定价模型给出的均衡状态下股票的预期收益率，才有可能正确地检验市场是否有效。毫不夸张地说，自20世纪70年代以来，学术界在实证资产定价方面的研究都是在联合假说的框架下进行的。

在多因子模型方面，Eugene Fama也是当仁不让的代表人物。首先来看看多因子模型的特例CAPM。在CAPM被提出之后，学术界进行了大量的工作来检验它。其中颇具代表性的一篇文章是Fama and MacBeth (1973)。该文章中的方法在检验CAPM时，巧妙地规避了收益率随机扰动项截面相关性的影响，得到了更加令人信服的结果并拒绝了CAPM模型。值得一提的是，Fama and MacBeth (1973) 虽然是为了检验CAPM，但提出的回归方法却得到了更广泛的传播，成了因子投资中的一个重要统计手段^[4]。

除了检验CAPM之外，Fama的另一个贡献是在Fama and French (1992) 一文中整合了学术界后来陆续发现的便宜股效应和小市值效应，它们均是CAPM无法解释的异象。面对诸多异象，人们再也无法无视不能被单一市场因子解释的系统性风险了，学术界也亟待一个新的定价模型的出现。这时Fama and French (1993) 横空出世。该文在市场因子的基础上加入了代表便宜股和小市值效应的HML和SMB两个因子，构成了一个三因子模型。时至今日，Fama–French三因子模型早已

成为全球各国股票市场上实证资产定价研究的首选，而学术界后来推出的主流多因子模型也都建立在该三因子模型之上^[2]。

Lars Peter Hansen对实证资产定价的贡献主要体现在计量经济学理论的突破。除多因子定价模型之外，其他最重要的资产定价模型大概要数基于消费的定价模型（Consumptionbased CAPM, CCAPM）。Hansen（1982）提出的广义矩估计（Generalized Method of Moments Estimator, GMM）在检验CCAPM模型时起到了巨大的作用。由于GMM非常强大，因此也常被用来检验多因子模型。关于这点，Cochrane（2005）有过详细的介绍。

与Eugene Fama认为市场是有效的截然不同，Robert Shiller则是行为金融学的代表性人物之一。Robert Shiller在其代表性论文Shiller（1984）中提出了著名的噪音交易者模型。这篇论文成了日后日益增长的行为金融学文献的起点。在Shiller（1984）的模型中，聪明投资者依股票的内在价值进行投资；而噪音交易者的存在造成了价格和内在价值出现了偏离；价格的过度波动来源于人们非理性行为造成的对基本面信息的过度反应。反观聪明交易者，虽然能对预期回报做出理性反映，但这种反映因受到自身财富的限制而并不充分。与传统金融学不同，行为金融学认为人并非完全理性的、会犯各种认知偏差，并从这个角度研究人的行为如何影响资产的预期收益率。如今，行为金融学也已经被广泛接受，不少学者从投资者行为的角度提出了错误定价因子（mispricing factors），构建了多因子模型。

自20世纪70年代以来到今天，学术界在实证资产定价方面的研究已经走过了近半个世纪。除了上述三位代表性的研究之外，更有大量的学者持续地在这个领域做出杰出的贡献，这些非凡的学术成果为因子投资打下了坚实的基础。这些发现可以归结为理论方面和实证方面两大类。理论方面包括各种统计学方法，用来检验异象和因子，或者比较多的因子模型。以后者为例，主流的统计手段包括Gibbons et al.（1989）检验（简称GRS检验）、Huberman and Kandel（1987）的均值—方差张成（mean-variance spanning）检验以及Barillas and Shanken（2018）的贝叶斯方法。此外，随着近几年机器学习的迅猛发展，也有越来越多的学者将其应用到因子模型检验上。本书的第2章将系统而严谨地介绍研究因子投资时用到的各种常见统计手段。在实证方面，学术成果则主要指的是各种因子以及多因子模型。本书的第3章和第4章将分别介绍这些因子和多因子模型，并针对A股市场进行实证分析。和理论推进的按部就班不同，学术界在实证方面挖因子的努力是十分激进和狂热的。正如一枚硬币有正、反两面一样，这些实证结果一方面的确丰富了人们对股票收益率背后真实驱动的理解，但另一方面很多所谓的发现却仅仅是样本内过拟合的产物，对因子投资并无实质的帮助。

1.2.2 研究现状

如今，学术界已经挖出了超过400个因子（包括定价因子和异象因子）。在发表偏差和多重假设检验（multiple hypothesis testing）的不良影响下，绝大多数因子仅仅是数据窥探（data snooping）的产物。数据窥探是统计学术语，指的是学者们在查看数据之后才去提出假设，进行统计检验，以此达到找到能够获得显著超额收益因子的目的。学术界对于挖因子的狂热和浮躁的态度已经引起很多学者的警惕。2011年John Cochrane在美国金融协会主席演讲时用“因子动物园”（factor zoo）

一词来描述当前学术界因子研究的现状（Cochrane 2011），并提出了三个至关重要的问题：（1）哪些因子是独立的？（2）哪些因子是重要的？（3）因子驱动资产价格的原因是什么？

Cochrane（2011）的提问引发了学术界的深刻反思。最近几年，越来越多发表于顶级期刊上的文章致力于回答这些问题。Harvey et al.（2016）研究了316个因子，提出了一种能够利用不同因子之间相关性的全新分析框架，把代表因子显著性的 t -值（ t -statistic）的阈值从2.0提高到了3.0。Chordia et al.（2020）则对标了前文，认为Harvey et al.（2016）考虑的316个因子仅仅是被发表的，因而仅仅是所有被研究过的因子的子集，所以基于它们的分析会低估多重假设检验的影响。该文使用模拟定量刻画了学术界挖因子的过程，并进一步将代表因子显著性的 t -值的阈值提升到了3.4以上^[3]。除此之外，Green et al.（2017）使用Fama–MacBeth回归同时分析了近100个因子，发现真正独立的因子少之又少。Harvey and Liu（2018）以“幸运因子”为题提出了一个基于正交化和自助法的分析框架，从一揽子因子中逐一识别对解释资产预期收益率截面差异有增量贡献的因子。

一个新因子被提出后，随着越来越多人使用，它在发表后样本外的效果就会变差。在这方面，McLean and Pontiff（2016）研究了97个因子在被发表之后的表现，发现因子的收益率比论文中的表现降低了50%以上。类似的，Linnainmaa and Roberts（2018）研究了36个财务因子在样本外的表现，绝大多数因子在样本外的表现令人失望。另外，在学术研究中通常没有对交易费用给予充分的考虑，造成对因子收益率的高估。由于因子投资组合一般都是多、空对冲的组合，因此如果不合理考虑做空限制，也会高估因子的收益。当然可喜的是，学者们意识到了交易费用对因子效果的冲击，并开始在论文中对其加以必要的惩罚。

学术界对因子研究的另一个关注点是使用多变量构建复合异象或因子，用它来选股并获取超额收益。在这方面的代表作包括Piotroski（2000）提出的F-Score以及Mohanram（2005）提出的G-Score。使用多指标选股构建的多、空对冲投资组合能够获得主流多因子模型无法解释的、显著大于零的 α 收益率。本书的第5章将会解读包括上述F-Score、G-Score在内的一些非常有意思的异象。它们可以极大地丰富读者进行因子投资的思路。此外，美国著名的对冲基金AQR Capital Management（以下简称AQR）提出的质量因子也是很好的例子（Asness et al.2019）。就构建符合异象或因子问题，Novy-Marx（2015a）指出，虽然这么做无可厚非，但由于变量数量增多导致过拟合风险加大，因此需要谨慎考虑每个变量对于提升收益率的增量贡献。

从式（1.3）可知，异象的超额收益说明资产的定价中存在错误。近年来，行为金融学得到了长足的发展。与传统金融学不同，它假设人是有限理性的，因此在不确定性下做决策是会出现预期中的偏差以及风险偏好中的偏差。这些偏差对错误定价产生的原因提供了有效的解释，并可以解释很多市场中观察到的异象以及因子。基于这些新的发现，学术界和业界都逐渐同意风险补偿并非因子背后的唯一原因，而人的各种偏差则也是因子成因的必要补充。近年来，越来越多的研究着眼于理解这些偏差以及投资者情绪和收益率之间的关系。这些研究不仅涉及市场整体收益率在时序上的可预测性，也包括不同资产预期收益率在截面上的差异性。

学术界的最后一个关注点可以归结为因子收益率和宏观经济的关系。在不同

的经济环境中，不同因子的表现大相径庭。研究经济环境和因子收益率的关系对于因子择时也很有帮助。在这方面，Claessens and Kose（2018）调研了学术界的大量相关文献，是一篇很好的综述。虽然理想很丰满，但是现实却很骨感，宏观经济和因子收益率之间的关系并没有人们预期的那么强，在这方面还有大量的研究要做。

[1]该方法被学术界称作Fama–MacBeth回归法。

[2]除了有效市场假说和多因子模型外，Eugene Fama对实证资产定价的另一个贡献是事件分析（Fama et al.1969）。此外，Eugene Fama在金融学的其他方面——比如公司金融——也有很多广为流传的成果。为了记录Fama对金融领域的贡献，John Cochrane和Tobias Moskowitz挑选了Fama最具代表性的论文编制了一本论文集，题为*The Fama Portfolio*（Cochrane and Moskowitz 2017）。

[3]Chordia et al.（2020）考虑了两种常见的检验方法，它们的 t -值阈值分别为3.4和3.8。

1.3 因子投资的业界发展

本节梳理因子投资在业界的发展脉络，以下行文又细分为基金管理人和普通投资者两个维度^[4]；而本书的第7章将会详细论述业界在实践因子投资时需要解决的各种问题。

对于主动型管理人来说，因子投资早已成为投资工具箱中的重要选择。一个因子在学术论文中的超额收益可能非常诱人，但是对于管理人来说，如何在现实中实现它是必须要解决的问题。在实际围绕某个因子构建投资组合时，必须要考虑可投资性的约束，成功的因子投资需要注重从理论到实践的每一个细节。对于普通投资者来说，了解每个因子背后的原因，选择适合自己风险偏好的因子，以及使用合适的金融工具（例如Smart Beta ETF）交易这些因子就成为重中之重。

1.3.1 因子投资和管理人

当管理人实施因子投资时，往往可以从以下几个角度考虑：（1）收益预测vs风险管理；（2）资金流入削弱因子收益率；（3）因子择时；（4）各类因子大行其道，如何鉴别能够获得超额收益的管理人；（5）创新的价值。

1. 收益预测vs风险管理

不同的管理人使用多因子模型的角度可能截然不同；有的管理人从截面的角度使用因子获取超额收益；而有的管理人从时序的角度使用多因子模型计算投资组合中股票之间的协方差矩阵，即风险管理；还有管理人使用多因子模型来同时进行收益预测和风险管理，并在这个基础上进行投资组合的优化。

2. 资金流入削弱因子收益率

当越来越多的资金涌入因子投资时就会造成因子拥挤（factor crowding）。而使用相似的指标排序、接近的调仓频率的因子投资则加剧了这种负面影响。任何投资策略想要持续赚钱都是利用了市场在某方面的非有效性。当使用的人越来越多，市场在这方面就会变得更有效，从而降低了该因子的预期收益率。

因子拥挤也会造成流动性冲击。一旦市场中发生冲击风格因子的事件后，持有相似头寸的管理人会竞相卖出手中的股票，由此产生的流动性危机会造成很大的亏损。2007年8月，很多美股量化基金经理在短时间内清理了相似的头寸，巨大的抛压对流动性造成了巨大的打击，使得这些股票的价格在短时间内大幅下降，导致市场上一些非常优秀的量化对冲基金在短时间内录得了巨大的亏损。如何定量计算因子拥挤度，并规避因子拥挤带来的负面影响就是管理人必须面对和解决

的问题。

3. 因子择时

因子择时是个大课题，其背后的动机简单而清晰：因子的表现都有周期性，有时能够获得超额收益，有时却持续亏损。因此如果能成功择时，将会极大提升因子投资的收益率。关于这个话题，Bender et al. (2018) 客观地综述了使用不同预测指标在不同时间尺度下对不同风格因子择时的效果。这些预测指标包括因子估值、因子动量、投资者情绪以及各类宏观经济指标。

在因子择时的众多方法中，最重要的两类方法是按因子估值（factor valuation）和按因子动量（factor momentum）择时两大立场。这两个立场来自美国两个著名且风格鲜明的对冲基金。前者的拥趸是来自美国西海岸的Research Affiliates，而后的支持者是来自美国东海岸的AQR。他们两家均认为自己的方法更好，认为对方的存在逻辑缺陷，并相互发论文，用实证分析证明对方的错误，颇为精彩。

4. 区分 α 和 β 收益

随着因子投资的盛行，一个新的挑战油然而生：主动管理人获取的超额收益到底来自异象，还是因为配置了某些因子？从多因子模型的定义以及式（1.3）中的数学符号可知，某个资产的高收益既可能来自高 α 也可能来自选择了收益率高的因子并且以高暴露（即高 β ）配置在这些因子上，因此业界（和学术界）通常把这两部分收益称为 α 和 β 收益。Bender et al. (2014) 指出，管理人获得的超额收益中，80%的部分可以由因子暴露解释，只有剩下20%才取决于获得 α 的能力。

由于不同风格因子的表现受到宏观经济的影响，主动配置因子 β （涉及因子择时和风险控制）从而获得超额收益也彰显管理人的能力。在一项最新的研究中，Duanmu et al. (2018) 比较了主动 β 和主动 α 管理人，并指出从长期来看，最顶尖的主动 β 型管理人能够获得比顶尖的主动 α 型管理人更优异的收益。

5. 创新的价值

对于任何研究领域，创新的作用都至关重要。对因子投资来说，创新意味着使用新的数据或者算法。随着大数据的普及，越来越多的管理人开始搜寻新的能够成为收益源的数据（比如舆情数据、专利数据、新闻数据等）。此外，机器学习算法也被广泛地应用于因子投资实践中，包括使用机器学习算法预测基本面财务指标，或挖掘因子和收益率之间的非线性关系。

考虑到数据和技术的革新对金融领域的影响，2019年 *Institutional Portfolio Research Journals* (IPR Journals) 为旗下的期刊系列增加了一位最新成员——*Journal of Financial Data Science*，旨在指导金融领域的实践者正确使用与日俱增的数据和日新月异的技术。当然，创新的机遇总是伴随着巨大的挑战。虽然机器学习在金融领域以及因子投资领域拥有乐观的前景，但由于金融数据信噪比很低，因此需要时刻保持冷静。Lopez de Prado (2018) 详细论述了在投资领域实践机器学习算法时可能遇到的各种坑，填补了理论和实践之间的空白。

1.3.2 因子投资和投资者

再来看看投资者视角下的因子投资。自20世纪80年代末以来，业界开始把学术界的研究成果很好地落地，形成了许多风格因子指数。举例来说，1987年罗素（Russell Investments）推出了最早的两个风格指数，分别对标价值和成长因子。在这之后，标准普尔指数公司（S&P）^[2]、明晟（MSCI）等机构也推出了它们各自的风格指数。最初，推出这些指数的目的是评估主动基金经理人的业绩。然而不久之后，复制这些指数的基金便应运而生，而这其中的代表则是大名鼎鼎的先锋集团（Vanguard）。1992年，先锋推出了第一支价值指数基金和第一支成长指数基金。这些指数基金一经推出便受到了追捧。人们把它们视作主动型价值投资的低成本替代品。鉴于价值因子取得的巨大成功，业界也开始把关注的重点转移到学术界发现的其他因子，并构造出了一系列Smart Beta ETF产品^[3]，造福了普通投资者。

2018年，《经济学人》估算因子ETFs^[4]的总规模超过6500亿美元。大量低成本的ETFs产品让普通投资者也能享受风格因子带来的收益——前提是能够选择适当的ETFs。然而，这并不容易。在一项针对大学本科生和研究生的实验中，Choi et al.（2009）让参与者从众多追踪标准普尔500指数（简称标普500指数，英文S&P 500）^[5]的被动型ETFs中挑出他们认为最好的。实验中为参与者提供了ETFs的费率以及它们在不同时期的收益表现情况。作为具备金融知识的理性投资者来说，鉴于这些ETFs均是被动地追踪标普500指数，因此应该把费用的高低作为唯一的标准来选择。然而，参与者被它们在不同时期的收益率差异所迷惑，绝大多数投资者完全忽视了费用的差异，而选择了收益率最高的ETF。另外，就算是排除了投资者本身的非理性行为，琳琅满目的ETF产品也足以让人迷失。以价值因子为例，表1.1总结了美股上主流的价值因子ETFs。

针对某个因子的不同ETFs，一般可以从因子敞口、费用、流动性等角度进行比较。即便表1.1中所列出ETFs都是为了捕捉价值因子的超额收益，但它们的风格差异却十分鲜明。这无疑对投资者的选择带来了巨大的挑战。进一步从表1.1列举的ETFs中挑选那些以标准普尔500指数成分股为标的构建的价值因子ETFs，表1.2统计了它们自2013/04/30至2020/03/31的表现。由于最近几年价值投资不太好使，这些因子均跑输标普500指数（SPX）本身，且它们之间也有明显的差异。从夏普比率的角度来说，最好的要数VTV，而垫底的则是RPV。不幸的是，对于投资者来说，事前就从琳琅满目的ETFs中挑出最好的并不容易。

Smart Beta ETFs的流行给投资者提供了越来越丰富的工具，也对投资者提出了更高的要求。在投资ETFs时，投资者需要首先明确自己的目标，比如是分散化风险还是获得相对于市场的超额收益。在明确目标之后，需要理解每个风格因子背后的逻辑和它代表的风险。唯有这样，才有可能享受这些标的带来的更高性价比的风险收益。

表1.1 美股上主流价值因子ETFs

代码	名称	成立日期	总规模 (亿美元)	持仓数量	管理费 (%)
VTV	先锋价值 ^①	2004/01/26	497	330	0.04
VBR	先锋小盘价值	2004/01/26	131	854	0.07
VOE	先锋中盘价值	2006/08/17	92	199	0.07
MGV	先锋超大盘价值	2007/12/17	25	154	0.07
IWD	iShares 罗素 1000 价值 ^②	2000/05/22	319.3	766	0.19
IVE	iShares 标普 500 价值	2000/05/22	145.6	395	0.18
IWS	iShares 罗素中盘价值	2001/07/17	83.7	632	0.24
IWN	iShares 罗素 2000 价值	2000/07/24	64.7	1393	0.24
IUSV	iShares 核心标普美国价值	2000/07/24	54.4	692	0.04
VLUE	iShares Edge MSCI 美国价值	2013/04/16	30.7	148	0.15
PRF	景顺 FTSE RAFI 美国 1000 ^③	2005/12/19	35.7	1010	0.29
RPV	景顺标普 500 纯价值	2006/03/01	4.5	108	0.35
SCHV	嘉信美国大盘价值 ^④	2009/12/11	54.8	415	0.04

①所有先锋ETFs的总规模和持仓数量为截至2020/02/29的数据。

②所有贝莱德（BlackRock）旗下的iShares ETFs的总规模为截至2020/04/06的数据，持仓数量为截至2020/04/03的数据。

③所有景顺ETFs的总规模为截至2020/04/06的数据，持仓数量为截至2020/04/03的数据。

④总规模和持仓数量为截至2020/04/06的数据。

表1.2 不同价值因子ETFs的风险收益特征

	SPX	VLUE	VTV	IVE	VOE	SCHV	RPV
年化收益率 (%)	7.35	3.16	4.30	3.77	2.71	3.03	0.41
夏普比率 (年化)	0.61	0.28	0.38	0.33	0.25	0.29	0.13
月度最大回撤 (%)	-22.28	-30.37	-26.37	-26.70	-33.01	-27.20	-42.41

数据来源：Wind。表现区间2013/04/30至2020/03/31。

[1]Cerniglia and Fabozzi（2018）一文从学术界、管理人和投资者三个视角探讨了因子投资，非常值得一读。本小节的部分内容参考了该文。

[2]在2012年7月，标准普尔指数公司与道琼斯指数公司合并，成为标普道琼斯指数有限公司（S&P Dow Jones Indices LLC）。在本书后文介绍相关因子指数的提供商时，将使用标普道琼斯指数公司。

[3]ETF是Exchange Traded Fund的首字母缩写，它的中文含义是“交易型开放式指数证券投资基金”，简称“交易型开放式指数基金”。ETF是一种跟踪标的指数变化，且在证券交易所上市交易的基金。投资人可以如买卖股票那么简单地买卖标的指数的ETF，获得与该指数基本相同的回报率。

[4]ETFs是ETF的复数形式。文中的表述遵循了英文语法惯例。当使用ETFs时，其含义为多个交易型开放式指数基金产品。

[5]标普500指数由标普道琼斯指数有限公司开发并维护。该指数包含了美股市场中市值最大的500支股票，被广泛认为是衡量美国大盘股市场的最好指标。