前 言

21 世纪以来，我国资本市场一直处在发展和变革之中。在与世界其他资本市场接轨的过程当中，我们需要不断地了解和学习国外相应的交易规则、产品、交易策略等知识。只有这样，我们才能适应不断变化的市场环境，并在未来的竞争中获得一席之地。

算法交易是近十年来国外流行起来的一种交易方式，主要是利用计算机系统通过一些规则化或智能化的交易策略进行自动的交易和投资活动。它的出现和流行对于资本市场的结构和运作方式都产生了很大的影响。套利交易策略主要出现在机构投资者和对冲基金当中，通过发现两种本质相同或类似的证券之间的价格差异来获取利润。本书主要介绍算法交易和一些套利交易的策略，以便于读者对相关方面的内容进行阅读和学习。

在本书的第一部分，我们回顾了投资学一些相关的基本内容。其中，前两章介绍了证券投资的收益和风险等特征，以及马可维茨的最优资产配置模型。第 3 章则介绍了股票投资分析当中常用的资本资产定价模型(CAPM)、套利定价模型(APT)，以及因素模型。然后，第 4、5 章分别讲到了金融证券估值模型、基本面分析和技术分析。

第二部分则着重于介绍算法交易的相关内容。第 6 章给出了算法交易的整体概况，其中包括算法交易的历史和概况，算法交易策略的简介，以及算法交易的优势和可能产生的影响。第 7 章介绍了算法交易相关的计算机和交易系统。算法交易出现的主要原因就在于买方和卖方机构追求更低的交易成本、更高的流动性，以及自动化的需要。第 8 到第 10 章从交易成本开始介绍了算法交易开发的目标、模型以及评价方式。其中，第 8 章全面分析了交易成

本的产生、分类，以及原因，然后介绍了交易成本管理的目标和方法。第 9 章从一个最优交易的模型出发，介绍 Implementation Shortfall、VWAP 等等的一些算法交易策略。第 10 章则探讨了算法交易策略的选择和交易后分析等相关内容。

本书的最后一部分介绍了一些套利交易策略。套利交易的思想是把握本质相同或者近似的两种证券之间的价格差异，当两种证券的价格出现偏离的时候，投资者可进行交易，并从中获利。“一价定律”或者“近似一价定律”是套利交易的理论基础。本书的第 11 到第 14 章深人详细地介绍了配对交易、可转债套利、并购套利和股指期货套利等几种具体的套利交易策略。随着我国市场上融资融券和股指期货业务的推出，可以预见，这些交易策略在未来几年将会逐渐出现并活跃在 A 股市场当中。第 15 章则以长期资本管理公司为例，总结了套利交易和对冲基金操作当中的一些经验教训。

参加本书写作的有：方意（第一章、第五章），郑玉宝（第二章、第四章第一、二节），陈正元（第三章），张雪莲（第四章第三、四节），武文超（第六章、第七章前三节），魏子哲（第九章，第七章后三节），李燕燕（第八章第一、二节），杨锐（第十章、第八章第三节），翟光宇（第十一章，第十四章），王昊彦（第十二章），赵胜民（第十三章，第十五章），全书由赵胜民最终定稿。

编者

2010 年 7 月

目 录

前 言

第一章
投资的收益与风险

1.1 概述

投资者通过把剩余的钱投资于金融市场上，以使其未来的消费最大化。然而在投资于金融市场之前，我们要对收益与风险有一个很好的了解。收益与风险评估是投资者设计并应用投资组合时所必须考虑的。对收益与风险的论述、计算与评估将贯穿本章。在需要时，我们还将应用这些概念的特殊形式。但是，在开始分析各种各样的证券之前，我们有必要了解收益与风险的基础知识。因此，我们必须思考这些概念并学会如何分析与计算。已实现的收益，顾名思义，就是过去的收益，或是已经或本应取得的收益。由于已实现的收益已经发生，因此可以运用恰当的数据进行计算。反之，预期的收益则是指投资者预期将在未来的某个时期内从其资产中取得的收益。显然，预期收益具有不确定性，尽管我们经常用已实现的收益（历史收益）来预测预期收益，但是二者有着本质区别。

1.2 收益

投资者的目标是在一定的约束与风险条件下，最大化预期收益。对于投资者而言，取得投资收益至关重要，收益是投资活动的全部意义之所在。计算已实现的收益是投资者评价其投资好坏或是其委托的投资管理者投资

好坏所必需的。并且，对历史收益的计算也会在评估预期收益中发挥很大的作用。

一般投资的收益由两个要素组成：

(1)收入。即在投资过程中，从投资中获得的定期的现金流。如债券的利息、股票的股利。收入的特点是发行人向资产持有人支付的现金。

(2)资本增值(损失)。另一个要素是资产价格的增值(或减值)，也即价格变动。在长期情况下，它是资产的购买价与可能的或已实现的售价之间的差额；在短期情况下，它是销售价格与该短期收盘价格之间的差额。

记 TR 为总收益，I 为收入，资本增值为 ΔP，则

$$TR=I+\Delta P \tag{1.1}$$

其中 $I,\Delta P$ 均可正可负。

【例 1.1】某投资者以 950 美元购买 10 年期的债券面值为 1 000 美元的 IBM 债券，其票面利率为 8%，半年付息一次。3 个月后他收到一次利息，且此时的债券价格为 1 050 美元。则可知其总收益为 1 000×4%＋(1 050－950)＝140 美元。

如果投资者从 t 到 $t+1$ 时刻持有一份资产，那么他所获得的简单收益率为：$R_t=\frac{P_{t+1}}{P_t}-1$

其中 P_t 为该资产在 t 时刻的价格。进一步地，投资者从 1 时刻开始，持有资产到 $t+1$ 时刻所获得的总收益率为：

$$\begin{aligned}R&=\frac{P_{t+1}}{P_1}-1=\frac{P_2}{P_1}\times\frac{P_3}{P_2}\times\cdots\times\frac{P_{t+1}}{P_t}-1\\&=(1+R_1)\times(1+R_2)\times\cdots\times(1+R_t)-1\end{aligned}$$

1.2.1 算术平均值

对大多数人来说，最好的统计方法就是计算算术平均值。因此，在提及平均收益时，除非特别指明，通常应认为是指算术平均值。算术平均值通常用 $\overline{R}$ 表示，是一系列数据的平均值。计算如下：

$$\overline{R}=\frac{\sum_{i=1}^{n}R_i}{n} \tag{1.2}$$

某投资者投资金融市场5年。5年收益率分别为20%,15%,-10%,-5%,30%,则他在这5年的平均收益率为$\bar{R}=\frac{0.2+0.15-0.1-0.05+0.3}{5}=10\%$。

1.2.2 几何平均值

算术平均收益是对特定时期内(如10年期的)收益数据分布的中心趋势的大致预测。然而,在一段时间内,收益以复合增长率变化,用算术平均值测算这些变化会产生误导。如某投资者投资某股票(假设无股利),买入时价格为10美元,一年后股票价格为20美元,再过一年后,股票价格为10美元。则收益率的算术平均值为[100%+(-50%)]/2=25%。显然这个不太合理,因为投资者最终一分钱没有赚,由开始的10美元变为最终的10美元,收益率为0,则平均收益率也为0。而几何平均值可以解决这个问题。

一组收益率的几何平均值由下式定义:

$$G=\left[\prod_{i=1}^{n}(1+R_i)\right]^{\frac{1}{n}}-1 \tag{1.3}$$

其中R_i为投资期间的每年收益率,$\prod_{i=1}^{n}$为n个数相乘。显然,上述的投资者的投资收益率的几何平均值为:

$$G=[(1+100\%)\times(1-50\%)]^{\frac{1}{2}}-1=0$$

几何收益率测算了一定时期内的复合增长率。

1.2.3 算术平均值与几何平均值的比较

一般而言,几何平均值都比算术平均值小,除非给定的值是相同的。两者的差异取决于给定区间的标准差;标准差越大,两者的差异就会越大。[$(a+b)/2\geqslant\sqrt{a\times b}$,当且仅当$a$,$b$相等时取等号]。

当我们需要描述金融资产的收益时,什么时候使用算术平均值,什么时候使用几何平均值具体应取决于投资者的目标:

· 在计算单个期间的平均业绩时,算术平均值较好。在预测下一期间的预期收益时,它是最好的预测指标。

· 在计算过去的总财富值变化时(多个期间),几何平均值较为适用。它

是一个后向视角，计量过去特定期间财富增长的现金复合收益率。

1.2.4 通货膨胀调整收益

前面讨论的所有收益都是名义收益，或称之为货币收益。通常，我们日常生活中见到的收益率都是以名义利率表示的收益率，如金融机构支付的利率，固定收益证券利率等。但是我们需要考虑收益的实际购买力。为达到这一目标，我们必须分析实际收益，也就是考虑通货膨胀的作用。

名义收益与实际收益的关系为：$R \approx r + \pi^e$，其中 R 为名义收益率，r 为实际收益率，π^e 为预期通货膨胀。

精确的表达式为：

$$(1+R)=(1+r)(1+\pi^e) \tag{1.4}$$

2004 年标准普尔 500 指数的总收益率为 10.8742%（假设每月的股息都用于再投资）。通胀率为 3.2556%。因此，在 2004 年用标准普尔 500 指数计量大盘股的实际总收益为：

$$r=(1+R)/(1+\pi^e)-1=(1+10.8742\%)/(1+3.25565\%)-1=7.378\%$$

1.2.5 期望收益

期望收益显然是衡量未来收益的一个指标，是投资者结合自己现在所获得的信息对未来收益的一个估计。上面我们讨论了投资者投资于金融资产多年后可以获得的平均收益。已实现收益非常重要是因为投资者需要知道他们的投资组合表现如何。已实现收益还能为投资者建立对未来收益的预期做出一个非常准确的估计。

投资者的投资收益是未知的，我们必须进行估计。未来的收益是一种预期的收益，可能实现也可能不能实现。一个投资者可能预期某证券未来一年的收益为 0.10，但是实际上这只是一个点估计。当投资者作出投资决策的时候，风险或者不利事件发生的可能性也就随之发生。投资者常对预期收益过度乐观。我们可以用“随机变量”这个词来描绘证券的预期收益，它是一个随机波动的不确定值。一般来说，随机变量服从一定的分布。投资者需要牢记尽管他们期望证券的收益为 10%，但是这只是整个可能范围内的一个点估计。由于投资者面临的未来是不确定的，因此各种结果都有可能发生。考虑

这些结果及其概率总的来说是一个收益的概率分布，包含可能发生的结果和这些结果发生的概率。概率分布中的概率总和必须等于1，因为它们必须完整地描述所有可能发生的结果。

概率分布可以是离散的，也可以是连续的。在离散的概率分布中，每个可能的结果都被指定一个概率。在连续的概率分布中，如图1.1，存在无限多种可能性。因为概率用图1.1中曲线的面积来衡量。

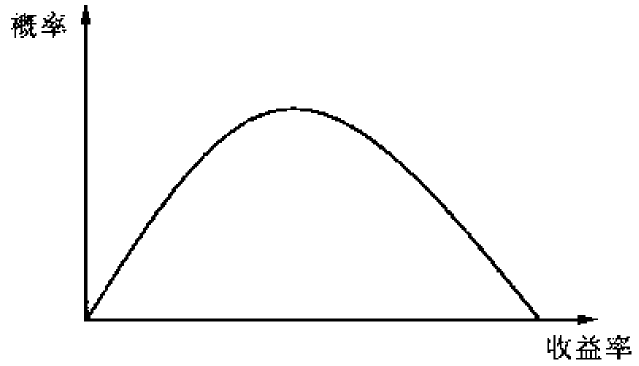


图 1.1

为了确定某个特定概率分布的一个可能性最大的结果，我们必须计算期望值，期望值是所有结果的加权平均，权重就是该种结果发生的概率。对于离散的随机变量，期望定义为

$$E[R]=\sum_{i=1}^{m}R_i p_i$$

其中 R 为随机变量，R_i 为随机变量 R 在第 i 个可能性发生时的取值，p_i 为第 i 个可能性发生时的概率。m 为出现所有结果可能性的个数。

1.2.6 资产组合的收益率

一个资产组合的简单收益率可以表示为它所包含的资产的简单收益率的加权平均，其中，权重就是每种资产在投资组合中所占有价值的比例。例如，对于投资组合 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^T$，其中 x_i 为资产 i 在投资组合总价值中所占有的比例。那么投资组合的总收益率为 $R_t=\sum_{i=1}^{n}x_iR_{it}$。

1.3 风险

对投资者而言，股票的平均收益仅仅说明了数据的中间值是多少，而并没有反映数据的离散程度。而这个数据的离散程度对于投资者来说，重要性不言而喻。这个可以用一个关于数学家的古老故事来说明。数学家认为平均值本身就可以完整地描述一个过程，结果他淹死于平均深度为 2 英寸的溪流中。

风险是度量投资的实际收益与预期收益之间差异的指标。风险越大，实际收益偏离预期收益的可能性越大，幅度也会越大。特别是，大多数投资者均不希望其实际收益小于预期收益。一般来说，只有当预期收益足以补偿其风险时，投资者才愿意购买该资产。但是投资者必须明白其对收益的预期可能不会实现，此时其实际收益将与预期收益有所出入。实际上，证券的实际收益会呈现出相当大的波动性，可能高于预期值，也可能低于预期值，甚至遭受无法承受的损失。

1.3.1 风险的来源

证券的风险来源大概可以分为两种：不可分散的风险（或者系统风险）和可分散风险（或者证券特有的风险）。

1. 不可分散风险。不可分散风险是直接与整个市场或经济的走势相关的证券总收益的波动。实际上所有证券（不论股票还是债券）都有某种程度的系统风险。它又可以具体地分为利率风险、市场风险、通胀风险、流动性风险、汇率风险、国家风险等。

（1）利率风险是由于利率水平的变化所导致的证券收益的波动。利率水平与收益呈负相关关系。虽然利率风险对债券的影响比对普通股更为直接，但其对两者的影响都是存在的，因而绝大多数投资者都应对其重点关注。

（2）市场风险是由于整个市场情况（整个股票市场）的变化所引起的收益波动而带来的风险。所有证券都面临市场风险，但其主要影响的还是普通股。

（3）通胀风险是影响所有证券的另一个因素。由于通胀的不确定性，即使名义收益是稳定的，实际收益也会存在风险。因为借款人需要提高利率来补

偿购买力的损失，所以通胀率上升时利率也会随之上升，因此，通胀风险与利率风险相关。

（4）流动性风险是投资者变现时的成本。流动性定义为投资者在不遭受损失的情况下迅速变现的能力。一般来说，市场上的活跃证券的存量越大，流动性越大。比如，国库券几乎没有流动性风险，而小型 OTC 股票由于存量小，显然面临着较大的流动性风险。而且，证券越简单，越容易被投资者了解，其流动性越高。那种定价特别复杂的证券，不可能期望其流动性很高，例如奇异期权。评价一项投资的流动性风险时，投资者必须考虑关于流动性的两个问题：①将该项投资转换为现金将需多长时间？②所获得价格的确定性如何？

（5）汇率风险可定义为汇率波动所引起的证券收益的波动。汇率风险也被称为货币风险。例如美国投资者购买了以欧元计价的德国股票，由于是美国投资者，他必须将该股票的收益兑为美元。如果汇率向不利于投资者的方向变化，由汇率变化所导致的损失部分地或完全地抵消原来的收益。

（6）国家风险，也称为政治风险，是当今投资者所面临的一种重要风险。随着越来越多的投资者直接地或间接地从事国际投资，一国政治与经济的稳定和波动，必须成为重要的考虑因素。例如投资者认为美国的国家风险最低，然后将其作为衡量其他国家的标准，并因此确定其他国家的国家风险。

2. 可分散风险。上面考虑的都是系统性风险。而可分散的风险，则是随着更多的证券加入到投资组合中，该组合的风险逐渐降低，而这降低的风险就是可分散的风险。它是公司特定的风险，或者是与某个公司相关的特定风险。它包括商业风险、经营风险、财务风险等。

（1）商业风险是在某个特定行业和地区进行商业活动所遇到的风险。例如，美国电报电话公司是传统的家用电话供应商，随着通信业日新月异的变化，该公司面临着极大的挑战。

（2）经营风险是由企业经营性质引起的收入流的不确定性。该企业收入流的不确定性越强，其投资者收入流的不确定性就越强。例如，一家食品零售公司在投资期内一般会呈现非常稳定的销售和收益增长，并且与汽车行业中的一家企业相比应具有低的经营风险；而汽车行业中的企业销售和收益在经济周期中显著波动，这隐含着高的经营风险。

（3）财务风险与公司使用的债务融资相关。在其他条件相同时，债务融资购买资产比重越大，收益的波动性也会越大。

1.3.2 风险的衡量

风险与可能出现的离散程度相关。直觉上，衡量结果偏离均值的一个直接方法是考虑两者的差值，也就是计算 $R_i-\overline{R}$。对每一个结果计算出这一差值后，可以通过计算这些差值的平均值来得到一个整体性的衡量指标。尽管这些方法在直觉上是合理的，它仍有一个问题。这些差值中有些为正数，有些为负数，它们彼此会抵消。彼此抵消的结果是，一个极不稳定的离差平均值可能比稳定收益的离差平均值小。如：1，－1，1，－1 用上述算法得出其离差平均值为 0，而这显然不合理。

解决这一问题有两种方法。第一种是对均值的差值取绝对值，也即 $\frac{1}{n}\sum_{i=1}^{n}|R_i-\overline{R}|$，去掉负号对平均离差的影响。但是在数学上对绝对值不是很好处理(要分区间，且其可导性也不是很好)，人们往往用第二种方法，即收益率的 2 阶矩来表示，也即离差平方和的均值 $\frac{1}{n}\sum_{i=1}^{n}(R_i-\overline{R})^2$。其也称为方差。对上面的公式可解释为资产收益率的每种收益率是等可能性的。如果不是等概率的，则要用到 $E(R_i-E(R_i))^2$，其中 E 表示期望。

在有些情况下，应用另一种描述离散程度的指标——标准差是方便的。标准差是方差的平方根，它以 σ 来表示。通常来说，由于标准差和期望收益是等量纲的，所以更常用。

描述离散程度还有一些其他方法。我们已经提到过一种，即离差绝对值的平均数。有一种方法只考虑均值以下数值的偏离。这一方法认为，高于均值的收益是理想的，而困扰投资者的收益是低于均值的收益。体现这一思路的值表示低于均值收益的离差平方和的均值。例如，在表 1.1 中，资产 1 唯一低于均值的收益是 3。3 与均值之差为 6，其平方为 36，其他两种收益均不低于均值，因而它们低于均值的离差为 0。(0)＋(0)＋(36)的均值为 12。这一指标为半方差(semivariance)。

半方差度量的是，相对于由期望收益决定的基准的下行风险。它是众多衡量下行风险的可能方法之一。更一般地讲，我们可以相对其他基准来考虑收益，这包括无风险收益基准和 0 收益基准。这些一般化的方法称为下偏矩(lower partial moment)。另一种衡量下行风险的方法即所谓的在险价值(VaR，value at risk)，银行普遍使用这一方法来衡量不利条件下的风险暴露。

表 1.1 不同投资的收益

市场情况	收益			
	资产 1	资产 2	资产 3	资产 4
好	15	16	1	16
一般	9	10	10	10
差	3	4	19	4
收益均值	9	10	10	10
方差	24	24	54	24
标准差	4.9	4.9	7.35	4.9

①每种资产的不同收益被假定为等可能性发生。

从直觉意义上讲，这些下行风险的替代性衡量方法具有合理性，并且一些投资组合理论也应用这些方法获得了进展，但是，当我们从单个资产转移到投资组合时，这些方法应用起来就非常困难。如果收益分布是对称的，均值方差空间内的投资组合排序，与均值半方差空间的或其他已经讨论的下行风险空间内的投资组合排序是一致的。因此在绝大多数投资组合文献中，方差或标准差被用来衡量离散程度。

1.3.3 风险溢价

风险溢价是指投资者由于承担了更多的风险而期望或已经收到的额外收益。它用来衡量各种风险所获得的补偿。风险是由系统风险和非系统风险构成的，风险溢价也即对投资者承担风险的回报指的是对系统风险的回报。因为随着投资组合中证券数目的增多，非系统性风险被抵消了，所以非系统风险在金融经济学意义上是不能获得回报的。在开放条件下，投资组合的系统风险从国家范围内扩展到了世界范围，世界范围的系统风险显然低于国家范围的系统风险，因此风险溢价会低于不开放条件下的风险溢价。

第二章

资产配置与均值方差模型

上一章，我们考虑了历史的收益率和方差，它为我们理解现代投资组合理论(MPT)打下了基础。我们知道在投资决策之前，首先要做的就是我们应当把我们的资金在现有的市场资产类型中做出合理的配置。我们首先考虑持有投资者最熟悉的两种主要资产类别的情形：所有投资者都持有股票或债券或者两者相结合的投资组合。这是因为债券是两种资产中相对安全的一种，许多投资者至少有一部分资金配置在债券上。根据历史数据，债券的收益低于股票，但是风险也相对低得多。首先考虑一个投资者由于其风险承受能力，他只能持有一个债券投资组合。该投资者明白，如果仅持有债券组合，其收益率将低于持有股票组合，但是他也清楚其风险也相对较低，因此是投资者的风险承受能力决定了这样的决策。然而，考虑一个持有100%债券组合20年的投资者，该投资者的复合年收益率为11.05%，风险水平为11.15%。但是现在我们注意一个投资组合含有50%的股票和50%的债券，其标准差相对较低，只有10.85%，其年收益率更高为11.9%。显然，股票和债券的资产配置确实为投资者带来了好处。除非你预期未来将与过去截然不同，否则你不能说只持有债券是合理的。

上面的例子引出了一个问题，就是我们该把多少的资金投资于债券，而又把多少的资金投资于股票呢？在解答之前，我们先看一下，我们可以投资的资产类别有些什么。

2.1 一些主要的资产类别

(1)股票。股票是最重要的资产类别，我们经常提到的投资通常就是指投

资于股票市场。

(2)国际投资。十多年前，个人投资者能够获得的投资工具只有在本国证券市场出售的股票和债券，而现在投资者可以投资于在世界范围内发行的各种证券。从历史来看，国际分散化可以提供一些风险降低的收益，因为各国之间资产收益的正相关性很低。许多研究表明，这种较低的相关性使得投资者更多地将国外投资作为一种资产类别。

(3)债券。作为可以加入投资组合的资产类别，债券是一个比较常见的选择。过去，资产分配就是指将投资组合的资金分配给股票、债券和国库券。

(4)房地产。房地产是风险分散化的另一个常用选择。房地产通常被称为与股票的相关性很小或为0的资产类别。投资者可以轻松地通过购买房地产投资信托(REITs)来持有房地产。

2.2 投资组合的一般特性

2.2.1 投资组合的期望收益率

投资组合的收益是组成资产收益的加权平均。每种收益的权重就是该资产在投资组合中的比重。如果 $E(R_j)$ 代表投资组合的第 j 种资产的期望收益，X_j 是投资者投资于第 j 种资产的比重，n 是资产种类，则有 $E(R_P)=\sum_{j=1}^{n}E(R_j)X_j$（$R_j$ 为随机变量，本章中无特别说明，均指随机变量）。

【例 2.1】表 2.1 给出了资产组合收益率的计算过程。

表 2.1

类别 收益与比重	资产 1	资产 2	资产 3
期望收益	10%	20%	30%
比　重	30%	40%	30%
资产组合期望收益率为 30%×10%+40%×20%+30%×30%=20%			

2.2.2 投资组合的方差

投资组合的方差的决定比投资组合的期望收益的决定要复杂一些。我们从两个资产构成的组合开始讨论。投资组合P的方差定义为σ_P^2，它是投资组合收益与期望收益差平方的期望值，即$\sigma_P^2=E(R_P-E(R_P))^2$，将投资组合的期望收益代入上式：

$$
\begin{aligned}
\sigma_P^2 &= E(R_P-E(R_P))^2 \\
&= E[X_1R_1+X_2R_2-(X_1E(R_1)+X_2E(R_2))]^2 \\
&= E[X_1(R_1-E(R_1))+X_2(R_2-E(R_2))]^2 \\
&= E[X_1^2(R_1-E(R_1))^2+X_2^2(R_2-E(R_2))^2+2X_1X_2(R_1-E(R_1)(R_2-E(R_2)] \\
&= X_1^2E(R_1-E(R_1))^2+X_2^2E(R_2-E(R_2))^2+2X_1X_2E[(R_1-E(R_1)(R_2- \\
&\quad E(R_2)] \\
&= X_1^2\sigma_1^2+X_2^2\sigma_2^2+2X_1X_2E[(R_1-E(R_1)(R_2-E(R_2)]
\end{aligned}
$$

（其中σ_1^2，σ_2^2为相应资产的方差）

$E[(R_1-E(R_1)(R_2-E(R_2)]$有一个特别的名字，叫协方差，它被定义为$\sigma_{12}$。用$\sigma_{12}$代替$E[(R_1-E(R_1)(R_2-E(R_2)]$，得$\sigma_P^2=X_1^2\sigma_1^2+X_2^2\sigma_2^2+2X_1X_2\sigma_{12}$。

协方差是考察多种资产的收益如何共同变动的指标。当两种资产收益变化方向相同时，它们的协方差为正，则$\sigma_{12}>0$；当两种资产收益变化方向相反时，则$\sigma_{12}<0$，两资产负相关；当两者的收益变化方向没有任何规律可循时，$\sigma_{12}=0$。

由于协方差和期望值量纲不一致，所以往往将协方差标准化。将协方差除以两种资产各自的标准差的乘积，就得到具有协方差相同性质的变量，且该变量处于-1到1之间（查看任何一本概率统计的书可知）。这一指标称为相关系数。以ρ_{ij}表示证券i和j的相关性，相关系数可定义为$\rho_{ij}=\frac{\sigma_{ij}}{\sigma_i\sigma_j}$（其中$\sigma_i$，$\sigma_j$分别为资产$i$，$j$的标准差）。

【例2.2】考虑股票A和B实现的总收益。两只股票的统计数据如表2.2。

表2.2

	股票A	股票B
期望值(%)	15.16	12.12
标准差(%)	25.97	21.58
相关系数	0.29	

假设我们投资于每只股票的资金相同，因此权重分别为0.5，0.5。

$$\begin{aligned}\sigma_P &= [X_1^2\sigma_1^2 + X_2^2\sigma_2^2 + 2X_1X_2\sigma_{12}]^{\frac{1}{2}} \\ &= [(0.5)^2\times(21.58)^2+(0.5)^2\times(25.97)^2+2\times0.5\times0.5\times0.29\times21.58\times \\ &\quad 25.97]^{\frac{1}{2}} \\ &= 19.14\end{aligned}$$

投资组合的方差受两只股票的相关性影响很大。在其他因素不变的情况下，投资组合风险随着相关系数从1.0的下降而降低。

【例2.3】 假设我们有一些关于两个公司A，B的数据，估计的期望收益率分别为26.3%和11.6%，标准差分别为37.3%和23.3%。它们收益的相关系数为0.15。为了显示相关系数变动的影响，我们假设权重为0.5，即每个股票的投资额为50%。本例中的数据总结为：

$\sigma_A=0.373$，$\sigma_B=0.233$，$X_A=X_B=0.5$，$\rho_{AB}=0.15$

使用这些数据，该投资组合的标准差或风险为σ_P，即$\sigma_P=[X_A^2\sigma_A^2+X_B^2\sigma_B^2+2X_AX_B\sigma_{AB}]^{\frac{1}{2}}=0.0435$。

投资组合的风险显然依赖于第三项的值，而第三项又取决于两只股票收益的相关系数。为了衡量相关性的潜在影响，考虑$\rho_{AB}=1, 0.5, 0.15, 0, -0.5, -1.0$。计算上述情况下投资组合的风险如下：

如果$\rho_{AB}=1$，$\sigma_P=0.303$；如果$\rho_{AB}=0.5$，$\sigma_P=0.265$；如果$\rho_{AB}=0.15$，$\sigma_P=0.234$。

如果$\rho_{AB}=0$，$\sigma_P=0.220$；如果$\rho_{AB}=-0.5$，$\sigma_P=0.160$；如果$\rho_{AB}=-1.0$，$\sigma_P=0.070$。

这些计算显示将相关性较低的证券组合在一起，投资组合的风险将会降低。相关性从1.0降到−1.0，投资组合的风险从0.303降到0.070。然而，当相关系数从1.0降到0的时候，投资组合的风险仅从0.303降低到0.22，而当相关系数降低到−0.5时，投资组合的风险仅减少了一半。

前面考虑的是两种资产组合的情形。这一公式可以扩展到n种资产的情形。

$$\sigma_P^2=\sum_{j=1}^{n}(X_j^2\sigma_j^2)+\sum_{j=1}^{n}\sum_{k=1,k\neq j}^{n}(X_jX_k\sigma_{jk}),$$

表示为矩阵的形式为

$\sigma_P^2=\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}$，其中$\boldsymbol{X}^T=(X_1,X_2,X_3,\cdots,X_n)$，$\boldsymbol{V}=(\sigma_{ij})_{n\times n}$为协方差矩阵。

2.3 现代投资组合理论

投资者投资于金融市场上，最终的目的是使得投资末期的期望效用最大化。期望效用函数为冯·诺依曼—摩根斯坦(Von Neumann-Morgenstern)的预期效用函数 $H(P)=\sum_{z\in Z}p(z)u(z)$，其中 $p(z)$ 为未来收益的分布，$u(z)$ 为每个状态的效用函数，Z 为未来收益的状态空间。马科维茨通过使用预期收益率和方差两个参数来代替期望效用函数得到了非常好的结论。除此假设外，他还假设：

(1)投资者是永不满足的，且规避风险的，即在一定的风险下要求最大的收益率，或者在一定的收益率下最小化风险。

(2)单一投资期间。即投资者初期有一禀赋 W，并将其分配在 n 个风险资产上至期末投资结束，不存在投资策略随时间的调整。

(3)没有交易成本。

2.3.1 不存在无风险资产的情形

在具有相同期望收益率的资产组合中，具有最小方差的资产组合称为前沿边界的资产组合。一个资产组合 P 是一个前沿边界的资产组合，当且仅当 n 维资产组合权重向量 $\boldsymbol{X}_P$ 是以下二次规划的解：

$$\min_{\{w\}}\frac{1}{2}\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X} \tag{2.1}$$

给定

$$\boldsymbol{X}^T\boldsymbol{e}=E[R_P] \tag{2.2}$$

$$\boldsymbol{X}^T\boldsymbol{1}=1 \tag{2.3}$$

其中，$\boldsymbol{e}$ 表示这 n 项资产的 n 维期望收益率向量，$E[R_P]$表示资产组合 P 的期望收益率，$\boldsymbol{1}$ 是 n 维单位向量，$\boldsymbol{V}$ 为 n 项资产的协方差矩阵。

上式的意思是在一定的期望收益率 $E[R_P]$下，最小化投资组合的方差 $\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}$，由于$\frac{1}{2}\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}$ 与 $\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}$ 取最小值时的资产组合 X 是相同的。注意卖空(即负的资产组合权重)是允许的。因此，可行的资产组合的期望收益率的

范围是无界的。

解上述二次规划，可得到$\dfrac{\sigma^2(R_P)}{\frac{1}{C}}-\dfrac{\left(E[R_P]-\frac{A}{C}\right)^2}{\frac{D}{C^2}}=1$（具体解答过程见附录2-1），其中$A=\mathbf{1}^T V^{-1} e, B=e^T V^{-1} e, C=\mathbf{1}^T V^{-1}\mathbf{1}, D=BC-A^2$，这个在$\sigma_P$-$E[R_P]$的空间中表现为双曲线，如图2.1。

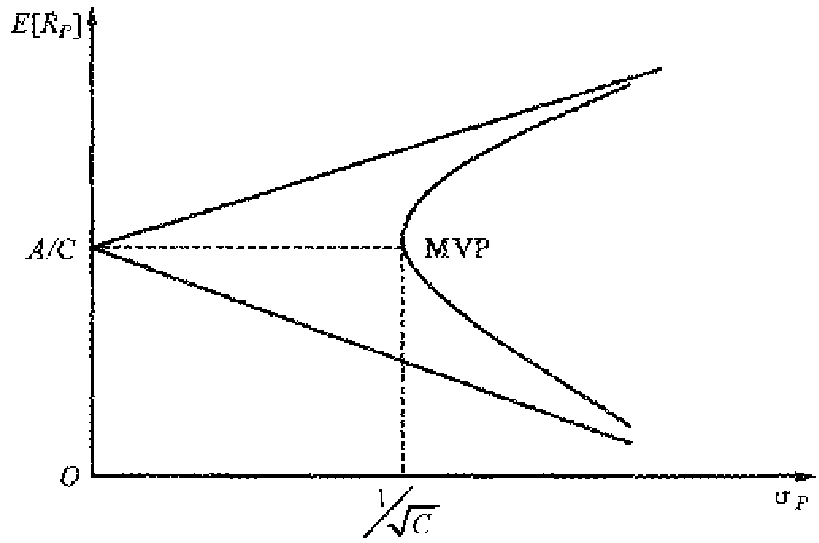


图 2.1　在σ_P-$E[R_P]$空间中的资产组合前沿

上图描述了在只存在n项风险资产时的资产组合前沿边界，且其中最小方差资产组合点为MVP，其对应的期望收益率为$\frac{A}{C}$，标准差为$\frac{1}{\sqrt{C}}$。

显然投资者是风险厌恶型的，可以得到只有MVP点以上的资产组合前沿边界才是理性投资者选择的，被称为效率边界。因为在MVP以下的前沿边界的每一点，在同样的方差的情形下，MVP上面的点的期望收益率高于MVP下面的点的期望收益率。

2.3.2 存在无风险资产的情形

令P是所有$n+1$个资产的前沿边界组合，$\boldsymbol{X}_P$表示风险资产在P中的资产组合向量，则$\boldsymbol{X}_P$是下述规划的解：

$$\min_{(X_P,\lambda)} \frac{1}{2}\boldsymbol{X}^T \boldsymbol{V}\boldsymbol{X}$$

$$\text{s.t } \boldsymbol{X}^T\boldsymbol{e}+(1-\boldsymbol{X}^T\boldsymbol{1})r_f=E[R_P]$$

其中，r_f 为无风险利率（在没有特别说明下，本书都用 r_f 代替无风险资产收益率）。

解上述二次规划，我们可以得到（详细过程见附录 2-2）：

$$\sigma_P=\begin{cases}\dfrac{E[R_P]-r_f}{\sqrt{H}} & \text{当 } E[R_P]\geqslant r_f\\ -\dfrac{E[R_P]-r_f}{\sqrt{H}} & \text{当 } E[R_P]<r_f\end{cases}$$

这就是说，所有资产的资产组合前沿边界在 σ_P-$E[R_P]$平面上是由两条从点$(0,r_f)$发出的斜率为$\sqrt{H}$和$-\sqrt{H}$的半直线组成的。

r_f 显然要小于最小方差资产组合点 MVP 的期望收益率，因为投资者为风险厌恶型的，在同样的方差的情形下，他更喜欢期望收益率更大的资产组合，在同样的期望收益率的情形下，他更倾向于方差更小的资产组合。最小方差资产组合 MVP 的方差大于无风险资产的方差，故其收益率要大于无风险资产组合，否则，没人会选择 MVP。

另外上面得到的新的前沿边界的上半支切于只有风险资产时的前沿边界。为验证这点，我们只需说明：

$$\frac{E[R_P]-r_f}{\sigma_P}=\sqrt{H}$$

其中 $H=(\boldsymbol{e}-r_f\boldsymbol{1})\boldsymbol{V}^{-1}(\boldsymbol{e}-r_f\boldsymbol{1})=B-2Ar_f+Cr_f^2$，此处我们取切点 p 的期望收益率为 $E[R_P]=\dfrac{A}{C}-\dfrac{\frac{D}{C^2}}{r_f-\frac{A}{C}}$（见附录 2-3）。利用关系式$\dfrac{\sigma^2(R_P)}{\frac{1}{C}}-\dfrac{(E[R_P]-\frac{A}{C})^2}{\frac{D}{C^2}}=1$，我们得到：

$$\frac{E[R_P]-r_f}{\sigma_P}=\left(\frac{A}{C}-\frac{\frac{D}{C^2}}{r_f-\frac{A}{C}}-r_f\right)\frac{-C\left(r_f-\frac{A}{C}\right)}{\sqrt{H}}=\sqrt{H}$$

新的效率边界　在线段$\overline{r_fP}$上的投资组合是资产组合 P 与无风险资产的一个凸组合，见图 2.2。即投资者把一部分资金投资于资产组合 P，剩下的部分投

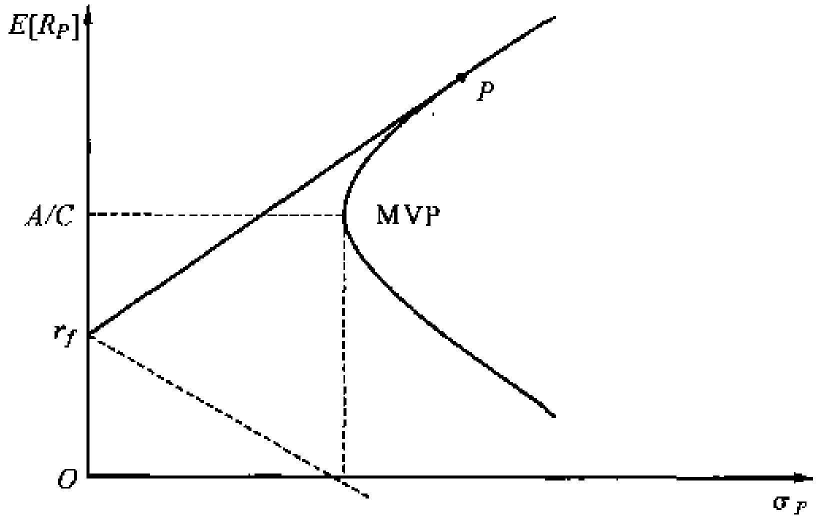


图 2.2 有无风险资产时的资产组合前沿边界

资于无风险资产上。任何在半直线 $r_f+\sqrt{H}\sigma_P$ 且不在 $\overline{r_fP}$ 上的资产组合为卖空无风险资产并将所得全部投资于资产组合 P 上。效率边界为高于 r_f 点的直线。

注意:通过引入无风险资产后,马科维茨的效率边界被扩展了。

2.3.3 选择一个风险资产的最优投资组合

一旦我们使用马科维茨模型确定了效率边界,投资者必须在这一系列的投资组合中选择最适合他们的一个。马科维茨模型并没有制定某一个最优的投资组合,相反,它产生了一系列的有效投资组合,这些都是最优的投资组合。

无差异曲线 我们使用无差异曲线来确定适合个人投资者偏好的期望收益—风险组合。

图 2.3 是一个风险厌恶投资者的无差异曲线,代表了投资者对风险和收益的偏好。每一条无差异曲线对投资者来说都代表偏好相同的风险与期望收益的组合。

显然无差异曲线有以下几个特征。首先,无差异曲线不会相交,且投资者可以有无数条无差异曲线。其次,风险厌恶投资者的无差异曲线的斜率是向上的,因为投资者是风险厌恶型且是单调偏好的,这样让其承受更大的风险必须要有更大的回报作为补偿。另外,曲线的形状与投资者个人的偏好有关。位置更高的无差异曲线优于位置较低的曲线。无差异曲线的斜率越大,投资者的风险规避程度越高。最后,离横轴越远的点代表越大的效用。

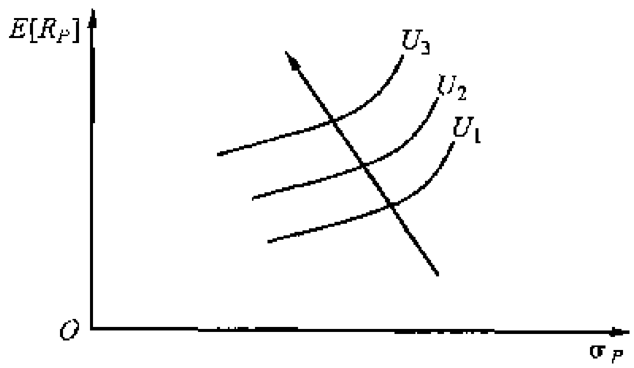


图 2.3　无差异曲线

选择最优的投资组合　风险规避投资者的最优投资组合是效率边界与投资者无差异曲线的切点。

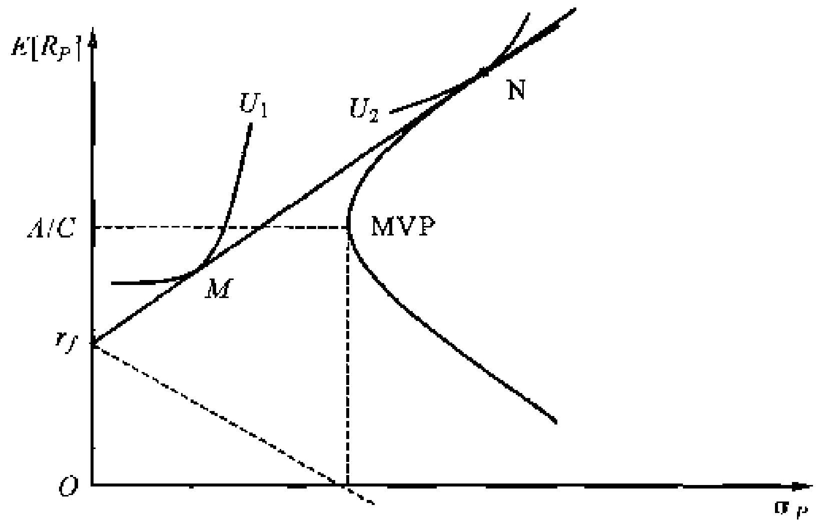


图 2.4　选择效率边界上的一个投资组合

在图 2.4 中，有两个风险投资者 A 和 B，其中 A 更厌恶风险，其无差异曲线为 U_1（图中只画出相切的那一条），B 的无差异曲线为 U_2。M，N 点分别为无差异曲线 U_1，U_2 与效率边界的切点。因为在相切时才能既满足风险资产组合边界所达到的投资组合，又能使得投资者的效用最大。这样 A 在 M 点达到最大的效用，B 在 N 点达到最大的效用。M 点无风险资产的比例要高于 N 点的无风险资产的比例。这说明风险厌恶程度高的投资者在无风险资产上投资的比例要高些。

2.4 输入数据时的考虑

在这一部分，我们将讨论在投资组合问题中影响输入数据选择的一些考虑。几乎所有的资产配置分析都是始于利用历史数据对组合选择过程中的输入数据进行估算的。分析师通常对这些数据进行修正，以使得它们能更好地反映对未来的想法。在关于指数模型的第 3 章，我们将讨论如何运用历史数据估计方差和相关系数，这比简单地采用历史值更精确。但在开展这一工作前，我们将讨论如何运用历史数据通常要考虑的问题。

2.4.1 优化问题中经通货膨胀调整的输入数据

有效边界技术被广泛地应用于确定长期资产配置的实践中，特别是养老资金配置。当投资期是以 10 年为单位衡量时，考虑货币购买力变化如何影响投资决策就非常重要。特别是，投资者可能更关注投资组合的未来实际购买力，即通过通货膨胀调整的价值，而非投资组合的未来名义价值。解决这一问题的一种可能方法是，在有效边界技术中应用通货膨胀调整后的收益。表2.3比较了美国股票、政府债券、国库券和通货膨胀的历史数据。注意，国库券的回报与通货膨胀是正相关的，当通胀较高时，国库券有较高收益，当通货膨胀较低时，国库券收益较低。这意味着，国库券可以部分地对通货膨胀套期保值。

表 2.3 没有通胀调整的收益率

	几何平均	标准差	相关系数			
			S&P	债券	国库券	通货膨胀
股 票	10.03	15.67	1.00			
债 券	5.73	11.14	0.22	1.00		
国库券	4.56	3.19	−0.65	−0.24	1.00	
通胀率	4.24	3.65	−0.36	−0.20	0.38	1.00

表 2.4 给出了经通货膨胀调整后的股票、债券、国库券的收益数据。在创建经通货膨胀调整的有效边界时，读者可以利用这些数据作为起点。尽管如国库券之类的证券能提供针对通货膨胀的部分套期保值功能，在上面的例子

中没有无风险资产——即使国库券也有一定程度暴露在通货膨胀风险之中。美国近期发展了一种新的证券，这种证券可以提供近似完美的通货膨胀套期保值，自 1997 年，美国发行了与通货膨胀挂钩的证券(TIP)，其价值部分取决于消费物价指数(CPI)的变化。这些债券的收益随通货膨胀而变化，并使得债券有很好的通货膨胀套期保值功能。有时，美国年通货膨胀率高于 10%，这意味着投资于与通货膨胀不挂钩的资产的购买力实际下降 10%。因而，与通货膨胀挂钩的证券在通货膨胀时期有保护投资者免受严重侵蚀的功能。

表 2.4　经通货膨胀调整的收益率

	几何平均	标准差	相关系数		
			S&P	债券	国库券
S&P	5.78	17.32	1.00		
债　券	1.49	12.39	0.37	1.00	
国库券	0.31	3.83	0.33	0.54	1.00

在进行投资分析时，分析师在考察名义收益率时，也一并考察经通货膨胀调整后的收益率，或用经通货膨胀调整后的收益率代替名义收益率。此外，与通货膨胀挂钩的证券越来越可能成为投资者组合优化中的一种重要资产。

2.4.2 输入数据估计值的不确定性

可靠地输入数据对在投资组合配置中合理应用均值-方差分析方法非常关键。常见的是，应用历史的风险、收益和相关系数作为起点来获取计算有效边界所需输入数据。

(1)如果收益特性不随时间改变，则有效数据的时期越长，均值的估计就越准确。设时间序列独立同分布，且每个独立的随机变量 R_i 的方差为 σ^2，则样本均值 $(\frac{1}{N}\sum_{i=1}^{N}R_i)$ 的方差为：$COV(\frac{1}{N}\sum_{i=1}^{N}R_i,\frac{1}{N}\sum_{i=1}^{N}R_i)=\frac{1}{N}\sigma^2$，其中 N 为样本容量。对观察到的独立收益序列而言，N 为历史数据观察时期数。因为在收益不相关的假设下，更多的历史数据能改进均值-方差模型中预期收益的估计，尽管改进作用是递减的。

为了展示这一问题对投资组合选择的重要性，假定投资者必须在两种投资间进行选择，这两种投资有相同的样本均值和方差。在其他条件相同的情

况下，标准方法将两者视为等价。如果考虑到额外的信息，即第一个样本的均值是基于1年数据之上，而第二个样本均值是基于10年数据之上，则通常的感觉是，第二种方法比第一种方法风险更小。

在这里注意一下，前面两个都谈到了方差，但是它们有本质的区别。第一个是每项资产在观察年段的方差，第二个是在观察年段内，所有风险资产的均值的方差。由于第一个样本是基于1年的，从而得到的每项均值的误差可能性更大，基于这种均值的多项风险资产的方差肯定更不准确了。

我们还可以假定投资者主要关注下个月的收益率，其平均收益率是 $E[R]$，其收益率的方差为 $\sigma_{\text{Pred}}^2=\sigma^2+\frac{\sigma^2}{T}$。

式中：

σ_{Pred}^2——预测方差；

σ^2——月收益率方差；

T——时期数。

表达式的第一部分反映的是收益率的内在风险（n 项风险资产之间的方差）。表达式的第二部分反映的是缺乏对真实收益率均值的了解而导致的不确定性（每项风险资产在观察期的对其本身均值的方差）。在贝叶斯分析中，这一方程右边两项表达式之和被称为收益预测分布的方差。注意，由于未来均值存在的不确定性，预测方差总是大于历史方差。

（2）证券收益特性总是随时间而改变。虽然使用长时间观察能改进估计，但是与此同时，证券特性会随时间而改变（如期望值会变），这给长时期观察带来潜在的不确定性。这样，投资分析师必须权衡观察时期数长短的利弊。由于这一冲突，多数投资分析师会修正历史数据，以反映他们关于目前情况与历史情况之间存在差异的看法。

2.5 对现代投资组合模型的评价

马科维茨模型的初衷就是为投资组合选择单个证券，然而该模型并不适用，如果投资组合包含很多证券的话，我们要估计的协方差的数目惊人。

【例 2.4】 如果一个分析师要考虑100个证券的投资组合，他就要估计（100×99）/2＝4 950个不同的协方差。如果是250个证券，那么就会产生（250×249）/2＝31 125个协方差。

做投资决策最重要的是要利用市场数据迅速找到套利的机会，在短时间估计出这么多协方差几乎是难以想象的难题，且投资分析师都是专注于几个行业的股票，不可能对每个行业都了如指掌。这样，马科维茨模型在实际中有极大的缺陷。后面章节中还会讲到这一问题。

马科维茨的理论假设也存在问题。该模型是建立在冯·诺依曼—摩根斯坦的预期效用函数上。投资者选择自己的投资策略（就是投资于每份资产上的收入份额）来使得自己的期望效用最大化。

个体的效用函数可以在期望财富附近泰勒展开：

$$u(W)=u(E[W])+u'(E[W])(W-E[W])+\frac{1}{2}u''(E[W])(W-E[W])^2+R_3$$

其中 $R_3=\sum_{n=3}^{\infty}\frac{1}{n!}u^{(n)}(E[W])(W-E[W])^n$，$W$ 为期末财富，由于风险资产的收益率为随机变量，故期末财富也为随机变量。

假设这个泰勒级数收敛，并且取期望和求和的过程可以互换，则个体期望效用可以表示为

$$E[u(W)]=u(E[W])+\frac{1}{2}u''(E[W])\sigma^2(W)+E[R_3]$$

其中 $E[R_3]=\sum_{n=3}^{\infty}\frac{1}{n!}u^{(n)}(E[W])m^n(W)$，其中 $m^n(W)$表示 W 的 n 阶中心矩。

上述关系式指出对于一个对期望收益偏好而对方差厌恶的投资者，除了期望和方差外，还有更高阶矩的项，它说明对于任意的分布和偏好，期望效用不能仅仅由财富的期望值和方差来确定。

解决上述难题的可以从两个方面来下手：收益率分布和效用函数。

首先看收益率分布（或财富分布）。现代投资组合模型在风险资产收益率服从多元正态分布的假设下有效。因为正态分布完全由它的两个参数：期望和方差来确定。在正态分布下 $E[R_3]=\sum_{n=3}^{\infty}\frac{1}{n!}u^{(n)}(E[W])m^n(W)$ 所含的三阶和更高阶矩可以由前两阶的方程来表示。尽管对数正态分布收益也可以由其均值和方差完全确定，但是它不是可加的。这样在假设单个证券的收益率满足对数正态分布的情况下，资产组合却不是对数正态分布。

其次看效用函数。假设效用函数为二次效用函数 $u(W)=W-\frac{b}{2}W^2$（b 为

参数，是风险厌恶度的一种度量）。二次效用函数的3阶和更高阶导数为零，因此 $E[R_3]=0$。所以个体的期望效用函数被他的期末财富的前两阶矩决定。

这样表面上，马科维茨的假设不存在任何问题，只要收益率分布为正态分布或者效用函数为二次效用函数就行。但是这样不是很合理。收益率分布为正态分布，而正态分布的两端是无穷的，这与有限责任和经济理论不一致，后者认为负的消费是没有意义的。针对二次效用函数，由于其为递增的绝对风险厌恶函数（绝对风险厌恶指标 $R_A(W)=-\frac{u''(W)}{u'(W)}$），表明投资者随着初始财富的增加投资于风险资产上的资金越少，而这显然违背常识。而且在财富达到 $\frac{1}{b}$ 时，投资者的效用随着财富的增加而减少这个也不合理。解决的办法可以使得 b 很小，乃至趋近于零，但是这又使得效用函数为1次了。

另外，冯·诺依曼—摩根斯坦的预期效用函数也并不是完美无瑕的，阿里亚斯悖论（Allais paradox）就推翻了预期效用函数建立的基础：替换公理或独立公理。

当然，任何一个模型不可能完美无缺。1990年诺贝尔经济学奖授给马科维茨，足以看出本模型的重要性。马科维茨把期望和方差引入资产组合管理，使得金融第一次脱离主观描述而变得精确起来，为现代的风险管理打下了坚实的基础。

概括起来，马科维茨对于现代金融理论的贡献在以下几个方面的命题上：

（1）金融经济学家传统上把预期收益的最大化看作投资组合的目标，实际上，分散投资行为常常与此目标相矛盾，但分散投资行为却与均值—方差的目标函数一致。

（2）均值—方差的目标函数与投资者追求预期效用最大化的目标一致。

（3）投资组合的方差是证券方差和协方差的函数。因此，单一证券对于投资组合风险的贡献取决于它与其他证券的相关性。

（4）投资者关心有效组合集——就是说，在期望收益率给定的情况下，选择方差最小的组合；在方差给定的情况下，选择期望收益率最大的组合。前者产生了最小方差集，后者产生了有效集。

尽管马科维茨模型在资产组合的应用上有很大的限制。但是在考虑资产类别上还是有很大的用处。资产分配决定是将投资组合资产分配到广泛的资产市场上，换句话说，投资组合的资金有多少投资于股票，多少投资于债券、货币市场资产等。每个权重都可以赋予从0～1.0。

附　录

附录 2.1：不存在无风险资产的情形的解法详细过程

由拉格朗日解法，X_P 是下式的解：

$$\min_{\{X,\lambda,\gamma\}} L=\frac{1}{2}\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}+\lambda(E[R_P]-\boldsymbol{X}^T\boldsymbol{e})+\gamma(1-\boldsymbol{X}^T\boldsymbol{1}) \tag{2.4}$$

其中，λ，γ 是两个正的常数。一阶条件是

$$\frac{\partial L}{\partial \boldsymbol{X}}=\boldsymbol{V}\boldsymbol{X}_P-\lambda\boldsymbol{e}-\gamma\boldsymbol{1}=0 \tag{2.5}$$

$$\frac{\partial L}{\partial \lambda}=E[R_P]-\boldsymbol{X}_P^T\boldsymbol{e}=0 \tag{2.6}$$

$$\frac{\partial L}{\partial \gamma}=1-\boldsymbol{X}_P^T\boldsymbol{1}=0 \tag{2.7}$$

其中，$\boldsymbol{0}$ 是 n 维零向量。由于 $\boldsymbol{V}$ 是正定矩阵（由方差恒大于等于 0 得到，$X^TVX\geqslant 0$，$\forall X\neq 0$），因此上述一阶条件也是全局最优的充要条件。

由(2.5)式知，

$$X_P=\lambda(\boldsymbol{V}^{-1}\boldsymbol{e})+\gamma(\boldsymbol{V}^{-1}\boldsymbol{1}) \tag{2.8}$$

将其代入到(2.6)和(2.7)知：

$$E(R_P)=\lambda(\boldsymbol{e}^T\boldsymbol{V}^{-1}\boldsymbol{e})+\gamma(\boldsymbol{e}^T\boldsymbol{V}^{-1}\boldsymbol{1}) \tag{2.9}$$

$$1=\lambda(\boldsymbol{1}^T\boldsymbol{V}^{-1}\boldsymbol{e})+\gamma(\boldsymbol{1}^T\boldsymbol{V}^{-1}\boldsymbol{1}) \tag{2.10}$$

由上两式可解得

$$\lambda=\frac{CE[R_P]-A}{D},\gamma=\frac{B-AE[R_P]}{D} \tag{2.11}$$

其中 $A=\boldsymbol{1}^T\boldsymbol{V}^{-1}\boldsymbol{e}$，$B=\boldsymbol{e}^T\boldsymbol{V}^{-1}\boldsymbol{e}$，$C=\boldsymbol{1}^T\boldsymbol{V}^{-1}\boldsymbol{1}$，$D=BC-A^2$

由于正定矩阵的逆矩阵仍然为正定的，所以 $B>0$，$C>0$。我们可以证明 $D>0$，因为：

$(A\boldsymbol{e}-B\boldsymbol{1})^T\boldsymbol{V}^{-1}(A\boldsymbol{e}-B\boldsymbol{1})=B(BC-A^2)$，而左边显然大于 0，故 $D>0$。将(2.11)代入到(2.8)得到

$$X_P=g+hE[R_P] \tag{2.12}$$

其中 $g=\frac{1}{D}[B(\boldsymbol{V}^{-1}\boldsymbol{1})-A(\boldsymbol{V}^{-1}\boldsymbol{e})]$，$h=\frac{1}{D}[C(\boldsymbol{V}^{-1}\boldsymbol{e})-A(\boldsymbol{V}^{-1}\boldsymbol{1})]$

由于(2.5),(2.6),(2.7)是 X_P 成为前沿边界的充要条件,所以任何前沿边界组合都可由(2.12)得到。

$$\frac{\sigma^2(R_P)}{\frac{1}{C}}-\frac{\left(E[R_P]-\frac{A}{C}\right)^2}{\frac{D}{C^2}}=1$$

此即正文中的表达式。

附录 2.2　存在无风险资产的情形的解法详细过程

用拉格朗日方法,我们知道 X_P 是下面式子的解

$$\min_{\{X_P,\lambda\}}\frac{1}{2}\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}+\lambda\{E[R_P]-\boldsymbol{X}^T\boldsymbol{e}-(1-\boldsymbol{X}^T\boldsymbol{1})r_f\}$$

X_P 是解的一阶充分必要条件是:

$$\boldsymbol{V}\boldsymbol{X}_P=\lambda(\boldsymbol{e}-\boldsymbol{1}r_f)\text{和 }r_f+X_P{}^T(\boldsymbol{e}-\boldsymbol{1}r_f)=E[R_P]$$

求解 X_P,我们有

$$\boldsymbol{X}_P=\boldsymbol{V}^{-1}(\boldsymbol{e}-r_f\boldsymbol{1})\frac{E[R_P]-r_f}{H}$$

其中 $H=(\boldsymbol{e}-r_f\boldsymbol{1})\boldsymbol{V}^{-1}(\boldsymbol{e}-r_f\boldsymbol{1})=B-2Ar_f+Cr_f^2$,$A,B,C$ 仍然沿用上述定义。可以容易地验证当 $A^2-BC<0$ 时,$H>0$。资产组合 P 的方差为

$$\sigma_P^2=\boldsymbol{X}_P^T\boldsymbol{V}\boldsymbol{X}=\frac{(E[R_P]-r_f)^2}{H}$$

等价地,我们可以写成:$\sigma_P=\begin{cases}\dfrac{E[R_P]-r_f}{\sqrt{H}} & \text{当 } E[R_P]\geqslant r_f\\ -\dfrac{E[R_P]-r_f}{\sqrt{H}} & \text{当 } E[R_P]<r_f\end{cases}$

此即正文中出现的表达式。

附录 2.3

对于资产组合前沿边界的一个重要性质是,对于前沿边界上的任何资产组合 P,除了最小方差资产组合,存在唯一的前沿边界资产组合,用 $zc(P)$ 表示,与 P 的协方差为 0。

给定关系式 $\sigma_{PQ}=\boldsymbol{X}_P^T\boldsymbol{V}\boldsymbol{X}_Q=\frac{C}{D}\{E([R_P]-\frac{A}{C}\}\{E([R_q]-\frac{A}{C}\}+\frac{1}{C}$,令资产

组合 p 与 $zc(P)$ 之间的协方差等于 0：

$$\mathrm{COV}(R_P,R_{zc(P)})=\frac{C}{D}\{(E[R_P]-A/C)(E[R_{zc(P)}]-A/C)+D/C^2\}=0$$

对 $zc(P)$ 的期望收益率求解，我们得到：

$E[R_{zc(P)}]=A/C-\dfrac{D/C^2}{E[R_P]-A/C}$，实际上上式定义了唯一的 $zc(P)$。从几何上，$zc(P)$ 的位置可以如下确定：在标准差—期望收益率空间中，经过与任何前沿边界资产组合(除了最小方差资产组合)相对应的点，与资产组合前沿边界相切的直线在期望收益率坐标轴上的截距是 $E[R_{zc(P)}]$。

为证明这点，对 $\dfrac{\sigma^2(R_P)}{\frac{1}{C}}-\dfrac{(E[R_P]-\frac{A}{C})^2}{\frac{D}{C^2}}=1$ 关于 σ_P，$E[R_P]$ 求一阶全微分得到：

$$\frac{\mathrm{d}E[R_P]}{\mathrm{d}\sigma_P}=\frac{D\sigma_P}{CE[R_P]-A}$$

这是资产组合前沿边界在 $(\sigma_P,E[R_P])$ 点处的斜率。切线在期望收益率坐标轴上的截距为：

$$\begin{aligned}E[R_P]-\frac{\mathrm{d}E[R_P]}{\mathrm{d}\sigma_P}\sigma_P&=E[R_P]-\frac{D\sigma_P}{CE[R_P]-A}\sigma_P\\&=A/C-\frac{D/C^2}{E[R_P]-A/C}\\&=E[R_{zc(P)}]\text{（由这个即可得正文里面的理论）}\end{aligned}$$

第三章

CAPM 和套利定价理论

前面两章，我们讲到了现代投资组合模型（MPT），它描述了投资者应如何选择风险证券的最优投资组合。这一章，我们首先来讨论资产定价模型。如果在理想条件下，所有投资者都利用马科维茨框架寻找风险证券的投资组合，将会如何？这将对证券的均衡价格和均衡收益有怎样的影响？换言之，最优多样化将如何影响证券的市场价格？在这些理想条件下，投资者将在风险和收益之间如何进行平衡？总的来说，我们想要一个能够解释证券的市场均衡价格的模型。这些就是资本资产定价模型，或称为风险资产的估价模型。

紧接着我们介绍一些关于 CAPM 经验检验的方法。由于经验检验与 CAPM 的理论阐述有较大的出入，所以有针对这种出入的各种 CAPM 扩展模型。

最后，我们将介绍一种全新的解释资产定价的方法——套利定价模型。该方法的理论机制如下：指定证券收益的产生过程，然后从套利论证中推导资产价格。套利定价模型对均衡本质进行了有趣的考察。然而，该理论的应用绝不是简单的事。这一领域的实证研究尚处于初级阶段。此外，在理论的应用方面提出了其他的方法。在介绍这些不同的方法之后，我们考察一下支持套利定价模型的证据，与均衡模型的资本资产定价模型的标准型或其他型相比，它是否必然不一致。最后我们将讨论套利定价模型的实用性和优势。

3.1 CAPM的推导

标准的资本资产定价模型是由夏普(1963,1964)和特里诺(Trenor,1961)分别独立建立的模型,后来莫辛(Mossin,1966)、林特纳(1965,1969)和布莱克(1972)又进一步发展了这个模型,故又称为夏普—莫辛—林特纳模型。该模型表明所有风险资产的均衡收益率是它们与市场投资组合协方差的函数。资本资产定价模型CAPM具有严格的假定条件:

(1)投资者是价格接受者,并且对于具有联合正态分布的资产收益有完全相同的预期。

(2)投资者是使其期末最终财富的预期效用最大化的风险厌恶者。

(3)资产数量是固定的。所有资产都可以任意买卖并且可以完全分割。

(4)存在无风险的资产,投资者可以在无风险利率的条件下借入或贷出任何数量的资金。

(5)不存在任何的市场不完善性。例如,不存在税收、管制或卖空。

(6)资产市场是无摩擦的,信息无代价并适用于所有的投资者。

CAPM的理论基础是马科维茨关于均值—方差选择理论。根据上一章,我们知道,如果存在市场均衡,那么必须对所有资产的价格调整直到资产为投资者持有,没有过剩需求。也就是说,资产价格要建立在总供给与总需求相等的基础上。在均衡状态下,市场投资组合是由所持有的全部可买卖的资产按其加权比重或比例组成的。每种资产在市场投资组合的均衡比例为:

$$W_i=\frac{\text{资产 } i \text{ 的市场价值}}{\text{所有资产的市场价值}} \tag{3.1}$$

需要指出的是,按照法马(1968)的定义,市场投资组合的概念是指包括市场中的每种证券(股票的)的总的结合,其中每种证券的组合权数,等于该种证券在市场上全部证券的总价值所占的比例。

设风险资产 i 和市场投资组合M分别以 x 和 $(1-x)$ 的比例形成组合P,那么分别得到该组合的期望收益率 $E[R_P]$ 和标准差 σ_P 为:

$$E[R_P]=xE[R_i]+(1-x)E[R_M] \tag{3.2}$$

$$\sigma_P=[x^2\sigma_i^2+(1-x)^2\sigma_M^2+2x(1-x)\sigma_{iM}]^{\frac{1}{2}} \tag{3.3}$$

需要解释一下，前沿边界就是由风险资产和市场投资组合的各种组合来组成的。那么，该组合期望收益率和标准差分别随 x 的变化情况，可求导如下：

$$\frac{\partial E[R_P]}{\partial x}=E[R_i]-E[R_M] \tag{3.4}$$

$$\frac{\partial \sigma_P}{\partial x}=[x^2\sigma_i^2+(1-x)^2\sigma_M^2+2x(1-x)\sigma_{iM}]^{-\frac{1}{2}}[x\sigma_i^2-(1-x)\sigma_M^2+(1-2x)\sigma_{iM}] \tag{3.5}$$

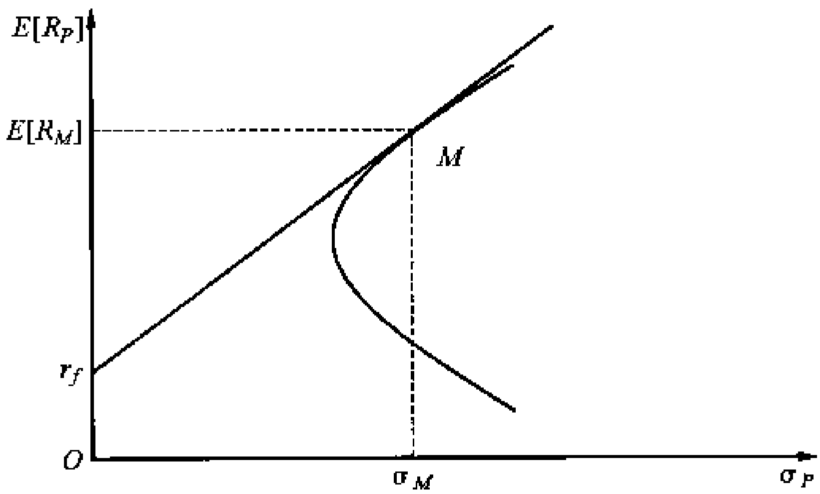


图 3.1　由风险资产 i 和市场投资组合 M 所组成的前沿边界组合

按照夏普和特里诺的看法：可以运用上面的事实来决定市场均衡状态价格。在均衡时，市场投资组合已经包括了风险价值权数，风险资产 i 的风险价值权数为 W_i。因此，上面方程中 x 衡量的个人对风险资产 i 的超额需求。按照均衡的条件，由于价格将调整到全部资产都被投资者持有的程度，对任何资产的超额需求都为零。

从而 $x=0$ 时，我们可以得到：

$$\left.\frac{\partial E[R_P]}{\partial x}\right|_{x=0}=E[R_i]-E[R_M] \tag{3.6}$$

$$\left.\frac{\partial \sigma_P}{\partial x}\right|_{x=0}=(\sigma_M^2)^{-\frac{1}{2}}(-\sigma_M^2+\sigma_{iM}) \tag{3.7}$$

在均衡时，前沿边界点 M 的切线斜率为：

$$\frac{\partial E[R_P]}{\partial x}\Big/\frac{\partial \sigma_P}{\partial x}=\frac{E[R_i]-E[R_M]}{\sigma_{iM}-\sigma_M^2}\sigma_M \tag{3.8}$$

由前面论述知，连接 r_f 和市场投资组合的直线切于由资产 i 和市场组合

M 组成的有效前沿边界(如图 3.1)。且把这条线叫做资本市场线(CML)。其斜率为$\frac{E[R_M]-r_f}{\sigma_M}$。这样针对同样的市场组合 M 点有两条切线斜率,这样二者必相等。

即:

$$\frac{E[R_i]-E[R_M]}{\sigma_{iM}-\sigma_M^2}\sigma_M=\frac{E[R_M]-r_f}{\sigma_M}$$

化简上式得到:

$$E[R_i]=r_f+[E[R_M]-r_f]\frac{\sigma_{iM}}{\sigma_M^2}$$

令 $\beta_i=\frac{\sigma_{iM}}{\sigma_M^2}$,则上式化为

$$E[R_i]=r_f+[E[R_M]-r_f]\beta_i \tag{3.9}$$

上式即为著名的资本资产定价模型——CAPM。

证券市场线(SML):资本市场线刻画了均衡的金融市场中风险和收益的比较关系。然而它只适用于有效投资组合,并且不能评估单个证券的均衡收益。由 CAPM 模型知道,所有投资者都持有市场投资组合,它是衡量其他一切投资组合的基础,单个证券对投资组合的风险有影响,这就决定了单个证券或者无效投资组合的均衡收益。证券市场线指风险资产的期望收益率与其 β 值之间关系的一条直线。见图 3.2。

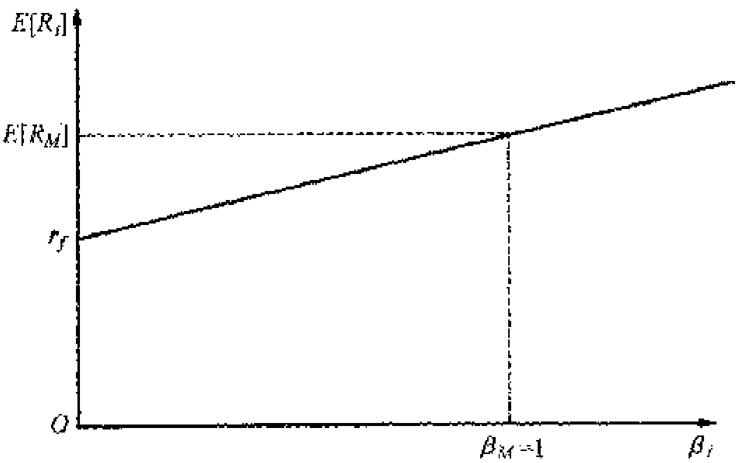


图 3.2 证券市场线

注意证券市场线和资本市场线的区别：资本市场线都是投资者最终选择投资的资产组合，是有效率的，而证券市场线并不是每点都是有效率的，只有$\beta_M=1$的市场组合的点才是有效率的。在公式(3.9)中，任何资产所要求的收益率都等于无风险收益率加上风险溢价。所谓风险溢价就是风险价格与风险数量的乘积。风险价格为市场投资组合的预期收益率与无风险收益率之差。以图形表示则是证券市场线的斜率，风险的数量则是β。

公式(3.9)通常为CAPM的事前形式。

由于无风险资产与市场投资组合的协方差为0，所以无风险资产的β为0；市场投资组合与其本身的协方差等于市场投资组合的方差，因此，市场投资组合的$\beta=1$。所以在现实证券市场上，相关系数多在0～1之间。

进一步推知：风险资产的期望收益率可以由两部分构成，即无风险收益率和因风险存在而增加的收益补偿；风险越大，风险溢价也越大，对该资产的期望收益率也越高。这就解释了前面提到的，风险越大，则人们要求的收益率也应越高。从几何意义上说，在均衡时，每种资产的定价恰好使其风险调整所要求的收益率落在证券市场线上，因为根据投资组合理论，投资者除了无法避免资产与市场投资组合的风险外，其余风险可通过增加证券投资的数目而分散掉。这样，资产的总风险可以分成两部分：系统风险和非系统性风险。而β表示的系统风险的数量。由CAPM的形式可知，对风险的补偿只是对系统性风险进行补偿，而不是对其总风险进行补偿。即如果某证券的总风险很大，但是绝大部分只是由其特有的风险组成，而其系统性风险很小，则人们对其要求的收益率并不会高。

分离定理　由前面假设可知，每个投资者对风险资产的预期收益率、方差和协方差有相同的看法，这就意味着新的有效边界（一条直线）对所有投资者来说是相同的。这样对于每个投资者来说，只需要考虑把手中的财富在无风险资产和相同的市场资产组合M进行分配就可以。至于把多少资金投资于市场组合，这取决于投资者对风险的回避程度（见图3-1），风险回避程度高的投资者将贷出更多的无风险资产，风险回避程度低的投资者将借入更多的资金投资于市场组合M。关于投资与融资分离的决策理论被称为分离定理。

证券价值的高估与低估　证券市场线对证券价格有重要的内涵。在均衡状态下，每种证券都应该在SML上，因为它的预期收益率应该补偿投资者面临的系统风险。如果投资者决定让一种证券不在SML上，将会如何？他们必须运用一种独立的方法对该证券的预期收益进行评估。

【例 3.1】 如图 3.3 所示，两种证券被划分在 SML 的两边。证券 X 在 SML 的上方，通过分析，它有较高的预期收益率，而证券 Y 有较低的预期收益率，它在 SML 的下方。哪种证券被低估了？

在 SML 上方的证券 X 被低估了，因为在既定的风险下，它的预期收益比投资者要求的高。投资者要求的预期收益率的最小值为 $E[R_x]$，而通过基本原理分析，证券 X 的预期收益率为 $E[R_x']$。如果投资者意识到这一点，他们将会：

购买证券 X，因为它提供高于必要的收益。随着越来越多的证券 X 被购买，这种需求将会使证券 X 的价格上涨。收益将会因此而下降，直到它达到 SML 的水平。

现在考虑证券 Y。根据 SML，投资者对证券 Y 要求 $E[R_y]$，但 Y 只提供了 $E[R_y']$。当投资者意识到这一点，他们将会卖出证券 Y（或卖空 Y）。这种供给将使 Y 的价格下降。由于现在以较低的价格支付股息，也从较低的价格开始增值，该股的收益将会上升。它的价格将会一直下降，直到预期收益上升至 SML 为止，于是该证券再次达到均衡。

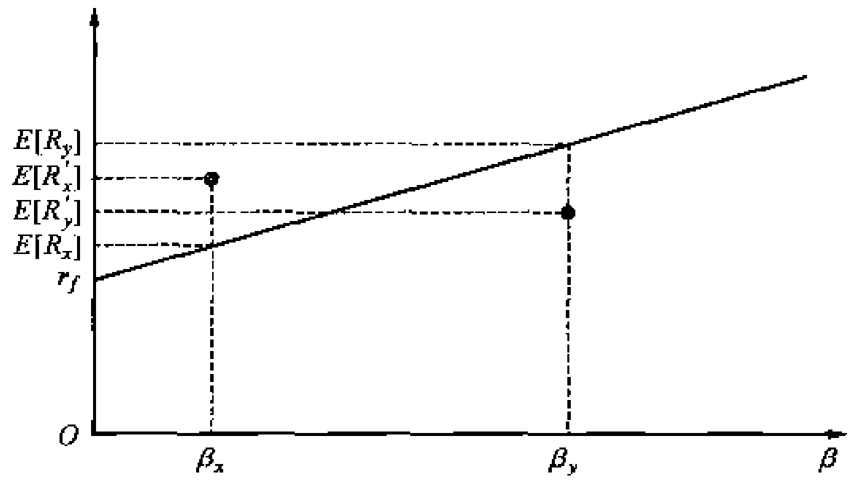


图 3.3　SML 图解证券价值的高估与低估

3.2 对 CAPM 的经验检验

CAPM 的结论是完全符合逻辑的，收益和风险是一定相关的——高风险应该获得高收益。一种证券的相关风险是它相对投资组合风险的大小。或者

说承受非市场风险不应有额外收益。由于资本市场理论的大部分假设都是不现实的，为了评估它和其他理论的有效性，我们必须采用经验检验。

一些早期的实证检验。大多数 CAPM 的早期实证检验都使用一个时间序列（首次）回归估计贝塔，然后使用一个横截面（二次）回归检验我们从 CAPM 模型中推出的假设。为了更具体，我们介绍一项早期由林特纳做出的 CAPM 的实证研究，道格拉斯（1968）重复了该研究。林特纳首先对他样本中的 301 只普通股票分别计算贝塔，方法是：用每只股票的年收益率对样本中所有股票的平均收益率进行回归，使用的数据是 1954—1963 年。首次回归的形式是

$$R_{it}=a_i+b_iR_{Mt}+\epsilon_{it}$$

式中，b_i（回归系数）是对股票 i 真实贝塔的估计。接着林特纳进行了二次横截面回归

$$E[R_i]=a_1+a_2b_i+a_3S_{ei}^2+\eta_i$$

式中，S_{ei}^2 是来自首次回归的残差，是股票的特有风险。这个模型中的每个参数都有一个理论值。根据所要检验的 CAPM 形式，$a_3=0$（特有风险不应得到报酬），$a_1=r_f$ 或者 $a_1=E[rc(M)]$（见下一节的零贝塔模型），$a_2=E[R_M]-r_f$ 或者 $a_2=E[R_M]-E[rc(M)]$。他得到的值是

$$a_1=0.108, a_2=0.063, a_3=0.237$$

这些结果看起来似乎违背了 CAPM。代表残差风险的项在统计上显著，并且为正。截距 a_1 看起来大于对 r_f 或 $E[rc(M)]$ 任何合理的估计值，a_2 虽然显著，但略低于我们的合理预期值。道格拉斯使用了类似的方法，得出了与林特纳相似的结果。

在这些文献中，大多数研究利用的是上市普通股票的月总收益（红利用于再投资）。经常使用的技术是依靠计算证券收益和通常等同于所有上市普通股票加权指数的市场指数间的协方差除以市场指数的方差来估计每种证券在 5 年持有期的 β 值。然后将证券按 β 排列，置于 N 个投资组合中（N 为 10，12 或者 20）。通过把所选证券置于大的投资组合中进行非系统风险的最大分散，有可能避免在估计单只股票 β 时出现的大部分测量误差。投资组合的 β 值和收益在下一个 5 年期和与 CAPM 经验市场线相似的回归计算中求得。

按照 CAPM 的事前形式：

$$E[R_i]=r_f+[E[R_M]-r_f]\beta_i$$

平均来看，资产实现的收益率等于预期收益率，按照收益正态分布可以计算出 CAPM 的事后形式：

$$R_j - r_f = (R_M - r_f)\beta_j + \varepsilon_j$$

这里所有变量都为事后值，并不为事前的随机变量。CAPM 的这一事后经验模型与事前理论模型的重要区别就是，前者可以有负斜率，而后者一定为正斜率。事实上，人们可能已经经历了市场收益率为负的自然状态。当这一状况发生时，经验证券市场线就会向下倾斜。而在理论上，CAPM 总是要求市场的事前预期收益率要高于无风险收益率。

当以经验检验 CAPM 时，通常以下列形式表示：

$$R'_{Pt} = \gamma_0 + \gamma_1\beta_P + \varepsilon_{Pt}$$

其中 $\gamma_1 = R_{Mt} - r_{ft}$；$R'_{Pt} = R_{Pt} - r_{ft}$，投资组合 P 的超额收益。通常称上式为经验市场线形式。

当我们估计上式时，γ_0 应该近似于平均无风险收益率，γ_1 应该近似于研究时段的平均市场风险溢价。

关于资本市场理论特别是 CAPM 的延伸文献，尽管有各种不同的证据，却没有令人信服的案例说明非系统性风险应获得风险溢价。换言之，投资者仅仅因承担系统风险而获得收益。普遍认为 CAPM 的斜率 γ_1 比理论中假定的平缓，截距项 γ_0 比 0 大。

证券市场线(SML)外观上是线性的，也就是说，预期(必需)收益和风险之间的相关关系是一条向上倾斜的直线。

检验资本市场理论的主要问题在于，它基于一个事先的基础被阐释，却只能依据事后来检测。我们永远不会明确地知道投资者的预期，因此，该模型的检验在某些时候产生相矛盾的结果以及检验结果与模型的假设相背离就不足为奇了。事实上，检验结果对基本的资本资产定价模型非常支持。基于多年来对数据的研究，我们发现，股票市场在系统风险和收益呈线性关系的基础上为有价证券定价，而非系统性风险在定价机制中几乎没有任何影响。

资本资产定价模型至今仍然没有被人证明，将来也不会。事实上，Roll 指出，市场投资组合中因包含所有资产而难以预测，因此资本资产定价模型是不可检验的。实际上，Roll 认为资本资产定价模型的检验本质上是对市场投资组合的平均偏差的效率的检验。尽管如此，资本资产定价模型仍然是考虑预期收益和风险的比较关系的一种逻辑方式，同时在金融领域中作为一种模型被广泛地应用。

3.3 修正的资本资产定价模型

如果前面设定的每个假设条件都成立，那么上面构建的CAPM模型就可以完全描述资本市场的行为。但是我们知道对CAPM的经验检验发现，CAPM并不是那么完美。上节所表明的截距项 γ_0 明显地不为0，斜率 γ_1 小于投资组合收益与无风险利率的差。这说明理论和经验检验有出入。而且多数个人和很多机构投资者持有的风险资产组合并不像理论中所阐述的那样，人人都持有市场组合与无风险资产的组合。

下面，在对以下各影响因素做出更现实的假设的条件下，我们来讨论一般均衡模型：不允许卖空；对无风险借贷的修正；个人税收；差别期望。

1. 不允许卖空

推导资本资产定价模型的假定是，投资者可以无限制卖空，并且卖空被定义为最为宽泛的形式，即投资者可以卖出任何证券（不论是否持有），而且用卖空所得购买其他任何证券。这是一个方便的假设，并且简化了推导，但它不是一个必然的假设。如果不允许卖空，也可以得到同样的结果。由于在均衡时没有投资者卖空任何证券，所以限制卖空并不会改变均衡。因此，无论是否允许卖空，推导出的CAPM的关系是一样的。

2. 无风险借贷的修正

CAPM的另一个假设是投资者可以以无风险利率无限制的借贷。这样的假设显然与现实不符。更现实的假设是投资者可以以无风险利率贷出资金，但不能以无风险利率借入资金。

(1)不存在无风险借贷的情形

首先我们假设不存在无风险借贷的情形。令P是一个最小方差资产组合之外的前沿边界资产组合，Q是任何资产组合，可以证明 $E[R_Q]=(1-\beta_{PQ})E[R_{zc(P)}]+\beta_{PQ}E[R_P]$，其中 $\beta_{PQ}=COV(R_Q,R_P)/\sigma^2[R_P]$。由于市场组合在前沿边界组合上，且市场组合不是最小方差组合，故我们有 $E[R_Q]=(1-\beta_{PM})E[R_{zc(M)}]+\beta_{PM}E[R_M]$。又由于对于任意一个风险资产来说，其本身就是可行的资产组合，所以有 $E[R_j]=(1-\beta_{jM})E[R_{zc(M)}]+\beta_{jM}E[R_M]$。

这被称为零 β 资本资产定价模型。

我们断言 $zc(M)$ 不可能是有效资产组合，或者说 $zc(M)$ 一定在MVP下面。下面我们证明这一点。

由于 MVP 在风险资产组合前沿边界上，而 M 和 $zc(M)$都在风险资产前沿边界上，且两者明显不重合，由第 2 章我们知道风险资产组合前沿边界上相异的两点的线性组合可以组成整个前沿边界。故可设 MVP 由市场组合 M 和对应的零贝塔资产组合 $zc(M)$的线性组合组成。$R_{MVP}=\alpha R_M+(1-\alpha)R_{zc(M)}$，这个组合的方差为：$\sigma_{MVP}^2=\alpha^2\sigma_M^2+(1-\alpha)^2\sigma_{zc(M)}^2$，由于 M 和 $zc(M)$的协方差为 0(由零贝塔得到)，故上式不含协方差项。为了找到最小方差的各组合所占比例，关于上式对 α 求导，并令其等于 0，得到：

$$\frac{d\sigma_{MVP}^2}{d\alpha}=2\alpha\sigma_M^2-2\sigma_M^2+2\alpha\sigma_{zc(M)}^2=0$$

解 α，得到 $\alpha=\frac{\sigma_M^2}{\sigma_M^2+\sigma_{zc(M)}^2}$，由于 σ_M^2 和 $\sigma_{zc(M)}^2$ 必然是正数，可能的最小方差组合包含的零贝塔组合和市场组合的权重必然是正数。由于 $E[R_{zc(M)}]<E[R_M]$，由 M 和 $zc(M)$构成的权重为正的期望收益必然高于 $E[R_{zc(M)}]$，所以 $zc(M)$在 MVP 之下，即零贝塔组合不可能是有效资产组合。

由于没有无风险借贷，每个人持有的资产组合都是高于 MVP 的风险资产前沿边界。而不是像简单的 CAPM 那样，仅仅持有市场组合和无风险资产的组合。类比前面，也有两基金分离定理：每个人可以持有市场组合基金和零贝塔组合基金，至于两个基金比例由每个投资者决定。

(2)存在无风险贷出但没有无风险借入的情形

如果我们允许无风险贷出，那么投资者的选择可以由图 3.4 表示，正如我们在第 2 章中所指出的，一个无风险资产和一个风险资产构成的所有组合都在连接该资产和组合的线上。最佳的组合位于经过无风险资产且与有效边界相切的直线上，即图中的 r_fT 直线。

注意我们把 T 画在 M 的下面，这不是巧合。由于位于 T 之下的风险资产组合不再会被投资者持有，因为在那种情景下，投资者会选择 T 组合和无风险资产的组合(r_fT 高于 T 之下的风险资产组合的前沿边界)。这样，投资者只会选择高于或等于 T 点的风险资产组合。对所有投资者进行加总，就得到市场组合高于 T 点。由于 M 点的斜率小于 T 点的斜率(因为有效前沿边界为凹的)，故 $E[R_{zc(M)}]>r_f$。这样我们解释了实证检验的第二条，即截距项大于 0，且由于零贝塔的斜率低于 r_fT 的斜率，我们也解释了实证检验的第三条(从图中可以看出切于市场组合的那条直线缓于 r_fT)。

新的有效边界为 r_fTMC。由于 T 点可由 M 和 $zc(M)$的线性组合来表示，这样 r_fT 就可以由无风险利率 r_f，市场组合 M 和 $zc(M)$的线性组合来表

示，而位于高于 T 点的纯粹风险资产组合，由前面论述，当然可以由 M 和 $zc(M)$ 的线性组合来表示。这样我们得到了三基金分离定理，即投资者只需要在无风险基金、市场组合基金、零贝塔组合基金中分配比率就行。

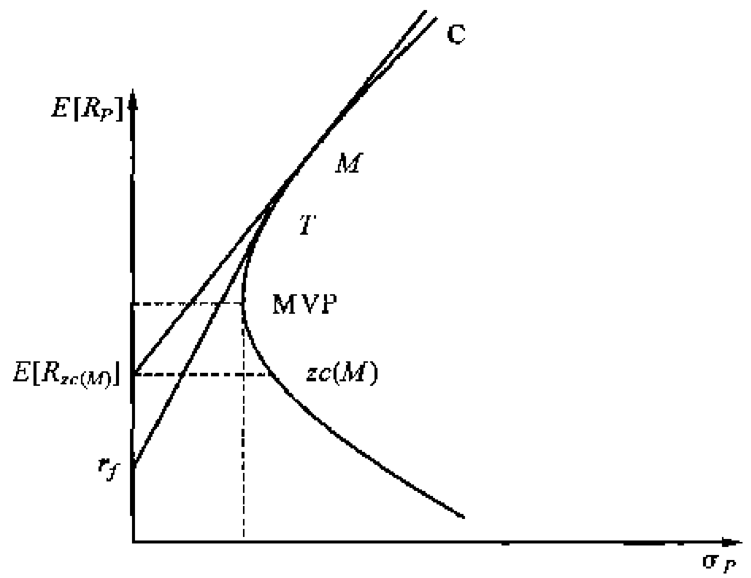


图 3.4　只有无风险贷出条件的情形的有效前沿边界

显然高于 T 点的证券市场线满足零贝塔 CAPM，故为直线。在图 3.5 中表示为 TC 的直线。在 T 点以下，由于新的前沿边界不再是风险资产前沿边界而是一条直线，故在图 3.5 中不可能和 TC 线在一条直线上。而风险资产与无风险资产的组合在期望收益—贝塔空间中为一条直线，我们已经知道 T 点、r_f 点，且已经证明了 r_f 在 $E[R_{zc(M)}]$ 的下面，故新的证券市场线可以表示为图 3.5。

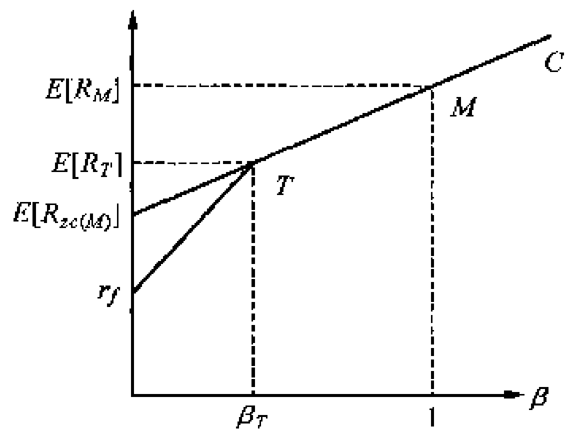


图 3.5　期望收益—贝塔空间中的证券市场线

(3)存在无风险贷出，也存在无风险借入，且借入利率高于贷出利率

现实生活中借入资金的成本往往高于贷出资金的收益。这样借入利率高于贷出利率。

如图3.6，借款利率为$E[R_B]$，贷款利率为$E[R_L]$，借款利率高于贷款利率。有效前沿边界为$E[R_L]L$直线加上LB之间的弧线段以及B之后的直线段。市场组合处在L和B之间。因为低于L点的风险资产组合不会被投资者选择(在那种情况下，投资者更愿意选择无风险资产和L的线性组合的资产组合)。同样高于B点的风险资产组合也不会被投资者选择。在L和B之间的风险资产组合的任意一点都可能被投资者选择。这样总的说来，市场组合一定在L和B之间。

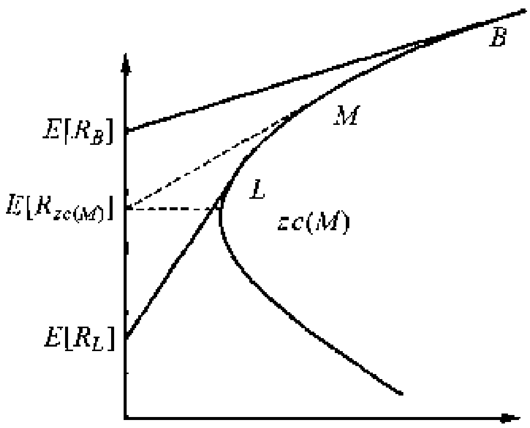


图3.6 借贷利率不同的前沿边界

四基金分离定理：风险资产组合的前沿边界均可以由市场组合及其零贝塔组合来表示。而$E[R_L]L$直线可以由L和无风险资产生成，而L可由市场组合和其零贝塔组合来生成。这样上述直线就可以由无风险资产、市场组合、零贝塔组合生成。LB之间的弧线段可以由市场组合及其零贝塔组合来表示。高于B点的直线为卖空无风险资产以及风险资产B组成，风险资产B由市场组合和其零贝塔组合来表示，这样它们也是由3个组合来表示。这样总的前沿边界就由贷款、市场组合、零贝塔组合、借款组成。

同存在无风险贷出但没有无风险借入的情形一样，我们也可以把上述思想表现在期望收益—贝塔空间中。

新的证券线为3段，低于L点为线段$E[R_L]L$；在L和B之间，为线段LMB；高于B点，为射线为BC。见图3.7。

3. 个人税收

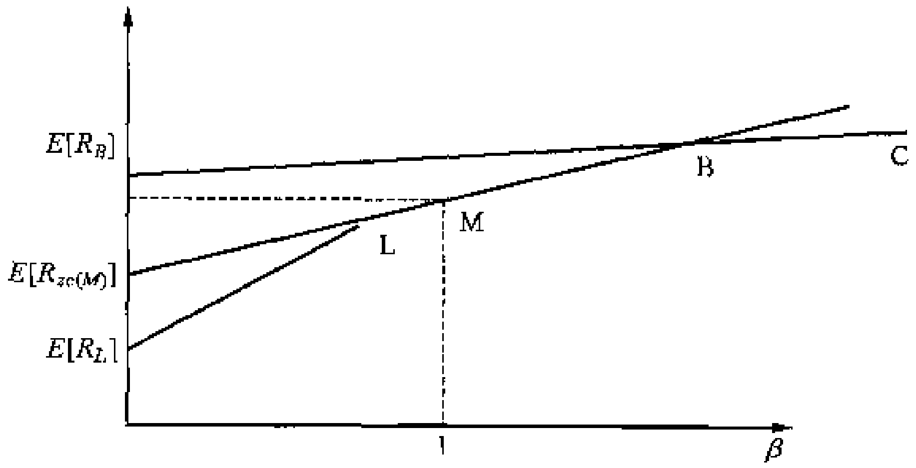


图 3.7 借贷利率不同的证券市场线

前面我们在讨论简单的 CAPM 模型时已经假设不存在税负,这个假设隐含了这样的一些意思:是以资本利得形式还是红利的形式得到收益,投资者对此是不关心的,并且所有投资者持有风险资产相同的投资组合。然而,税负是现实生活中的事实,它对于证券的定价是非常重要的。一般来说资本利得的税率要低于红利的税率。我们可以预料,税负不同的投资者将持有风险资产不同的投资组合。哪怕这些投资组合的税前期望回报率相同。

布伦南(Brennan)第一个研究了考虑资本利得与红利税负不一样时的资本资产定价模型。在建立税负调整的 CAPM 模型时,布伦南不仅使用了推导简单的 CAPM 模型时的一些常用假设,还假设红利收入是确定性的,考虑到税负不同的条件,资产或投资组合的回报率由下面的税负调整后的资本资产定价模型给出:

$$E[R_i]=r_f(1-T)+\beta_i[E[R_M]-r_f-T(D_M-r_f)]+TD_i$$

这里 $T=\dfrac{T_d-T_g}{1-T_g}$

其中:T_d 为红利的平均税率,T_g 为资本利得的平均税率,D_M 为市场组合的红利收益率,D_i 为股票 i 的红利收益率。

观察上式发现,当 $T_d=T_g$ 时,上式就变为简单的 CAPM。即简单的 CAPM 为上式中红利的税率与资本利得的税率相同时的特殊情形。现实中 $T_d>T_g$,这样 $T>0$,从而 D_i 的系数大于 0,即股票的红利收益率越高,$E[R_i]$ 越大,或者说股票的税前的必要收益率越大。

假设红利收益率大于无风险收益率,即 $D_M-r_f>0$ 和 $D_i>r_f$,则便得到:$E[R_M]-r_f-T(D_M-r_f)<E[R_M]-r_f$,$r_f(1-T)+TD_i=r_f+T(D_i-r_f)>$

r_f，这样新的资本资产定价模型的截距项比简单的 CAPM 高，而斜率要低于简单的 CAPM，这个和经验检验吻合。

如果证券的定价服从税负调整模式，那么投资者就会根据他们所处的所得税等级，对投资组合中持有或抛出高红利收益率的部分做出权衡。这就是说，投资者仍将持有像市场组合那样的充分多样化的投资组合，只是要向有比较优势的股票倾斜。这样每个人并不像简单的 CAPM 所指出的那样都持有市场组合。处于较高个人所得税等级的投资者应在投资组合中持有比市场组合比例更小的高红利股票，反之他们应持有更多的低红利高资本利得的股票，以实现他们税后回报率最大化。相应地，处于所得税等级较低的投资者将考虑在投资组合中倾向于高红利股票，这是因为这些股票的税负不利影响对他们来说比一般的持股者要小。

这样一种收益倾斜（yield tilt）策略在提高税后回报率方面有一定潜力，但是这种策略会带来额外的非系统性风险的成本。就是说，较之各种红利收益率水平充分多样化的投资组合而言，倾斜的投资组合很可能具有更大的非系统性风险。例如，许多高收益股票都是受管制的公用事业的股票，在整个股票市场层次上，它们的价格变动趋向一致。同样地，低收益的"成长性"股票变动趋势也一致。投资者要在收益倾斜带来的收益以及为此带来更大的未分散风险的成本之间做出合适的选择。

然而，证券定价中的税负效应的重要程度，甚至是否存在这种效应都存在争论。某些机构投资者的特定税收地位及投资者可用的平衡税负策略都有助于抵消税负对投资者回报率的影响，这样就减小了证券定价中的税负效应。

这些平衡力是否强有力到足以消除税负效应，实质上是一个实证性问题。一些有才能的研究者提出，这种效应存在；然而另一些同样有才能的研究者指出，在资本资产定价中，税负没有明显的影响。进一步说，即使是对税负效应持支持观点的研究工作也表明，这种效应的重要性非常有限，大约每年 30 个基点。执行收益倾斜策略好像主要依赖于投资者个人确信税负因素存在的强度。

4. 差别期望

在投资者期望不同的情形下，一些研究人员考察了一般均衡模型中解的存在性和特性。虽然所有这些模型导出的均衡定价方程的形式和简单的 CAPM 有相似之处，但它们也存在重要区别。均衡仍然可以用期望收益率、协方差和方差表示，但这些收益率、协方差和方差是不同投资者预期的复杂的加权平均。由于它们涉及投资者的效用函数的权重相当复杂，特别是它们包含关于投资者在期望收益和方差之间的权衡信息（边际替代率）。但多数效用

的权衡是一个对财富权衡的函数，因此是一个价格的函数。这意味着，价格会成为决定风险—收益权衡的因素，我们需要确定价格。因此一般来说，差别期望问题求出一个确定的解是不可能的。所以可以通过对投资者效用函数或者投资者面临的机会的特性加以额外的限制来简化这一问题。

3.4 单指数模型

上章我们讲到了马科维茨的投资组合理论，并分析了它的贡献以及不足。由于该模型需要估计的变量太多(主要是协方差的数目太多)，人们对它进行了各种简化。对它进行简化的第二个重要原因是大多数公司按照传统的行业分类划分它们的分析师关注的领域。一个证券分析师只跟踪一个行业的股票，但是投资组合分析要求这些证券分析师不仅要估计某一个行业的某一只股票，还要估计另一个行业的另一只股票。这样，均值方差模型所需要的相关性估计就成了重大问题。

多数金融机构需要跟踪 150～250 只股票。为了应用投资组合分析，机构需要估计 150～250 个期望收益以及 150～250 个方差。而估计的相关系数则为 11 175～31 125 个，这一数量让人瞠目结舌。所以对它的简化有其必然性。

3.4.1 单指数模型概述

随意观察股票价格就可以发现，当股市上涨时，大多数股票价格都会上涨，当股市下跌时大多数股票价格也倾向于下跌。这意味着，证券收益彼此相关的可能原因是对市场变动的共同反应。通过将股票收益和股票市场指数的收益联系起来，可以得到衡量相关性的有用指标。因此股票的收益可以写为：

$$R_i = a_i + \beta_i R_M \tag{3.10}$$

其中：a_i是证券 i 收益的一个组成部分，是独立于市场表现的随机变量；

R_M 为市场收益率，也是一个随机变量；

β_i 为一个常数，衡量 R_M 变化时 R_i 的期望收益变化。

这一等式将股票的收益简单地划分为两个部分：一部分来自市场的部分，另一部分独立于市场。β_i 衡量 R_i 对 R_M 的敏感程度。

a_i 代表证券 i 收益对市场不敏感的部分，可以将其划分为两个部分：α_i 代

表 a_i 的期望，ε_i 代表其受干扰项干扰的部分，显然 $E[\varepsilon_i]=0$。

故现在(3.10)可以改写为：

$$R_i=\alpha_i+\beta_i R_M+\varepsilon_i \tag{3.11}$$

令 R_M 和 ε_i 的标准差分别为 σ_M，σ_{ε_i}。到目前为止，我们还没有做出简单化的假设。为方便起见，我们令 R_M 和 ε_i 不相关。即

$$\mathrm{COV}(\varepsilon_i,R_M)=E[(\varepsilon_i-0)(R_M-E[R_M])]=0$$

上述假设意味着：证券收益的独特的风险因素(ε_i)独立于市场风险。α_i，β_i，σ_{ε_i} 是通过时间序列回归分析而得。到目前为止的所有单指数模型的特征都是定义的。单指数模型的关键假设是：对于所有的 i 和 j 而言，ε_i 和 ε_j 独立，即 $E[\varepsilon_i\varepsilon_j]=0$。这意味着，两股票同时变动的唯一原因是市场变动，除此之外，不存在其他原因导致两股票价格同时变动。但是，不同于 R_M 和 ε_i 不相关，没有任何用来估计 α_i，β_i，σ_{ε_i} 的回归方法能保证这一点。它只是简化现实的一个粗略假定。

在接下来的部分，我们推导在用单指数模型描述证券协同运动情况下的期望收益、标准差和协方差：

1. 收益均值为：$E[R_i]=\alpha_i+\beta_i E[R_M]$

2. 收益方差为：$\mathrm{VAR}[R_i]=E[\alpha_i+\beta_i R_M+\varepsilon_i-E[\alpha_i+\beta_i R_M+\varepsilon_i]]^2=\beta_i^2\sigma_M^2+\sigma_{\varepsilon_i}^2$

3. 证券 i 和 j 收益的协方差为 $\sigma_{ij}=\beta_i\beta_j\sigma_M^2$

注意：推导上述公式要用到上述的两个假设：R_M 和 ε_i 不相关，以及 ε_i 和 ε_j 独立，再根据方差协方差的性质易得出上述结果。

【例 3.2】已知某股票和市场收益的 5 个月份的数据如表 3.1 所示。

表 3.1

单位：%

月　份	股票收益	市场收益	ε_i
1	10	4	$10-2.6-1.2\times4=2.6$
2	3	2	$3-2.6-1.2\times2=-2$
3	12	8	$12-2.6-1.2\times8=-0.2$
4	9	6	$9-2.6-1.2\times6=-0.8$
5	3	0	$3-2.6-1.2\times0=0.4$
总收益	37	20	10.4
平均收益	7.4	4	0
收益方差	17.3	10	2.7

由上述图表知，R_i 的平均收益为 7.4%，R_M 的平均收益为 4%，而 $E[R_i]=2.6+1.2E[R_M]$，用图表算出来的数据和用回归线得到的吻合，满足公式 $E[R_i]=\alpha_i+\beta_i E[R_M]$。$\sigma_{\varepsilon_i}^2=2.7$，$\sigma_M^2=10$，$\mathrm{VAR}[R_i]=17.3$。

经用上述数据检验 $\mathrm{VAR}[R_i]=\beta_i^2\sigma_M^2+\sigma_{\varepsilon_i}^2$ 成立。

对上述表，可以画图如图 3.8 所示。

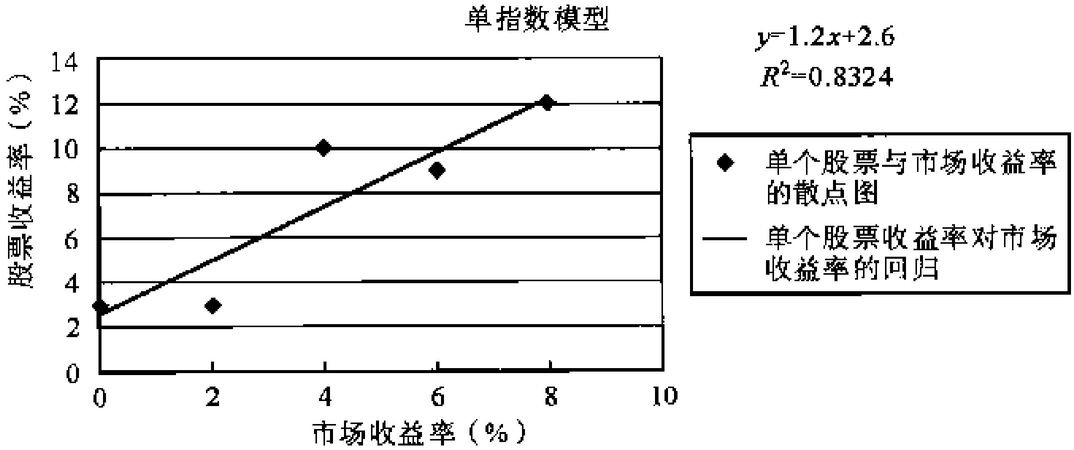


图 3.8

在介绍完这一简单例子后，在单指数模型成立下，我们可以转向计算任何投资组合的期望收益和方差。投资组合的期望为：

$$E[R_P]=\sum_{i=1}^{n}X_iE[R_i]$$

对上式用 $E[R_i]=\alpha_i+\beta_i E[R_M]$ 替换 $E[R_i]$，化简可得：

$$E[R_P]=\sum_{i=1}^{n}X_i\alpha_i+\sum_{i=1}^{n}X_i\beta_iE[R_M] \tag{3.12}$$

由前面章节，我们知道投资组合的方差为：

$$\sigma_P^2=\sum_{i=1}^{n}X_i^2\sigma_i^2+\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}X_iX_j\sigma_{ij}$$

用前面的方差、协方差公式代入上式得到：

$$\sigma_P^2=\sum_{i=1}^{n}X_i^2\beta_i^2\sigma_M^2+\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}X_iX_j\beta_i\beta_j\sigma_M^2+\sum_{i=1}^{n}X_i^2\sigma_{\varepsilon_i}^2 \tag{3.13}$$

从式(3.12)，(3.13)可知，如果我们能估计出每只股票的 α_i，β_i，$\sigma_{\varepsilon_i}^2$，以及市场的期望收益率 $E[R_M]$和风险 $E[\sigma_M^2]$，则投资组合的期望收益和风险就清

楚了。所以由上述，总共需要 $3n+2$ 个估计。前面讲到，一般投资机构都要跟踪 150～250 只股票，则用单指数模型需要 452～752 个估计。而在没有简化之前，则需要 11 175～31 125 个协方差的估计和 11 475～31 625 个总的估计。而且每个分析师只需要观察自己行业的股票和大盘指数的关系即可，而不需要观察其他行业的股票与本行业股票的关系。

3.4.2 单指数模型的特点

定义投资组合的贝塔为 β_P，它是投资组合中每只股票的 β_i 的加权平均，权重为每只股票在投资组合中所占比重，即：

$$\beta_P = \sum_{i=1}^{n}\beta_i X_i$$

同样可知 $\alpha_P = \sum_{i=1}^{n}\alpha_i X_i$，则对投资组合来说：

$$E[R_P]=\alpha_P+\beta_P E[R_M]$$

显然上式对市场组合也成立，代入市场组合 $E[R_M]$，可知 $\alpha_P=0$，$\beta_P=1$。这样推知市场组合的贝塔为 1。这条很重要。

一个特例：假设投资组合的每只股票的权重都相等，

$$\sigma_P^2 = \sum_{i=1}^{n}X_i^2\beta_i^2\sigma_M^2 + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}X_iX_j\beta_i\beta_j\sigma_M^2 + \sum_{i=1}^{n}X_i^2\sigma_{\varepsilon_i}^2\alpha_i$$，显然上式可以写成：

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j\beta_i\beta_j\sigma_M^2 + \sum_{i=1}^{n}X_i^2\sigma_{\varepsilon_i}^2$$

重新整理可得：

$$\sigma_P^2 = \left(\sum_{i=1}^{n}X_i\beta_i\right)\left(\sum_{j=1}^{n}X_j\beta_j\right)\sigma_M^2 + \sum_{i=1}^{n}X_i^2\sigma_{\varepsilon_i}^2$$

因而，投资组合的风险可以表示为：

$$\sigma_P^2 = \beta_P^2\sigma_M^2 + \sum_{i=1}^{n}X_i^2\sigma_{\varepsilon_i}^2 = \beta_P^2\sigma_M^2 + \frac{1}{n}\left(\sum_{i=1}^{n}\frac{1}{n}\sigma_{\varepsilon_i}^2\right)$$

看最后一项，它表示$\frac{1}{n}$乘以投资组合的平均残差。随着投资组合中股票数量的增加，平均残差的重要性剧烈下降。我们持有的越来越大的投资组合

中没有被消除的风险涉及 β_P 项。如果假定残余风险接近于0，则组合风险接近于

$$\sigma_P=\beta_P\sigma_M=\sigma_M\left[\sum_{i=1}^{n}X_i\beta_i\right]$$

σ_M 是相同的，则无论我们考察哪只股票，则该证券对大型投资组合风险的贡献为 β_i。

在这里我们可以看到前面叙述的观点。由于单个证券的风险为 $\beta_i{}^2\sigma_M^2+\sigma_{\epsilon_i}^2$，由于在投资组合变大时，$\sigma_{\epsilon_i}^2$ 对投资组合风险的影响趋于0，通常将 $\sigma_{\epsilon_i}^2$ 称为可分散风险。但是，当 n 变大时，$\beta_i{}^2\sigma_M^2\alpha_i\beta_{i2}=0.343+0.677\beta_{i1}\sigma_{\epsilon_i}^2$ 对投资组合的影响并不会消除。由于 σ_M^2 对所有证券都是常数，β_i 是衡量非分散化风险的指标。由于可分散风险可以通过持有大量证券而予以消除，β_i 通常被用来衡量证券风险。

3.4.3 贝塔估计

应用单指数模型要求估计潜在包含在投资组合中的每只股票的贝塔。证券分析师被要求提供每只证券或投资组合的主观贝塔估计。另一方面，对未来贝塔的估计首先利用历史数据估计历史贝塔，尔后再利用历史贝塔估计未来贝塔。有证据表明，历史贝塔能为未来贝塔提供有用信息。此外一些有趣的预测技术已经发展起来，它们可以增加从历史数据中获得的信息。

1. 估计历史贝塔

$R_i=\alpha_i+\beta_iR_M+\epsilon_i$ 被认为在任何时点上都成立，尽管 α_i，β_i，$\sigma_{\epsilon_i}^2$ 的值可能随时间而改变。当观察历史数据时，α_i，β_i，$\sigma_{\epsilon_i}^2$ 的值是不能被直接观察到的。我们能观察到的是证券和市场的收益。如果 α_i，β_i，$\sigma_{\epsilon_i}^2$ 在时间范围内被假定是常数，则同一方程在时间上的每一点都成立。在这一情况下，就存在直接估计 α_i，β_i，$\sigma_{\epsilon_i}^2$ 的程序。通常我们用回归分析方法来估计直线的位置。

2. 历史贝塔的精确度

从逻辑上讲，检查贝塔的第一步是，观察某一时期贝塔与相邻时期贝塔的关联。布鲁姆(Blume，1970)和莱维(Levy)都对不同时期贝塔的关联性做了大量实证研究。布鲁姆用月度时间序列数据回归了贝塔，回归期限是不交叉的7年。他分别估计出只含1只股票、2只股票、4只股票，直至50只股票的

投资组合的贝塔。对每种规模的组合，它都考察了一个时期的贝塔和下一时期贝塔的相关性。表 3.2 显示了结果。

从表 3.2 中，可以看到，较大的投资组合的贝塔包含了较多的这一组的未来贝塔信息。单个证券贝塔关于未来贝塔的信息则要少得多。为什么一个时期观察的贝塔与下个时期观察的贝塔有差异呢？其中一个原因可能是证券或组合的风险会改变；另一个原因是，每一时期的贝塔衡量有随机误差，随机误差越大，用一个时期的贝塔预测下一个时期的贝塔的能力就越弱。

证券贝塔的变化随证券不同而有所差异。有些会上升，有些会下降。这些变动在组合中倾向于相互抵消，因而我们发现，组合贝塔的变化要比单个证券的贝塔变化小。

而且，当证券被混合时，单个证券估计时的误差会相互抵消，因而，组合贝塔的误差要小一些。由于组合贝塔的误差要小一些，且组合贝塔的变化比单个证券贝塔的变化小一些，所以在预测未来贝塔的能力方面，组合的历史贝塔要比单个证券的历史贝塔强一些。

表 3.2　不同时期贝塔的关联

投资组合中的证券数量	相关系数	决定系数
1	0.60	0.36
2	0.73	0.53
4	0.84	0.71
7	0.88	0.77
10	0.92	0.85
20	0.97	0.95
35	0.97	0.95
50	0.98	0.96

3.调整历史估计

由上述可知，连续两期的贝塔存在相关性，但是不是完全相关，且大证券组合的相关性要高于小证券组合的相关性，这样，估计出历史贝塔必然要对其调整才行。我们发现对所有股票而言，用历史贝塔预测未来贝塔要逊色于用 1 来预测所有贝塔。现在假设不同的股票有不同的贝塔。我们估计的贝塔一部分是真实贝塔的函数，一部分是样本误差的函数。如果我们对一只股票计

算出非常低的贝塔，则负样本误差的可能性在增大。如果这种情况是正确的，则我们应该发现，平均来看，贝塔在连续的时间里向1收敛。实际上，这正是布鲁姆(1975)和莱维(1971)的研究结果。

4. 布鲁姆技术

因为预测期的贝塔比根据历史数据得到的估计值更接近于1，下一步显然是通过修正过去贝塔来体现这一趋势。布鲁姆(1975)第一个提出了这样做的计划。他是通过直接度量这种向1的调整，并假定一个时期的调整是下一时期调整的良好估计，来修正历史贝塔。

5. 瓦西切克(Vasicek)技术

前文已经说明，预测期的实际贝塔比根据历史数据得到的估计值更接近于平均贝塔。这样我们可以通过直接调整历史贝塔使其向平均贝塔靠拢。例如，取历史贝塔的一半加上平均贝塔的一半，能使历史贝塔向平均贝塔部分调整。

然而，并不是所有股票都以相同量向平均值调整，调整应按贝塔的不确定性(样本误差)的大小进行。样本误差越大，与平均值相差悬殊的可能性越大，出于抽样误差所需的调整就越大。瓦希切克(Vasicek ，1973)提出了以下具有这样特性的调整计划：如果我们以$\overline{\beta_1}$代表历史时期样本股票的平均贝塔，则瓦西切克就是想将$\overline{\beta_1}$和证券i的历史贝塔进行加权平均。令$\sigma^2_{\overline{\beta}_1}$代表样本股票的历史估计贝塔分布的方差。令$\sigma^2_{\beta_{i1}}$代表股票$i$历史估计贝塔分布的方差。

6. 调整贝塔的准确性

我们可以来考察布鲁姆和瓦西切克调整技术用于预测的效果。克莱姆考斯基和马丁(Klemkosky and Mardin ，1975)检验了这些技术在3个5年期里对包含1只股票和10只股票的投资组合的预测能力。如同所预想的，在所有例子中，布鲁姆和瓦西切克调整技术比未经调整的历史贝塔更准确。当采用一种调整技术后，预测贝塔的平均误差平方通常减少了一半。他们使用了一种有趣的方法来寻找预测误差的来源。具体而言，误差来源分为错误估计平均贝塔水平，对高贝塔高估和低贝塔低估的倾向，以及前两种影响都不能解释的部分。如同我们所预料的，当将布鲁姆和贝叶斯调整技术与未调整贝塔比较时，几乎所有的误差缩减都来源于降低对高贝塔高估和低贝塔低估。这并不令人奇怪，因为这就是两种技术设计的初衷。他们发现，瓦西切克技术要优于布鲁姆技术，但两者差异很小。

3.4.4 市场模型

尽管单指数模型是为了帮助资产管理而发展出来的，但一种限制更少的形式——市场模型在金融中的使用正在增加。除了没有假定 $COV(\varepsilon_i,\varepsilon_j)=0$，市场模型与单指数模型是相同的。

由于它没有假定股价的协方差源于一个共同的对市场的协方差，所以并不能得出单指数模型那样的投资组合风险的简单表达式。

3.5 多指数模型

单指数模型背后的假设是，股票价格一起变动的唯一原因是随市场共同变动。一些研究者发现，除了市场之外，还有其他影响因素导致股票价格一起变动。多指数模型试图抓住引起协同运动的非市场因素。寻找非市场影响因素就是寻找一组经济因素，它们说明了市场指数没有说明的股票价格的共同变动。找到一套在任何时期内与非市场影响相联系的指数并不困难，但是，如同我们将看到的，找出能成功预测与市场无关的协方差就另当别论。

3.5.1 一般多指数模型

任何证券间协方差的额外来源都可以纳入风险与收益的方程，只需将这些额外影响因素加到一般收益方程中。让我们假设任何股票的收益都是市场收益、利率变动和一系列工业指数的函数。如果 R_i 是股票 i 的收益，那么股票 i 的收益可以与其有影响的因素以下列方式联系：

$$R_i=a_i^*+b_{i1}^*I_1^*+b_{i2}^*I_2^*+\cdots+b_{iL}^*I_L^*+c_i$$

在这一方程中，I_j^* 是指数 j 的实际水平，b_{ij}^* 是度量股票 i 的收益对指数 j 的变动的反应指标。因而，b_{ij}^* 的含义与单指数模型中的 β_i 是相同的。如同单指数模型的情况，证券收益与指数不相关的部分可以分为两部分：a_i^*，c_i。a_i^* 是特有收益期望值，这与其在单指数模型中的含义相同。c_i 是特有收益的随机部分，其均值为 0，方差表示为 $\sigma_{c_i}^2$。

尽管本模型可以直接应用，但是如果指数是不相关的（正交），则模型有更

多方便的特性。这能让我们简化风险的计算和优化投资组合的选择。幸运的是，由于任何一组相关指数都可以转化为一系列不相关的指数，这不会带来理论问题。用这一方法，方程可以被重写为：

$$R_i = a_i + b_{i1} I_1 + b_{i2} I_2 + \cdots + b_{iL} I_L + c_i$$

其中 I_j 彼此无关。新的指数仍然具有经济含义。假定 I_1^* 是股票市场指数，I_2^* 是利率指数。I_2 现在变为实际利率与给定股票收益率(I_1)时的利率的差。

令残差与每个指数不相关也是方便的。正式地讲，这意味着 $E[c_i(I_j - \overline{I_j})] = 0$。这样构建的含义是，上式描述了任何证券收益的能力独立于假定指数的任意取值。

多指数模型的假定是 $E[c_i c_j] = 0$，其中 $i = 1, 2, 3, \cdots, n$，$j = 1, 2, 3, \cdots, n$。这一假定意味着，股票一同变动的唯一原因是，与模型中指定的一套指数共同变动。在这些指数以外，没有任何原因能说明股票之间的协同变动。这是一种简化，它代表对现实的近似。模型的表现取决于近似的好坏，而这又取决于我们选择的代表协同运动的指数是否能真正捕捉到证券间的协同运动模式。

当使用多指数模型描述证券结构时，我们可以给出收益期望值、方差和证券间协方差的表达式（推导过程同单指数模型相同，要用到期望收益、方差和协方差的性质，以及 $E[c_i(I_j - \overline{I_j})] = 0$ 和 $E[c_i c_j] = 0$)，它们等于：

1. 期望收益为：$E[R_i] = a_i + b_{i1} E[I_1] + b_{i2} E[I_2] + \cdots + b_{iL} E[I_L]$
2. 收益方差为：$\sigma_i^2 = b_{i1}^2 \sigma_{I1}^2 + b_{i2}^2 \sigma_{I2}^2 + \cdots + b_{iL}^2 \sigma_{IL}^2 + \sigma_{c_i}^2$
3. 证券 i 和 j 的协方差为：$\sigma_{ij} = b_{i1} b_{j1} \sigma_{I1}^2 + b_{i2} b_{j2} \sigma_{I2}^2 + \cdots + b_{iL} b_{jL} \sigma_{IL}^2 + \sigma_{c_i}^2$

从上面三式可知，如果已经估计出了每只股票的 a_i，以及关于每一因素的 b_{ik}，每只股票特有的 $\sigma_{c_i}^2$ 以及每个指数的均值 $E[I_j]$和方差 σ_{Ij}^2，则我们显然就能估计任何投资组合的期望收益和风险。这总共是 $2n + 2L + Ln$ 个估计。对应用 10 个指数跟踪 150～250 只股票的机构而言，这需要 1 820～3 020 个输入数据。这大于单指数模型要求的输入数据数量，但远远小于没有简化假设情况下的输入数据。注意，证券分析师现在必须估计他们所跟踪的股票对几个行业的变化的反应程度。

还有一种多指数模型引起了广泛关注。这类模型将注意力限定在市场和行业影响。不同的行业模型源于它们对收益行为的假设不同，因而，它们在需要输入数据的类型和数量方面有所不同。我们现在考察这些模型。

3.5.2 行业指数模型

几位学者处理多指数模型时是从基本的单指数模型开始的，尔后加上捕捉行业影响的指数。

如果我们假定证券间的相关性取决于市场效应和行业效应，我们的一般多指数模型可以写为：

$$R_i = a_i + b_{iM}I_M + b_{i1}I_1 + b_{i2}I_2 + \cdots + b_{iL}I_L + c_i$$

式中I_M——市场指数；

I_j——行业指数，它与市场不相关，且彼此不相关。

这一模型背后的假定是，公司的收益受市场和几个行业的影响。对一些公司而言，这看起来时合适的，因为它们的业务跨越了几个传统行业。但是，有些公司的主要利润来源于一个行业，也许更重要的是，它们被投资者视为某一特定行业。在这一情况下，它们所不属于的行业的指数对公司收益影响就较小，而包含这些指数会带来更多的随机噪音，这多于它们提供的信息。这促使研究人员建议使用一种更简单形式的多指数模型：假设公司收益只受一个市场指数和一个行业指数的影响。此外，该模型还假定每个行业指数被构建为与市场无关，且与其他行业指数无关。对在行业 j 的公司 i 而言，收益方程可以写为：

$$R_i = a_i + b_{iM}I_M + b_{ij}I_j + c_i$$

如果公司处于同一行业，则证券 i 和证券 k 的协方差可以写为：

$$b_{iM}b_{kM}\sigma_M^2 + b_{ij}b_{kj}\sigma_{Ij}^2$$

对于不同行业的公司，则

$$b_{iM}b_{kM}\sigma_M^2$$

可以看出来，投资组合选择所需的输入数据数量被消减至 $4n+2L+2$。

3.6 APT 的推导

APT 是对资产定价的一种新颖而不同以往的方法。该模型的基础是一价定律：相同的两种物品不能以不同的价格出售。推导 CAPM 时，对效

用函数所做的强假设(递增的和严格凹的,即投资者贪得无厌和厌恶风险)不再是必要条件。事实上,与 CAPM 中只简单化地受到均值和方差影响这一假定相比,套利定价所描述的均衡更加一般化,但共同预期这一假设是必要的。证券收益产生过程的假设取代了假设投资者使用均值—方差的分析框架。

APT 要求任何股票的收益过程为:

$$R_i = E[R_i] + b_{i1}I_1 + b_{i2}I_2 + \cdots + b_kI_k + \varepsilon_i \tag{3.14}$$

其中 ε_i 为均值为 0,方差为 $\sigma^2_{\varepsilon_i}$ 的随机误差项。I_j 是均值为 0 的指数。

为了使该模型完全地描述证券收益产生的过程,有:

$\mathrm{COV}(\varepsilon_i,\varepsilon_j)=0, \forall i, \forall j$,且 $i \neq j$,$\mathrm{COV}(I_i,\varepsilon_i)=0, \forall i$。

APT 的贡献就是描述了如何(以及在什么条件下)由多指数模型推导出均衡状态。APT 的一般假设条件是完全竞争的无摩擦的资本市场,进一步地说,个人被假定有相同的信念:所考虑的资产集的随机收益由公式(3.14)给出。理论上要求所考虑资产数目 n 要比指数数目 k 更大。

APT 利用均衡中不存在套利机会而给出证券的定价。什么是无套利机会?通俗地说,就是如果一项投资成本为 0,且没有任何风险,则最终的收益也为 0。否则,假设最终的收益为正的,则大家都会发现,并去套利,结果最终使低估的证券价格上升,高估的价格证券价格下降,最终收益变为 0。

设初始 n 份资产的资产组合 P 的权重为 X_i,由于初始所花成本为 0,即

$$\sum_{i=1}^{n} X_i = 0 \tag{3.15}$$

$$\begin{aligned} R_P &= \sum_{i=1}^{n} X_iR_i \\ &= \sum_{i=1}^{n} X_iE[R_i] + I_1\sum_{i=1}^{n} X_ib_{i1} + I_2\sum_{i=1}^{n} X_ib_{i2} + \cdots + I_k\sum_{i=1}^{n} X_ib_{ik} + \sum_{i=1}^{n} X_i\varepsilon_i \end{aligned} \tag{3.16}$$

可以证明

$$E[R_i] = \lambda_0 + \lambda_1 b_{i1} + \cdots + \lambda_k b_{ik}$$ (3.17)(见附录 3-1)

b_{ij} 是第 i 个证券的收益对于第 j 个指数的"敏感度"。如果有无风险收益率为 r_f 的无风险资产,那么有 $r_f = \lambda_0 + \lambda_1 \times 0 + \cdots + \lambda_k \times 0$,这样有 $r_f = \lambda_0$。

图 3.9 说明了公式(3.17)假定的只有单一随机要素 k 的套利定价关系。在均衡中,所有的资产必须落在套利定价线上。

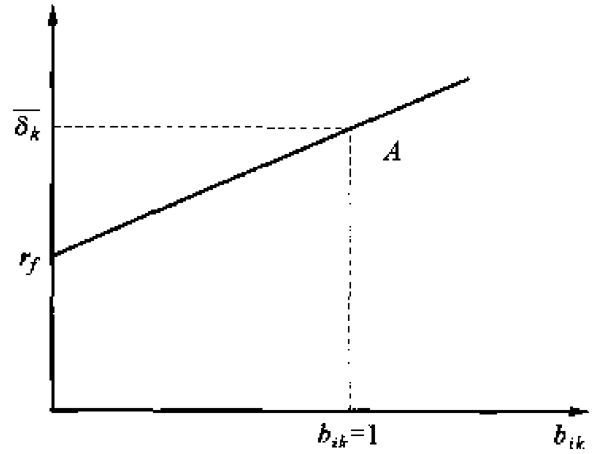


图 3.9 套利定价线

令$\overline{\delta_j}$表示第 j 个指数敏感程度为 1,其他指数敏感程度为 0 的资产的期望收益。

$\overline{\delta_j}=r_f+\lambda_j\times 1$,这样 $\lambda_j=\overline{\delta_j}-r_f$,即 λ_j 代表上述资产的超额收益,或者说是第 j 个指数的超额收益。

这样公式 $E[R_i]-r_f=\lambda_1 b_{ij}+\cdots+\lambda_k b_{ik}$ 可以表示为:

$$E[R_i]-r_f=(\overline{\delta_1}-r_f)b_{i1}+(\overline{\delta_2}-r_f)b_{i2}+\cdots+(\overline{\delta_k}-r_f)b_{ik} \tag{3.18}$$

如果(3.18)被解释为线性回归方程(假设收益的向量有联合正态分布,而且指数已经被线性变换,以至于被变换后的指数是正交的),那么系数 b_{ij} 可如同 CAPM 的 β 一样的方式定义为:

$$b_{ij}=\frac{\mathrm{COV}(R_i,\delta_j)}{\mathrm{VAR}(\delta_j)}$$

式中,$\mathrm{COV}(R_i,\delta_j)$为第 i 个资产收益和第 j 个指数的协方差,$\mathrm{VAR}(\delta_j)$为第 j 个指数的方差。

因此,CAPM 被看作 APT 的特例(资产收益被假定为联合正态)。APT 是 CAPM 的拓展,这是因为:

1. APT 没有对资产收益的经验分布做任何假设。

2. APT 没有对个人效用函数做任何假设(至少没有比贪婪和风险厌恶更强的假设)。

3. APT 允许资产收益取决于许多要素,而非只是一个要素。

4. APT 对任何资产子集的相对定价进行了说明，因此，不必为了检验理论而在整个资产范围进行计量。

5. 在 APT 中，市场投资组合没有什么特殊地位，而 CAPM 要求市场投资组合必须有效。

6. APT 很容易扩展到多种因素框架。

由此可见，APT 是比 CAPM 更具有一般性，更容易为人所接受的资本市场均衡理论。

我们需要强调一下，使用罗尔和罗斯的程序并找到显著不为零的 λ_j 不止一个，并不能充分地拒绝资本资产定价模型。下面以双指数模型来说明：

$$R_i = a_i + b_{i1} I_1 + b_{i2} I_2 + \varepsilon_i$$

在存在无风险资产的条件下，基于上述双因素收益产生过程的套利定价模型的均衡模型为：

$$E[R_i] = r_f + b_{i1}\lambda_1 + b_{i2}\lambda_2 + \varepsilon_i$$

回忆一下，如果资本资产定价模型是均衡模型，它适用于所有的证券，同样也适用于所有的证券组合。假设指数可以通过证券组合加以表示。事实上，我们已经看到 λ_j 是组合的超额收益，该组合的一个指数的 b_{ij} 为 1 而其他指数的 b_{ij} 为 0。如果资本资产定价模型成立，则对每一个均衡收益可由资本资产定价模型给定如下式

$$\lambda_1 = \beta_{\lambda 1}(E[R_M] - r_f)$$
$$\lambda_2 = \beta_{\lambda 2}(E[R_M] - r_f)$$

把上面两式子代入 $E[R_i] = r_f + b_{i1}\lambda_1 + b_{i2}\lambda_2 + \varepsilon_i$ 得到：

$$\begin{aligned} E[R_i] &= r_f + b_{i1}\beta_{\lambda 1}(E[R_M] - r_f) + b_{i2}\beta_{\lambda 2}(E[R_M] - r_f) \\ &= r_f + (b_{i1}\beta_{\lambda 1} + b_{i2}\beta_{\lambda 2})(E[R_M] - r_f) \end{aligned}$$

定义 β_i 为 $(b_{i1}\beta_{\lambda 1} + b_{i2}\beta_{\lambda 2})$，这样 $E[R_i] = r_f + \beta_i(E[R_M] - r_f)$

所以如果 λ_j 并不显著的不同于 $\beta_{\lambda j}(E[R_M] - r_f)$，这一实证结果可以与资本资产定价模型的 CAPM 形式完全一致。即完全有可能出现这种情况，有不止一个指数可以解释证券收益间的协方差，但资本资产定价模型仍然成立。

3.7 多指数模型和套利定价模型的应用

在选择证券、管理和评估组合这些领域中，多指数模型和多指数均衡模型

(套利定价模型)的使用迅速增加。许多公司、金融机构和金融咨询公司开发出自己的多指数模型,用以辅助投资过程。这些模型越来越受欢迎,因为这些模型使得风险控制更加严格,使得投资者避开自身敏感的具体风险或者对某些风险下具体的赌注。

在这一小节中,我们将讨论套利定价模型和多指数模型在辅助消极管理和积极管理中的应用。在此之前,我们首先简单地回顾一下多指数模型和套利定价模型,给出一个简单的套利定价模型的例子,在本节讨论中将用该模型说明一些现象。

我们在本章前面部分介绍的收益产生过程为:

$$R_i = E[R_i] + b_{i1}I_1 + b_{i2}I_2 + \cdots + b_{ik}I_k + \varepsilon_i \tag{3.19}$$

其中 ε_i 为均值为 0,方差为 $\sigma_{\varepsilon_i}^2$ 的随机误差项。I_j 是均值为 0 的指数。

我们可以看到可由上式得出期望收益表达式:

$$E[R_i] = r_f + \lambda_1 b_{i1} + \lambda_2 b_{i2} + \cdots + \lambda_k b_{ik} \tag{3.20}$$

将(3.19)代入到(3.20)得到:

$$R_i = r_f + \lambda_1 b_{i1} + \lambda_2 b_{i2} + \cdots + \lambda_k b_{ik} + b_{i1}I_1 + b_{i2}I_2 + \cdots + b_{ik}I_k + \varepsilon_i \tag{3.21}$$

有若干种方法确定式(3.21)中的 I_j,b_{ij} 和 λ_j,但我们只以其中一个简单方法为例来说明这种模型的使用。

我们假设识别了收益产生过程(3.19)中的 4 种影响如下:

I_1——通货膨胀率的非预期变化;

I_2——总销售额的非预期变化;

I_3——石油价格的非预期变化;

I_4——排除上述 3 种影响以后的标准普尔指数的收益。

进一步,假设石油价格没有被定价($\lambda_3 = 0$)。式(3.20)变为:

$$E[R_i] = r_f + \lambda_1 b_{i1} + \lambda_2 b_{i2} + \lambda_4 b_{i4}$$

同时(3.21)变为:

$$R_i = r_f + \lambda_1 b_{i1} + \lambda_2 b_{i2} + \lambda_4 b_{i4} + b_{i1}I_1 + b_{i2}I_2 + b_{i3}I_3 + b_{i4}I_4 + \varepsilon_i$$

将模型参数化,可以使我们清晰地看到任何因素在决定其对标准普尔指数的超额期望收益时的重要性。为将模型参数化,只需简单地将与该因素相关的 b 乘以相应的风险价格(λ)。

表 3.3 说明了标准普尔超额期望收益为 8.09%。销售增长对标准普尔

超额期望收益的贡献为2.54%。

表3.3

因素	b	λ(%)	对标准普尔超额收益的贡献(%)
通胀	-0.37	-4.32	1.59
经济增长	1.71	1.49	2.54
石油价格	0.00	0.00	0.00
市场	1.00	3.96	3.96
标准普尔指数股票组合的超额期望收益			8.09

这种类型的分析可用于考察风险来源对于任何证券或组合的超额期望收益的重要性。例如：b，λ 和对增长型股票组合超额期望收益的贡献率，如表3.4所示。

表3.4

因素	b	λ(%)	对增长型超额收益的贡献(%)
通胀	-0.50	-4.32	2.16
经济增长	2.75	1.49	4.10
石油价格	-1.00	0.00	0.00
市场	1.30	3.96	5.15
增长型股票组合的超额期望收益			11.41

注意，增长型股票组合的超额期望收益(11.41%)高于对标准普尔指数的(8.09%)。这并不奇怪，因为增长型股票组合与标准普尔组合相比，对每一个指数都具有更高的风险。对于增长型股票组合和标准普尔指数而言，单个影响(指数)对超额期望收益的绝对和相对贡献是不相同的。例如，销售增长对超额期望收益的贡献为4.10%。销售增长的贡献占到增长型股票超额期望收益的35.9%。但增长型股票对所有重要指标都非常敏感，而且一般来说是这样的，所以这才更加令人感到惊讶，虽然对销售增长的敏感度对超额期望收益的作用最大，但是所有影响的变化导致了更大的超额收益。

现在，我们转向讨论这一模型在投资和组合管理方面的应用。组合经理人可分为消极经理人和积极经理人。消极经理人认为不能识别出错误定价的

证券，因而他试图持有并模仿某种股票组合。最常见的消极管理的方法就是，选定某一个指数，并持有一个锁定该指数的组合。积极管理则是根据对一个或多个证券错误定价的判断来设计组合，并在这些证券或组合上进行投机。

3.7.1 消极管理

多指数模型对优化消极管理有重要作用，它可用于追踪一个指数或者设计适合特定客户的消极组合。

多指数模型的最简单应用就是紧密追踪一个指数以创造一个股票组合。也许大家会说，构建一个指数基金有什么难的，只需要按照其指数比例所持有相同比例的股票就行了，为什么还需要多指数模型？话虽如此，但是事实上，由于各种原因（下面会介绍），许多指数基金并没有简单地根据指数中的比例来持有指数中的全部股票，而只是持有少量股票来复制出指数。一个指数包含的股票越多，公司在指数中所占的比例就越小，该指数中股票的流动性就越小，以指数比例购买股票就越来越接近该指数。显然，一旦涉及追踪某一代表市场相当大部分的指数，使用完全相等的比例就越来越不合适。可以使用单指数模型创造指数基金，只需找到能够使所需指数的贝塔为 1，规模给定的组合的残差风险最小的组合。

不采用单指数模型，而采用多指数模型，可以创造出更加匹配目标指数的一个指数基金。由于多指数模型找到了导致收益波动的更多的原因，所以构建一个良好的多指数模型能更好地匹配指数。而单指数模型只是匹配市场收益风险。另外，仅仅与市场风险保持匹配，可能仍会有这样的情况：组合和指数在除去市场收益后关于影响二者的其他共同因素有不同的敏感性（b_{ij}）。例如，它们关于通货膨胀的敏感度不同。

一般而言，一个组合包含的股票越少，它与目标指数匹配程度就越低，而使用多因素组合相对于单指数组合的优势就更加突出。之所以会如此，是因为当关于这些缺失指数的敏感性不能保持不变时，缺失指数的非预期变化对未来时期的残差风险的影响是不同的。

构建的组合常常是包含少量股票的组合的原因有：

(1)公司常构建组合作为套利组合，用于指数期权或期货交易。而这时，公司需要建立一个小篮子（25 或 30 只）股票，以便在公司期权或期货头寸变化时进行积极的交易。篮内股票数量少的原因是因为它们经常被买卖。在这种情况下，多指数模型的使用就显得极其关键。

(2)消极管理中经常遇到的另一问题是,因为种种原因某些股票被禁止纳入投资组合,而投资者还是需要构建投资组合匹配目标指数。例如,在过去的10年中,养老基金宣布将不会持有烟草公司股票或者投机股票。很有可能某一市场板块如烟草股关于通货膨胀的敏感度不同于平均股票。如果使用单指数模型来构建一个排除烟草股的指数基金,那么该基金关于市场收益的敏感度能做到匹配,但是关于其他重要因素的敏感度很可能不能匹配。

(3)同样,投资者可能需要建立一个必须包含某种股票的投资组合。例如在日本,这种情况很常见,股票持有的原因是因为公司之间的业务关系,在美国,投资者持有或增加某个组合可能出于业务原因,或者是投资者出于税收考虑或报告考虑而不希望实现那些累积的尚未实现的资本损失或利得。现在,问题就变为,找到一个全面组合,其中包含一组已定义的股票而且能十分接近匹配指数。而这些股票可能对所要匹配的指数之外的一些重要影响因素具有相当的敏感度,所以应该在每个关键风险因素上都达到明确的匹配。

3.7.2 积极管理

积极管理中的多指数模型的使用与消极管理中的使用相似。按照与前面讨论的顺序相反的顺序展开讨论会更加简便。多指数模型弥补了单指数模型不能实现的功能是,允许投资者对某些因素进行投机。如果你相信非预期通货膨胀以高于市场预期的速率加速($I_1>0$),那么你就可以增大对通货膨胀的暴露度(b)而投机。可以持有关于通货膨胀的敏感度比标准普尔指数大的组合实现。

例如所罗门兄弟公司(1989)近来发展的收益产生过程模型,通过它,投资者可以就每种因素进行投机。它所包含的因素有:

(1)经济增长。作为对经济增长趋势的一个代理,它适用总工业产出的年变化。这一系列衡量了经济的总体福利。

(2)商业周期。他们认为,经济的短商业周期行为可以用投资级公司债券的债券收益与美国国库券的利差来捕捉。他们使用的是20年到期期限。他们认为,两种工具的利差能抓住违约风险。

(3)长期利率。他们认为,长期利率反映了金融资产相对吸引力的变化,这应引起投资组合的变化。该模型用10年政府债券收益率的变动作为无风险债券吸引力的显示指标。

(4)短期利率。类似的道理,短期利率的变动会改变较长期投资工具的供

给，如股票和债券。该模型使用1个月期国库券收益率的变动作为收益率短期变化的显示指标。

(5)通货膨胀。消费价格指数被用来衡量通货膨胀。股票因素是以实际通货膨胀和预期通货膨胀的差异来衡量的。

(6)美元汇率。货币汇率波动对股票市场的影响以15个国家的贸易加权货币篮子的变化来衡量。所罗门兄弟公司发现，股票收益和货币波动存在统计意义上的稳定关系。

(7)市场指数中与前面不相关的部分。

返回到我们正在讨论的简单模型上来，假设标准普尔指数是合适的基准，且分析师相信销售的增长将比市场预期的高出1%。分析师可能会将关于销售的b值，从由标准普尔指数得到的1.71提高到2.21。根据套利定价模型，确认关于销售的$\lambda=1.49\%$，销售敏感度增加了0.5(2.21－1.71＝0.5)，将导致期望收益增加$0.5\times1.49\%=0.745\%$，这些作为投资者额外风险的补偿已经足够了。但是如果销售实际增长了1%，那么组合收益将额外增加2.21%(由上述知实际的R_i将增加$b_2\Delta I_2$，而$\Delta I_2=1\%$，$b_2=2.21$)。在这2.21%的增长中，包含了0.745%是由于销售敏感度的增加，而$2.21\%-0.745\%=1.465\%$，则是因为b值保持在标准普尔指数的水平。增加的0.745%被称为超额风险调整收益，来自比市场更好的对因素的预测能力。

与单指数模型和资本资产定价模型一样，多指数模型和套利定价模型也可用来对基于单个证券表现的估计来建立最优组合。其中证券间的协方差是由多指数模型产生的，而期望收益和方差则是分析师的预测和历史数据相结合得出的。

套利定价模型的另一个应用就是确定被高估或低估价值的股票。在这一程序中，分析师给出股票收益的预测值。然后套利定价模型与因素的敏感度的估计值一起使用，以计算出一个理论上的股票必要收益率。如果股票实际回报率高于上述理论上必要收益率，就买入；如果股票实际回报率低于上述理论上必要收益率，就卖出。

这是资本资产定价模型而不是套利定价模型作为均衡模型时的分析的扩展。回忆前面的内容，资本资产定价模型在期望收益—贝塔空间中是一条直线，如果一个公司的股票在证券市场线上面，则其所给出的收益高于同样风险下的必要收益率，应该买入该公司的股票；如果一个公司的股票在证券市场线下面，则其所给出的收益低于同样风险下的必要收益率，应该卖出该公司的股票。使用套利定价模型进行分析可得出相同的逻辑。考虑二因素套利模型。

类似于CAPM,我们也可以以几何图形来反映APT。这样均衡收益率是三维空间中的一个平面,这个三维空间的坐标分别为两个因素的敏感度和期望收益。仿照上面,如果一个公司的股票在上述平面的上面,则其所给出的收益高于同样风险下的必要收益率,应该买入该公司的股票;如果一个公司的股票在上述平面的下面,则其所给出的收益低于同样风险下的必要收益率,应该卖出该公司的股票。

套利定价模型最广泛的应用是建立一个股票组合,该组合紧密追踪指数的同时还产生对该指数的超额收益。执行这一程序的方法之一就是,简单使用本章前面内容的指标匹配程序,但只能从分析中已经确定为表现出色的一组股票中进行选择。其他方法要么使用股票的离散数字式排序,要么使用股票的期望收益,在使用多指数模型来尽可能精确追踪指数的同时尝试产生股票对某一指数的超额收益。这种方法设计出的组合被称之为研究指数基金(research titled index fund)。虽然引入了一些额外的风险(当从严格限定的股票组合中进行选择时,不能严密地追踪指数),使用这项技术的投资者会发现追踪指数能力方面轻微的损失却可以得到额外的收益。多指数模型相对于单指数模型的优势在于,目标指数可以被紧紧地追踪,因为多指数模型明确地引入了风险的不同来源。

如果投资者具有超强能力可以识别股票,其收益超过或低于基于套利定价风险调整基础的平均收益,那么使用套利定价模型就可以建立可提供超额收益且对任何因素都是敏感程度为0的组合。设α_i为超额收益,则式(3.21)可以写为:

$$R_i = r_f + \alpha_i + \lambda_1 b_{i1} + \lambda_2 b_{i2} + \cdots + b_{i1} I_1 + b_{i2} I_2 + \cdots + \varepsilon_i$$

考虑一下这两个组合:组合L是多头头寸的组合,组合S是一个具有空头头寸的组合。且两组合组成的风险敏感程度都为0,即$\forall j, b_{Lj} + b_{Sj} = 0$。那么结合上面的方程可知,我们得到风险中性的组合N,其期望收益为:$R_N = r_f + \alpha_L + \alpha_S$,而风险为:$\varepsilon_L + \varepsilon_S$。

波麦斯特,罗尔和罗斯(1994)根据1991年4月至1992年3月的时期数据建立了上述模型,对收益进行了考察,假设α可以被正确识别。结果发现,在这一时期,标准普尔指数的年收益率为11.57%,标准差为18.08%。他们的风险组合年收益为30.04%,标准差为6.26%。虽然这些都是乐观数据,因为他们假设预见性很准确,但它们确实可以降低风险,并且如果存在预测能力的话,收益也会增加。

尽管也可以使用单因素模型进行同样的分析，但组合的总体风险会增大，而且投资者会发现他是在对因素（通货膨胀、利率等）进行投机，而不是对纯证券进行投机。因为他们只看到了市场贝塔，全然不管其他因素。

当然，多指数模型和套利定价模型的应用领域的最后一个考察是组合绩效的评估，为了使本书结构更加合理，我们将在后面章节中进行讨论。

附　录

附录 3.1

为了得到无风险的套利投资组合，把可分散的风险（ε_i）和不可分散的风险（I_j）风险去掉。必须有：

$$\sum_{i=1}^{n} X_i b_{ij} = 0, \forall j = 1,2,\cdots,k \tag{3.22}$$

$$\sum_{i=1}^{n} X_i \varepsilon_i = 0 \tag{3.23}$$

将（3.22），（3.23）代入（3.16）得到 $R_P = \sum_{i=1}^{n} X_i E[R_i]$

因为 ε_i 是独立的，大数定律保证随着证券数目 n 越来越大，$\sum_{i=1}^{n} X_i \varepsilon_i$ 将趋近于 0。

由于 R_P 为无风险，且初始投入为 0，故 $R_P=0$，这样 $\sum_{i=1}^{n} X_i E[R_i]=0$。

这样由 $\sum_{i=1}^{n} X_i = 0$，与 $\sum_{i=1}^{n} X_i b_{ij} = 0, \forall j = 1,2,\cdots,k$，便可以推出 $\sum_{i=1}^{n} X_i E[R_i]=0$。

令 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^T$，$\boldsymbol{b}_j=(b_{1j},b_{2j},\cdots,b_{nj})^T$，$\boldsymbol{E}[\boldsymbol{R}]=(E[R_1],E[R_2],\cdots,E[R_n])^T$，

$\boldsymbol{e}=(1,1,\cdots,1)^T$

这样上面所述可以变为：

已知 $\boldsymbol{X}^T\boldsymbol{e}=0$，以及 $\boldsymbol{X}^T\boldsymbol{b}_j=0, j=1,2,\cdots,k$ 可以推出 $\boldsymbol{X}^T\boldsymbol{E}[\boldsymbol{R}]=0$。

即 $\boldsymbol{X}$ 与 $\boldsymbol{e}$，$\boldsymbol{b}_j (j=1,2,\cdots,k)$ 正交，则 $\boldsymbol{X}$ 与 $\boldsymbol{E}[\boldsymbol{R}]$ 正交。由线性代数上理论可知 $\boldsymbol{E}[\boldsymbol{R}]$ 可由 $\boldsymbol{e}$，$\boldsymbol{b}_j (j=1,2,\cdots,k)$ 线性表示，即存在不全为 0 的实数 $\lambda_0,\lambda_1,\cdots,\lambda_k$ 满足：

$$\boldsymbol{E}[\boldsymbol{R}]=\lambda_0\boldsymbol{e}+\lambda_1\boldsymbol{b}_1+\cdots+\lambda_k\boldsymbol{b}_k$$

表示成分量即为：

$$E[R_i]=\lambda_0+\lambda_1 b_{i1}+\cdots+\lambda_k b_{ik} \tag{3.24}$$

因此 $E[R_i]=\lambda_0+\lambda_1 b_{i1}+\cdots+\lambda_k b_{ik}$ 可以由超额收益的形式重新写为：

$$E[R_i]-r_f=\lambda_1 b_{i_1}+\cdots+\lambda_k b_{ik} \tag{3.25}$$

此即正文中出现的表达式。

第四章

金融计算基础及对债券的估值

4.1 金融计算基础

由于货币具有在一定利率水平下进行投资的机会，因此货币是有时间价值的。时间价值是分析任何金融工具都要用到的基础概念之一，时间价值的概念在进行投资决策和衡量投资收益的过程中都有着广泛的运用。

4.1.1 终值的计算

资金的终值(FV)是当投资者按一定利率投资直至到期，并以复利计算到期时的价值。任意数量货币的终值可以用下面的公式来计算：

$$FV=PV(1+r)^N$$

式中，FV 为终值；PV 为期初本金或者现值；N 为期限数，与利息支付周期有关，一般为 1 年；r 为每期的利率。

例如：某投资者期初用 1 000 元投资于某种 8 年期债券，年率为 4%，利息按年支付，并以 4% 的利率进行再投资，这笔投资到期时的终值将是 $FV=1\ 000\times(1+4\%)^8=1\ 368.6$ 元。

当然一般情况下，付息频率为一年，当付息频率不为一年时，上式中的 N 和 r 都要调整如下：

N=期限数×付息频率

r=年利率/付息频率

从而更一般的终值计算公式修改为：

$$FV=PV(1+r/n)^{N\times n}$$

显然，在其他条件相同时，n 越大，也即付息频率越大时，投资的终值越大，这是因为付息频率越大，将所获利息进行投资的机会将更大。下面考虑一种极限情况，即 $n\to\infty$ 时的终值：

$$\lim_{n\to\infty}PV(1+r/n)^{N\times n}$$

由高等数学的知识，我们知道上述极限为 $PVe^{N\times r}$，通常把这种情况称为连续复利。

上面介绍了最简单的终值计算方法，下面介绍一种比较复杂的终值计算方法。当投资者不是一次将所有资金全部进行投资，而是分期进行等额投资时，就称为年金，从第一期期末进行第一笔投资时称为普通年金，普通年金是最常见的年金。普通年金的计算公式为：

$$FV_O=A(1+r)^{N-1}+A(1+r)^{N-2}+\cdots+A=A\sum_{t=0}^{N-1}(1+r)^t$$

式中 A 为年金额，由中学的等比数列的知识，我们可以得到这一公式的简化形式：

$$FV_O=A\frac{(1+r)^N-1}{r}$$

如果年金的投资周期每年不止一次，则对 N 和 r 也要进行相应调整，调整方法同一般的终值计算完全相同。

【例 4.1】 有一种 20 年期的债券，面值为 1 000 元，年利率为 8%，每半年付息一次，所得利息按 12%的年利率进行再投资，则期末可以获得的利息总额是多少呢？每次的利息支付都是期末发生的，而且金额相同，因此可以看作是一种普通年金，年金金额为：1 000×8%/2=40，代入上面的公式，可以得到期末获得的利息总额(不包括本金)为：

$$40\times\frac{(1+4\%)^{40}-1}{4\%}=3\ 801(\text{元})$$

需要注意的，这里的再投资期限为 40 期，而不是 20 期。

除了普通年金外，当每次的现金流都是在期初发生时，这种年金被称为预付年金，预付年金的终值计算方法可以通过普通年金的变形得到：

$$FV_D = FV_O \times (1+r)$$

这是因为预付年金和普通年金的唯一区别就是每次现金流的发生都比普通年金早一期，因而每期年金的再投资期限也多一期，因此只要在计算普通年金时多乘1次$(1+r)$就可以得到预付年金的终值。

4.1.2 现值的计算

现值的计算和终值的计算相反，现值是为了在投资末期得到一定金额的货币量而在期初投入的货币量。

现值的计算公式为：

$$PV = \frac{FV}{(1+r/n)^{N\times n}}$$

式中，PV为现值；FV为终值；r为贴现率（一般是年利率）；N为期限数（一般是年数）；n为每个贴现率周期内发生的现金流次数。

计算现值的过程也叫贴现，其中最重要的是确定贴现率。贴现率有时同证券的利率是相同的，但证券发行后，随着市场的变化，更多的时候贴现率同证券的利率并不相同。

上面所说的是一种存续期间不发生现金流的证券，金融市场上还存在着很多证券，在它们的存续期间，现金流并不是一次发生的，而是按期分次发生的，在这种情况下，证券的现值就等于各期现金流现值之和：

$$PV = \sum_{t=1}^{N\times n} \frac{C_t}{(1+r/n)^t}$$

式中C_t为每期发生的现金流；r为贴现率（一般是年利率）；n为每个贴现率周期内发生的现金流次数；N为期限数。

【例4.2】一种10年期国债，面值为1 000元，票面利率为4%，每年年底付息一次，到期还本，这段时间市场贴现率为5%，则此债券整个存续期间所有现金流的现值如下：

$$\sum_{t=1}^{10} \frac{40}{(1+5\%)^t} + \frac{1\ 000}{(1+5\%)^{10}} = 922.78(\text{元})$$

特别地，由于年金有着比较特别的现金流，因此在计算年金现值的时候，也可以使用下面的简化公式。

普通年金为：

$$PV_O=\sum_{t=1}^{N\times n}\frac{A}{(1+r/n)^t}=A\left[\frac{1-\frac{1}{(1+r/n)^{N\times n}}}{r/n}\right]$$

根据前面所说的普通年金和预付年金的关系，可以得到预付年金现值的计算公式为 $PV_D=PV_O(1+r/N)$，因为预付年金为期初付，所以它比普通年金少贴现一次。

下面介绍一种特殊的年金：永续年金，即这种年金没有到期日，即 $N\to\infty$，这样由上述公式可得：

$$PV_P=\frac{A}{r/n}$$

4.1.3 贴现率的确定

上面都是把贴现率作为已知量给出，但是现实中，贴现率是一个需要确定的变量。一般来说贴现率是一个反映货币时间价值的指标。但在不同的领域中，贴现率有着丰富的具体含义。在证券投资中，贴现率可以看作投资者的机会成本，即投资者投资于某证券而损失的其他投资机会所获得的最高收益率，因此，投资者投资于某一证券的条件就是该证券的收益率至少不低于其他投资对象。基于这种考虑，我们通常用要求的回报率来作为它的贴现率。要求的回报率指在考虑了证券的风险后投资者所要求的最低收益率，另一方面，对公司来说，投资者要求的回报率就是他们的股权成本。在分析债券价值时，我们通常用可类比的债券收益率作为要求的回报率，这里可类比债券的意思是指具有相同信用等级且到期日相同的不可赎回债券。此外，还可以通过构建收益率曲线等方法来比较精确地计算要求的回报率，这将在后面的内容中进行详细的介绍。本部分内容主要介绍分析股票价值时贴现率的确定方法。

股票价值的分析方法大都是建立在现金流贴现基础(DCF)之上的，因此，贴现率的确定在分析股票价值时是至关重要的。一般来讲，确定股票贴现率或要求回报率的方法有三种：通过资产定价模型来计算，债券收益加上风险溢价及风险加成法。下面将对每一种方法进行详细介绍。

(1)用前面章节的资产定价模型来确定贴现率。资产定价模型主要是资本资产定价模型和套利定价模型。这些模型都认为资产的收益率等于无风险

收益加上该资产相对于某一(或某些)要素的风险溢价。其中,CAPM 的形式如下:

$$E[R_i]=r_f+\beta_i(E[R_M]-r_f)$$

使用 DCF 法分析股票价值时,我们通常使用这里的 $E[R_i]$ 作为贴现率,即 $E[R_i]=r$。

【例 4.3】 假设某公司股票的贝塔值为 0.8,同期市场风险溢价为 6%,无风险收益率为 5%,那么该公司要求的回报率为 5%+0.8×6%=9.8%。

通过观察公式可以看出,在使用公式 $E[R_i]=r_f+\beta_i(E[R_M]-r_f)$ 计算某一证券的预期收益率时,关键是要确定无风险收益率 r_f 和市场风险溢价 $E[R_M]-r_f$。一般来讲,我们可以用短期政府债券或长期国债的收益率来作为无风险收益率。其中可交易的长期国债收益率更加常用,因为股票是不偿还的,因此也可以看作是久期(后面将会讲到)很长的债券,因此根据久期匹配的原则,长期国债的收益率更适合作为计算股票收益率时的无风险收益率,我们一般使用 20 年期或 10 年期可流通国债的收益率作为无风险收益率。因此,在计算贴现率时,上述公式可以重新定义为:

$$\text{贴现率}=\text{长期国债收益率}+\text{该股票}\,\beta\times\text{预计股票市场对长期国债的溢价}$$

因此接下来就只剩下市场风险溢价一项了。通常计算市场风险溢价的方法有两种。第一种是使用股票市场对国债市场风险溢价的历史平均数,另一种方法则是基于对公司的预测数据。在使用历史平均法时,既可以使用算术平均也可以使用几何平均来计算过去市场风险溢价的平均数。但是,这种用历史数据平均得出的溢价很可能是不准确的,因为股票市场经常会有业绩不佳的股票退市,使得使用历史平均法算出的市场风险溢价很可能是被高估的,因此有必要对这个数值进行必要的调整。

用预测法来计算市场风险溢价是戈登模型(Gordon model)(这将在下一章股票价值模型中详细介绍)的一个应用,计算方法如下:

股票市场风险溢价=预期明年的市场平均股利收益率(股利/股票价格)+公认的上市公司长期盈利增长率-当期长期国债收益率

如果某股票市场预计明年的股利收益率为 2%,且市场一致认为整个市场上的上市公司未来 5 年平均长期盈利增长率为 4.8%,20 年期国债的当前收益率为 3%,那么该市场的风险溢价为:2%+4.8%-3%=3.8%。

预测数据有着较大的主观性,因此一般不建议使用预测法来计算市场风

险溢价，除非市场刚刚建立，历史数据太少。预测法主要适用于那些新兴市场。

除了 CAPM 模型外，我们还可以使用 APT 模型来计算股票的要求回报率。由于两者使用方法非常相似，因此此处就不再详述。

很明显使用资产定价模型来计算贴现率存在着三个明显的不确定性：模型的不确定性（模型不一定正确）、输入数据的不确定性（比如 CAPM 计算中的市场风险溢价）和敏感度 β 的不确定性。

（2）用债券收益率加风险溢价来计算股票贴现率。对于有发行可交易长期债券的公司，我们还可以使用债券收益率加风险溢价的方法来计算股票贴现率。

股权成本（贴现率）＝公司发行的长期债券的到期收益率＋风险溢价

风险溢价的确定没有一个确定的计算方法，而更多地是一个经验估算。在美国市场上，这一溢价一般为 3%～4%。由于中国企业很少发行企业债，因此这一方法在中国并不适用。

（3）风险加成法。对于非上市公司来说，它们的公开数据很少，我们无法计算它们的 β 或因素敏感系数，也就无法用 CAPM 或 APT 模型来计算它的贴现率。此时，我们可以用风险加成法来确定它们的贴现率。

股权成本（贴现率）＝无风险收益率＋风险溢价

与 CAPM 和 APT 不同的是，这里的风险溢价是一个主观的估计数字，因此使用这种方法一般要求分析者有丰富的投资经验。

4.2 债券价值分析

在第一节我们介绍了金融领域的一些比较基础的计算，下面我们就要将这些计算方法运用到实际的证券投资分析中去。本节主要介绍关于债券投资的一些基本分析方法。

4.2.1 收入资本化法与债券价值分析

根据收入资本化法，任何资产的内在价值都等于投资者对持有该资产预期的未来现金流的现值。根据资产的内在价值与市场价格是否一致，可以判

断该资产是否被低估或高估，从而帮助投资者进行正确的投资决策。所以，如何决定债券的内在价值成为债券价值分析的核心。下面将对不同的债券种类分别使用收入资本化法进行价值分析。

(1)贴现债券

贴现债券，又称零息票债券(zero-coupon bond)，或贴现债券，是一种以低于面值的贴现方式发行，不支付利息，到期按债券面值偿还的债券。债券发行价格与面值之间的差额就是投资者的利息收入。由于面值是投资者未来唯一的现金流，所以贴现债券的内在价值由以下公式决定：

$$P=\frac{F}{(1+r)^T} \tag{4.1}$$

其中，P 代表贴现债券的内在价值，F 代表面值，r 是市场利率，T 是债券到期时间。

假定某种贴现债券的面值为 1 000 元，期限为 20 年，利率为 8%，那么它的内在价值应该是：

$$P=\frac{1\ 000}{1.08^{20}}=214.55(\text{元})$$

(2)息票债券

息票债券，是一种按照票面金额计算利息，票面上可附有作为定期支付利息凭证的息票，也可不附息票的债券。投资者不仅可以在债券期满时收回本金，而且还可定期获得固定的利息收入。所以，投资者未来的现金流包括了两部分：本金与利息。息票债券的内在价值为：

$$P=\frac{C}{1+r}+\frac{C}{(1+r)^2}+\cdots+\frac{C}{(1+r)^T}+\frac{F}{(1+r)^T} \tag{4.2}$$

其中，C 是债券每期支付的利息，其他变量与贴现债券的公式相同。

当债券清算日距离下次付息日不满一年时，我们就应该对公式(4.2)进行一些改动

$$P=\sum_{t=1}^{n}\frac{C}{(1+r)^d\,(1+r)^{t-1}}+\frac{F}{(1+r)^d\,(1+r_n)^n}$$

其中，d 为清算日距离下次付息日的天数/365。

下面介绍一个小技巧。

当息票利率高于实际利率时，债券价格高于面值；

当息票利率等于实际利率时，债券价格等于面值；

当息票利率低于实际利率时，债券价格低于面值。

(3)统一债券

统一债券是一种没有到期日的特殊的定息债券，它和上面讲述的永续年金相同。最典型的统一公债是英格兰银行在18世纪发行的英国统一公债，英格兰银行保证对该公债的投资者永久地支付固定的利息。直至如今，在伦敦的证券市场仍然可以买卖这种公债。历史上美国政府为巴拿马运河融资时也曾发行类似的统一公债。但是，由于美国政府在该种债券发行时还附有赎回条款，所以美国的统一公债已经退出了流通。在现代企业中，优先股的股东可以无限期地获得固定的股息，所以优先股实际上也是一种统一公债。统一公债的内在价值的计算公式如下：

$$P=\frac{C}{1+r}+\frac{C}{(1+r)^2}+\cdots=\frac{C}{r} \tag{4.3}$$

例如，某种统一公债每年的固定利息是60美元，假定市场利率水平为12%，那么，该债券的内在价值为500美元，即：

$$P=\frac{60}{12\%}=500(\text{美元})$$

4.2.2 收益率曲线

(1)收益率曲线及其构建。前面我们介绍了单一贴现率下债券收益率和价格的计算，前面所有的计算都是基于这样的一个假设的前提上的，那就是假设所有的再投资利率都与到期收益率相同。但是现实中，不同期限投资其收益率往往是不同的，比如一年期国债的年收益率一般低于二年期或更长期限的国债，而对在不同时间产生的现金流，也应该用相应期限的收益率来对其进行贴现，如对第一年年底发放的利息用一年期的收益率进行贴现，对第二年的利息则用二年期的即期利率(即期利率是只对投资者支付一次现金流的贷款或债券的到期收益率)来进行贴现……因此，当放开收益率不变的假设之后，我们就需要按照现实情况对计算方法进行相应的改进。此时债券的价格公式为

$$P=\sum_{t=1}^{n}\frac{C}{(1+r_t)^t}+\frac{F}{(1+r_n)^n} \tag{4.4}$$

利用这种方法计算债券价格最重要的就是要知道任何期限投资的收益

率，比如如果一种债券离到期日还剩下 5 年零 1 天，下一次利息支付发生在一天后，每年支付一次利息，那我们就需要知道 1 天期，1 年零 1 天……直到 5 年零 1 天的收益率，而市场上并不一定会存在有相同剩余期限的同类债券，因此就需要一种方法来估算任何期限的收益率。

我们通过构建收益率曲线来估算任何期限的收益率。收益率曲线是反映即期利率同剩余到期时间函数关系的曲线（即期利率关于时间的函数）。我们假定即期利率和期限存在着某种函数关系，然后通过市场现有债券的数据来推导出两者之间的函数关系。当然市场不可能有每个时间点上的利率，这样我们只能通过一些离散点来拟合出一条光滑的曲线。当然本节讲述的推导出的收益率曲线只是比较简单的拟合，用两个点之间的直线来代替真实的收益率曲线，且不考虑连接两条直线之间点的光滑性（用到一阶导数）。

下面我们通过构建 2005 年 3 月 21 日中国国债市场（见表 4.1）的理论收益率曲线来具体说明这种构建方法。下面主要用的是拉格朗日插值法的线性插值法。首先要选择不同期限的国债，由于市场未清偿债券的剩余期限一般不会正好是整年，因此我们要尽量选择剩余期限接近整年的债券，而且每种剩余年份的债券尽量都要有。

表 4.1　2005 年 3 月 24 日部分国债的相关资料

国债名称	剩余期限（年）	全价（元）	票面利率（%）	还本付息方式
05 国债 02	0.975	98.040	NA	贴现式
04 国债 11	1.729	101.561	2.98	年
99 国债 5	2.408	102.854	3.28	年
21 国债 3	3.085	102.87	3.27	年
97 国债 4	2.449	122.35	9.78	年

此外，央行一天超额准备金利率为 0.99%，由于期限只有 1 天，我们将这一利率近似作为期限为 0 的国债收益率。

这样收益率曲线的计算过程可以描述如下：

①0.975 年期的年收益率为：

$$r_{0.975}=\left(\frac{100}{98.04}\right)^{\frac{1}{0.975}}-1=2.05\%$$

设 0.975 年以下的收益率曲线符合抛物线方程：

$r=a\sqrt{t}+0.99\%$，因为期限为 0 的国债收益率为 0.99%，这样上述曲线的截距就为 0.99%。

将点(0.975，2.05%)代入上面式子，得到 $a=0.0107$。

这样抛物线方程为：

$$r=0.0107\sqrt{t}+0.99\% \qquad t\in(0,0.975] \tag{4.5}$$

②计算 1.729 年期的收益率，$r_{1.729}$ 满足如下条件：

$$\frac{2.98}{(1+r_{0.729})^{0.729}}+\frac{102.98}{(1+r_{1.729})^{1.729}}=101.561$$

将 $t=0.729$ 代入上述抛物线方程得到 $r_{0.729}=1.90\%$，将这一数据代入到上面的条件解得 $r_{1.729}=2.53\%$。由于是直线拟合，故只需要两点，前面已经知道 $r_{0.975}$，加上现在的 $r_{1.729}$，我们用这两点可以得到这两点之间的线段的方程为：

$$0.0048t-0.754r+0.010777=0 \qquad t\in(0.975,1.729] \tag{4.6}$$

③计算 2.408 年期的收益率，$r_{2.408}$ 满足如下等式：

$$\frac{3.28}{(1+r_{0.408})^{0.408}}+\frac{3.28}{(1+r_{1.408})^{1.408}}+\frac{103.28}{(1+r_{2.408})^{2.408}}=102.854$$

$r_{0.408}$ 可由(4.5)得出为 1.67%，而 $r_{1.408}$ 可由(4.6)得出为 2.33%，将这两个数据代入到上式可以得到 $r_{2.408}=2.90\%$。由上面的 $r_{1.729}$ 和这次的 $r_{2.408}$ 可得到两点之间的直线方程为

$$0.0037t-0.679r+0.0107814=0 \qquad t\in(1.729,2.408] \tag{4.7}$$

④计算 3.085 年期的收益率。由于其满足以下方程：

$$\frac{3.27}{(1+r_{0.085})^{0.085}}+\frac{3.27}{(1+r_{1.085})^{1.085}}+\frac{3.27}{(1+r_{2.085})^{2.085}}+\frac{103.27}{(1+r_{3.085})^{3.085}}=102.87$$

通过(4.5)，(4.6)，(4.7)，可以分别求得 $r_{0.085}$，$r_{1.085}$，$r_{2.085}$，这样把上述三个数据代入到上式中可以得到直线方程：

$$0.0044t-0.677r+0.0090378=0 \qquad t\in(2.408,3.085] \tag{4.8}$$

⑤计算 97 国债 4 的理论价格：

利用前面几个方程可以计算得到 $r_{0.449}=1.72\%$，$r_{1.449}=2.38\%$，$r_{2.449}=2.93\%$。

这样国债价格为$\frac{9.78}{(1+r_{0.449})^{0.085}}+\frac{9.78}{(1+r_{1.449})^{1.085}}+\frac{9.78}{(1+r_{2.449})^{2.085}}=121.44$(元)

而该债券市场价格为 122.35 元,误差并不很大。

⑥这样根据①~④的计算,可以得到 2005 年 3 月 24 日中国国债市场的即期利率曲线,如图 4.1 所示。

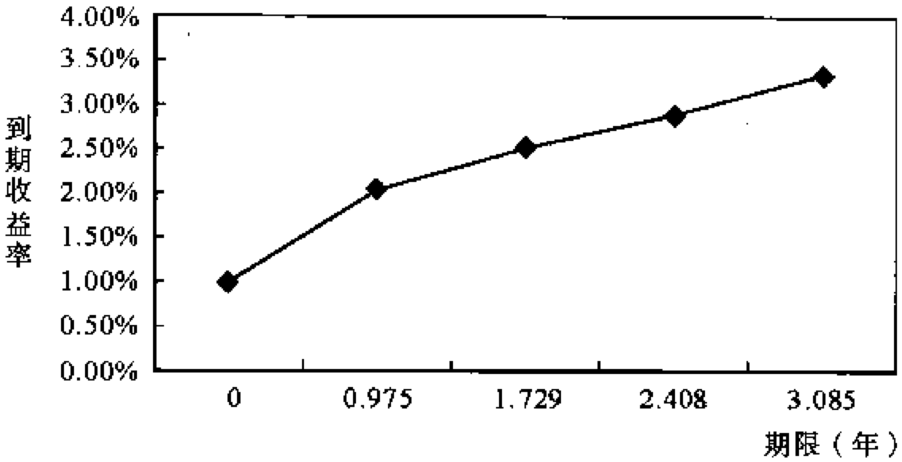


图 4.1 2005 年 3 月 24 日中国国债市场收益率曲线

上面说的是即期收益率的估算,下面介绍一种新的利率:远期利率。远期利率是承诺日和资金贷放日不同情况下的债券利率。对于两年期的情形,假设第一年的即期利率为 i_{01},一年后的一年期的远期合约的远期利率在现在的价格为 i_{12},且现在两年期的即期利率为 i_{02}。则远期利率必满足$(1+i_{01})(1+i_{12})=(1+i_{02})^2$。针对一般形式有:现在时刻为 0 时刻,$n$ 年期的即期利率为 i_{0n},m 年期的即期利率为 i_{0m},不妨假设 $n<m$。则 0 时刻的 n 年后的 $m-n$ 年期的远期利率为 i_{mn},则远期利率必满足$(1+i_{mn})^{m-n}(1+i_{0n})^n=(1+i_{0m})^m$。

4.2.3 利率的期限结构理论

即期利率曲线有各种形状,有向上倾斜的、向下倾斜的、水平的等等。为什么会有各种形状,且大部分都是向上倾斜的呢?而且收益率曲线一般会整体移动,即当短期利率上升时,长期利率也会上升。针对上述现象,债券的期限结构理论有四种解释,下面逐一介绍。

(1)纯预期理论(pure expectation theory)

它假定投资者对于风险无特定偏好(或者风险中性),只要给予高的期望

收益即可。或者对于期望收益相同的不同证券，投资者没有觉得它们有什么不同。

下面以两年期为例。假设第一年的即期利率为 i_{01}，预期一年后的即期利率为 i_{12}^e，则由预期理论可以知道，二年期的即期利率 i_{02} 必满足下式：

$$(1+i_{01})(1+i_{12}^e)=(1+i_{02})^2$$

因为投资者风险中性，这样对于他们来说投资一年后再投资一年和现在投资两年，只要收益率完全一样，他们就会觉得无差异。这样只有满足上式的即期利率才能是均衡利率。如果不满足的话，就会有套利出现。比如 $(1+i_{01})(1+i_{12}^e)>(1+i_{02})^2$，则投资者必然选择卖空两年期的贴现债券，这样买一个一年期的贴现债券后到期时再买一个一年期的贴现债券进行套利。套利的结果就使得两年期的贴现债券价格上升，一年期的贴现债券价格下降。最终两年期的即期利率上升到等式成立(一年期的即期利率也下降)。另一种情况可类似分析。

如果预期利率上升，即 $i_{12}^e>i_{01}$，则由上述等式可得 $i_{02}>i_{01}$，这样收益率曲线就向上倾斜。

如果预期利率下降，即 $i_{12}^e<i_{01}$，则由上述等式可得 $i_{02}<i_{01}$，这样收益率曲线就向下倾斜。

如果预期利率不变，即 $i_{12}^e=i_{01}$，则由上述等式可得 $i_{02}=i_{01}$，这样收益率曲线就为水平线。

必须注意的是，一年后的预期即期利率 i_{12}^e 和现在时刻一年后的远期利率 i_{12} 在概念上是两种不同的利率：一个是人们现在预期的一年后的即期利率，一个是市场上远期协议的利率。预期理论认为两者必然相等。

这样我们就解释了收益率曲线的三种形状。且它解释了收益率曲线的整体移动。因为假设现在的短期利率上升，则由上述等式可知，在其他条件不变的情况下，两年期的即期利率也会上升。

尽管上面以两年期为例，但是完全可以推广到多年期的情形。

纯预期理论很好地解释了收益率曲线，但是它并没有解释为什么收益率曲线一般向上倾斜。

(2)流动性贴水理论(liquidity premium theory)

该理论基本上承认期限结构的纯预期理论，但是对它有一个重大的修正。流动性贴水理论认为，长期债券比短期债券的流动性差，必须要给予投资者更高的期望收益，才能使得他们愿意持有长期债券。因为长期债券比短期债券

担负着更大的市场风险——价格波动和难以变现的风险。这种由于增加的市场风险而产生的长期债券额外的报酬称为流动性贴水。当然，对于债券发行者而言，由于金融市场上长期投资者少而短期投资者多，因此，长期资本的筹集所付出的成本较高。

例如，一年期即期利率 i_{01} 为 10%，预计下一年即期利率 i_{12}^e 为 16%，则由纯预期理论，两年期的贴现债券的到期收益率为 13%，但在流动性贴水理论中认为，两年期的到期收益率要高于 13%，才能使得那些原来持有一年期债券的人愿意持有两年期的债券，也就是说，除了 13%的收益外，尚需再加上一个流动性贴水。

流动性贴水理论认为，不论对未来的预期是多少，流动性贴水理论预测的长期收益将高于纯预期理论。当预期利率不变时（在纯预期理论中为水平的收益曲线），流动性贴水理论的收益曲线则为一条稍微向上倾斜的曲线；当预期利率上升时（在纯预期理论中为向上的收益曲线），流动性贴水理论的收益曲线的斜率更大，曲线更陡；当预期利率下降时（在纯预期理论中为向下倾斜的收益曲线），流动性贴水理论的收益曲线则为稍微向下倾斜的曲线或几乎是一条平坦的或稍微向上倾斜的曲线。

流动性贴水理论还认为，贴水随着债券到期期限的延长而变大，例如，三年期债券利率贴水要大于两年期债券利率贴水。

这样流动性贴水理论更能说明向上倾斜的收益曲线存在的原因。

(3)市场分割理论(market segmentation theory)

该理论认为一些投资者和债券发行人对债券的某一期限有特殊偏好，而对不同期限债券不感兴趣，换句话说，高利率也不能吸引他们。

首先考虑一种较长期限的债务。从保险公司的角度来考察期限选择问题。人寿保险公司提供的保单，可能长期都不会要求任何支付。例如，一份出售给一位 25 岁个人的保单，保险公司预期第一笔支付可能发生在 25 年甚至更久之后。保费的多少部分由预期的利率决定。如果保险公司投资于长期债券，债券的利息已知，而且如果利息超出保险合同承诺的水平，这就会大大降低保险公司的风险。但仍然存在某些风险，因为获得的利息需要被再投资，而再投资利率是未知的。但是，本金的投资利率始终是已知的，这大大降低了风险。或者保险公司为了满足长期债务，可以购买一系列 1 年期债券。但这样一来，1 年以后的所得都是不确定的。如果利率下降至低于保险合同的预期水平，公司可能难以满足其债务。不仅债券票面利息的再投资回报率不确定，本金的回报率也不确定。因此，很多保险公司投资于长期债券，即使短期利率

明显高于长期利率。

我们再从长期债券的发行者的角度来考虑期限选择的问题。建设一个厂房或仓库或其他设备需要公司支出一大笔资金。这些都属于长期资产。公司一般希望对它们的支付也在较长的一段时间内进行。公司可以通过发行长期债券来实现这种现金流模式，或者通过发行短期债券并在较长时间内不断重复发行。如果公司发行的是长期债券，公司的成本事前就已经确定，不存在再投资的利率风险。这就说明，公司通常会发行长期债券来满足这类债务。

同理可推知短期债券的情况。公司都有很多需要定期支付的短期债务，税收和工资是其中的两个例子。公司通常备有一定的现金来满足这些支付要求。如果公司购买贴现债券，债券的到期日恰好是支付日，公司将来的可用资金数额的风险为0(不考虑违约风险)。如果公司购买长期债券，财务上就会面临未来利率上升、债券价格下降，未来可用资金数额小于支付的风险。商业银行持有大量的短期证券。商业银行从事短期贷款以使资产和债务在期限上匹配。

市场分割理论认为投资者是风险厌恶的(这个和前面两种理论有明显的区别)，只在他们所希望的期限区间内经营。收益率差异不能导致他们改变期限。所以，决定长期利率的只是长期资金的供给和需求。同样，短期利率也只由短期资金的供求决定。相信市场分割理论的人研究这些市场的资金流入流出情况，以预测收益率曲线的变化。

根据该理论，如果当前企业和政府主要发行长期债券，那么长期利率将高于短期利率，此时收益率曲线向上倾斜；如果当前企业和政府主要发行短期债券，那么短期利率将高于长期利率，此时收益曲线向下倾斜。

同预期理论一样，它不能解释收益率曲线往往向上倾斜，且不能解释同步变动。

(4)期限偏好理论(term preference theory)

期限偏好理论是前面几种理论的折中。该理论认为投资者对特定期限都有很强的偏好，因而收益曲线不会严格地服从纯预期理论和流动性贴水理论的预测。投资者之所以如此，是因为在债券投资过程中，投资者的资产周期和负债周期正确匹配会使其处于最低风险状态。例如银行和货币市场共同基金一般购买短期债券，而人寿保险公司则偏好于购买长期债券。但是，如果不符合投资者偏好的期限的预期额外收益变大时，实际上投资者将修正原来偏好的期限。或者说只要给投资者足够的利益补偿，投资者也会考虑改变以前投资的品种。比如，当长期债券的预期收益远远超过短期债券的收益时，银行和

货币市场共同基金将增加该期限的资产，也就是说，他们将购买长期债券。如果购买短期债券的预期收益变大，人寿保险公司将暂时取消只投资长期债券的规定，并在他们的资产组合中加入适当的短期债券。换句话说，如果投资者已经使自己处于某种期限偏好状态，则要使其离开原偏好状态，就必须提供额外贴水 P 作为增加风险的补偿。

例如，在纯预期理论中，对两年期债券来说，有：

$$(1+i_{01})(1+i_{12}^e)=(1+i_{02})^2$$

在期限偏好理论中，则应是：

$$(1+i_{01})(1+i_{12}^e+P)=(1+i_{02})^2$$

从上式中，我们可以看出，若短期资金供给较多，则投资者偏好于购买短期债券，要使其购买长期债券，必须提供风险贴水，也就是说，$P>0$；相反，若长期资金供给较多，则投资者偏好于购买长期债券，要使其购买短期债券，必须提供风险贴水，也就是说，$P\leqslant 0$。

期限偏好理论以实际的观念为基础，即投资者为预期的额外收益而承担额外的风险。在接受市场分割理论和纯预期理论部分观点的同时，也剔除了二者极端的观点。该理论能较近似地解释现实世界的现象。

4.3 债券属性与价值分析

债券的价值与债券以下六方面的属性密切相关。这些属性分别是：到期时间（期限）长短、息票率、可赎回条款、税收待遇、流通性以及违约风险。其中任何一种属性的变化都会改变债券的到期收益率水平，从而影响债券的价格。下面将采用局部静态均衡的方法，即在假定其他属性不变的条件下，分析某一种属性的变化对债券价格的影响。

4.3.1 到期时间

从第 2 节可以发现：当市场利率 r 和债券的到期收益率 r^* 上升时，债券的市场价格和内在价值都将下降。当其他条件完全一致时，债券的到期时间越长，市场利率变化引起的债券价格波动幅度越大。后面久期可以用来解释这一原因，因为到期时间越长，久期会越大，这样市场利率变化引起的债券价

格波动幅度越大。但是当到期时间变化时，市场利率变化引起的债券的边际价格变动率递减。

【例 4.4】假定存在 4 种期限分别为 1 年、10 年、20 年和 30 年的债券，它们的息票率都是 5%，面值均为 100 元，其他属性也完全一样。如果期初的市场利率为 5%，根据内在价值的计算公式可知这 4 种债券的内在价值都是 100 元。如果相应的市场利率上升或下降，这 4 种债券的内在价值的变化如表 4.2 所示。

表 4.2　内在价值与期限之间的关系

期限	相应的市场利率下的内在价值(元)			内在价值变化率(%)	
	3%	5%	7%	5%→3%	5%→7%
1	101.94	100	98.13	1.94	−1.87
10	117.06	100	85.95	17.06	−14.05
20	129.75	100	78.81	29.75	−21.19
30	139.20	100	75.18	39.20	−24.82

表 4.2 反映了当市场利率由现在的 5%下降到 3%，4 种期限债券的内在价值分别上升 1.94%，17.06%，29.75%，39.20%(由于利率为 5%时的内在价值为 100 元)；而当市场利率由 5%上升到 7%，4 种期限的债券的内在价值分别下降 1.87%，14.05%，21.19%，24.82%。

可以看出当市场利率变动到 3%时，期限由 1 到 10 年时债券价格变动 15.12元，而由 10 年到 20 年时债券价格变动 12.69 元，20 年变动到 30 年时债券价格变动 9.45 元。当然利率上升也会有这种规律：市场利率变化引起的债券的边际价格变动率递减。

大家还可以发现市场利率上升导致债券价格的波动幅度不如同样幅度利率下降导致债券价格的波动幅度。这个涉及债券的凸度，后面将解释这种情况的原因。

4.3.2 息票率

债券的到期时间决定了债券的投资者取得未来现金流的时间，而息票率决定了未来现金流的大小。在其他属性不变的条件下，债券的息票率越低，市场利率变化引起的债券价格的波动幅度越大。后面将用久期来解释这种情形，下面只举例来说明这种现象。

【例 4.5】设有 4 种债券，期限相同均为 10 年，面值均为 100 元。它们之间唯一的区别在于息票率，且它们的息票率分别为 3%，5%，6%，7%。假设初始的市场利率水平为 5%，那么可以利用第二节的公式分别计算出它们的初始内在价值。如果市场利率发生了变化(上升到 6%，或下降到 4%)，相应地可以计算出这 4 种债券的新的内在价值，具体结果见表 4.3。

表 4.3　内在价值变化与息票率之间的关系

息票率	相应的市场利率下的内在价值			内在价值变化率(%)	
	4%	5%	6%	5%→4%	5%→6%
3%	91.89	84.56	77.92	8.67	−7.85
5%	108.11	100	92.64	8.11	−7.36
6%	116.22	107.72	100	7.89	−7.17
7%	124.33	115.44	107.36	7.70	−7.00

从表 4.3 中可以发现，面对同样的市场利率变动，无论是市场利率上升还是下降，4 种债券中息票率最低的债券的内在价值波动幅度最大，而随着息票率的提高，4 种债券的内在价值的变化幅度逐渐降低。所以，债券的息票率越低，市场利率变化引起的债券价格的波动幅度越大。

4.3.3 可赎回条款

许多债券在发行时含有可赎回条款，即在一定时间内发行人有权赎回债券。这是有利于发行人的条款，因为，当市场利率下降并低于债券的息票率时，债券的发行人能够以更低的成本筹到资金。所以，发行人可以行使赎回权，将债券从投资者手中收回。尽管债券的赎回价格高于面值，但是，赎回价格的存在制约了债券市场价格的上升空间，并且增加了投资者的交易成本，因此降低了投资者的收益率。为此，可赎回债券往往规定了赎回保护期是发行后的 5 至 10 年。

【例 4.6】假设一种 10 年期债券，息票率为 10%，面值为 1 000 元，赎回价格为 1 110 元，且赎回保护期为 5 年，假设必要报酬率为 9%，则未行使赎回权时，债券的内在价值为：

$$P=\frac{100}{1+8\%}+\frac{100}{(1+8\%)^2}+\cdots+\frac{100}{(1+8\%)^{10}}+\frac{1\ 000}{(1+8\%)^{10}}=1\ 134.2016(\text{元})$$

当发行者行使赎回权时，债券的内在价值为：

$$P=\frac{100}{1+8\%}+\frac{100}{(1+8\%)^2}+\cdots+\frac{100}{(1+8\%)^5}+\frac{1\,050}{(1+8\%)^5}=1\,113.8834(\text{元})$$

我们可以看到当公司行使赎回权时，债券的内在价值降低了。可见，可赎回权的存在，降低了该类债券的内在价值，并且降低了投资者的实际收益率。(可以想象得到，可赎回债券相当于投资者卖出了一个关于公司的看涨期权，从其名字就可以看出：callable bond，这样，可赎回债券的价值要低于相同条件下普通债券的价值)一般而言，息票率越高，发行人行使赎回权的概率越大，即投资债券的实际收益率与债券承诺的收益率之间的差额越大。

4.3.4 税收待遇

在不同的国家之间，由于实行的法律不同，不仅不同种类的债券可能享受不同的税收待遇，而且同种债券在不同的国家也可能享受不同的税收待遇。债券的税收待遇的关键，在于债券的利息收入是否需要纳税。由于利息收入纳税与否直接影响着投资的实际收益率，所以，税收待遇成为影响债券的市场价格和收益率的一个重要因素。例如，美国法律规定，地方政府债券的利息收入可以免交联邦收入所得税，所以地方政府债券的名义到期收益率往往比类似的但没有免税待遇的债券要低20%至40%。此外，税收待遇对债券价格和收益率的影响还表现在贴现债券的价值分析中。贴现债券一般具有延缓利息税收支付的优势，但对于美国地方政府债券的投资者来说，由于贴现的地方政府债券可以免交联邦收入所得税，使得贴现债券的税收优势不复存在，所以，在美国地方政府债券市场上，贴现债券品种并不流行，对于(息票率低的)贴现债券的内在价值而言，由于具有延缓利息税收支付的待遇，它们的税前收益率水平往往低于类似的但没有免税待遇的(息票率高的)其他债券，所以，享受免税待遇的债券的内在价值一般略高于没有免税待遇的债券。

4.3.5 流通性

债券的流通性或者流动性，是指债券投资者将手中的债券变现的能力。如果变现的速度很快，并且没有遭受变现所可能带来的损失，那么这种债券的流通性就比较高；反之，如果变现速度很慢，或者为了迅速变现必须承担额外的损失，那么，这些债券的流动性就比较低。

通常用债券的买卖差价的大小反映债券的流动性大小。买卖差价较小的债券的流动性比较高;反之,流动性较低。这是因为绝大多数的债券的交易发生在债券的经纪人市场,对于经纪人来说,买卖流动性高的债券的风险低于买卖流动性低的债券,故前者的买卖差价小于后者。所以,在其他条件不变的情况下,债券的流动性与债券的名义到期收益率之间是反比关系,即流动性高的债券的到期收益率比较低,反之亦然。相应地,债券的流动性与债券的内在价值呈正比关系。

4.3.6 违约风险

债券的违约风险是指债券发行人未按照契约的规定支付债券的本金和利息,给债券投资者带来损失的可能性。债券评级是反映债券违约风险的重要指标。美国是目前世界上债券市场最发达的国家,所拥有的债券评级机构也最多,其中最著名的两家是标准普尔公司(Standard &Poor's,S&P)和穆迪投资者服务公司(Moody's Investors Services)。尽管这两家公司的债券评级分类有所不同,但是基本上都将债券分成两类:投资级和投机级,投资级的债券被评定为最高的四个级别。例如:标准普尔公司和穆迪投资者服务公司分别将AAA、AA、A、BBB和Aaa、Aa、A、Baa四个级别的债券定义为投资级债券,将BB级以下(包括BB级)和Ba级以下(包括Ba级)的债券定义为投机级债券。有时人们将投机级债券称为垃圾债券(junk bonds),将由发行时的投资级转变为投机级的债券形象地称为"堕落天使"(fallen angels)。标准普尔公司的债券评级标准详见表4.4。在政府债券和公司债券之间,包括AAA级在内的公司债券的违约风险高于政府债券;在政府债券内部,中央政府债券的违约风险低于地方政府的债券;在公司债券内部,AAA级的债券的违约风险最小,并随着评级的降低,违约风险不断上升。

表 4.4 标准普尔公司的债券评级标准

级别	评级标准
AAA	是标准普尔公司评定的债券的最高级别,说明支付利息和偿还本金的能力最高
AA	说明支付利息和偿还本金的能力很强,与最高级别相比稍逊一点
A	尽管A说明环境变更和经济条件变更比上述两种级别更易引起负面影响,但其支付利息和偿还本金的能力依然相当强

续表

级别	评级标准
BBB	被定义为BBB级的债券被认为有足够的能力支付利息和偿还本金。尽管在通常情况下其能得到足够的保护，但与前几级相比，变化的环境更可能削弱该级债券的还本付息能力
BB~C	定级为BB、B、CCC、CC和C的债券被认为还本付息有明显的投机特征。BB表示最低程度的投资性，而C则表示最高程度的投机性。尽管这种债券很可能质量尚可，并且有某些保护性条款，但是其不确定性和可能受不利条件影响的程度则更为严重
CI	是为没有利息收入的收入债券(income bonds)准备的
D	被定为D级的债券现在已经处于违约状态

那么，债券的违约风险与债券的收益率之间存在什么关系呢？既然债券存在着违约风险，投资者必然要求获得相应的风险补偿，即较高的投资收益率。所以，违约风险越高，投资收益率也应该越高。在美国债券市场上，联邦政府债券的违约风险最低，地方政府债券的违约风险次低，AAA级的公司债券的违约风险较高，D级的公司债券的违约风险最高。相应地，上述债券的收益率从低到高排列。但是，由于地方政府债券的利息收入可以免缴联邦政府收入所得税，所以，美国地方政府债券的投资收益率低于联邦政府债券的收益率，而联邦政府债券的投资收益率又低于AAA级的公司债券的收益率。在公司债券中，投资级债券的投资收益率低于投机级债券的收益率。

表4.5是对本节内容的总结，综合了上述六方面的债券属性与债券价值分析之间的关系。

表4.5　债券属性与债券收益率(价格)

债券属性	与债券收益率(价格)的关系
期限	当市场利率调整时，期限越长，债券的价格波动幅度越大，但是，当期限延长时，单位期限的债券价格的波动幅度递减
息票率	当市场利率调整时，息票率越低，债券价格波动幅度越大
可赎回条款	当债券被赎回时，投资收益率降低。所以，作为补偿，易被赎回的债券的名义收益率比较高，不易被赎回的债券的名义收益率比较低
税收待遇	享受税收优惠待遇的债券的收益率比较低，无税收优惠待遇的债券的收益率比较高
流动性	流动性高的债券的收益率比较低，流动性低的债券的收益率比较高
违约风险	违约风险高的债券的收益率比较高，违约风险低的债券的收益率比较低

4.4 债券组合管理

我们在前面几节讲述了影响债券收益和价值的债券性质，在本节，我们将讨论债券的组合管理。现代组合理论对债券管理的影响要远远小于对普通股管理的影响，而且债券管理中使用的一些组合管理技术只适用于债券领域，也并不源于现代组合理论。本节我们要介绍债券领域特有的技术及一般组合理论在债券领域中的应用。

本节分为四部分，第一部分讨论债券管理者面临的主要风险来源——收益率曲线的变动，以及测度债券对风险来源敏感性的指标。第二部分，我们将讨论构建一个与上述风险隔绝的债券组合的方法。这通常被称为消极的组合策略，但我们也将看到，它们也包括积极的组合调整。第三部分是积极的债券管理，既包括积极债券管理的特有技术，也包括现代组合中的债券管理技术，即首先估计期望收益率，然后估计方差、协方差。最后一部分是关于债券和利率互换。

债券收益包括两个部分：利息收入和价格变动导致的资本损益。价格发生变动可能是由于时间的推移或收益率曲线的移动。在下面的分析中，为方便起见，我们假设利息是每年支付一次。而且，我们还假定收益率曲线是平的，即所有的即期利率都等于 r。

4.4.1 价格因时间推移而发生变动

首先考虑价格因时间推移而发生变动的情况。假定收益率曲线是平的，利率为 8%。现在，我们看一个距离到期日还有 4 年的贴现债券。到期日支付 100 元，3 年期贴现债券的价格为：

$$P_4=\frac{100}{(1+8\%)^4}=73.50(\text{元})$$

假定即期利率在第一年保持不变。因为在一年后，该债券一定与一个二年期债券的收益率相等，所以它的价格变为：

$$P_3=\frac{100}{(1+8\%)^3}=79.38(\text{元})$$

这一价格的变化发生在第一年，等于 $P_3-P_4=5.88$ 元，所以第一年的收益率为：

$$\frac{P_3-P_4}{P_4}=\frac{5.88}{73.50}=8\%$$

时间推移对债券价格变动的影响对于贴现债券不难理解。因为贴现债券不支付利息，所有的收益都源于价格变化。息票债券的价格也可能因时间推移而发生预期变化。大量的债券除了票面利率不同外，在其他方面都是可比的。这些债券必须为投资者提供相同的收益率。所以，这些债券会有预期价格变化。例如，一个票面利率为 4%的债券将折价出售，如果一个类似债券的当期利率是 10%，预期价格会上升。大多数债券将预期价格变动作为收益的一部分。

4.4.2 非预期价格变化

价格变动的另一个原因是对未来利率的预期发生变化（非预期的收益率曲线变化）。假定收益曲线移动，所有期限的新利率是 14%，再假定移动是瞬间发生的。在这种情况下，4 年期贴现债券的价格变为：

$$P_4'=\frac{100}{(1+14\%)^4}=59.21(\text{元})$$

导致的价格变化为 $P_4'-P_4=-14.29$ 元。

如果一段时间内收益率曲线保持不变或预期不变，因时间推移引起的价格变动容易计算，但因非预期的收益率曲线变化导致的价格变动就不一样了。

如果我们知道对于未来利率的预期是如何随时间变化的，而其他人不知道，那么我们可以计算出每一债券的价格变化，并将所有的资金投入到收益最高的债券中去。但是，这是不可能的，我们最多只能计算出每一债券对收益率曲线移动的敏感性。

4.4.3 对收益率曲线移动的敏感性

在前面的章节中，我们用贝塔来测度一个普通股对某个指数的敏感性。债券也有一个类似的指标，称为久期(duration)。久期测度债券价格对利率

变动的敏感性。用符号表示如下：

$$D=\sum_{i=1}^{n}\frac{iC}{P\ (1+r)^{i}}+n\frac{F}{P\ (1+r)^{n}}$$

D 表示久期，P 为债券价格。由上式可知，D 为债券现金流量产生时刻的加权时间，权重为现金流的现值与债券价格的比值。由于所有现金流的现值之和为债券价格，故权重之和为 1。

$$P=\sum_{i=1}^{n}\frac{C}{(1+r)^{i}}+\frac{F}{(1+r)^{n}}$$

对上式关于 r 求导得到：

$$\frac{\partial P}{\partial r}=\sum_{i=1}^{n}\frac{C(-i)}{(1+r)^{i+1}}+\frac{F(-n)}{(1+r)^{n+1}}=\frac{1}{1+r}(-D)P$$

其中第二个等号要用到久期的定义。

对上式变形可得：

$$\frac{\partial P}{P}=\frac{1}{1+r}(-D)\partial r$$

上式说明了一个债券的久期越大，同样的利率变动导致的债券价格变动的幅度越大，且利率变动方向与价格变动方向相反。

由于贴现债券只有到期时才有现金流，故现金流产生时刻的加权时间为它的期限。因此，如果假设收益率曲线是平的，一个贴现债券对收益率曲线变化的敏感性就直接与它的期限成比例。

由于 $D=\sum_{i=1}^{n}\frac{iC}{P\ (1+r)^{i}}+n\frac{F}{P\ (1+r)^{n}}<\sum_{i=1}^{n}\frac{nC}{P\ (1+r)^{i}}+n\frac{F}{P\ (1+r)^{n}}$ $=n$，所以贴现债券的久期小于它的期限。到此为止，我们都是假设收益率曲线是平的，而且发生了移动。

表 4.6 给出了不同期限和不同票面利率的债券的久期。要注意，债券的久期要远远小于它的期限，尤其是那些期限很长的债券。

表 4.6 不同期限和票面利率的债券的久期

单位:年

票面利率(%)	距离到期日的年数		
	3	5	10
4	2.88	4.57	7.95
6	2.82	4.41	7.42
8	2.78	4.28	7.04
10	2.74	4.17	6.76
12	2.70	4.07	6.54
14	2.66	3.99	6.36

仔细观察表 4.6 我们可以发现以下规律:

(1)票面利率越高,久期越小。原因不难理解,随着票面利率提高,前期现金流的现值相对于最终现金流的现值增加。这样就提高了前期现金流的权重,从而减小了久期。而这正好解释了上节的问题:票面利率越高,债券价格受利率波动的影响越小。

(2)一般来说,期限越长,久期也越大。这个也解释了上节的疑问。

4.4.4 凸性(convexity)

近年来,人们越来越认识到在收益率曲线出现微小变化时,久期可以很好地解释价格的变化,但如果收益率曲线变换幅度很大,其解释就差强人意了。下面予以解释。

$$P=\sum_{i=1}^{n}\frac{C}{(1+r)^i}+\frac{F}{(1+r)^n}$$

当 r 变为 $r+\Delta r$ 时,则新的价格:

$$P'=\sum_{i=1}^{n}\frac{C}{(1+r+\Delta r)^i}+\frac{F}{(1+r+\Delta r)^n}$$

对上式在 r 处展开得到:

$$P'=\sum_{i=1}^{n}\left(\frac{C}{(1+r)^i}+\frac{-i}{1+r}\frac{C}{(1+r)^i}\Delta r+\frac{1}{2}\frac{i(i+1)}{(1+r)^2}\frac{C}{(1+r)^i}(\Delta r)^2+\cdots+\right.$$

$$\frac{1}{n!}\frac{(-1)^n i(i+1)\cdots(i+n-1)}{(1+r)^n}\frac{C}{(1+r)^i}(\Delta r)^n+\cdots\Big)+$$

$$\left(\frac{F}{(1+r)^n}+\frac{-n}{1+r}\frac{F}{(1+r)^n}\Delta r+\frac{1}{2}\frac{n(n+1)}{(1+r)^2}\frac{F}{(1+r)^n}(\Delta r)^2+\cdots+\right.$$

$$\left.\frac{1}{n!}\frac{(-1)^n n(n+1)\cdots(n+n-1)}{(1+r)^n}\frac{F}{(1+r)^n}(\Delta r)^n+\cdots\right)$$

括号里面只取前两项并化简有：

$$P'\approx P+\frac{-DP}{(1+r)}\Delta r+\frac{C}{(1+r)^2}P(\Delta r)^2$$

其中 $C=\frac{1}{2}\left[\sum_{i=1}^{n}\left(i(i+1)\frac{C}{(1+r)^i P}\right)+n(n+1)\frac{F}{(1+r)^n P}\right]$

可以将上式进一步变形得到：

$$\frac{P'-P}{P}=\frac{-D}{(1+r)}\Delta r+\frac{C}{(1+r)^2}(\Delta r)^2$$

这样可以看出，久期只是一阶泰勒展开而已，新定义的 C 我们称之为凸度。

为了说明凸性的应用，我们来看一个例子。5 年期的息票债券，面值为 1 000元，息票利率为 10%。假设利率变化由 10%变为 12%。当利率为 10%时，5 年期息票债券的价格是 1 000 元。当利率为 12%时，债券价格为 927.90 元。

利率由 10%上升至 12%时，价格变化比率为：

$$\frac{P_{12}-P_{10}}{P_{10}}=-7.21\%$$

用 Excel 计算债券的久期为 4.135482。

这样用$\frac{\partial P}{P}=\frac{1}{1+r}(-D)\partial r$：

我们得到：

$$\frac{P_{12}-P_{10}}{P_{10}}=\frac{1}{1+10\%}\times(-4.135482)\times 2\%=-7.5191\%$$

同样可以计算凸度是 10.09003，这样把上述数据代人上面公式可得

$$\frac{P_{12}-P_{10}}{P_{10}}=-7.5191\%-\frac{10.09003}{(1.10)^2}\times(2\%)^2=-7.1578\%$$

这样显然可以看出加入凸度之后，结果更精确了。

4.4.5 债券凸性与久期之间关系

现在，我们讨论债券的凸性与久期的关系。从上面的分析可以发现它们都涉及债券收益率变动与债券价格变动之间的联系。

图 4.2 中的曲线是真实的债券价格与利率之间的关系。假设市场利率为 B，则债券的价格为 PB。作为久期计算债券价格与利率之间的关系，就是过曲线中 M(坐标为(B,P_B))点作一条切于上述曲线的直线。现在假设市场利率上升至 C，则真实的债券价格为 P_{C1}，而由久期算出的价格为 P_{C2}。而 $PC1<P_{C2}$。当市场利率下降到 A 时，真实的债券价格为 P_{A1}，而由久期算出的价格为 P_{A2}。而 $P_{A1}>P_{A2}$。而凸度可以更接近于真实的债券价格。即含凸度的表达式可以更近似地接近那条曲线。大家可以看到凸性是个好东西。因为当利率上升时，它使得债券的价格下降得比只用久期表示的更少；而当利率下降时，它使得债券的价格上升得比只用久期表示的更多。

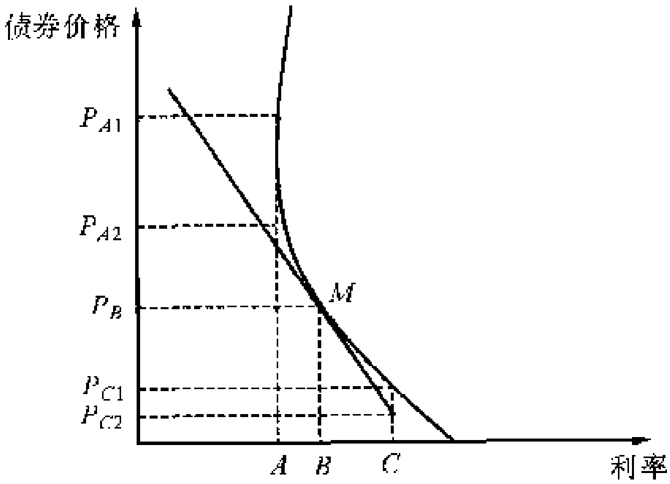


图 4.2 债券的久期与凸度

4.4.6 防范期限结构变动

大多数管理者将利率期限结构的变动视为债券组合的主要风险来源。就像市场移动对所有股价都产生系统性影响一样，期限结构的变动也影响所有债券的价格。

有两种技术用来将组合与期限结构的变动隔绝，即精确匹配和免疫化。

(1)精确匹配

精确匹配就是找到成本最低的投资组合，该组合的现金流能精确匹配由该组合予以资金支持的现金流出。见表 4.7 中的例子。在这个例子中，我们假设必须在未来 3 年满足现金流出 100 美元、1 000 美元和 2 000 美元。这些现金流可能是为了支付养老金。债券组合是被用来满足这些债务的投资。一个精确匹配的计划是确定一个由 1 年期、2 年期和 3 年期债券组成的组合，使支付的利息和本金正好与 3 笔现金流匹配。

表 4.7 中的组合 A 就是现金流匹配的组合。大多数投资机构还考虑前期有多余现金流可以用于满足后期现金流需求的组合，也达到现金流匹配的效果。表 4.7 中的组合 B 说明了这一点。在这个例子中，第 1 期流入的 195 美元中的 100 美元用于满足第一期的债务 100 美元，多出的 95 美元进行投资直到第 2 期，用于弥补第二期缺少的 100 美元。只要第一期的剩余资金投资可获得 5 美元的利息，组合就可以保证现金流匹配。

表 4.7 现金流匹配的组合

	时期		
	1	2	3
债务	100	1 000	2 000
组合 A	100	1 000	2 000
组合 B	195	900	2 000

精确匹配计划有两个风险。第一，由于债券违约或被赎回，现金流可能不能实现。第二，如果计划含有现金结转(表 4.7 中的组合 B)，就会存在结转资金收益率不足的风险。但是，即使收益率曲线发生移动，管理者还是在相当程度上保证了满足债务需求。

(2)免疫化

第二种防范利率移动的技术是免疫化计划。前面我们已经介绍了久期，它用于测度一个债券或一个债券组合对利率移动的敏感性。免疫化利率通过匹配资产和负债的久期，来消除组合对期限结构移动的敏感性。因此，如果久期确实是一个测度对利率移动敏感性的指标，那么期限结构的移动对资产和负债现值的影响是相同的，使得组合满足债务的能力保持不变。如果利率上升，资产和负债的现值等量下降。同理，如果利率下降，资产和负债的现值等量增加。也许用贝塔来做类比有助于理解。如果一个债务的贝塔值是 1.5，

那么购入一个贝塔是 1.5 的资产就构成了零贝塔组合。因为债务是现金流出，所以贝塔是－1.5. 由于－1.5 贝塔的债务和＋1.5 贝塔的资产构成的零贝塔组合对市场变动无敏感性。

为了进一步说明，考虑一个 5 年后支付 100 美元的债务。投资计划的目标是满足这一债务。如果购入一个 5 年期债券，投资者可以确定到期时的债务价值，但不能确定支付的利息再投资的回报率。如果利率上升，债务得到超额满足，因为支付的利息再投资的回报率高于预期。但是，如果利率下降，债务将得不到满足，因为支付的利息再投资的回报率低于预期。如果投资者购入期限长于 5 年的债券，投资者将不能确定第 5 年时债券的价值。假定利率上升，到期时支付的利息的总价值将高于预期，因为支付的利息再投资回报率上升。但是，债券到期日的价值因为利率的上升而降低。这两种影响相反。如果恰当地选择债务，这两种影响正好平衡。同理，如果利率下降，支付的利息的再投资回报率低于预期，到期时支付的利息的总价值将下降，但下降的利率使得债券价格上升。同样地，或许可以找到某种期限的债券使得这两种影响正好相互抵消。

这里讨论的原因能够说明为什么免疫化可以奏效。在等于资产久期的时点，再投资收入的变化正好与债券价值的变化相匹配。表 4.8 说明了这一点。假定所有期限的利率现在为 11%，再假定债券每年支付一次利息，利率为 13.52%，期限为 5 年。表 4.8 的第 2 列列出了这些现金流。这一债券的久期是 4 年。如果利率保持在 11%，到第 4 期末债券的价值是 165.946 美元(这里前期现金流也要拿来投资，这里的价值不同于市场上债券的价格，我们可以把它等价于一个期限为 4 年的贴现债券，且终期支付 165.946 美元)。

表 4.8 不同利率条件下的债券价值

时期	现金流	第 4 期末的价值		
		11%	10%	12%
1	13.52	$13.52\times(1.11)^3$	$13.52\times(1.10)^3$	$13.52\times(1.12)^3$
2	13.52	$13.52\times(1.11)^2$	$13.52\times(1.10)^2$	$13.52\times(1.12)^2$
3	13.52	$13.52\times(1.11)$	$13.52\times(1.10)$	$13.52\times(1.12)$
4	13.52	13.52	13.52	13.52
5	113.52	$113.52\times(1.11^{-1})$	$113.52\times(1.10^{-1})$	$113.52\times(1.12^{-1})$
		165.946	165.946	165.974

如果利率下降到10%，到第4期末债券价值是165.946美元。债券价值不变，原因在于支付的利息减少了0.930美元，第5期支付的113.52美元在第4期末的价值增加了0.930美元，两者正好抵消。如果利率上升到12%，到第4期末支付的利息增加，而第5期支付的113.52美元在第4期末的价值减少。尽管两者没有完全抵消，但非常接近。这个例子说明了免疫化的思想。如果我们在第4期有一个债务，我们可以购买足够多的债券来满足这一债务。例如，一个995美元的债务可以用6(995/165.946≈6)个债券来匹配。无论利率是升还是降，债务都可以得到满足。

为什么表4.8中的债券具备这些性质呢？选择表4.8中的债券是因为久期是4年。贴现债券的久期等于其期限。一个4年期的贴现债券（终值为165.946美元）的久期也是4年。我们已经说明久期是测度对利率敏感性的指标。在利率为11%时，第4年末的息票债券的价值等于这个贴现债券。当利率变动时，第4年末贴现债券的价值仍然为165.946美元。由于它们对利率敏感性相同，所以即使利率变动，息票债券在第4年末的价值仍然等于贴现债券的到期价值，这样，就解释了在第4年末，息票债券的价值永远在165.946美元附近。

上面我们讨论增加凸性后可以提高价格变化估计值与实际值的近似程度。很多采用免疫化策略的管理者在凸性和久期上都进行匹配。他们担心债务和资产的凸性相差太大，单独使用久期来近似可能导致比较大的误差。添加凸性的效果是两面的。一方面，添加后能够对期限结构的移动进行更好的防范。但是另一方面，在久期和凸性两方面都能匹配的组合很少。因此，在两个指标上都进行匹配可能导致组合的成本比较高。

免疫化策略的使用非常广泛，其目的就是减轻利率变动的影响。对于如何设计免疫化组合已经进行了大量的研究。现在，我们讨论的是这些研究的一些含义。一个债券组合的久期是构成组合的各资产的久期的加权平均。用X_i表示债券i在组合中的比例，D_i表示资产i的久期，D_P是由N个债券构成的组合的久期。

则 $D_P=\sum_{i=1}^{N}X_iD_i$

下面对上面进行解释：

因为$\frac{\partial P}{P}=\frac{1}{1+r}(-D)\partial r$，设总资产的价值为$P$，则资产$i$的价值$P_i$为$X_iP$，由于$\frac{\partial P_i}{P_i}=\frac{1}{1+r}(-D_i)\partial r$，其中$D_i$为第$i$项资产的久期。而$\partial P=$

$\sum_{i=1}^{N} X_i \partial P_i, P = \sum_{i=1}^{N} X_i P_i$。

这样 $\partial P = P \frac{1}{1+r} \partial r \sum_{i=1}^{N} X_i D_i$。所以 $D_P = \sum_{i=1}^{N} X_i D_i$。

显然，要构造一个具有一定久期的组合有多种方式。例如，假定一个债券组合的久期为10年。再假定有3个债券，久期分别为10，12，8，6年。只有久期为10年的债券显然可以满足条件。或者可以将1/6的资金投资于久期为6年的债券，1/4的资金投资于久期为8年的债券，剩余的7/12投资于久期为12年的债券。这样可以使组合的久期为10年，因为：$\left(\frac{1}{6}\right)\times 6+\left(\frac{1}{4}\right)\times 8+\left(\frac{7}{12}\right)\times 12=10$，且 $\left(\frac{1}{6}\right)+\left(\frac{1}{4}\right)+\left(\frac{7}{12}\right)=1$。

我们可以采用两种方法来实施免疫策略，它们是杠铃策略和集中型策略。集中型策略是找到一组债券，每一债券的久期都与债务的久期接近。例如，如果债务是10年，那么债券的久期可以在9～11年之间。债券组合久期集中在债务久期附近。杠铃策略选用的债券的久期差异很大，例如，5年和15年。要满足久期为10年的要求，可以是一半久期为5年的债券和一半久期为15年的债券。杠铃策略是不必为满足每一个债务构建单独的债券组合。因为针对每一个债务都只需要这些债券，而集中型策略必须针对不同久期的债务选在这个久期附近的债券。

这两种策略中的哪一种可以更好地使资产和债务对利率变动的敏感性相等？这一问题的实证研究对集中型策略有所支持。这正所谓有利必有弊。理由如下：所有久期指标都是对利率形态实际移动的影响的近似。当单个资产和债务的久期类似时，这些误差也是类似的。但当组合中的单个资产的久期与债务不同时，即使组合的久期与债务相同，误差的模式也会有很大的不同。这正是杠杆策略的误差形式。因此久期估计不准确部分解释了为什么实证倾向于集中型策略。

最后，免疫化策略还有一方面值得讨论。免疫化策略经常被看作消极策略——购买一组债券持有至到期日。这种印象是不正确的。久期是针对某一收益率曲线计算的。随着收益率曲线的移动，久期发生变化，资产和债务的久期可能不相等。如果两者的差异足够大，就要进行重组了。此外，即使收益率曲线保持不变，资产和负债的久期也会出现差异，除非资产和债务的现金流模式相同。这时也要进行重组。所以，免疫化是一种积极策略。

免疫化策略有哪些风险？主要的风险是选择了错误的久期指标。每种

久期指标都是基于对收益率曲线移动模式的不同假定。组合免疫化通常只使用一种指标。例如,用来准确衡量收益率曲线平行移动对价格影响的指标,无法准确衡量收益率曲线变陡(长期利率升高幅度大于短期利率)对价格的影响。

为了避免读者过于担心,我们有必要再次指出即使是本章讨论的最简单的指标,其效果也很好。免疫化策略的第二个风险是当组合没有做到免疫时收益率曲线发生移动。如前所述,时间推移或收益率曲线的微小变化都会使组合不免疫。组合产生的现金流被用于购买债券,以重新平衡组合,使资产的久期和债务的久期相近。债券的买卖(重新平衡)也可以用来对组合进行精确的免疫化。但是债券的互换需要支付成本,所以管理者允许资产的久期偏离债务的久期,而不是在所有时点都做到免疫化。风险在于管理者进行债券互换,以调整久期使组合免疫之前,利率可能发生剧烈的变化。

现金流匹配(也就是精确匹配)的组合当然是免疫化的。由于它的免疫化来自现金流匹配,而不是某一指标的准确性,所以它的风险通常较小。因此,机构要采用免疫化,就要求免疫化组合的成本低于现金流匹配的组合成本。通常采用部分组合是现金流匹配,其余的采用免疫化。

在这一小节,我们介绍了防范利率移动的各种技术。接下来,我们将讨论构建组合的技术,假定评价的是每年的业绩表现。

4.4.7 债券组合的年收益率管理

上面我们讨论了如何设计出对收益率曲线不敏感的债券组合。这些组合的收益率各期都有可能发生很大的波动,因为其主要是为满足将来的债务,而不是最大化每期的收益率。有的管理者感兴趣的不是将来的债务能否得到满足,而是组合的各年收益率。债券基金和很多养老金的管理者关心的是每年收益率的变动。

本小节分成两部分,第一部分讨论指数化。指数化是对每年收益率感兴趣的管理者所使用的消极策略。第二部分讨论积极的债券管理技术。

(1)指数化

债券管理者青睐的另一种消极策略是复制指数。债券中复制指数的主要原因是业绩。积极管理的基金中的极少数可以超越主要债券指数的业绩。鉴于这种情况,很多养老基金管理者将他们的资产指数化。债券的指数化操作与普通股不同。资本市场上有大量的公司债券,其中很多是不活跃的。因此,

按指数的比例持有债券是不可行的。指数化一般是通过单元匹配来实现。一个债券的主要特性包括类型(政府、公司、公用事业等)、久期、票面利率和债券等级,然后对所有的特性确定组合中符合某种特性的比例。例如,组合中,级别为 Baa 级、久期在 4～5 年之间、票面利率为 8%～9%的公司债券的比例是多少？对债券特性的所有可能组合都要计算出这些比例,然后构建一个债券组合,每一单元的比例与债券近似。这种复制指数在匹配指数的业绩方面非常成功。

(2)积极的债券管理

债券领域有四种积极投资策略,即总利率预测、板块选择、板块轮换以及债券价格错位。

总利率预测　对一个管理者来说,每年收益率变动的主要原因是收益率曲线的非预期移动。在大多数年份里,非预期收益率的绝对值远远高于预期收益率。例如,尽管中期和长期债券的期望收益率几乎是一样的,但 1993 年长期债券的收益率是 16.38%,中期债券的收益率是 7.91%。同样地,在 20 世纪 80 年代,长期债券收益率在－3%～42%之间变化,而期望收益率的变化要小得多。从对久期的讨论中我们知道利率上升是预期之外的,久期短的债券的损失要小于久期长的债券;如果利率非预期性下降,久期短的债券要比久期长的债券盈利小。

因此,管理人遵循的一种投资策略是,当他们预期利率上升幅度高于市场预期时(由收益率曲线反映),就会缩短久期。当预期利率下降幅度大于市场预期时,就会延长久期。债券管理者要为这种时机把握有所付出。大多数债券并不像普通股流动性那么强。那么市场规模大而且可以在短期内进行交易的债券主要是某些期限的政府债券。如果只限于购买这些债券就会导致期望收益率低于购买高期望收益的公司债券,或者低于购买价格错位的债券。此外,一些实证指出,用于时机选择的债券由于其市场流动性小,它们的收益率比政府债券低。最后,为了时机选择集中于一些政府债券会使得组合的分散化程度相对较低。

没有人能准确预测所有的时机。对未来利率的预测必须是准确、与市场共识不同的,才能产生效果,因为市场共识已经反映在现有的利率中了。如果一个预测者准确预测利率方向的几率为 60%,对市场的预测就是极为出色的。市场时机选择涉及每期对未来利率的估计。因为即使是优秀的预测者也会经常犯错,所以一个具备时机选择能力的管理者在有可能取得高收益前将会经历长期磨炼。

有些管理者在进行免疫化的同时也进行市场时机选择。对这些管理者来说，一般的策略是使资产和债务的久期相等。如果他们预期利率上升的幅度高出市场预期，资产的久期就会小于债务的久期；如果他们预期利率下降的幅度大于市场预期，资产的久期就会大于债务的久期。采用组合免疫化策略的管理者在运用市场时机选择上也会非常谨慎。

板块选择 管理者之所以进行板块选择是因为他们相信某些板块在长期会有更好的表现。板块选择最常见的形式是降低组合的平均信用等级。例如，一个管理者可能认为垃圾债券提供的风险溢价高于任何风险差异所能解释的水平，而且这一溢价将会长期存在，垃圾债券的表现会超出其他类型的债券，这个管理者就会总是投资垃圾债券。

板块轮换 板块轮换可以运用前面讨论的任何债券性质来进行。板块轮换与板块选择相关。板块轮换需要加重某一板块的比重，因为管理者相信板块在下一期的表现相对出色。例如，平价出售的AAA级公司债券的到期收益率要高于平价出售的同样期限的政府债券。这一差异部分是违约溢价，部分是风险溢价。如果投资者相信违约风险会增加，这一价差将会扩大。如果一个管理者相信市场会对预期的风险增加反应溢价过度，他就会转向于AAA级债券。如果管理者的判断是正确的，他就可以在那一期获得超出正常水平的风险溢价，如果在很多投资者意识到反应过度后违约溢价随之减少，管理者还可以获得额外的收益。例如，假定政府债券和公司债券的正常违约溢价为0.125%，而违约溢价扩大到0.25%。如果价差还原到0.125%，那么AAA级公司债券的收益率相对于政府债券来说下降，而其价格会相对上升。

如前所述，板块轮换可以根据影响债券价格的任意因素来进行。我们再来看第二个例子，假定市场低估了利率的波动性。利率波动性越大，利率可能发生变化的幅度就越大。利率变动幅度越大，就越可能出现极低的利率水平，公司就越应该赎回债券。于是，投资者如果认为市场低估了利率波动性，就会认为可赎回债券的吸引力相对较低，从而会将可赎回债券板块从组合中转出。

债券价格错位 债券选择程序通常有两种。一种是认为债券分类是准确的(如AAA或AA)，并试图在某一给定的类别中找到最有吸引力的债券。例如，一个公司可能将所有8～10年AA级不可赎回的债券归为风险等同的一类，然后公司可以考察所有的满足这一标准的债券，从中选择最有吸引力的。提供债券服务的公司如Barra或Gifford Fong用实际价格和理论价格的差异

来衡量吸引力。理论价格是将未来的现金流以估计的即期利率折现，并经过期权价值调整之后的结果。

公司进行债券选择的另一种程序是寻找分类错误的债券。这在级别低的债券中尤其普遍。债券评级中隐含着违约的可能性及违约发生时的期望损失。使用这种方法的公司考察发行债券公司的特性，并找出违约可能性或期望损失不同于债券级别所隐含的水平的债券。那些具有更强的吸引力特性的债券将被选中。

下面我们将讨论类似于在股票市场中使用的用于选择债券的技术。

4.4.8 运用现代组合理论积极选择债券

现代组合理论可以用于债券和股票管理。这里，我们将讨论这是如何进行的。

(1)估计期望收益

我们先考虑最简单的一类债券：美国联邦政府发行的不可赎回债券，接着我们再探讨非政府债券的期望收益问题，以及可赎回性和税收因素的影响。

虽然可以使用任何一种利率期限结构理论来估计债券的期望收益，我们开始还是选择最简单的理论——纯预期理论来说明方法。根据纯预期理论，所有债券某一期限的收益率必须相等。因此，下一期的期望收益率就是 1 期的即期利率。

如果我们认识到债券价格错位，这一结论就要经过修正。如果存在价格错位，那么期望收益率就是价格错位和 1 期即期利率的函数。为了理解这一点对期望收益的影响，我们有必要对市场需要经过多长时间才能修正价格作出假设。我们假定价格在 1 个时期内(一年内)调整为均衡水平。这一假设是大多数商业服务公司的隐含假设，看起来与实证也是一致的。为计算一个债券的期望收益率，我们需要计算 1 年后的期望均衡价值。然后，根据这一期内预期支付的利息和资本利得，我们可以计算期望收益率。

为了计算 1 期以后债券的价格，我们需要对未来时点的即期利率或远期利率进行预期。在前面，我们学习了如何根据即期利率得出远期利率。如果预期理论成立，预计远期利率不会随着时间而改变。表 4.9 是一组假设的利率。根据这些假设的利率，我们来考察一个 5 年以后到期的债券的收益率，该债券每期支付利息 8 美元。该债券本金为 100 美元，当前价格为 82 美元。

表 4.9　一组假设利率

时期	当前 1 期的远期利率(%)	一年后的期望远期利率(%)
1	10	
2	11	11
3	12	12
4	13	13
5	14	14

如果在期初该债券价格是均衡水平，它的价格为

$$P_0=\frac{8}{1.10}+\frac{8}{1.10\times1.11}+\frac{8}{1.10\times1.11\times1.12}+\frac{8}{1.10\times1.11\times1.12\times1.13}+\frac{108}{1.10\times1.11\times1.12\times1.13\times1.14}$$
$$=86.16\text{ 美元}$$

1 期后的期望价格为：

$$P_1=\frac{8}{1.11}+\frac{8}{1.11\times1.12}+\frac{8}{1.11\times1.12\times1.13}+\frac{108}{1.11\times1.12\times1.13\times1.14}$$
$$=86.77\text{ 美元}$$

要注意如果债券价格在时点 0 是均衡价格，1 期的现金流就是 8 美元的利息加上 61 美分的资本利得，总收益率为 8.61/86.16，即 10%。当然，10%就是第一期的即期利率。如果债券可以以 82 美元的价格买入，总收益就等于 8 美元的利息加上 4.77 美元的资本利得，收益率为 15.57%。这一收益率可以分为 3 部分：来自利息的 9.76%(8/82)、来自债券均衡价值变化的 0.74%((86.77−86.16)/82)以及来自价格错位的 5.07%((86.16−82)/82)。

如果另一种期限结构理论可以更好地贴近现实，期望收益率还有另外的影响因素。但即便是这样，同样的技术也适用。以流动性溢价理论为例。根据流动性溢价理论，期望收益率是 1 期的即期利率加上使债券价格均衡的各种调整，再加上流动性溢价。同使用预期理论一样，可以采用同样的程序对债券进行估价，但是必须纳入流动性溢价变动的效应。见表 4.10 中的例子。

表 4.10 假定的远期利率

单位：%

时期 1	当期			下一期		
	远期利率	流动性溢价	远期利率（不含流动性溢价）	远期利率（不含流动性溢价）	流动性溢价	远期利率
1	10		10			
2	11	0.1	10.9	10.9		10.9
3	12	0.2	11.8	11.8	0.1	11.9
4	13	0.3	12.7	12.7	0.2	12.9
5	14	0.4	13.6	13.6	0.3	13.9

表 4.10 分为两个部分：与当期相关的计算和与未来 1 期相关的计算。表中给出了当期的远期利率。这些 1 期利率可以用上一节讨论的技术根据即期利率求得。第 2 列是一组假定的流动性溢价，等于远期利率减去第 4 列不含流动性溢价的远期利率。假定这些利率不变。所以，第 5 列与第 4 列相等。有变化的一列是流动性溢价。流动性不变，但每一溢价移向未来 1 期。所以，期初 2 年期流动性溢价 0.1%出现在第 3 期，而不是出现在第 2 期，因为在第 1、2 期以后是第 3 期。

假定还是前面讨论的那种债券：票面利率 8%，本金支付 100 美元。再假定它以均衡价格出售。表 4.10 中截止到当期的远期利率与表 4.9 中的相同，所以均衡价格不变，即

$$P_0 = 86.16 \text{ 美元}$$

根据表 4.10 中的利率，1 期以后的均衡价格为

$$P_1 = \frac{8}{1.109} + \frac{8}{1.109 \times 1.119} + \frac{8}{1.109 \times 1.119 \times 1.129} + \frac{108}{1.109 \times 1.119 \times 1.129 \times 1.139}$$
$$= 87.05 \text{ 美元}$$

如果不假设流动性，1 期以后的均衡价格是 86.77 美元。87.05 美元与 86.77 美元之间的差异就是承担期限风险额外获得的资本利得。总期望现金流是利息收入 8 美元、没有流动性溢价的期初资本利得 86.77 美元－86.16 美元即 61 美分，以及作用于不同现金流的流动性溢价的效应为 87.05 美元－

86.77 美元即 28 美分。总收益率就是(8+0.61+0.28)除以 86.77，等于 10.32%。多出的 0.32%就是流动性溢价效应。对于可能的价格错位可以用前面讨论预期理论时的方法处理。

(2)指数模型

在前面章节中，我们讨论了如何计算普通股收益的方差—协方差矩阵。普通股适用的一般原则同样也适用于债券。但是，债券具有特殊性质，所以有必要进行某些修正以及重新加以说明。

单指数模型 这一部分我们将讨论在债券组合管理中如何运用单指数模型。首先考虑应用于不可赎回的、没有特殊税收效应的政府债券。政府债券的收益率可以分为两个部分：预期收益率和非预期收益率，非预期收益率源于收益率结构变化和相对于收益率结构的债券定价变化的共同作用或单一作用。如前所述，如果预期理论是正确的，且债券定价是公平的，那么所有债券在第一期的期望收益率是相等的。如果其他理论是正确的或者存在价格错位，那么债券可能具有不同的收益率，收益率取决于债券的期限。我们将在假定预期理论成立的前提下推导单指数模型。

非预期收益有两个来源：收益率曲线的变化或相对于收益率曲线的债券价格的变化。在本章已经指出收益率曲线移动导致债券价格变动等于负的久期与一个利率变动指标的乘积。我们还强调久期指标是基于对收益率曲线非预期移动的一个非常简单化的假设。假定在久期推导中假设以外的影响是随机的，再假定相对于收益率曲线的债券收益率的变动是随机的，在这样的假设情况下，对收益率的这两种影响是随机的，以 e_i 表示，e_i 的期望值是 0，方差为 $\sigma_{e_i}^2$。

将这些加总起来，表示如下：

总收益率=期望收益率+源于收益率曲线移动的收益率+对收益率的随机作用

$$R_i = E[R_i] - D_i\Delta + e_i$$

式中：R_i 为债券 i 的收益率；

$E[R_i]$为债券 i 的期望收益率；

D_i 为债券 i 的久期；

Δ 为利率变动除以 1 与利率之和；

e_i 为随机作用，期望值为 0，方差为 $\sigma_{e_i}^2$。

在前面章节中，我们用一个股票指数表示单指数模型。同样我们也可以用一个债券指数来表示债券的收益率。令 X_i 表示债券 i 占债券指数的比例。指数的收益率 R_M 就等于 $R_M = \sum_i X_i R_i = \sum_i X_i E[R_i] - \sum_i X_i D_i \Delta + \sum_i X_i e_i$

如果一个债券指数含有大量的债券，$\sum_i X_i e_i$ 应近似等于 0。这是因为假定 e_i 之间相互独立。将 $\sum_i X_i D_i$ 定义为 D_M，即债券指数的久期。替换后，得到

$$R_M = E[R_M] - D_M \Delta$$

求解上式中的 Δ 并代入式 $R_i = E[R_i] - D_i \Delta + e_i$ 得到：

$$R_i = E[R_i] + \frac{D_i}{D_M}(R_M - E[R_M]) + e_i$$

将 $\frac{D_i}{D_M}$ 定义为 β_i。因为假定 e_i 与债券指数独立，β_i 与前面章节中的含义相同，即 β_i 是 R_i 与 R_M 的协方差除以 R_M 的方差。但是没有理由要用历史数据或修正后的历史数据来估计 β_i，而是可以直接以久期的比率来度量。

上式与普通股的单指数模型类似。如果对单指数模型做出假设：对 $i \neq j$，$E(e_i e_j) = 0$，那么我们会得出：

$$\mathrm{COV}(R_i, R_j) = \frac{D_i D_j}{D_M^2} \sigma_M^2$$

$$\mathrm{VAR}(R_i) = \frac{D_i^2}{D_M^2} \sigma_M^2$$

这并不令人惊讶，因为我们已经说明 $\beta_i = \frac{D_i}{D_M}$。单指数模型在股票选择中使用非常普遍，但在债券管理中对于其使用的经验却很少，直到 20 世纪 80 年代才将单指数模型商业化应用于债券。同样，对于单指数模型在债券管理上的适用性的学术研究也非常少。这与在普通股领域开展的广泛研究形成鲜明对比。

多指数模型　多指数模型比单指数模型更重要有很多原因（在前面的内容中有些已经讨论过）：

（1）可更准确地测度利率变动的效应。

（2）可反映某一风险级别债券与政府债券之间的收益率价差变化引起的变动性。

（3）可反映不同板块——政府、金融机构和公司之间的债券收益率价差变化引起的变动性。

（4）可反映可赎回价值变化引起的变动性。

（5）可反映税收重要性变化引起的变动性。

上面这些因素中的任何一种都可能是非常重要的，使得多指数模型比单

指数模型能更好地反映协方差结构。

很多研究都表明为捕捉期限结构变化必须有两个因素。例如，研究人员曾使用的两个因素是长期利率的变动，以及长期与短期利率之差的变动。举例来说，看下面的两因素模型

$$R_{it}=E[R_i]+\beta_{i1}F_{1t}+\beta_{i2}F_{2t}+e_{it}$$

式中：R_{it}为债券 i 在第 t 期的收益率；

$E[R_i]$为债券 i 的期望收益率；

β_{ij}为债券 i 对 j 因素的敏感性；

F_{jt}为因素 j 在第 t 期的值；

e_{it}为随机误差项。

使用两个因素看起来显著提高了这类收益产生过程的解释力。为了更具体地说明这些因素，我们来看一个例子。第 t 期的因素 1 就是 10 年期政府债券的利率从第 t 期到第 $t+1$ 期的变动。利率的变动度量了期限结构的上下移动。可以认为 β_{i_1} 是负数，这样，当利率上升时，债券价格将下跌，因收益率曲线向上移动导致非预期收益率也将是负数。有些研究人员把短期利率的变动作为因素 2，有的则使用长期和短期利率之差的变动作为因素 2。例如，利差可以是 10 年期与 1 年期利率的差。从第 t 期到第 $t+1$ 期这个利差的变动就是第二个因素的值。如果长期利率保持不变，长短期利差扩大意味着短期利率的下降。这会导致短期债券有一个正的收益率，所以 β_{i2} 应该是一个正数。

估计债券收益产生过程中的敏感性比普通股要难得多。对普通股而言，收益率对因素的时间序列回归往往是估计过程的起点。债券的期限随着时间推移而逐步缩短。一般认为敏感性与期限相关。例如，在单因素模型中，当敏感性与久期相关时，随着期限缩短，久期也缩短，因此敏感性发生变化。要使时间序列回归是估计敏感性的合适方式，敏感性必须是不随时间而改变的。所以用时间序列回归估计敏感性对单个债券来说可能是不合适的。

可以应用的对象是期限固定的纯贴现债券。因为息票债券可以被看作纯贴现债券的组合，而且一个组合的敏感性等于构成该组合的债券的敏感性的加权平均，所以可以用来估计一个债券的敏感性。例如，计算出各因素的每月收益率，以及 10 年期纯贴现债券的每月收益率。当然，10 年期纯贴现债券每月收益率都发生变化。将 10 年期纯贴现债券的收益率对两个因素进行回归。任何一个息票债券都可以看作贴现债券的一个组合。一个组合的敏感性等于组合元素的敏感性的加权平均，权重是各元素占整体的比例。例如，定义

(1)b_{t1}、b_{t2}分别是 t 期贴现债券对因素 1 和因素 2 的敏感性。

(2)$PV(Cf_{it})$是债券 i 在第 t 期现金流的现值。

(3)P_i 是债券 i 的价格。

债券 i 的敏感性是纯贴现债券敏感性的加权平均,即

$$\beta_{i1} = \sum_t \frac{PV(Cf_{it})}{P_i} b_{t1}$$

$$\beta_{i2} = \sum_t \frac{PV(Cf_{it})}{P_i} b_{t2}$$

还有其他的方法用来估计敏感性。对单因素模型来说,我们可以用久期推导出敏感性的一个理论值。还有双参数久期模型,其对敏感性的推导类似于双因素模型。最后,还有研究人员将久期作为第一个因素,将凸性作为第二个因素。这两个因素都可以直接计算可得。

4.4.9 互换

近年来,互换日益成为债券管理的一个重要组成部分。债券管理者可以进行债券互换,或者进行利率互换。

(1)债券互换

替代互换 替代互换是交换性质相同但价格不同的两个债券。假定有两个 10 年期政府债券;两个债券的票面利率都是 8%,但其中一个的价格低于另一个。替代互换就是卖出价格较高的债券,并买入价格较低的债券。如果存在替代互换的机会,那么肯定是一价定律被违背。将一个债券替换为另一个债券,替代互换获利的情况是非常少见的。获利更可能发生在涉及大量债券的复杂组合的情况下。

收益率提升互换 收益率提升互换是将一个到期收益率较低的债券换为具有同样风险和期限,但到期收益率较高的债券。正如我们在前面讲到的组合的收益率并不等于构成组合的债券的到期收益率的加权平均。所以,将一个债券换为收益率较高的债券实际上可能降低组合的收益率。此外,如果价格的决定是以即期利率来折现现金流,到期收益率较高的债券可能定价过高,而到期收益率较低的债券可能是公平定价的。因此,尽管收益率提升互换经常被提及,但其逻辑性并不强。

税收互换 在包括美国在内的很多国家,个人要为实现的资本利得或损失缴税。税收互换就是造成一个资本损失,或是为了抵消资本利得,或是在一

定程度上降低一般收入。假定一个投资者持有一个债券，其卖出价格低于当初的买入价格，但投资者希望持有一个与该债券性质相同的债券。这个投资者可以卖出价格下跌的债券，造成资本损失，并购买另一个具有相同性质的债券。这一买卖行为称为税收互换。

在美国，国税局是不允许个人在买卖同一证券（虚假抛售）的情况下申报资本损失的。但对债券来说，通常很容易找到利息、期限和风险方面都与第一个债券相似的另一个债券。税收互换对市政债券特别有利。假定投资者持有市政债券，而且利率上升了，投资者的债券价值下跌。因为政府债券的利息是免税的（至少在联邦政府层面），卖出价格已下跌的市政债券，并买入一个平价的类似的新市政债券，就造成资本损失，而购入的债券没有纳税义务。

（2）利率互换

固定收益管理中使用的一个重要投资工具是利率互换。利率互换是交换利息而不交换证券。利率互换的最基本类型是固定利率与浮动利率的互换。在这类互换中，一方向另一方支付一个固定的利息，以换取一个可变的利息。例如，A 方可能同意在未来 5 年每半年向 B 方支付 6%的利息，作为交换，A 方得到一个浮动利率，等于每 6 个月期初时的 6 月期国库券利率。双方不仅在利率上达成一致，对于利率使用的本金也要达成一致，称为名义本金。如果名义本金是 1 000 万美元，其现金流见表 4.11。

表 4.11　名义本金为 1 000 万美元的固定对浮动利率互换的现金流

单位：美元

时期（以半年计）	6 月期国库券利率（年利率）	B 向 A 支付	A 向 B 支付
1	6%	300 000	300 000
2	4%	200 000	300 000
3	7%	350 000	300 000
4	6%	300 000	300 000
5	5%	250 000	300 000
6	7%	350 000	300 000

所有主要的经纪公司都安排利率互换。进行互换的双方可能知道对方，也可能不知道对方。互换是直接卖出的一个替代方式。投资者有兴趣将一个长期债券交换成一系列 6 月期国库券，他可以卖出长期债券，再买进一系列 6 月期国库券。尽管交换并不涉及实际出售，但可以达到同样的目的。为什么采用互换呢？

第一，互换相对比较便宜。互换利率可能比卖出长期债券并买入短期债券更便宜。第二，一方或多方可能不愿意出售资产。例如，储蓄贷款机构的资产多是它们社区内的长期抵押贷款。它们主要的债务是短期储蓄。为了防范利率期限结构移动，它们希望资产和负债的久期匹配。储蓄贷款机构可能觉得为维持长期的地方形象，它们需要持有长期抵押贷款。固定对浮动的互换可以使久期匹配，而不需要实际出售资产。运用互换的第三个原因是比较优势。一般认为信用等级低的公司发行固定利率债券时必须支付的风险溢价要高于它们发行可变利率债券时的水平。此外，由于利率互换不涉及本金，只是利息流，一方破产对另一方造成的损失只是没有进行更好的利率互换的机会成本。因此，我们认为高信用等级和低信用等级的公司可以互换受益：一个高信用等级的公司希望以浮动利率借入资金，该公司可以以固定利率借入资金，然后通过与希望以固定利率借入资金的低信用等级公司进行固定对浮动的互换来实现。

第五章

股票基础分析

5.1 基础分析方法概述

证券投资基础分析法是分析影响证券未来收益的基本经济要素的相互关系和发展趋势，据此预测证券的收益和风险，并最终判断证券内在价值的一种分析方法。基础分析方法的产生可以追溯到 20 世纪 30 年代，其标志是 1934 年本杰明·格雷厄姆和大卫·多德的《证券分析》一书的出版。

本章我们先讲述基础分析的第一个层次：宏观经济分析和行业分析；然后我们讲股票基础分析的第二个层次：公司价值分析。股票价值分析模型主要有绝对价值模型和相对估值模型两类。绝对价值模型是确定资产内在价值的模型。这种模型可以提供价值的估计，且可与股票市场价格比较，以确定股价是否存在低估或高估。同债券价值模型一样，股票的绝对价值模型也是基于现金流贴现思想之上的。绝对价值模型都认为股票的内在价值等于预期未来现金流的现值。尽管绝对价值模型都是预期现金流的贴现，但是对现金流的不同理解，就会有不同的现金流定义，也就有着不同的绝对价值模型。因此，绝对价值模型可以分为股利贴现模型(Discounted Dividend Model;DDM)、自由现金流模型(Free Cash Flow to Firm;FCFF 或 Free Cash Flow to Equity;FCFE)。这些内容都在本章进行详细介绍。相对估值模型描述了一种股票相对于另一种股票的价值。相对估值模型的思想基础是一价定律，即类似的股票应该以类似的价格出售。本章主要介绍市盈率模型。

5.2 宏观政治经济和行业分析

5.2.1 政治因素分析

政治不但是经济的集中表现，而且还深刻影响着经济。一国的政局是否稳定对证券市场有着直接影响。一般而言，政局稳定则证券市场稳定运行；相反，政局不稳则常常引起证券市场价格下跌。政治因素对证券价格带来的影响往往具有突破性，它们来得突然，变化迅速，很难预测。政治因素包括的内容十分广泛，诸如政府更迭、国内战争、民族冲突、国内罢工、政治丑闻、重要政府官员的更换等。

5.2.2 经济因素分析

宏观经济因素对证券市场的影响具有根本性、全局性和长期性。所以，要成功地进行证券投资，首先必须认真研究宏观经济状况及其走向。影响证券市场的宏观经济因素主要有利率、通货膨胀率和汇率等。

(1)利率

利率是货币资金的价格，反映了市场上资金的供求状况，因此证券价格对利率波动十分敏感。在宏观经济因素中，利率对证券市场的作用最为直接，影响也最大。当利率升高时，公司借款成本增加，利润率下降，股票价格自然下跌，同时利率上升使债券和股票投资的机会成本增加，吸引部分资金从证券市场特别是股票市场转向银行储蓄，导致证券需求下降，证券价格下跌。特别重要的是，市场基础利率水平决定股票"内在价值"，二者呈反比关系。

影响利率变动的因素有很多，对宏观经济的分析可以为预测利率提供基础，从而为判断证券市场的价格走势提供依据。

(2)通货膨胀率

通货膨胀对股票价格走势的影响比较复杂，既有刺激股票价格上涨的作用，也有抑制股票价格的作用。由于股票代表对企业的所有权，企业中的实物资产会随着通货膨胀而升值；另外，企业还可以通过提高产品的售价来弥补原

材料的价格上升造成的损失，这样企业的利润就不会受到通货膨胀的影响。所以，一般来说，在适度通货膨胀的情况下，股票具有一定的保值功能。适度的通货膨胀还可以造成有效需求增加，从而刺激生产的发展和证券投资的活跃。但是，通货膨胀达到一定限度就会损害经济的发展，严重的通货膨胀会导致货币加速贬值，人们将资金用于囤积商品保值，这时人们对经济发展的前景不会乐观，对政府提高利率以抑制通货膨胀的预期增强，许多证券投资者可能退出证券市场，这样就导致证券市价下跌。同时，企业成本上升，盈利水平下降，企业破产数量增多，经济形势进一步恶化，导致社会恐慌心理加重，从而加深了证券市场不景气的状况。

(3)汇率

由于世界经济一体化趋势逐步增强，包括证券市场在内的各国金融市场的相互影响日益加深，一国汇率的波动也会影响其证券市场价格。一方面，汇率上升，本币贬值，将导致资本流出本国，于是本国证券市场需求减少，价格下跌。另一方面，汇率上升，本币贬值，本国产品的竞争力增强，出口型企业将受益，因而此类公司的证券价格就会上扬；相反，进口型企业将因成本增加而受损，此类公司的证券价格就会下跌。但是，这种影响对国际性程度较低的证券市场来说比较小。

5.2.3 经济周期分析

宏观经济周期一般经历四个阶段，即复苏、繁荣、衰退、萧条。证券市场综合了人们对于经济形势的预期，这种预期较全面地反映了有关经济发展过程中表现出来的有关信息，特别是经济中人的切身感受，这种预期又必然反映到投资者的投资行为中，从而影响证券市场的价格。从证券市场的情况来看，证券价格的变动大体和经济周期一致。一般来说，经济繁荣，证券价格上涨；经济衰退，证券价格下跌。但是，不同行业受经济周期影响的程度会有差异，有些行业(如钢铁、能源、耐用消费品)受经济周期影响比较明显，而有些行业(如公用事业、生活必需品行业等)受经济周期影响较小。

5.2.4 经济政策分析

政府对经济的干预主要是通过货币政策和财政政策来实现的。不同性质、不同类型的政策手段对证券市场价格变动有着不同的影响。

(1)财政政策分析

财政政策是指政府的支出和税收行为,它是需求管理的一部分。财政政策可能是刺激或缓减经济增长的最直接方式。财政政策分为短期、中期、长期财政政策,并各有目标,其中短期目标是促进经济稳定增长,而中长期目标是实现资源的合理分配,并实现收入的公平分配和社会和谐发展。

在财政政策的几种手段中,财政预算、税收和国债是最主要的,这三方面的相关动向在分析时必须密切注意。财政政策对宏观经济的影响有“相机抉择”和“自动均衡”两个方面,而在进行正常分析时主要关注前者。从总体来看,不管是扩大支出、减税、减发国债,宽松的财政政策主要会通过增加社会需求来刺激证券价格上涨。例如,减税会增加居民的可支配收入和企业的投资积极性,供需提高使企业的股票和债券的价格上扬;减发国债首先会通过降低债券市场供给来提高债券价格,并通过货币供给效应和证券联动效应来刺激证券价格。反之,紧缩的财政政策会使证券价格下跌。

(2)货币政策分析

货币政策是指政府为实现一定的宏观经济目标所制定的关于货币供应和货币流通组织管理的基本方针和基本准则,一般由一国的货币当局实施。

中央银行主要通过三大货币政策工具来实现对宏观经济的调控,即存款准备金率、再贴现率和公开市场操作。货币政策对证券市场的影响是通过投资者和上市公司两方面因素来实现的。对于投资者来说,当增加货币供应量时,一方面证券市场的资金增多,另一方面通货膨胀也使人们为了保值而购买证券,从而推动证券价格上扬;反之,当减少货币供应量时,证券市场的资金减少,价格的回落又使人们对购买证券保值的欲望降低,从而使证券市场价格呈回落的趋势。对上市公司来说,宽松的货币政策一方面为企业发展提供了充足的资金,另一方面扩大了社会总需求,刺激了生产发展,提高了上市公司的业绩,证券市场价格上升;反之,紧缩的货币政策使上市公司的运营成本上升,社会总需求不足,上市公司业绩下降,证券市场价格也随之下跌。从具体的政策手段来看,中央银行对再贴现率的调整将直接影响市场基准利率,对证券市场的影响最为显著。

5.2.5 行业分析

就如宏观分析和公司分析一样,行业分析的目的也在于寻找更好的投资机会,或者说是寻找更好的收益-风险组合来满足投资人的需要。正如在糟糕的宏观经济环境下各行业难以取得较好的业绩一样,在一个不景气行业中

的公司的收益状况也往往令人担忧。寻找有潜力的行业与选择具有较高成长性的公司同样重要。而行业分析的重要任务之一就是要挖掘最具投资潜力的行业，进而在此基础上选出最有投资价值的公司。

上面虽然提到了行业分析的重要性，但其有效性却往往受人质疑。要明确行业分析的重要意义，还必须弄清楚以下四个问题。首先是必须明确在特定的时期内不同行业间的收益率是否有明显差距。其次，行业分析的有效性是以不同时期行业业绩的相关性为前提的。再次，要分析各行业内公司的收益是否有一致性。最后，还必须对行业的风险有明确的认识。

(1)行业生命周期分析

一个行业的发展过程可以被划分为起步、增长等几个时期，这也就是行业生命周期。一般可以由以下五个方面来判断行业所处的实际生命周期的阶段。①行业规模变化趋势，行业的市场容量和行业资产规模总会经历一个"小——大——小"的阶段。②产出增长率，该指标在产业成长期高而在成熟期和衰退期较低，经验数据一般以15%为界。③利润水平，该指标是一个行业兴衰过程的综合反映，在整个生命周期中，行业的利润水平会经历一个"低——高——稳定——低——亏损"的过程。④技术进步率、技术熟练程度和开工率，随着行业的兴衰，行业的创新能力有一个从高增长到逐步衰弱的过程，技术成熟程度有一个"低——高——老化"的过程，而开工率的高低与行业发展景气程度正相关。⑤资本进退，行业生命周期中的每个阶段都会有企业的进退发生。在成熟期以前，进入行业的企业数量及资本量要大于退出量；而进入成熟期以后，进入量和退出量有一个均衡的过程；在衰退期，退出量明显超过进入量，整个行业开始萎缩、转产，倒闭多有发生。

(2)行业竞争性分析

竞争决定了一个行业的利润率。竞争规律体现为五种竞争的作用力：新的竞争对手入侵，替代品的威胁，客户的议价能力，以及现存竞争对手之间的竞争。这五种竞争作用力综合起来决定了某行业中的企业获取超额收益率的能力。这五种作用力的作用随行业的不同而不同，随着行业的变化而变化，所以不同行业的内在盈利能力并不一致。

(3)影响行业发展的其他非经济因素

行业生命周期勾画了一个行业发展的基本轨迹，这也是行业发展的内在规律，但是一个行业的发展很大程度上也取决于其所处的环境。通常所指的行业环境，不仅包括经济环境，还有社会环境、技术环境和政策环境。理解这些因素对行业发展的影响将有助于我们对行业进行更全面的分析，并得出更

可靠的结论。行业分析中经常用到的 PETS 分析方法，就是通过对行业以上四种环境的分析，来预测行业的发展趋势的。

技术进步是影响行业发展的最主要因素，它一方面推动现有行业的技术升级，甚至可以使处于衰退期的行业焕发出新的生命力，另一方面，技术更新也决定了新行业的兴起和旧行业的衰亡。技术进步不仅仅使企业生产出新的产品和提供新服务，也使企业的生产流程得到改善。

对行业发展产生影响的社会环境变化主要来自人口结构的变化和社会习惯的改变。各年龄层次人口的比例情况称为人口结构，处于不同年龄层次的人有不同的消费需求、储蓄习惯甚至是业余爱好。社会习惯对国民经济构成中的消费、储蓄、投资、贸易等方面都有较大的影响，这自然也影响到行业的发展和行业结构的演进。

行业所处的政策环境指的是行业所受到的政府干预情况和行业政策影响。就政府干预经济的效果，不同经济学家有不同的看法，但是从经济发展的历史来看，无论是奉行自由主义还是强调政府干预的国家，都对经济有着不同程度的干预。政府对不同行业的干预主要是通过补贴、税收、关税、信贷和价格等手段实现的，除此之外的手段还有市场准入、企业规模限制、环保标准限制，甚至是政府直接干预。

5.3 股利贴现模型

股利贴现模型(DDM)认为未来现金流应该是股票未来所发放的全部股利，因为股利是投资者可以从股票投资中获得的唯一的现金收入。

股利贴现模型是最早出现的，也是最简单的股票价值分析模型，但并不影响其成为一个重要的股票价值分析工具。根据对外来股利发放情况的预测不同，DDM 可以分为不变增长 DDM 和可变增长 DDM，而其中可变增长 DDM 又可以按增长方式的不同分为两阶段模型、H 阶段和三阶段模型。本节首先介绍 DDM 最一般的形式，然后再对 DDM 的各种具体的形式进行详细介绍。

5.3.1 股利贴现模型的一般形式

同其他基于现金流贴现思想的价值模型一样，股利贴现模型认为股票的价值等于股票所有未来股利的现值，用公式表达如下：

$$P_0=\frac{D_1}{1+r}+\frac{D_2}{(1+r)^2}+\cdots+\frac{D_n}{(1+r)^n}+\cdots$$

式中，P_0 为股票第 0 期的价值（当前价值），D_n 为股票第 n 期的股利，r 为股票要求的回报率。

如果计算得出的股票当前价值 P_0 与股票的当前价格不同，那么说明当前股票价格被高估或低估了，这时，投资者可以通过做空或做多来获得超额回报。

5.3.2 不变增长股利贴现模型

(1)零增长模型

显然，上述公式在实际中根本无法应用。因为这要求对以后每一期股利都做出预测，但要精确地预测十几年后的公司盈利和股利发放情况几乎是不可能的。因此，要将 DDM 模型运用在投资实践中，就要对未来股利的增长方式进行一定的假设。假如用 g_t 来代表每期的股利增长率的话，那么 g_t 的数学表达式为：

$$g_t=\frac{D_t-D_{t-1}}{D_{t-1}}$$

由于这是戈登（Gordon）首次提出的，因此也被称为戈登增长模型。显然，在所有的股利增长方式中，最简单的当然就是假设未来公司的股利以一个固定的比例增长。

我们先来看戈登模型的一个特例。那就是 $g_t=0$，即零增长模型，根据等比数列求和公式，零增长模型可以化简为

$$P_0=\frac{D_0}{r}$$

很明显，零增长模型大大简化了 DDM。但对于股利零增长的假设却很不合理。因为一个公司的股利发放通常是同公司的盈利相关的，若股利一直不增长，就说明公司的业绩一直停滞不前，在竞争如此激烈的社会，一公司如果长期不发展就很可能遭到淘汰。

(2)戈登模型

如果在戈登模型中假设股票今后的股利增长率是一个保持不变的常数 g，那么未来任意一期的股利 D_t 就可以表示为 $D_t=D_0(1+g)^t$。

从而戈登模型在数学上就可以表示如下：

$$P_0=\frac{D_0(1+g)}{1+r}+\frac{D_0(1+g)^2}{(1+r)^2}+\cdots=\frac{D_0(1+g)}{r-g}$$

值得注意的是上述后面的形式只有在 $r>g$ 时才成立，因为上式其实是一个无穷等比数列的求和过程。如果 $r=g$ 或 $r<g$，那么就会使 P_0 无穷大，而这两种情况是没有意义的。

由于股利发放额等于公司的净利润同股利发放率的乘积，因此当公司的股利发放率保持不变时，预期的股利增长率就等于公司的预期利润增长率。这样，戈登模型就将股票价值的估计同公司经营状况很好地结合在了一起。

戈登模型是最常用的简化价值分析模型之一。但在使用戈登模型的过程中，有几个问题是需要注意的。首先，既然戈登模型通过股利发放率将公司盈利和股票价值联系到了一起，那么根据戈登模型的形式，要使用其来进行价值分析的话，就要求公司的净利润也是以一个固定的比例增长的。显然，并不是所有的公司的盈利都是以这种方式增长的。对于那些新兴行业的公司来说，它们的盈利增长率开始可能会很高，然后慢慢下降，最后才保持一个比较稳定的增长率。因此，戈登模型主要适用于那些成熟公司或者公用事业公司。我们一般认为一个公司的盈利增长率大大高于名义 GDP 增长率的公司，戈登模型是不适用的，而应该使用后面将要介绍的多阶段模型。在使用戈登模型时还有一个要注意的问题，那就是模型计算的结果对要求的回报率和股利增长率都非常敏感。

此外，戈登模型与股票的市盈率之间也有着紧密的联系。市盈率指的是股票市场价格同公司的每股净利润的比值，这是一个非常重要的相对指标，在股票价值分析中也有着非常重要的作用，在本章后面将要讲述的相对估值模型中，我们将对市盈率模型进行详细的介绍。在这里，我们就可以通过戈登模型计算市盈率：

$$\frac{P_0}{E_0}=\frac{D_0(1+g)}{E_0(r-g)}=\frac{b(1+g)}{r-g}$$

式中，E_0 为当期每股净利润，$b=\frac{D_0}{E_0}$ 为股利发放率，$1-b$ 为收益留存率。

市盈率也可以用来判断股票的市场价格是否合理，因此只要估计出合理的市盈率，就能判断股票价格是否存在高估和低估。上述公式正好给出了一个估计合理市盈率的方法。

5.3.3 两阶段模型

戈登模型虽然是最常用的股利贴现模型，但是前面已经提到过，戈登模型只适用于增长比较稳定的成熟型企业，但对于那些还处于上升阶段的公司，假设股利只以一个不变的速度增长是不合理的。

对于大部分企业来说，它们都要经历增长期、过渡期和成熟期三个阶段。在增长期，公司主要的特点是每股盈利超常增长，但由于迅速的发展需要大量的资金，因此在这个阶段公司的股利发放率很低；在过渡期，公司盈利增长有所减缓，但仍然高于平均水平，而且由于投资机会的减少，公司的股利发放率也有所提高；公司最后进入成熟阶段之后，盈利的增长率和股利发放率都相对稳定在一个长期水平上，戈登模型就是主要适用于处在这一阶段的公司。

多阶段股利贴现模型假定了股利增长模式会随着时间的推移而发生改变，从而在分析股票价值时更符合那些尚未进入成熟期公司的现实。根据对股利增长模式的不同假设，多阶段股利贴现模型又可以分为两阶段模型、H模型、三阶段模型等。本部分内容首先对两阶段模型进行详细的介绍。

两阶段模型假设公司的股利增长经历了两个阶段，在第一阶段，股利每年以一个较高（或较低）的速度增长，一段时间后，股利增长稳定在了一个较低（较高）的水平上。这里假设第一阶段股利增长率高于第二阶段，那么这个过程就可以用图5.1来形象地表示。

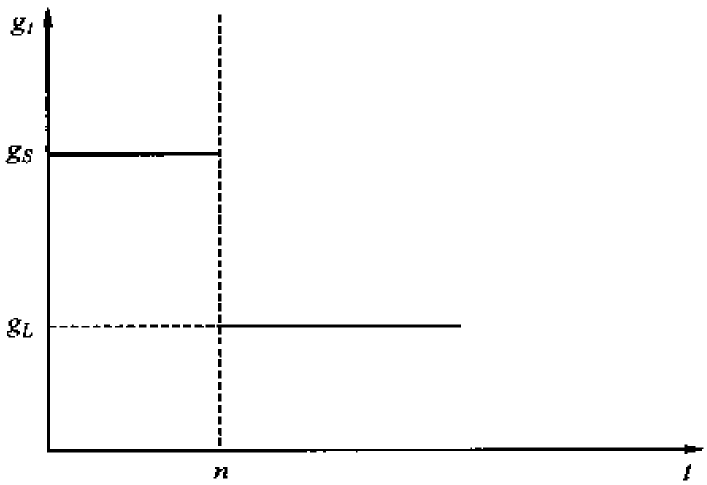


图 5.1　两阶段股利增长模型

两阶段模型的数学表达式为

$$P_0=\sum_{t=1}^{n}\frac{D_t}{(1+r)^t}+\frac{P_n}{(1+r)^n}$$

在上述公式中，n 为第一阶段的年数，P_n 被称为终值，它是股票在第一阶段结束时的价值。由于第二阶段本身可以看成是一个从第 n 年开始的不变增长的阶段，那么 P_n 就可以用戈登模型来计算，即

$$P_n=\frac{D_{n+1}}{r-g}=\frac{D_0\ (1+g_S)^n(1+g_L)}{r-g}$$

此时，两阶段模型的具体形式为

$$P_0=\sum_{t=1}^{n}\frac{D_0\ (1+g_S)^t}{r-g}+\frac{D_0\ (1+g_S)^n(1+g_L)}{(1+r)^n(r-g)}$$

式中，g_S 为快速增长阶段的股利增长率，g_L 为低速增长阶段的股利增长率。

对于戈登模型来说，由于其假设股利增长模式经历了一个阶段，那么当股票现在没有发放股利时，戈登模型就无法用来估计股票价值。但是，两阶段模型却可以很好地解决这个问题。我们可以将不发放股利的阶段看成是两阶段模型中的第一阶段，开始发放股利后就进入第二阶段。

5.3.4 H 模型

通过观察图 5.1，可以看到从第 n 年开始，股利突然从 g_S 下降到了 g_L，但在现实中，这种情况是很罕见的，因为股利如此大幅度变动很可能对公司的股价造成剧烈的影响，这是公司董事会不愿意见到的情况。因此股利更多的可能是连续慢慢下降的。

因此，就出现了模型假设第一阶段的股利增长率是变化的。这就是 Fuller 和 Hsia 1984 年提出的 H 模型。H 模型也假设股利增长经历了两个阶段，但同两阶段模型不同，股利增长率在第一阶段是呈线性下降的，不是不变的。如图 5.2 所示。这里假设起始股利增长率大于最终的增长率，H 模型的计算公式如下：

$$P_0=\sum_{t=1}^{n}\frac{D_0\prod_{i=1}^{t}g_i}{(1+r)^t}+\frac{D_0\prod_{t=1}^{n}g_t(1+g_L)}{(1+r)^n(r-g_L)}$$

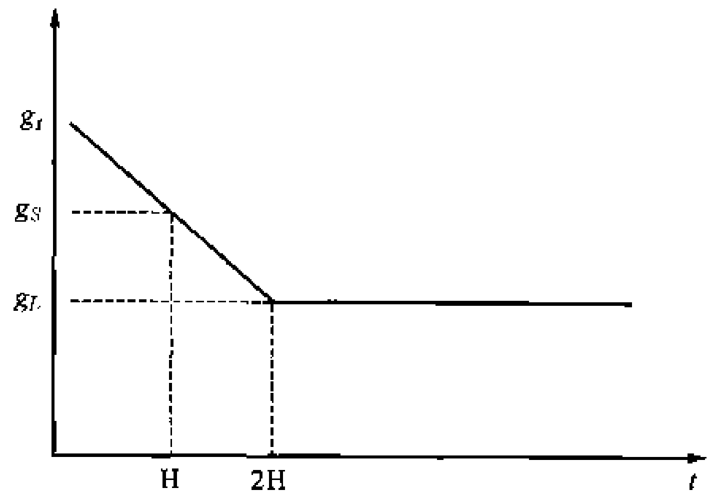


图 5.2 H 模型

有时 H 模型也许能更好地贴近现实，但是这个模型计算起来却非常复杂，因为在股利增长率下降到 g_L 以前，每年的 g 都是变化的，如果第一阶段持续的时间较长，那计算就会变得非常复杂，因此有必要对其进行一定的简化计算。H 模型的近似计算公式为

$$P_0=\frac{D_0(1+g_L)}{r-g_L}+\frac{D_0H(g_S-g_L)}{r-g_L}\text{或}P_0=D_0\frac{(1+g_L)+H(g_S-g_L)}{(r-g_L)}$$

式中，g_S 为初始的股利增长率，g_L 为最终的股利增长率，H 为第一阶段期限的一半。

【例 5.1】某公司去年每股股利为 1 元，预计今年股利增长率为 15%，在今后 16 年里，这个增长率按线性方式递减，16 年后，股利增长率稳定在 8%。市场无风险利率为 5%，市场风险溢价为 6%，该公司股票对于市场的贝塔值为1.25，该公司股票现在的价值为多少？

首先，由 CAPM 计算该股票的要求回报率为 5%+1.25×6%=12.5%。

将上述数据代入到上述近似公式可以算出

$$P_0=1\times\frac{(1+15\%)+8\times(15\%-8\%)}{12.5\%-5\%}=22.6(\text{元})$$

公式虽然是一个近似公式，但大部分时候它的计算和用精确公式计算出来的结果是很接近的。但是当第一阶段时间特别长或者 g_S 和 g_L 相差很大时，这个公式就会变得误差很大，此时就要用精确的计算公式来计算，也就是计算出第一阶段每一年的股利增长率，并进而计算出这期间每一年的股利，再进行贴现。这个过程更多地要借助计算机程序来完成。

5.3.5 三阶段模型

为了进一步贴近现实，股票分析师们可能会将股利的增长模式设计得更复杂，这就产生了三阶段模型。

三阶段模型有两种主要形式。第一种形式同两阶段模型类似，股利增长经历了三个独立的不变增长阶段，见图 5.3 中左图；第二种形式同 H 模型比较相似，只是在三阶段模型的第一阶段还有一个股利增长率不变的时期，见图 5.3 中右图。

对于第一种形式，由于股利增长率在每个阶段是不变的，因此价值计算的方法可以同两阶段增长模型一样，只要对每阶段分别进行计算就可以了。虽然如此，但由于多了一个阶段，计算公式仍然很复杂：

$$P_0 = \sum_{t=1}^{T_1} \frac{D_0 (1+g_S)^t}{(1+r)^t (r-g_L)} + \sum_{t=1}^{T_2-T_1} \frac{D_0 (1+g_S)^{T_1} (1+g_M)^t}{(1-r)^{T+T_1}} + \frac{D_0 (1+g_S)^{T_1} (1+g_M)^{T_2-T_1} (1+g_L)}{(1+r)^{T_1+T_2} (r-g_L)}$$

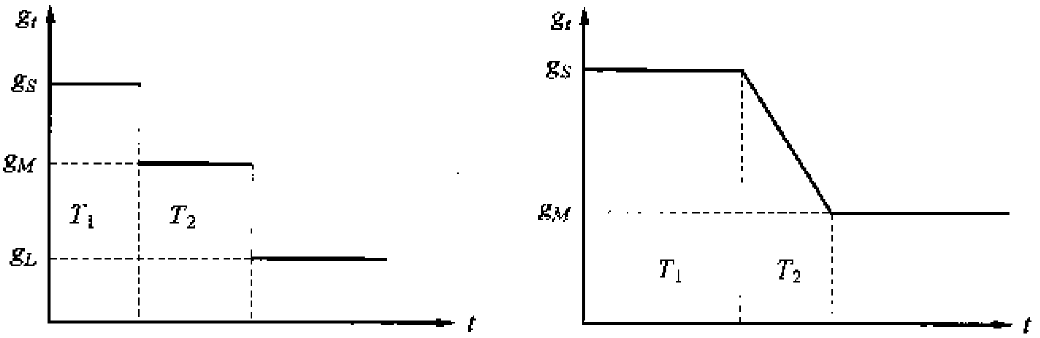


图 5.3 两种形式的三阶段增长模型

5.3.6 多元增长模型

虽然股利贴现模型拓展三阶段模型已经变得非常复杂，但是有时候分析师们为了分析得更加精确，可能会对公司未来的股利增长情况进行更复杂的假设，如假设公司成长经历了初期、平稳期、转折期、稳定期和衰退期等很多阶段，并对每个阶段的股利增长率分别进行预测，对于这种模型，我们只能通过表格工具或者建立计算机程序来分析了。

5.4 自由现金流模型

在本章的导言中已经介绍过，在使用现金流贴现模型分析该股票价值的时候，对于“现金流”的不同看法会产生不同模型。股利贴现模型将股票的股利看作全部的“现金流”。可以说，这是一种就股票论股票的分析方法，当投资者只是希望通过投资股票获利时，用DDM进行分析当然不错，但是如果投资者购买股票的目的是为了控制公司或者实现股东价值最大化呢？此时投资者更关注的恐怕是整个公司的价值而非股票的价值。此外，当公司不发放股利时，股利贴现模型也就无法用来分析股票价值了。为了解决这些问题，我们就要求寻找新的“现金流”来对公司的价值进行分析。在本节中我们先介绍自由现金流模型。

5.4.1 自由现金流的计算和自由现金流模型

(1)自由现金流的计算

自由现金流，顾名思义，当然是指公司能自由分配的现金。自由现金流可以分为公司层面和股东权益层面两个层次，两者分别被称为公司自由现金流(FCFF)和股权自由现金流(FCFE)。FCFF是指在不影响公司资本投资时可以自由向债权人和股东(即所有的资金供给者)提供的资金；FCFE则是指在不影响公司资本投资时可以自由分配给股东的资金，它等于FCFF减去所有对债权人的支付，FCFE可以用来衡量公司支付股利的能力。

自由现金流并不是在公司财务报表中直接披露的，我们只能通过财务报表的各种数据计算出自由现金流。在财务学中，衡量公司盈利的指标主要有净利润(net income; NI)、经营活动的净现金流(cash flow from operations; CFO)、息税前收益(earnings before interest and tax; EBIT)，以及息税折旧摊销前收益(earnings before interest, tax, depreciation and amortization; EBITDA)。由于自由现金流也可以看成是一种盈利指标，因此，我们主要用上面这些盈利指标来计算自由现金流。计算现金流的方法很多，主要有用净利润、现金流量表、息税前收益和息税折旧摊销前收益等来计算自由现金流，由于计算自由现金流属于财务分析的内容，超出了本书的范围，为了帮助读者理解，我们在这里简单介绍了如何用净利润计算自由现金流，对计算现金流感

兴趣的读者可以参阅财务分析方面的书籍来学习其他计算方法。

净利润是公司在权责发生制下的净收入，而自由现金流则考察的是现金的流量，同时净利润中已经扣除了利息支出，而公司自由现金流则还要考虑对债权人的自由现金流，此外，净利润并不反映公司的投资情况，而自由现金流则要扣除公司固定资产和资本的投资。综合来说，要用净利润计算公司自由现金流，要对净利润做出如下调整：

净利润＋净非现金支出＋利息支出×(1－税率)－本期固定资产投资－本期营运资本投资＝公司自由现金流

用公式表达即为

$$FCFF=NI+NCC+Int(1-t)-FI-WI$$

式中，NI 为净利润，t 为所得税税率，NCC 为净非现金支出，FI 为固定资产投资，Int 为利息支出，WI 为营运资本投资。

下面对上述公式中的每个项目进行说明。净利润是公司扣除折旧、摊销、利息支出、所得税及优先股股利等支出之后的收入。但在这些扣除的项目中，有一些并不是以现金支出的。因此为了保证最后得到的是现金流量，就要在净利润中加上非现金支出的增加，减去非现金支出的减少，两者合并就成了净非现金支出。最常见的非现金支出是折旧，这是因为固定资产的折旧仅仅是减少固定资产的金额，并不发生公司现金的变动。同样，长期成本和其他长期资产的摊销也属于非现金支出。

由于公司自由现金流是可以自由向所有资金供给者提供的现金流，因此要将净利润所扣除的利息支出加回去。但是，这个利息支出并不是公司利润分配表上所反映的利息支出的额，而是税后的利息额。学过公司财务的读者应该知道，债务的利息是有税盾效应的，这是因为利息支出是税前列支的，而利息支出可以减少税前利润从而减少所得税。综合利息支出和税收支出的减少，加回的就是税后利息支出。同样，优先股股利支出也是可以向资金供给者提供的资金，但在计算净利润时这一项也是被扣除的，因此在计算公司自由现金流时也要把这一项加上。与利息支出不同的是优先股股利是税后列支的，因此调整时不用考虑税收的因素。

公司为了维持经营，必须进行各种固定资产的更新和购买，因此进行这些投资支出的现金并不是可以自由分配的，应该把它从净利润中扣除。固定资产投资主要包括机器、设备和厂房投资、无形资产购买以及对其他公司的现金收购。但是，有时公司在进行固定资产投资的同时，也会对一些已有

的固定资产进行处理，此时处理固定资产所得到的现金要和固定资产的投资支出进行抵扣，也就是说，在净利润中扣除的固定资产投资是一个净值。还有一点要说明的是，有时候公司可能并不会用现金来进行投资，而是用债务或股票来交换所要购买的资产，这样虽然不会影响当前的自由现金流量，但债务在将来是要偿还的，这种情况在预测未来的 FCFF 时是一定要考虑的。

最后一个调整项目是营运资本的净投资。通常来说，营运资本是指流动资产和流动负债的差额。但在计算 FCFF 时，营运资本中不包括现金和短期债务。不考虑现金和现金等价物是因为它们正是我们要得到的项目，而去除短期债务是因为这些一年内到期的长期债务实际上属于融资活动，而不是经营活动。

由于 FCFF 是公司层面的自由现金流，它所对应的是公司的价值，而我们在做证券投资分析时更注重的是股票的价值，而在自由现金流模型中，股票价值是由 FCFE 决定的。因此，在得到 FCFF 以后我们还要计算 FCFE。根据本节开始的时候所给出的定义，FCFF 是可以自由向债权人和股东分配的资金，而 FCFE 则是可以自由向股东分配的资金，可见两者的差别就在于必须向债权人支付的资金，这主要是利息支出；此外，向债权人借得的资金由于不需要立即偿还，因此这部分资金是可以分配给股东的，但同时公司也会用现金偿还以前的借款，因此借款和还款相互抵消之后的净额（借款净额，NB）就是新增的 FCFE。因此，只要在 FCFF 的基础上减掉税后利息支出再加上借款净额，就可以得到 FCFE。

$$FCFE = FCFF - Int(1-t) + NB$$

(2)自由现金流价值模型

在计算出 FCFF 和 FCFE 之后，我们就可以用自由现金流模型来分析公司和股票的价值了。但在分析公司价值的时候，所用的贴现率并不是股票的要求回报率，而是公司的加权平均资本成本（weighed average capital cost; WACC）。公司价值等于未来所有的 FCFF 的贴现值，即

$$V_f = \sum_{t=1}^{\infty} \frac{FCFF_t}{(1+WACC_t)^t} \tag{5.1}$$

式中，V_f为公司价值。

同时，由于衡量股权价值时不需要考虑债务的成本，因此自由现金流股权公式的贴现率同股利贴现模型一样是股票的要求回报率。

$$V_e = \sum_{t=1}^{\infty} \frac{FCFE_t}{(1+r_E)^t} \tag{5.2}$$

式中，V_e为公司的股权的总价值。

与前文有所不同，在自由现金流模型和下一节的剩余收益模型中，我们都将用 r_E 来表示股票的要求回报率，以此明确区分股权成本和债务成本。公式(5.1)中的 WACC 是计算公司价值时一个关键要素，在公司价值模型中有着广泛的应用，WACC 等于公司所有股权资本成本和债务资本成本的加权平均值。

$$WACC = \frac{MV_d}{MV_d + MV_e} r_d (1-t) + \frac{MV_e}{MV_d + MV_e} r_E \tag{5.3}$$

式中，MV_d为当前所有债务资本的市场价值，MV_e 为当前所有股权资本的市场价值，r_d 为债务资本的加权平均成本，t 为所得税税率，r_E 为股权成本。

值得注意的是，虽然在用 FCFF 分析公司内在价值时要用 WACC 来作为贴现率，但在用 FCFE 分析股票价值时，我们仍然使用股票的要求回报率 r_E 来作为贴现率。

同第一节介绍的股利贴现模型一样，按照对未来自由现金流的不同预期，自由现金流价值模型也可以分为不变增长模型、两阶段模型和三阶段模型。

(1)不变增长自由现金流模型

不变增长模型假设公司未来的现金流以不变的速度增长，因此它的数学表达式同戈登模型几乎完全一样，只是公式中的一些变量发生了变化。不变增长公司价值模型的公式如下：

$$V_f = \frac{FCFF_0(1+g)}{WACC-g}$$

式中，g 为预期未来公司自由现金流增长率。

同样，公司股权的总价值可用如下公式计算：

$$V_e = \frac{FCFE_0(1+g)}{r_E-g}$$

(2)多阶段自由现金流模型

由于自由现金流模型的原理同股利贴现模型一样，既然股利贴现模型有两阶段、H 模型和三阶段模型，那么自由现金流模型自然也不例外。我们只要将股利贴现模型中相应的公式稍作修改，就可以得到自由现金流的两阶段公司价值模型和股权价值模型：

$$V_f = \sum_{t=1}^{n} \frac{FCFF_0 (1+g_S)^t}{(1+WACC)^t} + \frac{FCFF_0 (1+g_S)^n (1+g_L)}{(1+WACC)^n (WACC - g_L)}$$

$$V_e = \sum_{t=1}^{n} \frac{FCFE_0 (1+g_S)^t}{(1+r_E)^t} + \frac{FCFE_0 (1+g_S)^n (1+g_L)}{(1+r_E)^n (r_E - g_L)}$$

这些模型的应用方法同 DDM 模型完全一样，因此这里就不再举例说明了。

5.4.2 股利贴现模型、自由现金流模型二者的比较

股利贴现模型、自由现金流模型是两种最常见的现金流贴现模型，但二者由于对现金流的定义不同，因此最终得出的结果也往往不同。此时分析师们就面临着这样的选择：对于某种股票来说，到底用哪个模型才能更准确地分析它的价值呢？下面我们对这两个模型做一个简单的比较，并列出它们各自适用的情况。

股利贴现模型是最简单、最基本的价值模型，但这并不意味着它不如其他两个模型。事实上，股利的发放一般都比较稳定，不会受到短期因素的影响。而自由现金流是从公司盈利演化而来的，公司盈利很容易受到短期因素的影响而产生剧烈波动，根据非正常状态下的盈利计算和预测出来的自由现金流很可能会使分析结果变得不可信。因此分析师通常认为用 DDM 计算出来的价值更能反映股票的长期内在价值。但是股利贴现模型也有着比较明显的缺陷，比如在公司不发放股利，或者投资者投资股票的目的是控制公司时，DDM 显然不能作为合适的工具进行价值分析。当公司同时满足以下条件时，用 DDM 来分析股票价值是比较适合的：

(1)公司是发放股利的；

(2)公司的股利发放同公司盈利情况有稳定的关系；

(3)投资者投资股票的目的仅仅是为了获利。

自由现金流模型既可以分析公司价值，也可以分析股票价值，因此这是近来一个比较受欢迎的模型。特别是当公司没有发放股利或者发放的股利与公司盈利情况不协调时，我们一般就会用自由现金流模型来代替 DDM；此外，当投资者希望控制公司时，他们也更多地用自由现金流模型来分析公司的价值。但是自由现金流模型同样存在着一些问题，比如公司自由现金流长期为负值时，自由现金流模型就无法用来分析公司和股票价值了。一般来讲自由现金流模型适合在下列情况下使用：

(1)公司不发放股利；

(2)公司发放的股利同自由现金流偏差太大；

(3)预期未来公司的 FCFE 同盈利情况保持一致；

(4)投资者购买股票的目的是希望控制公司。

总而言之，两个模型各有长短，我们在具体分析中，只有根据它们各自的特点来选择最合适的模型，才能准确地分析出股票的内在价值。

5.5 相对估值模型

5.5.1 市盈率的计算

在所有的价格倍数中，市盈率(price/earning ratio；P/E)是最常见的股票价格指标之一，而市盈率模型也是最常用的相对估值模型。

从市盈率的英文缩写 P/E 可以看出，市盈率是公司股票价格和每股盈利的比值，其中 P/E 的分子 P 是股票的价格，显然，这是很容易确定的；然而分母中的每股盈利却因对市盈率的不同定义有不同的取值。一般来讲，市盈率可以分为当前市盈率和远期市盈率。当前市盈率是指当前股价同最近四个季度每股盈利总和的比值；远期市盈率是指当前股价同下一年度每股盈利的比值。要进行比较，就必须设立比较的标准，因此我们在计算 P/E 前首先要把每股盈利“标准化”。在公司财务报表公布的利润中，往往会包含一些短期的或者非重复性的项目，这些会使公司的盈利偏离正常的情况；采用的会计处理方法的不同也会弱化比较的基础；商业周期对公司的盈利也会产生短期的影响从而使公司盈利偏离正常状况。此外，当公司存在未到期的期权或可转换权的时候，普通股总数就变得不再确定了。针对这些情况，我们在计算每股盈利时要对净利润进行相应的调整，从而使不同时期或不同公司间的市盈率有充分的可比性。

5.5.2 市盈率模型的应用

用市盈率来分析股票价值的过程一般是：首先确定一个合理市盈率，即股价不存在高估和低估时的市盈率，然后再用根据当前股价计算的市盈率同这

个合理市盈率比较，如果高于合理市盈率则说明股价被高估了，反之则说明股价被低估了。这个过程中最关键的就是合理市盈率的确定。一般来讲，确定合理市盈率方法有两种：一种是根据绝对价值模型计算股票的内在价值，将用这个内在价值计算的市盈率作为合理市盈率。第二种方法则是选择一种与被考察公司的市盈率可类比的市盈率作为合理市盈率。实际上，所有相对估值模型的分析过程都与此相同，只是所选择的分析指标不同而已。本部分将对上述的两种方法进行说明。

第一种方法可以说是结合绝对价值模型和相对估值模型。前面所介绍的股利贴现模型、自由现金流模型和剩余收益模型都可以用来计算市盈率中的合理股价。下面我们用戈登模型来具体说明这种方法的具体使用。我们知道：

$$\frac{P_0}{E_0}=\frac{D_0(1+g)}{E_0(r-g)}=\frac{b(1+g)}{r-g}$$

我们也可以用同样的方法对远期市盈率进行类似的分解：

$$\frac{P_0}{E_0}=\frac{D_1}{E_1(r-g)}=\frac{b}{r-g}$$

根据上述公式，我们可以看到合理市盈率同稳定的股利发放率成正比，而同股票的要求回报率成反比。我们不需要预测未来的股利金额，而只要预测股利增长率、股利发放率和要求回报率等变量就可以得到当前和远期的合理市盈率。

当然，戈登模型是所有绝对价值模型中最简单的一个，如果对未来现金流预测的结果不允许我们使用戈登模型，那么我们就无法得到上述公式的简单关系式了，这时我们只能老老实实地先根据预测计算出股票的内在价值，然后再计算市盈率了。

另一种方法是类比法。类比法的关键是寻找可以“类比”的基准。对于这个基准，一般我们可以有以下五种选择：

(1)与被考察股票各方面都相似的个股的市盈率。

(2)被考察公司所处行业内所有公司市盈率的平均值或中值。

(3)同一子行业的市盈率的平均值或中值。

(4)有代表性的股票指数的市盈率。

(5)被考察股票历史平均的市盈率。

由于选择个股的市盈率来进行分析往往会出现偏差，而一组股票的平均

市盈率可以使这些偏差相互抵消，因此，我们通常选择后面四种基准市盈率来作价值分析的基础。

5.5.3 市盈率模型的优缺点

与股利贴现模型等绝对价值模型相比，市盈率模型的历史更为悠久。在运用当中，市盈率模型具有以下几方面的优点：

(1)由于市盈率是股票价格与每股收益的比率，即单位收益的价格，所以，市盈率模型可以直接应用于不同收益水平的股票的价格之间的比较；

(2)对于那些不能使用绝对价值模型的股票，比如在某段时期内不发放股利或自由现金流为负值的股票，市盈率模型仍然适用。

(3)虽然市盈率模型同样需要对有关变量进行预测，但是所涉及的变量预测比绝对价值模型要简单。

市盈率模型也有缺点：

(1)市盈率模型的理论基础较为薄弱，而绝对价值模型的逻辑性较为严密；

(2)在进行股票之间的比较时，市盈率模型只能决定股票市盈率的相对大小，却不能决定股票绝对的市盈率水平；

(3)公司的每股盈利同样可能为负值，此时市盈率就没有意义了；

(4)在计算每股盈利时要对净利润进行一系列调整，有时候对于非重复性项目的调整是很复杂的；

(5)在用类比法进行价值分析时，都隐含假定了所选择的行业平均市盈率、大盘市盈率都是合理的，但这是不一定的，整个行业或者市场的股价被高估或低估的情况经常发生。

(6)如果公司报表上的盈利并不真实，也会使市盈率模型分析的结果不正确。

第六章 算法交易简介

1971年著名的金融经济学家Fischer Black在他的一篇论文"Toward a Fully Automated Stock Exchange"中曾经预言过一个由计算机进行自动化交易的交易市场。在自动化的市场当中，计算机程序将会取代经纪商和交易员的工作，实现证券交易系统的计算机化，这些会极大地提高金融市场的处理速度和交易效率。

在早期的证券交易所里面，人们大多采用人工喊价交易的方式，即所谓的有形市场。因此，传统的证券市场都设有交易大厅。会员在交易大厅里面派驻交易代表，就是俗称的"红马甲"。投资者的买卖订单由"红马甲"代理完成。在过去的十几年里，随着计算机技术和通信技术的发展，交易所在会员营业柜台和交易所的计算机系统之间开始通过电子通信，并进一步实现了电子化的交易。金融行业信息化的发展速度超过了任何人的预测。现在，自动化和新技术已经极大地改变了金融业交易活动。交易员和其他市场参与者们不断地利用新的方式，通过通信和计算机技术来提高证券信息传递和交易活动的效率，并为金融活动带来了更低的交易费用。

在当今的华尔街，经纪商已经不可避免地需要利用电子交易的方式进行订单的交易。在过去，人们只有很少的机会能够使用技术手段处理交易，或者直接同交易所和金融市场上其他的参与者进行实时的交流。但是，通过引入计算机技术，金融市场已经发生了很大的变化。投资者可以直接将买卖订单发送到交易所的交易撮合系统当中。投资者可以坐在电脑前，通过计算机输入一些参数和指令进行交易和投资活动。传统的交易大厅将逐渐退出市场，"红马甲"也将随之消失。电子化的交易方式减少了交易流程中报价、清算、结算等等方面的人工因素的干预。这不但减少了出错的几

率，而且减少了交易成本，缩短了交易时间，使得市场的运作效率大大地提高了。

在交易和信息访问速度以及交易成本方面的竞争迫使许多公司大量投资于自动化交易的软硬件设备。在华尔街，电子和算法交易的扩张速度非常的迅速，交易过程中自动化的趋势在不断地加速。投资和交易流程中越来越多的工作可以通过计算机系统来完成，人们可以更好地实现自己的投资策略，提高投资效率，而金融市场也逐渐变得越来越现代化。

6.1 什么是算法交易?

算法交易(algorithmic trading)是指在金融市场中，投资者通过计算机程序来下达交易订单，并由计算机算法来确定交易订单的交易时机、价格、下单的数量等等的交易方式。在过去几年当中，算法交易主要是由机构投资者所使用，其中包括对冲基金、养老基金、共同基金以及经纪商等等。随着算法交易不断地发展和流行，中小投资者也能够通过经纪商的服务等方式利用交易算法进行投资操作。

具体说来，金融界对算法交易的内容和范围似乎还没有明确的界定，其中包括了很多不同的自动化交易方式和功能。订单智能路由、程序交易、订单分割、基于规则的自动化交易等等都会在不同场合被称为算法交易。

通常意义的算法交易定位于传统投资过程中的交易操作部分，其中包括直接入场技术、订单智能路由以及订单分割等等技术。算法交易通过计算机将大额的交易分解为若干笔小额的订单，然后自动地为订单选择最优的交易市场和最优价格，以便更好地降低市场冲击成本、机会成本、风险，以及发现流动性和实现交易的隐蔽性。截止到 2006 年，伦敦股票交易所有超过 40%的订单是通过算法交易执行的，而在 2007 年达到了 70%。作为金融市场的发达地区，美国市场上算法交易的比例则更高。

现代资产组合理论为人们如何配置资产组合提供了非常有意义的指导。通过均值方差模型的启发，投资者可以通过分散投资的方式很好地管理投资组合的收益和风险。但是对于机构投资者来说，在资产大量的买进和卖出的过程中不可避免地会产生市场冲击，很难在目标价格实现交易，因此提高了交易的成本，降低了投资的回报。所以，机构投资者会专门雇佣交易员负责交易的实施，尽可能地减少买进和卖出带来的交易成本，获得更加有利的交易价

格。对于减少交易成本，他们很典型的一种方法就是将大额的订单分割，减缓交易速度，利用较小的份额逐步地进行交易。算法交易的出现很大程度上替代了交易员在这一方面的工作。交易员不再需要盯着交易平台，不断地进行手工操作，而可以通过输入包括交易速度、时间、选择算法和参数等等，利用计算机算法进行自动的交易操作。

对于美国的金融市场，除了比较大的纽约证交所和纳斯达克交易所以外，还存在着众多的地区交易所、电子通信网络、另类交易系统等等交易市场，所以不同的市场上对于同一只股票会同时存在着不同的报价，而且不同市场上流动性条件也有所不同。另外，由于一些市场采用做市商制度，所以不同市场上还存在着不同的买卖价差。因此，通过直接入场技术和订单智能路由技术，投资者可以随时利用计算机系统监控不同市场上的交易条件和环境，发现最优的买卖双边报价，同时更好地利用不同市场上的流动性进行交易。

算法交易虽然是计算机在不受人管理的状态下传递交易意图和策略，但是实际的买卖决策还是由人制定的，而不是计算机。人作为进行最终的交易决策者，负责制定并实施交易当中需要的参数，由计算机通过制定好的算法进行计算，生成交易订单，并选择最有效的路径发送订单。典型的方式是，由基金经理做出投资决策，制定好一份交易列表，交给交易员进行操作。然后，交易员利用算法交易平台选择合适的算法，并制定交易参数，例如交易对象、数量、交易速度、时机等等，通过算法自动地执行交易操作。这样一方面可以节省人力，使交易员将精力花费到更能够产生附加价值的工作上面，另一方面也能够避免因为交易员的主观情绪对交易表现的影响。

除此之外，自动交易(automated trading)、机器交易(robot trading)、黑盒交易(black-box trading)等等有时也被称为算法交易。这些交易策略往往注重于投资过程中投资决策的制定，而前面提到的算法交易则是在实现投资策略的交易过程中实现自动化。这些交易策略不存在一些确定的方法和手段，投资者可以利用计算机实现他们任何的投资策略和想法，并进行投资交易，其手段可以包括技术分析、跨市套利、期现套利，甚至包括一些行为金融的思想等等。

有效市场假说认为金融市场是有效率的，股票的价格信息反映了公司的价值。价格随着新的信息进行动态的调整，市场不存在套利的机会。进一步地，一旦市场出现套利的机会，有信息的投资者会进行套利交易，将股票价格带回到合理的水平。自动化的交易策略正是希望利用市场上的一些无效率的

机会进行交易，实现获利。

市场上确实存在着许多这样的投资者，例如对冲基金等，他们一直在尝试发现市场无效率的部分，通过一些特殊的交易策略进行交易。他们有时可以利用计算机系统进行交易，通过电子方式接收的信息流启动交易指令，在投资决策和执行的任何一个阶段，算法交易信号都能够提供良好的技术支持，甚至整个投资决策和执行可以完全依靠算法交易自动运行。

在自动化的投资策略当中，计算机系统扮演着很重要的角色。首先，计算机系统接收和处理信息的能力是不能替代的。很多复杂的交易策略离开计算机的辅助会变得很难实现，例如指数套利这种计算量很大的交易策略，依靠人工计算几乎不可能实现。另外，有时候计算机甚至会在人工下单的交易者来不及对信息做出反应的时候完成交易，例如外汇市场上的三角套利，一两秒间差别都会很大。在这些情况下，计算机算法就会体现出其无与伦比的优势。

此外，我们还要对电子交易（electronic trading）和程序交易（program trading）做出一些说明：

电子交易：就是指通过电子化的方式交易股票、债券、外汇和衍生品等等的交易方式。它利用信息技术将买卖双方联系在一起，通过电子技术作为媒介创造出虚拟的交易市场。纳斯达克（NASDAQ）就是最著名的电子交易市场。在美国，实施电子化交易的交易所被统称为电子通信网络（Electronic Communications Networks，ECN）。电子交易被广泛地认为比传统的交易方式更加可靠，而且有着更好的价格发现机制，提高了市场的效率。

程序交易：纽约证券交易所（NYSE）将程序交易定义为“包含买进和卖出超过 15 只股票，总市值超过 100 万美元的组合交易策略”。其中甚至都没有直接提到使用计算机。其中使用“程序”这个词可能是因为在早期，一些大额的组合交易都需要进行一系列预先安排的步骤，所以叫作程序交易。20 世纪七八十年代，计算机技术逐渐在华尔街变得重要起来，使得投资者可以很快地执行大额的股票和期货组合交易，这也被认为是 1987 年美国股市大跌的原因之一。

本书中，我们将首先介绍相对狭义范围内的算法交易，也就是降低交易成本和发现流动性的订单分割算法。然后，在后面的章节介绍关于自动化的套利交易策略的一些相关的内容。

6.2 背景及历史回顾

6.2.1 交易系统的自动化

在金融市场广泛地应用信息技术之前，证券交易所都采用人工喊价制度。这种机制目前在纽约证券交易所、香港期货交易所等老牌交易所还能见到。这些交易活动在指定的交易大厅(trading floor)进行，代客买卖的交易商面对面地公开叫喊，以特定价格进行特定数量的交易。在这种制度下，交易商以口头和手势表达交易信息，各交易所都有各自的手势规定。这种情况下，成交的选择是随机的，成交机会的取得要依靠交易商积极争取，而通过计算机自动交易是指在整个交易过程中依靠计算机自动交易系统来辅助或执行买卖竞价，一般称之为“电子屏幕交易”(electronic screen trading)或“自动交易”(automated trading)。

金融市场的交易信息化开始于20世纪70年代早期，其标志是纽约证券交易所(New York Stock Exchange，NYSE)引入订单转送和成交回报系统(Designated Order Turnaround，DOT，及后来的Super DOT)，以及开盘自动报告服务系统(Opening Automated Reporting System，OARS)。DOT系统直接把交易所会员单位的盘房与交易席位联系起来，直接通过电子方式将订单发送到交易席位，然后由人工加以执行，而OARS系统可以辅助专家决定开盘结算价。

各证券交易所基于种种因素的考虑，所采用的交易系统各不相同，自动化的程度也有所差异。Domowitz将全球各交易所(含期货交易所)的交易系统按价格发现机制的自动化程度，分为由低到高的七个层级：

第一级，系统成交价由另一相关主要市场决定，没有独立的价格发现功能，称为被动的定价系统。如费城交易所PHLE系统、太平洋交易所MAX-OTC系统的证券价格完全由全国性交易所的市场价格决定。在做市商市场中(如NASDAQ)，有的系统的价格由做市商提供的最佳报价成交，也完全缺乏自动价格发现机能。

第二级，系统价格参考另外一些市场来决定，即价格可从主要市场获取，但有额外价格改进程序。较典型的是美国中西部证券交易所的MAX系统，

参考数个市场的行情，从中得到统一的最佳买卖价来执行委托(consolidated best bid or offer，简称 CBO)。该交易所对市价委托规定：如果以 CBO 价格撮合，比主要市场最近成交价上升或下跌不足两个升降单位时，则以 CBO 价格成交；如果超过两个升降单位，则暂停撮合。对暂停撮合的买进委托，如果下一次主要市场成交价不大于最近成交价时，则以最近成交价撮合；如果下一次主要市场成交价比最近成交价下跌两个升降单位或下降后平盘，则以最近成交价加 1/8 点的价格成交；若大于最近成交价，则以下一次主要市场成交价撮合；若低于 CBO 价格时，则以 CBO 价格撮合。对被停止撮合的卖出委托与上述规定对称。这种系统有能力对价位变动进行评估并建议是否暂停执行委托以等待新价格。

第三级，系统中存在部分议价功能。如巴黎交易所 CAC 系统对巨额委托有议价功能，一般对外公开委托数量但不公开委托价格，以吸引对方议价，一旦价格达成协议，该巨额委托即予以电脑撮合。再例如路透社拥有的 Instinet，该系统既可执行屏幕上的报价，也可在不具名基础上与其他委托方议价。这种公开招标方式会使部分已进入限价委托汇总表的委托转向招标交易，从而降低自动化撮合的程度。

第四级，屏幕选择定价成交。如法兰克福交易所的 IBIS 系统及辛辛那提交易所的 NSTS 系统，具有“hit a bid”或“take an offer”功能。对显示屏上的委托价格，交易者仅按一个键即可参与交易。如果申报价等于当时显示的最佳价格，在经过一次竞价后，无论是否全部或部分未成交，未成交部分马上被取消，如果多个投资者按下该键，则以时间优先原则撮合。这种委托并不会储存于系统中，属于立即成交或取消委托。这种交易不是在汇总的基础上，而是在一对一的基础上完成的。

第五级，自动连续的双边竞价成交。一般交易所都采用这种系统，买卖价连续报出，一旦指令吻合就进行交易。系统一般根据价格、时间、数量、委托类型、交易者分类设计优先顺序。

第六级，集中竞价成交。如台湾的 FASTS 系统和吉隆坡交易所的 SCORE 系统。委托要累积一段时间，然后所有交易在某一时点以同一价格成交，如同分盘交易。亚利桑那交易所采用的就是一个完全自动化的批量交易系统，第五级系统中的开盘价一般也采用单一价格竞价。

第七级，有定价模型的自动连续的双边竞价。当交易系统要计算那些与基础现货资产价格相关的衍生品价格时，会使用此系统，一般在有期货期权交易的交易所中使用，如全球期货交易系统 GLOBEX。

北美市场的电子系统多属第一级，一般从 NYSE 和 NASDAQ 引入价格进行交易，本身没有定价功能；亚洲市场双边竞价系统占绝对主导地位，沪深股市就采用第五级；而欧洲市场在更新系统时也更多地采用第五级模式，新近建立的系统也较多实行双边竞价，可以说自动连续双边竞价的交易系统是交易系统自动化发展的大趋势。

6.2.2 程序交易和 1987 年股市崩盘

程序交易(program trading)起源于 1975 年美国出现的“股票组合转让与交易”，即专业投资经理和经纪商可以直接通过计算机与证券交易所连接，以实现股票组合的一次性买卖交易。纽约证券交易所(NYSE)将程序交易定义为“包含买进和卖出超过 15 只股票，总市值超过 100 万美元的组合交易策略”。

程序交易主要是大型投资机构的交易工具，用于同时买进或卖出整个股票组合。程序交易的对象通常是纽约证交所的股票，以及其对应的芝加哥商品交易所的期权和芝加哥商业交易所的标准普尔 500 指数期货合约。程序交易完全基于股票组合价格和其他目标(如指数期货)的价格差异进行交易，而不涉及任何公司基本面信息和宏观经济信息。在实际操作中，这意味着所有的程序交易都是在计算机的辅助下完成的。进入 80 年代，程序交易已经被广泛应用于股票与期货的跨市场指数套利交易中。程序交易通常包括 3 种核心的交易策略：

1. 久期平均(duration average)

这种策略通常要确定股票或组合的一个合理的价格范围，当股票或组合价格超过这个范围时卖出，而在价格低于这个范围时买入，从而通过低买高卖来实现收益。要衡量股票组合的价格是低还是高，就需要将股票组合的合理价格和持有成本等影响价格敏感性的因素考虑进去，即用久期来衡量。也就是说，当股票组合价格维持在一个特定的价格区域时，这一策略才是有效的。一旦价格下跌低于该价格区域的下限，投资者就会亏损，同时投资者还丧失了当价格上涨超过该价格区域的上限时的获利机会。

久期平均交易策略是在价格下跌时买进，在价格上涨时卖出，因此该策略对价格波动没有负面影响，相反，还起到了减少价格波动幅度等平抑价格波动的作用。

2. 组合保险/动态对冲(portfolio insurance/dynamic hedging)

组合保险策略是指当市场处于下跌市中，对股票组合最小价值的一个保全措施安排，而当价格上涨时，股票组合仍不失去盈利的机会。也就是说，为股票组合确保了一个最低的收益率，同时股票组合又不失去从市场有利变动中获利的机会。

例如，组合保险者可以买入标准普尔 500 指数的看跌期权，即以预先确定的价格在某一个价格水平上卖出该指数的权利。如果该指数下跌低于先前设定的价格水平，投资者就可以行权，将损失锁定在预期范围内。如果指数上涨了，投资者所有的损失只是看跌期权的价格。组合保险最初是由养老金基金开发和应用的。组合保险的操作一定程度上增加了股价的波动性，在价格下跌时增加了卖出压力，在价格上涨时也增加了买入压力。但从整体看，由于这类交易的价格不受有关股价的基本面信息的影响，因此对股价波动性的影响也不是实质性的。

3. 指数套利(index arbitrage)

指数套利交易策略一般发生在股票指数的现货市场和与其相对应的股票指数期货市场。当股票指数现货与股票指数期货的价差大到足以超过无风险利率，并能够抵补所有的交易费用时，就可以进行指数套利。

例如，标准普尔 500 股票指数期货合约的价格与其所对应的标准普尔 500 股票指数现货价格存在一个差额，这种差额表现为一个价格区域，如在 0 到 10 美元之间。该价格区域会随着期货合约到期日的临近而日趋收窄，即 Spread 日趋缩小，最后回归到合理价格水平上，即期货价格最后收敛于现货价格。

指数套利交易策略就是，当实际期货价格超过合理价格时，则卖出期货，买入现货；当实际期货价格低于合理价格时，则卖出现货，买入期货。而程序交易在这一过程中所扮演的角色，就是运用精确的数量模型来构建并计算股票组合，并借助于计算机程序使交易得以快速实现。一般来讲，构建股票组合的买进和卖出过程中要发生交易成本，同时还要尽量接近股票现货指数。但有时与股票现货指数也并非完全一致，可能会采用一些优化组合的措施。这样，就产生了额外的交易成本。交易成本的存在就导致了一个无套利区域的产生，即只要期货价格落在这个区域里，就不可能有套利盈利。也就是说，当期货价格上涨高出这个区域时，套利者卖出期货买入现货才会获利；当期货价格下跌到这个区域以下时，套利者卖出现货买入期货才会获利。

在 20 世纪 80 年代，每次当股价快速移动时，尤其是快速下跌时，程序交易就成为罪魁祸首。1987 年的股市大跌在一开始被认为纯粹是由程序交易造成的，即使是 SEC 的专家在开始也这么认为。今天，大部分金融经济学家认为当时的市场存在过度反应，而且不止一个原因造成了 1987 年的大跌。

1987 年 10 月 19 日，星期一，道琼斯工业平均指数下跌了 508.32 点，收于 1 738.40 点。当天，标准普尔 500 指数下跌了 20%，成为史上最大跌幅，并超过了 1929 年 10 月 29 日 12%的跌幅，而后者被认为是大萧条的信号。程序交易在大跌之后马上受到了批评，在股市下跌时，程序交易员卖掉股票，来从下跌的指数期货中套利，会进一步加大卖出压力。但是，不应该因为 1987 年的大跌来责怪程序交易。那个黑色星期一股票报价更新非常快，程序交易根本来不及反应，因为它需要不断地对市场信息进行更新。

那次股市突然的大跌并不像是由美国或其他国家的任何基本面的新闻造成的，很多专家将责任归咎于组合保险。组合保险的交易员当天卖掉的股票几乎没有进行对冲的操作。他们在当天开盘时卖掉了之前一周未执行的卖单，加速了股市的下跌。每当市场要反弹的时候，大的卖单总会满足买人的需求，进而抑制反弹。一些没有组合保险的职业交易员也预期到被抑制的卖出需求，并预先卖出存货，而另外一些投资者误读了卖出的信息，认为它传递了市场将要发布的利空消息，而没有认识到组合保险交易只是对之前的市场变动做出相应的反应，并不包含基本面信息。另一种解释可能包含投资者对市场风险的察觉。由于一些不确定的原因，股票市场的风险在那个星期大幅增加，并导致了崩盘。厌恶风险的投资者开始离场，转向债券市场。价格过高的股票一直下跌到有足够的预期收益来补偿增加的风险，这被称为风险效应。标准普尔 500 指数在三天内下跌了 10%，导致了当时的崩盘，同时当月的波动性显著增加了。市场下跌引起的财富损失导致了更大的风险厌恶度。当市场产生利空消息，会导致价格下跌，投资者用更大的卖单对其反应，如此反复，这被称为财富效应。在市场下滑的时候，各只股票之间的相关性也可能增加了，提高了风险和风险厌恶度，进而减少了分散投资的益处。投资者普遍需要流动性供他们卖掉头寸，以减少风险敞口。无论如何，在 10 月的大跌期间，股票交易的买卖价差和市场冲击大幅地上升，直到很多重要股票同时停止交易。

总而言之，有四个原因导致高波动性期间股票价格的下滑和上涨：

• **风险效应**　高波动性导致更高的风险，反应到市场，导致资产价格的下跌。

• **财富效应** 较低的价格减少了财富，进而导致风险厌恶度的增加，导致更大的波动。

• **分散化效应** 市场下跌时，相关性增加，进而增加波动性并减少分散化的机会。

• **流动性效应** 波动性大的市场流动性会减少，导致风险厌恶的交易员以更低的价格卖出。

从 1987 年的崩盘以后，主要的股票和商品交易所都出台规定，限制在市场急剧下跌时的大量或恐慌性卖单，并通过中断器限制市场波动。这些机制被称为衣领规则(the collar rule)，或价格限制。中断器决定是暂时中断还是完全停止交易。证券和期货市场用中断器来提供信念，并协调大跌时市场间交易的中断。在美国，市场的大跌由道琼斯工业平均指数(Dow Jones Industrial Average, DJIA)单日的跌幅衡量。有三个终端限制：10%，20%和 30%，每一刻钟开始时对点位计算一次。纽约证交所的 80A 规则规定：如果道琼斯工业平均指数相比前一个收盘价浮动了 2%，程序交易作为指数套利的一部分，将会买入或卖出标准普尔 500 的股票来稳定股票价格。如果 DJIA 回到前一收盘价的浮动 1%范围以内，衣领限制就会解除。期货交易所设置价格限制来减少一些合约(如股指期货)价格的剧烈波动。价格限制不会阻止期货的交易，但是禁止在市场下跌时以低于限价交易。如果交易过程超过了预先设定的长度，比如 10 分钟，单日内的价格限制就会解除。但是，每天的价格限制还存在于该交易过程。具体的价格限制由交易所对每个股指期货合约进行设定。美国的股指期权、股票期权和股票没有价格限制。

1988 年，中断器开始实施，以防止期货市场在市场下滑时的暴跌。很多批评认为中断器只会增加波动性，而不是减少。中断器在三个层面实施。前两个层面被称为衣领。这个计划是为了在道琼斯指数相比前一日收盘价上涨或下降 50%时，限制计算机程序交易往纽约证交所下单。第二层面的计划是在道琼斯指数下跌超过 96 点或标准普尔 500 指数下跌超过 12 点时对程序交易进行延迟。这个层面限制交易员利用计算机程序处理大订单。第三个中断器是在道琼斯指数单日下跌超过 250 点时，美国所有主要交易所中断交易 1 个小时。1 小时之后，交易将会继续进行，但是，如果道琼斯指数持续下跌超过 150 点，市场将会关闭 2 个小时或更久。中断器主要是为了防止股票市场剧烈变化。它的有用性通常受到质疑，因为，为了防止市场的剧烈变动，价值的变化必须能够被反映。

6.2.3 交易系统的十进制化

纳斯达克市场在2001年开始应用十进制报价。在以前,美国股市上最小的价格增量(tick size)为1/16美元,而十进制化(decimalization)把每次交易的最小价格增量从1/16(0.0625)美元变为0.01美元,也就是1/100美元。这种转变的目的是要降低买卖价差,降低交易成本,并使股票价格更容易被投资者所理解。十进制化改变了市场的微观结构,使买卖价差可以变得更小,遏制了做市商的交易优势,因此也降低了市场上每个价位的流动性,但这个改革却可能促进了算法交易的发展。十进制化在1997年由美国国会在股票定价法案(The Common Cents Stock Pricing Act)中提议,于2000年1月由SEC在34—42360号文件中批准。从1997年开始,美国股市是世界上主要股票市场中唯一一个实行小数定价和报价的。考虑到市场的容量和质量,为了减少对市场的扰动,十进制的引入通过3个阶段执行。

第一阶段 2001年3月12日,14只非纳斯达克100的股票进行十进制化。

第二阶段 2001年3月26日,其他代表174家公司的194只证券进行十进制化。

第三阶段 所有剩余的纳斯达克证券在2001年4月9日转换为以美分递增。

十进制化通过更小的价格增量降低了交易成本,特别是对于中小投资者来说。十进制化导致了做市商利润的显著降低,使很多做市商退出了市场。根据纳斯达克对SEC关于十进制化的报告,通过引入以美分为单位的价格增量,大部分交易活跃的股票报价差下降了1.9美分到6.6美分。

通过引入十进制,更有利于投资机构将大的订单分割为较小的订单流进行交易。买方交易员有两种操作的选择。他们可以通过机构经纪商的交易平台直接发送订单,让做市商为他们下单;也可以自己通过ECN下单。机构投资者对于十进制交易操作反应不一。一些买方[①]交易员相信,成交量加权的

① 在金融交易活动当中,买方和卖方都是从证券交易服务的角度区分的。买方是指购买这些服务的市场参与者,卖方是指提供服务的公司。卖方通常是在交易所注册的经纪商;而常见的买方公司包括:共同基金、信托公司、对冲基金、养老基金和私营的交易平台等等。

执行价格不会增长，做市商的保证金不会变化，不需要将订单分割执行。另外一些交易员则认为十进制化的影响是负面的，他们认为经纪商处理大订单的劳动量会增加，而且为买方返回交易结果的过程会变得更长，所以经纪商的佣金会上升。但是，值得指出的是，十进制化加剧了金融市场关于减少交易成本的竞争，推动了算法交易的发展。

6.2.4 金融数据服务的发展

1981 年，后来成为纽约市市长的 Michael Bloomberg 创建并以他的名字命名的一家金融资讯公司 Bloomberg，中文译名叫做“彭博集团”。当时，全球经济正处在一个关键的转型时期，人们对于资讯的及时性和准确性的需求越来越强烈，服务业的比重也变得越来越大。随着互联网的腾飞，计算机的使用将所有的信息电子化，尔后通过网络以最为简便的方式传输给用户，实现了人类经济生活的一项巨大的变革。现在 Bloomberg 已经成为年收入达 54 亿美元的全球最大的财经信息系统和服务供应商之一，其他的如 Thompson Rueters，道琼斯等等这些金融信息供应商的发展成为金融行业信息化的标志之一。大量的历史和实时数据为金融业交易技术和自动化的发展提供了强大的动力和支持。投资者可以利用大量的交易数据进行复杂的交易分析，识别市场交易模式，制定和改进自身的交易策略。

6.2.5 算法交易发展的推动力

日益增长的算法交易应用背后有着很多推动力，主要包括了改变买方卖方之间关系的发展，来自买方交易员的日益复杂的交易需求等等。

买方和卖方交易员之间不断变化的关系是算法交易发展的一个重要因素。随着算法交易被更广泛地接受，以及其应用的增长，买方会要求增加与他们保持联系的经纪商的数量，以降低交易成本，他们希望经纪商只提供交易执行的服务。市场的一个常见的现象是，如果经纪商不改善他们在订单操作上所使用的算法模型，买方交易员将会对他们失去兴趣。这进一步加剧了经纪商之间的竞争，并重新定义了买卖双方合作关系发展和维持的基础。在过去，经纪商获取利润的关键是保持住客户资源，也就是维持长期忠实的客户关系，而现在，这已经不再被认为是市场上竞争的主要立足点。少数大型经纪商一直领导着算法交易的发展。在 2004 年，还只有少数经纪商提供交易算法，而

现在，许多公司的交易平台上已经有很多可以选择的交易算法。随着越来越多公司进入这个领域，算法交易市场的竞争会变得越来越激烈，更多的策略被开发出来，买方对算法交易的应用会不断增长，高端业务将变得越来越有挑战性。

由于买方不断地采取更有针对性的交易操作方式，下单的方式和方向将会根据买方需求所制定的速度和流动性进行改变。算法交易模型让交易员更好地根据自己的目标执行交易。随着算法变得越来越复杂，越来越多地被应用，买方交易员逐渐地开始更多地研究自动交易策略。为了满足买方对未来算法交易的潜在和实际需求，经纪商需要看到客户对定制化的需求，包括算法交易对市场条件变化做出正确反应的能力。

6.3 算法交易策略简介

在金融市场上，不同的机构投资者根据自身不同的交易需求会开发出不同的算法。正因为如此，市场上的算法可谓是多种多样，五花八门，而且有时候会被冠以一些很奇怪的名字。下面我们根据的交易的目的和方式对算法交易策略做一个简单的介绍。

6.3.1 算法交易策略

算法交易主要的功能包括两个方面，即订单智能路由和订单分割策略。另外，作为投资自动化辅助工具的订单管理系统还起到对投资的整合作用，其中包括交易前后数据的分析和整合、多种交易资产的支持等等。

1. 订单分割策略

当机构投资者要进行一个大订单的交易时，会在短时间内改变市场上供需双方的平衡。这时，市场会产生两种情况：

• 交易价格向不利的方向移动，也就是买单的价格升高，或者是卖单的价格下降，进而提升了交易成本。这也被称为市场冲击成本。

• 交易对手方发现大订单后，会期待更有利的价格，从而减缓交易的速度，进一步降低市场上的流动性。这会使得交易不能够按计划顺利完成，进而产生机会成本。

针对这种情况，机构投资者通常的做法是将大的交易订单进行分割，利用

较小的订单，逐渐发送到市场，从而达到降低交易成本和发现流动性的目的。这些交易策略根据交易日的和基准的不同会被赋予不同的名字。例如比较流行的 VWAP(交易量加权的平均价格)和 TWAP(时间加权的平均价格)。它们的目的就是使交易成本尽可能地达到，或者是超过交易的基准，也就是 VWAP 以及 TWAP。还有些算法根据所关注的交易成本的目标的不同进行命名，比如 implementation shortfall，有时简称 IS，就是要尽可能地减少交易的执行落差(implementation shortfall)。另外还有，例如按照设定的交易速率进行交易，按照特定的交易量进行交易的方式等等。总而言之，这些分割订单的策略都是为了减少交易操作中的成本，进而改善投资表现。

此外，订单分割策略还有利于隐藏交易目的和意图。当一些投资者发现投资机会时，例如利好和利空消息、被低估或高估的股票等等，必然会通过交易操作的方式实现利润。但是，如果投资者交易数额很大的话，就会很容易被市场上的其他参与者发现，进而识别其交易动机。这时，其他投资者就会跟进，导致投资机会的盈利的幅度和可能性降低。因此投资者对于交易的隐蔽性很重视，不愿意向市场透漏交易信息。这种情况下订单分割策略就可以在一定程度上满足投资者的交易需要。例如有种交易策略被称为“冰山一角”(iceberging)，这个策略的目的就是通过订单分割的方式，限制每一时段交易的最大数量，这样可以隐藏或部分隐藏交易的动机。策略的名字“冰山一角”形象地表达了其操作方式和目的，就像冰山一样，露出水面的永远只是一小部分。

2. 订单智能路由(smart routing)

美国的金融市场上存在着很多的交易所、ECN，以及其他的交易市场。因此，同一只股票在这些市场上可能会同时存在着不同的价格和流动性条件。订单智能路由系统的目的就是持续监视不同市场上的交易环境，试图获得最好的买卖报价和流动性。智能路由系统给交易员带来的益处是不言而喻的，它可以减少交易成本和买卖价差，更好地利用流动性，甚至是减少市场冲击。例如瑞士信贷开发的游击队(guerrilla)和狙击手(sniper)算法。

游击队算法可以实时地不断接收不同交易市场上的公开报价，并进行判断。其目标是发现能够进行交易，而且造成价格移动的可能性较小的报价，以避免影响所交易股票的交易模式。

一些交易所和交叉网络会提供一些隐藏的流动性，并不在传统的公开平台上交易，而只在大的机构投资者和经纪商之间交易。这些流动性通常被称为流动性暗池(dark pools of liquidity)。瑞士信贷开发的一种交易算法叫狙击手(sniper)，就专门用来发现这类隐藏的流动性。

3. 交易自动化

交易自动化技术的发展使得经纪商的一些行为也能够进行自动化操作。例如做市和公司股票的回购。其中，做市是指经纪商持有某些股票或债券或其他金融产品的存货，不断地向公众投资者报出某些特定证券的买卖价格（即双向报价），并在该价位上接受公众投资者的买卖要求，以其自有资金和证券与投资者进行证券交易。经纪商通过这种不断买卖来维持市场的流动性，满足公众投资者的投资需求。经纪商可以利用买卖报价的差异（bid-ask spread）从中获取利润。花旗集团在 2007 年 7 月购买的自动化交易平台（automated trading desk）就是一个活跃的做市商系统，它占到了纳斯达克和纽约证券交易所总成交量的 6%。

4. 多资产支持

随着金融市场的发展，机构投资者可以利用订单管理系统将多种资产整合到一个平台当中，其中可以包括股票、期权、期货、固定收益产品和外汇产品等。这样的整合有利于投资者降低总的交易成本和风险，更有利于投资者从总体上把握投资的头寸、风险和盈亏状况。

6.3.2 自动化投资策略

1. 套利交易

套利交易就是通过发现两种本质相同的证券之间价格的差异进行交易，从而获取利润。下面介绍几种常见套利交易的方式：

（1）三角套利

在外汇市场上，不同货币间以不同方式进行兑换时，兑换比例应该是相等的。例如，如果 1 美元兑换 6.8 元人民币，1 元人民币兑换 1.1 港元，那么 1 美元应该兑换 $6.8\times1.1=7.48$ 港元。但是在某些情况下，市场会出现异常。有时会出现 1 美元兑换 7.5 港元的情况，那么投资者可以利用不同兑换方式的差异进行套利交易。但是，在外汇市场上，这种机会往往很少，而且转瞬即逝，所以一些机构投资者会开发专门的计算机程序，实时地监控外汇市场的价格，进行三角套利。

（2）指数套利

指数套利顾名思义就是指就股票指数产品和成分股票之间的差异进行套利。在指数套利当中，ETF 套利在我国引起过比较大的影响。

ETF 是 exchange traded fund 的英文缩写，即“交易型开放式指数基金”，

又称交易所交易基金。ETF是一种在交易所上市交易的开放式证券投资基金产品，交易手续与股票完全相同。ETF管理的资产是一揽子股票组合，这一组合中的股票种类与某一特定指数，如上证50指数，包含的成分股票相同，每只股票的数量与该指数的成分股构成比例一致。ETF交易价格取决于它所包含的一揽子股票的价值，即单位基金资产净值。但是，有些时候ETF基金的价值会和其成分股票之间的价值出现一定的差异，这时就给投资者带来了套利的机会。要进行这样的套利交易，投资者需要能够及时地更新指数基金和成分股票的价格，并及时做出计算，所以指数套利也离不开计算机的辅助。

(3)股指期货套利

股票指数期货是期货交易所里交易的重要的衍生产品之一。它产生于20世纪80年代早期，现在在发达国家和发展中国家非常普遍。股指期货套利是利用目标指数期货和成分股票组合等产品之间的价差进行套利交易。有下面几种套利方式：

①跨期套利是指利用两个不同交割月份的股指期货合约之间的价差进行的套利交易。一般来说，相同标的指数的股指期货在市场上会有不同交割月的若干合约同时交易。由于同时交易的不同交割月合约均是基于同一标的指数，所以，在市场预期相对稳定的情况下，不同交割日期合约间的价差应该是稳定的，一旦价差发生了变化，则会产生跨期套利机会。

②期现套利是指股指期货与股指现货之间的套利，是利用股指期货合约与其对应的现货指数之间的定价偏差进行的套利交易。即在买入(卖出)某个月份的股指期货合约的同时卖出(买入)相同价值的标的指数的现货股票组合，并在未来某个时间对两笔头寸同时进行平仓的一种套利交易方式。

③跨市套利。如果同一标的指数在两个以上的市场上均有股指期货，那么各个市场的同一股指期货的价值应该接近。如果出现了较大的偏离，就可以进行跨市套利。

2. 指数跟踪

指数跟踪策略类似于指数套利策略，其方式就是利用一些股票的组合来模拟指数。如果投资者持有的组合价格高于指数基金产品，那么卖出股票组合，买入指数基金；相反，如果投资者持有的组合价格低于指数基金产品，那么买入股票组合，卖出指数基金。指数跟踪的策略可以利用因素模型等对投资组合进行模拟和优化，然后用协整方法进行检验，进而利用两者之间的波动进行交易操作。

指数跟踪和指数套利策略是不同的。指数套利是在本质相同的两个投资组合之间套利，风险远小于指数跟踪。但是，指数和成分股票之间的价格出现差异的机会可能比较小，相对来说指数跟踪策略的投资机会比较多。这里就需要投资者能够很好地复制指数，尽可能地减少跟踪误差，以获得较好的业绩。

3. 动量交易和均值反转

动量交易(momentum trading)假定股票价格会跟随趋势，也就是说，如果前一期股票价格上涨，那么下一期股票价格也将会上涨，那么就选择买进；相反如果前一期股票价格下跌，那么下一期股票价格将持续下跌，那么就选择卖出。

均值反转(mean-reverting)策略与动量交易策略恰恰相反，它认为股票价格会在某一个数值上下波动。具体地说，就是如果前一期股票价格上涨，那么下一期则会下跌；而如果前一期股票价格下跌，那么下一期股票价格则会上涨。

在两种情况下，投资者都可以制定能够获利的交易策略。如果投资者认为股票价格会出现趋势跟随，那么可以选择动量交易策略，相反则可以选择均值反转策略。如果股票价格表现为随机游走，那么就很难做出预测。

学术界有很多研究认为股票价格接近于随机游走，但是这并不意味着不会出现趋势跟随和均值反转的情况。例如人们常说的牛市和熊市，就是一段时间股票价格持续的上涨和下跌。这就是趋势跟随的一个明显的例子。此外，股票分析中常常提到的"震荡调整"也许就是均值反转的交易机会。选择趋势跟随还是均值反转需要投资者根据具体情况而定。投资者可以根据不同的基本面和市场状况，以及不同的目标投资期限来确定选择动量交易或是均值反转策略。

趋势跟随的现象一般认为有三个原因。第一，投资信息的逐渐扩散。当一个公司出现利好和利空消息的时候，也许公司员工先得到消息，然后是金融和新闻从业人员，之后扩散到大众。他们交易时机的不同会导致股票收益率的自相关。这也被称为异步交易。第二，大订单的分割操作。当机构投资者发现一个被低估的股票时，由于交易量很大，他们很有可能将订单分割到几天，甚至更长的时间进行操作，那么这段时间内股票价格会表现出持续的上涨。第三，羊群效应，俗称"跟风"效应。一只股票的上涨有时会被大众误读为新的投资机会，那么大量投资者的涌入会导致股票价格持续的上涨。

均值反转策略的基本观点就是股票价格最终会反映公司的基本面价值。

如果股票价格出现意外的上涨或下跌，那么在未来还会被拉回到内在价值的水平。因此，在没有新的利好或利空消息的情况下，股票价格一般会表现出均值反转。

动量交易和均值反转策略本身并不能为投资者提供很有帮助的指导，但是这并不意味着设计相应的交易策略会很难。例如，Khandani 和 Andrew Lo 在 2007 年发现一个很简单的均值反转策略，就是在一只股票出现最大涨幅之后卖出，在出现最大跌幅之后买入。如果不考虑交易成本，那么这个策略从 1995 年开始就有着很好的表现，在 2006 年其夏普比率高达 4.47。另外，一些技术分析者可以根据股票价格的 30 日均线设定动量或均值反转策略，例如如果股价超过均线，那么意味着价格的上涨，则可以选择动量交易；如果认为回归均线，则可以选择均值反转策略。

4. 配对交易（pairs trading）

股票市场投资不变的获利方式就是买入低估的股票，卖出高估的股票。在无法准确估计股票价值的时候，很难分清哪只股票是低估的，哪只股票是高估的。配对交易选择规避这一问题，这种交易方法采用相对估值的方法。具体说来，如果投资者选择两只类似的股票，则两只股票的价差应维持一个稳定数值。当价差偏离稳定值较远时，价差倾向于回归到稳定值。那么投资者可以在价差较大时买进价格较低的股票，同时卖出价格较高的股票，而在价差较小时进行相反的操作。因此，从本质上讲，配对交易也是一种均值反转的交易策略。

对于这一思想背后的理论基础存在两种观点：一是遵从 APT 理论。根据 APT 理论，当两只股票有几乎相同的风险因素时，则这两只股票的预期回报应该大致相同，所以由噪音或个别事件引起的回报偏差最终会消失；另一种观点认为，这一现象是基于资金的轮动。这一观点认为，资金在相似的股票间是轮动的，当一只股票相对于一只相似的股票价格上涨过大，则资金持有者会认为上涨的股票相对被高估，因此会抛售上涨的股票并购买相对价格较低的相似股票，使得价差回归稳定数值。

另外，模式识别和很多复杂的预测技术都能用来设计交易和投资策略。总之，算法交易可谓是多种多样，五花八门，而且还在不断地发展和变化。投资者会试图不断地发现证券市场上的某种价格变动或交易模式，设计新的交易策略，以达到降低成本、获取超额的收益等目的。麻省理工学院金融工程实验室主任 Andrew Lo 说："现在算法交易开始成为一场军备竞赛，每个人都在设计更复杂的算法，而且竞争越多，利润空间越小。"

6.4 算法交易的优势和影响

随着金融市场的不断发展，在欧美市场，算法交易逐渐成为主要的交易方式之一。算法交易之所以拥有这样强大的渗透力，也是因为它有着传统的人工交易方式无法比拟的一些特点和优势：

第一，计算机系统能够实现一些人工交易手段力不能及的交易策略和交易机会。

计算机系统有着非常快的反应速度，以及强大的计算能力。金融市场上有些交易机会稍纵即逝，例如外汇市场上的三角套利等，有时可获利的机会往往在几秒之内，这样的机会依靠人工手段很难把握得到；还有订单智能路由技术能够在不同的市场上获得最优的交易报价。此外，随着金融交易技术的发展，有很多交易策略的计算量是人力所不能及的。例如，指数套利需要在非常短的时间内对指数基金的价值做出计算，以发现指数基金或期货产品跟成分股票价格之间的差异；模式识别等高级的预测技术通过人工操作是不可能实现的。通过算法交易系统，可以将一些程序化的交易策略编制到计算机系统当中，实现快速的反应和强大的计算能力的结合。

第二，算法交易有利于减少交易操作过程中的成本。

通过算法交易，投资者可以在一定程度上减少交易操作当中的成本。订单智能路由技术可以使投资者获得最优的交易价格。直接入场技术可以减少交易过程中的延迟，而降低交易风险。订单分割策略可以通过交易优化技术减少交易操作当中的市场冲击、机会成本等等因素，尽可能地获得更好的交易平均价格，减少交易成本。

第三，算法交易系统能够节省人力。

金融市场的交易操作中，很多交易策略的概念和想法并不复杂，例如简单的套利交易和订单分割策略等。通过对类似交易行为的自动化，可以为金融机构节省很多的人力。这样，交易员们可以把精力更多地放在能为公司带来更多附加值的工作上，例如研究，以及开发新的产品和交易策略等等。

第四，算法交易能够减少人为因素对交易的影响。

在实际的交易操作中，有时交易员会违反公司的指示，导致公司承担一些不可预料的风险。1995 年，巴林银行在新加坡的一名交易员 Nick Leeson，受公司的指示对日经 225 指数期货在大阪交易所和新加坡交易所的价格差异进

行套利。但是随着交易经验的增长，他在公司的管理层不知情的情况下，逐渐从一个套利者变为了投机者。当他的交易发生了亏损的时候，起初他能够进行掩饰，但是为了挽回亏损，他开始进行更大仓位的投机操作。当 Leeson 最后被发现的时候，他带来的损失已经接近 10 亿美元。最终，巴林银行这家全球最古老的银行之一破产了，曾经是英国贵族最为信赖的金融机构，200 多年优异的经营历史，没能逃过破产的结局，事件震惊世界。

如果通过使用算法交易进行套利操作，可以避免类似的情况出现，计算机系统能够表现出绝对的纪律性。进一步的，利用算法交易和相应的监管措施，交易系统更能够按照事先设定的交易策略进行操作，而不受交易员主观情绪的影响。

第五，计算机系统可以实现不同交易步骤和资产之间的更加一体化的管理。

通过利用计算机系统，机构投资者可以将不同的金融资产（例如股票、外汇、期货、债券和衍生品）整合到一个统一的交易系统当中去，进而实现对总的收益和风险的统一的管理和监控。另外，通过机构内部信息系统纵向的整合，可以将前台部门和后台部门更紧密地结合在一起。这样，交易前的分析、交易的操作和交易后的评价机制将能够表现得更加一致，更有利于投资过程的管理。

在看到这些优势的同时，我们也应该意识到，不能够对算法交易的作用感到过分乐观。夏普在 1991 年的一篇文章当中提到，积极性基金总是一个负和的博弈。原因很简单，股票市场总价值的增长率既可以分解为成分股票增长率的加权平均，也能够分解为不同投资者的收益率的加权平均。投资者总的收益率是和股票市场总的增长率相一致的。一些投资者获得超额收益的同时，是以牺牲另外一些投资者的收益而换来的。算法交易不可能为所有的投资者都带来超额的收益，所以我们也可以预见，金融市场上对于新的交易算法和策略开发的竞争会愈演愈烈。

另外，从 1987 年美国股票市场的大跌也说明，计算机化的程序交易在某些极端情况下的快速反应也会给整个市场带来预见不到的巨大风险。因此，这也为证券市场的监管者们带来了严峻的挑战。

在亚洲金融市场，采用算法交易较多的主要是相对发达的东京证券交易所、香港联交所和新加坡交易所。与欧美市场相比，亚洲市场的股票价差更大、流动性更差、更难成交，因此算法交易也将更有用武之地。根据 Aite Group 的报告，2007 年，亚洲地区算法交易的使用比例为 4%，预计到 2010

年，会达到16%。作为相对落后的金融市场，预计在未来亚洲市场上算法交易将会以更快的速度发展。

在国内，随着金融行业的不断发展和国际化的提高，以及股指期货、融资融券规则的推出，我国证券市场单边交易和相对封闭、发展滞后的情况会有所改变，并逐渐赶上世界先进的证券市场。因此，算法交易在未来一定会呈现出快速发展的趋势。它不但有利于投资者减少交易成本，投资手段和策略的丰富和创新，而且能够促使市场更加的规范和高效。进一步的，我们可以预计，算法交易也将给国内的金融市场带来新的机遇和挑战，券商、基金、期货公司等金融机构也在积极地在算法交易上进行研究和投入。我们相信，能够在算法交易的创新和发展过程中获得优势的机构将在我国未来的金融市场上获得强大的竞争优势。

第七章

算法交易系统的结构

7.1 交易流程简介

随着市场的不断发展,金融行业已经很大程度地实现了交易流程的自动化。通过网络连接,人们可以实时地获得大量的交易和价格信息。计算机系统可以帮助投资者对市场进行大量复杂的分析工作,使投资行为更加地科学和有效。投资者还可以利用实时的网络连接进行下单、交易,这极大地提高了交易活动的速度和效率。每一个人能够同样地利用互联网访问市场,也提高了交易市场上的公平性。此外,海量的历史和实时交易数据也为投资和交易的评价工作提供了基础,使评价过程更加公平和有理有据。

在过去,当一个交易员进行交易操作时,必须有其他一系列相关的工作来保证交易双方之间的证券和资金的有效支付。前台和后台部门必须进行紧密地合作,这样才能防止交易过程中出现意外的差错。其中,前台部门负责投资管理、公司的投融资服务,以及销售,而后台部门则负责公司内部的运营。前台和后台责任的分离可以尽可能地减少欺诈行为,比如欺骗、贪污,以及违反规则的行为。交易流程中各个部分(例如处理、确认、交割)的独立能够实现操作的完整性。后台部门只负责几个最重要的部分:记录和确认完成的交易,提供内部控制机制来实现责任分离等等。后台部门能够保证金融机构内部操作的完整性,并减少操作、交割和法律的风险。前台和后台的联系可以是完全手工操作的,也可以是完全计算机化的。

交易过程需要达成一个协议,以便于在前台开始交易后进行交易数据的

确认。交易契约的副本用于交割和账目记录。一旦交易由前台部门执行以后,交易过程剩余的部分就留给了后台。后台部门负责各种证券、商品和书面合约的支付、发送和接收,以及确认合约所确定的支付对象和数量。

在一个交易协议达成以后,交易双方需要向对手方发送一个确认信息。后台部门跟着就需要提交确认信息,并跟踪和管理对手方的确认信息。严格的过程控制能够有效地防止欺诈性交易。例如,交易员可能进行虚假的交易,或者在达成一份合约并发送原始确认信息后毁掉副本。这些可能使交易员建立一个基金经理所不知情的交易头寸。在平仓的时候,交易员可以为之前毁掉的合约补一份记录,然后和抵消的合约一起上交,这样头寸就被抵消掉。如果由一个独立的部门来进行确认信息的接受和审核,就能够马上发现这种欺诈行为。

在一个买进或卖出交易完成以后,交易必须通过后台部门和一个清算代理机构的沟通进行清算。在结算日,买卖双方的资金和金融产品互相进行交换,并更新交易账簿条目。当买方或买方代理收到或者发送证券,以及卖方履行支付后,交易结算完成。经纪商会为这些任务指定一个专门的机构,比如清算中心,但是,经纪商仍然有责任保证为客户正确地处理交易。如果对手方没有进行支付,交易一方可能就会产生损失。有些情况下,清算代理人和经纪商对完成交易中的任何问题都需要负责。所以,应该通过一个信用部门对资金和债券流动进行持续监控,进而控制结算过程中的风险。

后台部门应该及时地进行对账,以确保机构的交易流程和规定保持一致。负责对账的人员应该和负责输入交易数据的人员保持互相独立。对账中,应该确认前台部门持有的头寸,以及提供审计跟踪报告,需要审核的报告包括:交易员的头寸、监管报告、收支报表。

交易系统和流程的自动化,或者进行外包,可以有效地减少后台部门当中的人工操作。这样能够为金融机构减少人力的成本,而且能够更加有效地对交易流程进行监控。人工操作的减少不但能够提高后台的运行效率,还可以减少人工操作带来的一些错误,并进一步地减低交易流程中的风险。

一般来讲,现在的机构投资者的交易流程大致可以分为三个阶段:交易前、交易和交易后三部分:

• 交易前部分负责投资组合的管理、投资的分析研究等等。引进 IT 技术对交易前分析有很大帮助,通过提供大量的数据资源和强大的计算能力,基金经理和交易员的分析工作变得更加的方便、简单和高效。

• 交易部门负责订单的管理,下达交易指令,以及交易头寸的实时动态管

理等等。简单地说，交易部门就是负责交易的操作，尽可能减少交易操作当中可能产生的成本和意外的风险。

• 交易后部门则负责交易的确认、审计、运营工作，以及交易评价。交易后部门负责确保交易的完成，并对交易表现做出恰当的评价。

如图 7.1 所示，交易前后的分析和交易的操作形成了一个交易的周期。

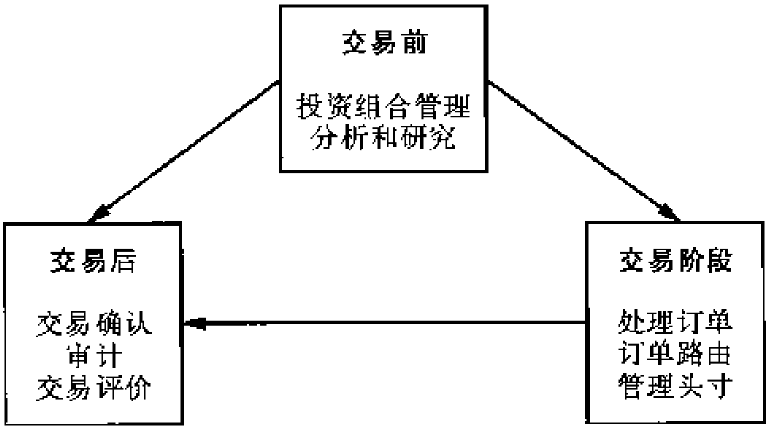


图 7.1 交易流程图示

交易前分析通过对投资策略的研究，可以使交易员充分了解当前所持有的投资组合的风险和交易模式等等特征。交易前分析还需要提供交易相关的许多信息，其中包括新的市场信息、流动性的特征和预测、过去一段时间的市场环境情况、交易当天的市场条件预测等等。这些信息对交易操作来说十分重要，它们可以使交易员充分地了解市场、交易目标、持有头寸等各方面的特征。通过交易前分析，交易员能够更好地贯彻基金经理的投资意图，也能够更好地进行交易操作，充分降低交易成本和风险，进而改善整体的投资表现。

交易操作过程中主要的目标就是降低交易操作给投资带来的成本，其关键主要包括三点：一是尽可能降低交易中的延迟，减少数据传输过程中的风险；其次是尽可能获得最优的买卖交易报价；第三，通过订单分割，尽可能减少交易当中的市场冲击和价格移动带来的意外风险。算法交易在这三个方面都做出了改善。通过计算机系统的整合，投资的分析活动和交易活动能够在一个系统中完成，不但有利于交易与投资目标的一致，而且能够减少不同部门之间信息传递的延迟，降低决策延迟为交易带来的风险。直接入场技术（DMA）能够帮助交易员直接访问到市场的交易系统，减少网络传输和交易环节上的延迟。而订单智能路由（smart routing）技术能够实时地监控不同市场上的交易报价和流动性情况，可以帮助交易员获得最优的交易价格。订单分割策略

也可以通过算法交易来实现。利用一些优化技术，计算机系统通过分割订单可以有效地平衡交易当中的市场冲击和时间风险，有效地改善交易操作的表现。

交易后分析用于确保交易完成的可靠性，另一方面还需要分析交易成本产生和大小，以及对交易操作做出合理的评价。分析交易成本是进行交易操作时决策的基础。只有了解交易成本产生的原因和大小，才能够做出正确的交易决策。此外，通过交易评价，基金经理可以更清楚地了解每个交易员和交易策略的表现和特点。这有助于在以后的交易当中选择适合的交易员和交易策略进行特定的交易。因此可以说，交易后分析是改善交易操作质量的第一步。

交易活动的三个阶段是互相联系和促进的一个整体，它们共同构成了一个交易活动的周期。交易操作的改善是一个长期的、反复的、循环的过程，而通过计算机系统的整合，整个交易流程变得更加协调、紧密和一致。此外，强大的计算机和数据库系统可以储存和处理大量的历史和实时交易数据，也成为交易分析和决策的有效工具。

7.2 直通式处理和算法交易

交易后台部门活动的自动化的发展，以及使用计算机记录和分析交易历史促进了交易前分析的发展。市场信息的数量在不断地增加，而 Bloomberg、道琼斯等就是金融信息服务这一领域的先驱。它们整合了市场数据、证券信息和交易分析，并使自己成为金融信息服务业的著名企业。同时，这些发展也使得前台和后台之间能够通过计算机技术进行联系和管理。从此，金融业的公司开始设计方法对两个原先不同的部门的数据流进行整合，产生了直通式处理(straight-through processing, STP)。

7.2.1 直通式处理

直通式处理使得资本市场的整个交易流程能够通过电子方式进行，而不需要人工干预。直通式处理的优点包括减少运营成本，缩短整个交易的周期，减少结算风险等。此外，通过直通式处理，交易流程全部在计算机系统中进行，也更便于进行监管部门和投资机构内部的管理。

直通式处理的一大进步就是为交易过程建立一个标准化的协议。通过标准化的协议，不同的机构和部门之间的交易管理系统可以轻松地实现连接和沟通。这有助于整合整个的交易流程，改善信息流处理和传递的速度，并缩短了交易结算时间。其次，电子交易数据的一致性通过手工操作很难达到，而直通式处理很方便地就能够实现这一点。另外，通过交易前工作的自动化，市场参与者可以进行交易前策略模型的构建、分析和头寸管理。

现在，交易员既可以通过电子系统，也可以通过电话沟通来完成交易。由于交易环境的变化非常快，信息有时会被错误地输入、传播或解释。因为交易信息通常不是在交易的同时输入，所以数据有时会出现丢失，或错误地录入或读取。电子交易可以解决其中的很多问题。如果使用集成化较高的系统，数据从创建一开始就能在交易流程当中保持一致性。

如果可以实现全面的直通式处理，那么投资管理机构、经纪商、客户、监管机构和其他金融业服务机构将会实现很大的便利。一些业界分析者认为，百分之百的直通式处理，也就是全面的自动化，是很难实现的。相对而言，金融机构在内部实现直通式交易可以鼓励不同部门之间的协作，改善交易操作和信息处理的质量。另外，进行交易的双方之间的直通式交易也将成为业界的主流。

7.2.2 算法交易

电子交易和直通式处理引发的一个关键性的技术进步就是算法交易。从出现以来，算法交易在华尔街就成为投资机构和经纪商用于降低交易成本的主要手段之一。算法交易系统通常分为四部分：数据管理、策略开发和实施、订单管理系统(order management system, OMS)，以及订单路由。

1. 数据管理

毫无疑问，现代的金融业是一个依靠信息和技术优势进行竞争的行业，尤其在算法交易兴起后，历史和实时数据成为十分重要的竞争优势。在收集和传播大量市场数据的基础上，交易员才能够通过计算机系统执行交易。处理市场数据的速度可能会意味着交易是否成功。1 毫秒的差距可能使一个公司丧失获利的机会。根据 Securities Industry Automation Corp(SIAC)的统计，市场数据流信息增长很快。下面是 2004—2006 年这 3 年中 11 月份信息量最高的 1 分钟内每秒的消息数：

- 2004 年 11 月：56 000 条每秒
- 2005 年 11 月：121 000 条每秒

• 2006 年 11 月：200 000 条每秒

自动化的交易模型每天可以执行数千次交易，并成为买方和卖方交易员普遍的交易手段。电子化的交易依赖于市场实时数据分析。每个使用算法交易的公司都在寻求减少信息传输延迟或零延迟的方法。数据供应商不断地为更高效的数据流工作。大量金融信息服务公司的涌现使得金融机构积极改进它们的数据设备。金融机构需要通过跟踪能力、延迟、执行质量等许多指标来评估它们的交易环境，以便于改进交易流程。精确评估一个公司的运作能够帮助它们提供更好的服务，更低的成本，以及减少交易活动当中的摩擦。许多买方公司已经应用实时的系统对算法交易进行评估和监控，持续地依据交易目标评估算法的表现，同时还关注后台的交易流程，包括订单发送比率，执行和完成订单的情况。当产生问题时，比如没有完成订单、丢失订单、模型的表现变差，就可以马上对交易算法和策略进行调整。

2. 策略开发和实施

客户利用数据库和策略分析工具分析大量市场数据，以便于开发交易算法用于交易操作。这些平台用于交易前和交易后的实时或历史数据分析，通过分析历史数据能够帮助交易决策，例如：管理订单流、大宗交易和净交易计算、流动性特征、低价值附加操作，以及交易成本分析等等。其中，交易分析是算法交易系统的核心部分之一。

当交易员进行交易时，他们需要决定向哪个交易市场发送订单，同时还需要考虑交易股票的流动性、供需状况等等因素。交易员发送订单时需要进行一些分析，他们通常会利用经纪商对交易的分析和观点。经纪商在交易环境、观点和操作策略上的研究一般都是有利用价值的。经纪商一般会提供关于交易成本及操作的详细建议。相对来说，大的基金经理更加重视经纪商的研究报告。在订单输入以后，交易员需要对订单交易方式进行选择：是进行大宗交易，还是和同类型以及相同操作指令的订单一起进行结算？订单发送以后，交易员的工作就是要增加交易的价值，通过分析来确定证券是否能够在当时市场条件下以最优的方式执行。在价差较大时、流动性较低时、订单交易量较大时，交易员更能够发挥个人的业务特长，使交易分析带来更多的附加价值。

随着算法交易成为主流，很多的交易分析和策略都能够通过计算机系统实现，下面给出几个例子：

• 相关性和波动性分析

• 识别交易机会

• 确定最优的交易时机和数量

• 根据交易的基准评价交易操作

通过策略开发和实施的平台，投资者可以很容易地利用数据发现交易机会，实现和修改交易的目标和策略，进而减少交易的成本，改善交易表现，以获取更高的回报。

3.订单管理系统

订单管理系统是负责下达、修改和处理订单的计算机系统。随着市场的发展，交易员为管理交易工作流程需要更好的工具，因此订单管理系统也在不断改进。在实际交易流程中，订单管理系统从各个基金经理那里收集订单和指令，把订单集中在一起形成大订单，同时收集信息，对数据库进行更新，并形成报告。

(1)特性和功能

订单管理系统一般有如下几个特性和功能：

• **交易记录** 交易记录作为中心，能够让交易员管理订单，应用各种基准进行评价，跟踪当前的头寸、交易数据和实时损益状况。

• **实施算法交易策略** 交易员进行交易操作的时候，通常需要选择进行交易中的算法和参数等等。然后，订单管理系统会自动利用算法对订单进行分割、交易等操作。算法需要保证一定的灵活性，以方便交易员根据需求进行修改和定制。

• **交易前和交易后分析** 交易前分析提供市场条件的分析，例如交易量、波动性等等。它能帮助交易员决定最适合特定情况的算法，并估计给定交易的成本。交易后分析则负责依据一定的基准和其他参数估计以往交易的表现。

• **交易市场的连接** 通信连接是算法交易系统的生命线。它使得买方交易员和经纪商能够进行电子通信。这样，订单管理系统才能够实时地通过算法做出交易决策。

• **处理多种资产** 算法交易系统应该不止支持股票的交易，而应该支持更多的金融产品，例如固定收益、衍生品、外汇等等。

• **提供合规报告** 类似于单独的股票和大宗交易，订单管理系统必须能够不断地根据证券业管理部门的监管环境进行定制和修改，通过基于规则的触发器和灵活的报告能力进行管理。

(2)举例

下面是一个通过订单管理系统进行交易的例子：

①基金经理决定买进 300 000 股 IBM 的股票。

②交易员接到交易指令，然后决定往哪个市场发送订单。

③当交易员发现他们需要买进 300 000 股 IBM 的股票时，通过 ECN 聚集器观察多个 ECN 和交易所的报价。

④买方交易员决定往哪儿发送对 IBM 的交易，选择包括：

• 通过经纪商进行大宗交易；

• 使用算法交易，比如利用 VWAP，并让算法识别交易模式，通过智能路由的功能，系统将会发现每次下单时最好的价格。

⑤交易执行的平台会把交易信息反馈给交易员，然后订单管理系统会将交易数据提交给数据系统，用以评价交易质量。

4. 订单路由

在美国股票市场上，除了较大的纽约证券交易所(NYSE)和 NASDAQ 交易所之外，还有许多的地区性股票交易所，以及另类交易系统(alternative execution venue)。一些公司的股票同时在几个市场上市，或者同时在不同的市场进行交易。由于不同市场上的流动性差异，以及报价信息延迟等等因素，一只股票在不同的市场上可能会存在着不同的买卖报价和流动性条件。例如，流动性差的市场上股票就可能会产生较大的买卖价差。这种情况下，交易员就希望能够在这些市场上获得最优的交易价格。订单智能路由技术就用于解决这一问题。订单路由一般通过直接入场技术(direct market access, DMA)实现。交易系统需要持续地监控不同的电子交易市场上的报价和流动性条件。当交易员发起交易时，交易系统识别订单的类型，并在满足预先设定交易参数的情况下进行交易。由于计算机系统可以同时获得不同市场的价格，通过比较不同的报价，系统自动会把订单发送到给出最优报价的市场，以实现最优交易。此外，交易系统也可以比较不同市场的流动性，选择流动性比较好的市场进行交易，这样可以尽可能地减少交易产生的市场冲击。例如瑞士信贷的游击队算法就是基于这个思路所开发的。总的来说，订单路由技术有利于改进市场上的价格发现机制，使投资者能够获取最优的交易价格、降低交易成本，同时也有利于改进股票市场的公平和效率。

7.3 FIX 协议

FIX 的全称是金融信息交换协议(Financial Information Exchange Protocol)，是国际上统一的一种电子通信协议，用于证券市场和交易信息的实时交换和通信。在最初的 1992 年，FIX 是作为富达投资(Fidelity Investment)

和所罗门兄弟(Salomon Brothers)之间股票交易的通信框架被创建的。多年以后,FIX在事实上已经成为全球股票市场上交易前和交易过程中通信的标准协议,而且现在还在快速地向交易后系统领域扩张,并能够支持直通式处理过程。基于这些基础,FIX已经获得了强大的推动力,在向外汇交易、固定收益和衍生品交易市场扩张。

FIX协议是一系列用于金融交易的电子通信消息包的规格说明。它是在全世界众多的银行、证券经纪商、交易所、机构投资者、其他金融业的部门协会,以及一些信息技术公司的合作下制定的。在日常事务当中,这些市场参与者都希望金融产品的自动化交易拥有一个统一的语言。因此,可以说是他们共同创造了FIX协议。

FIX是由金融业发起的用于应对全球金融服务业变化的通信标准协议。公司可以使用FIX协议进行透明化的、低成本和延迟的电子化交易。FIX是开源和免费的,但是它不是一种软件,而是一系列通信数据的规格标准。软件开发商可以基于FIX协议来开发自己的商业化或者开源软件。作为领先的交易通信协议,FIX已经被用于很多的订单管理和交易系统当中。因此,用户不需要了解FIX协议的细节,就可以从中受益。可以说,FIX协议是金融机构间对话的"语言"。

图7.2中显示了一个用FIX链接的交易系统的结构。交易双方的交易员通过订单管理系统进行订单的下达、修改和撤销等操作。在交易指令下达以后,订单管理系统将交易指令发送到基于FIX协议的交易系统当中。在这个系统,交易指令将以FIX协议的形式进行表示,集中负责处理交易信息和其他的业务信息,例如市场行情、利好/利空消息等等。然后,双方的交易系统通过TCP/IP网络实现交易的连接。整个流程中,交易员和订单管理系统,以及系统中各个模块之间都是基于FIX协议进行通信的。

2007年,Tower Group在美国发起的一项关于FIX的调查发现,全球金融服务领域中已经大量地使用了FIX协议。在被调查的对象中,75%的买方公司和80%的卖方公司在电子交易中使用了FIX协议,而且他们都表示要继续增加对FIX的使用,并使得FIX能够支持更多的资产类别。这项调查也显示,FIX在交易后的系统中成为重要的一部分。此外,FIX也受到了交易所的关注,超过3/4的交易所支持FIX的接口,大部分交易所通过FIX协议处理的交易量都超过了25%。在美国,几乎所有主要的股票交易所、投资银行、最大的共同基金,以及很多小的投资公司都使用FIX来进行电子化交易。许多期货交易所都提供FIX的链接,主要的债券经纪商也使用或正在部署FIX。

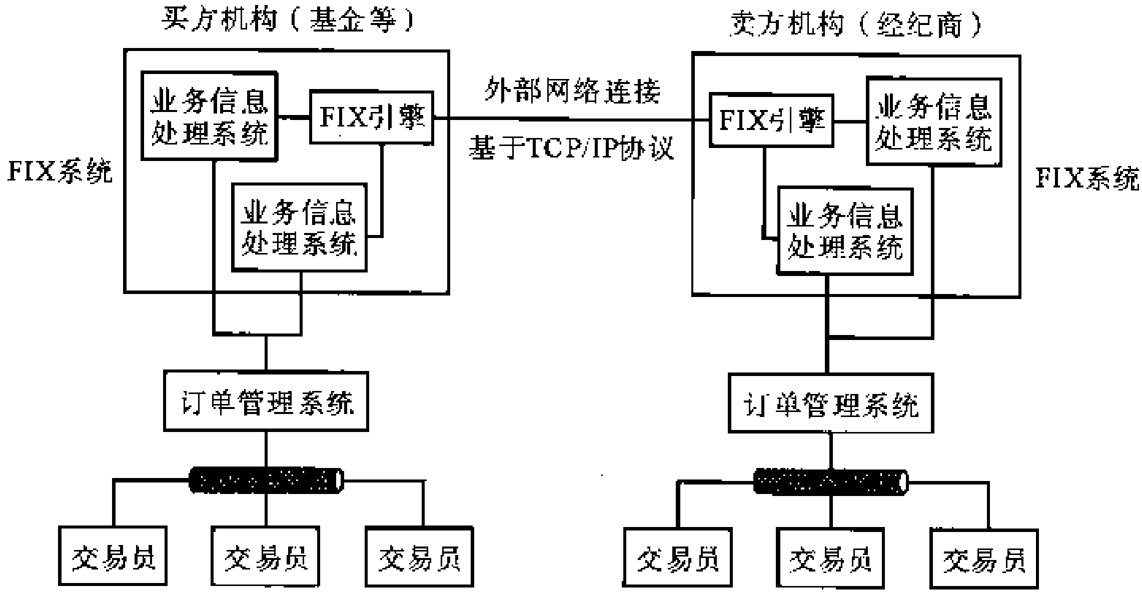


图 7.2 FIX 通信的构架

7.4 我国金融业的通信标准

在我国证券市场快速发展的环境下，市场对业务和技术创新的需求不断提高，急需建立一个统一高效的交易技术体系。然而在以前，沪深证券交易所、主要的期货交易所、券商和其他机构间都采用各自设计的非标准化的接口。数据信息交换模式、编码方式、接口、业务数据流程等都不统一，存在着对业务创新的适应性较差、适应成本高，不同市场间难以有效交换信息等问题。

2005 年，我国证券业期待多年的标准化工作取得了进展，公布了证券交易技术的八大标准。标准的制订和应用是我国证券市场高效、规范发展不可缺少的基础设施工作之一，会对我国证券行业产生广泛和深远的影响。其中，《证券交易数据交换协议》(Securities Trading Exchange Protocol，简称 STEP)是我国金融行业的通信标准。STEP 协议以 FIX 协议为基础进行开发，是我国证券交易技术的一大创新。它的实施和应用将提升两大交易所和所有证券市场参与者的技术和业务水平，促进我国证券行业标准化、国际化，为提升我国在国际证券市场的竞争力打下坚实基础。

STEP 协议主要应用于证券交易所和券商间的数据连接，规定了证券交

易所交易系统与市场参与者系统之间进行证券交易所需的数据交换规范。该协议不依赖任何物理网络，也不依赖特定的底层通信协议。因此，STEP 协议完全可以支持目前国内的地面网络(如 DDN)和卫星网络。作为证券交易技术标准，该协议规定了应用环境、会话机制、消息格式、安全与加密、数据完整性、扩展方式、消息定义、数据字典等内容。

STEP 的推出，适应了我国证券市场信息技术发展的迫切需求，迈出了走向统一证券技术体系的重要一步，为国内证券交易标准化接口提供了统一的规范，将积极促进市场参与者面对多个交易所、多通道、多协议接口问题的解决。国内证券市场的繁荣发展带来了业务技术的多样性和复杂性，市场参与者的技术投资越来越大，技术风险越来越得到重视，统一的接口标准会降低技术风险，从而减少整个行业的运行风险。各证券交易所和其他市场参与服务机构采用该标准后，将会给行业带来巨大的效益，对于广大市场参与者来说，采用符合标准的接口系统，除降低技术风险之外，也将会降低系统开发和维护成本。

随着 QFII 的实施，与国际业务交往日益增加，我国市场参与者将面临国际合作与竞争。FIX 协议是国际上普遍采用的接口标准，STEP 正是以 FIX 协议为基础，结合中国实际情况而制定的。在制定过程中，与 FIX 国际组织进行了充分的交流，确保 STEP 与 FIX 的兼容性，同时，也对 FIX 的完善和发展产生了积极影响，满足了我国证券业国际化的需求。简而言之，STEP 的推出对于降低行业交易成本、减少运行风险、提高市场效率具有重要意义。

7.5 电子通信网络

7.5.1 另类交易系统

另类交易系统(alternative trading systems，ATS)是美国证券交易协会(SEC)通过 300 号规则所确定的非交易所的证券交易市场。它们对公众在交易所以外获得交易流动性起到了重要的作用。SEC 在 300 号规则中将另类交易系统定义为符合下列条件的任何组织、协会、个人、团体和系统：

• 将买家和卖家聚集到一起，并组建、提供或维持一个市场，或者像规则所规定的股票交易所一样运行。

• 同时做到不出现以下行为：

1. 设定规则限制会员在该组织、协会、个人、团体和系统上的交易以外的行为。

2. 惩罚会员的方式不超过禁止交易。

典型的另类交易系统包括 ECN、交叉网络等等。它们不但为投资者提供了额外的交易场所和流动性，而且使投资者可以在交易所收盘之后的时间进行股票的交易活动。另类交易系统的出现为证券交易市场之间带来了竞争。证券交易所和 ECN 等市场开始更多地改进网络连接方式和服务质量，以及降低交易成本。

2000 年 5 月，美国证券交易委员会废除了 390 号规则。该规则禁止 1979 年 4 月以前在纽约证交所上市公司股票在全国性的证券交易所以外的市场进行交易。390 号规则的废除进一步促进了另类交易系统的发展，以及 DMA、订单路由等算法交易技术的发展。

另类交易系统在欧美的股票市场上获得了很大的成功，极大地改变了美国股票市场的结构。但是另类交易系统没有能够渗透到亚洲市场，一方面是亚洲市场没有充分的经验来实施另类交易系统，另一方面是亚洲大部分国家都只有单一的垄断性的股票交易所。不过，在 Aite Group 公司 2009 年的一份报告中显示，近几年亚洲市场的另类交易系统正在出现增长，他们预计在 2012 年会占有 20%的市场份额。

7.5.2 电子通信网络

电子通信网络(electronic communication network，简称 ECN)是指金融交易所、股票经纪商和它们的客户之间构建的电子化交易平台，用于在交易所外的金融产品交易。ECN 主要支持的金融产品包括股票和外汇。1998 年，美国证券交易委员会(SEC)批准建立了 ECN，它用于增加交易公司之间的竞争，以达到降低交易成本，使客户充分地访问订单簿，以及提供传统交易时间以外的撮合交易等目的。

ECN 能够使市场参与者更好地发现流动性，并帮助买方交易员更好地实现自动化。ECN 能够进行实时的价格发现，使买家和卖家在最少的中介的情况下相对便宜地进行交易。SEC 将电子通信网络定义为："在特定价位，自动地撮合买卖订单的电子交易系统。"ECN 是现代证券市场的一个发展。ECN 的自动交流和匹配系统为金融市场带来了更低的交易成本。

美国市场上早期主要的 ECN 包括：Instinet(INET)，Bloomberg(TradeBook)，Archipelago(ArcaEx)，SunGard(Brut)和 NASDAQ 的 SuperMontage。这些 ECN 都能够为市场提供流动性，并拥有自己的订单簿。在早期，在一个 ECN 将交易订单转发到其他 ECN 之前，会先在其内部寻找流动性。这意味着，虽然其他 ECN 可能会有更有利的成交机会，但是投资者不一定能够利用这样的机会，因为订单要先在 ECN 内部进行交易的撮合。这导致了美国股票市场之间的分化，不同交易市场之间的交流很少。为了抑制这种分化，金融业的公司和技术供应商开发了聚集流动性的工具、订单路由，以及 DMA 技术。

传统上，经纪商是证券交易市场的看门人，投资者需要一个经纪商作为他们在交易所进行交易的渠道或中介。交易所并不限制 ECN 的访问，而 ECN 能够为广泛的投资者提供交易服务。ECN 能够直接匹配买卖双方，从而省略了人工中介，减少了经纪商的利润。ECN 能够比现有的市场交易平台提供更高效的操作。ECN 可以让投资者看到限价订单簿，为投资者带来了更加清晰和完整的价格信息。尽管电子交易系统有其独特优势，但是很多交易员仍不欢迎 ECN。有些交易员认为，在 ECN 上不能有效地交易大订单，通过 ECN 交易不够直接和迅速，而且不方便在预期市场大变动的时候做出反应。但是事实上，ECN 能够像经纪商和做市商一样通过分割策略有效地完成大订单，同时还能实现隐蔽性。很显然，买方和卖方交易员都希望订单在市场上能够匿名操作。在传统的交易中，公司的身份、规模和交易活动都被买方所选中的中介所知，而 ECN 是以匿名交易而著称的。ECN 能够只显示订单大小和价格，为交易员和投资者提供匿名和隐蔽性。当一个 ECN 能够发现一个内部的买卖匹配，就能立即发生交易。如果不能发现内部匹配，ECN 能够为提交订单者提供选择：是撤销限价单还是路由到其他市场。通常，ECN 提供的信息包括：

- 证券的标识(ID)
- 交易方向，即买单还是卖单
- 交易价格
- 交易日期
- 订单指令(比如，市价、限价还是交叉订单)
- 机构的类别
- 经纪商身份

总的来说，ECN 的优点包括以下几点：

• 自动化。当提交一个订单之后，交易在不存在人工干预的情况下根据价格和时间的优先级进行操作，而不像传统的市场一样，由做市商持有。

• 匿名操作。交易者的身份不公开，这对于一些交易者来说很重要。

• 较低的交易成本。ECN 对于市场上的订单收取大约每股 3 美分的费用，同时 ECN 还为提供流动性的订单支付一些费用。

• 交易速度很快，操作和确认都通过电子化进行，可以在数秒内完成。

• 通过 ECN，交易者可以实施程序化的复杂交易策略。例如，某只股票和指数价格发生特定变化时，进行交易。

7.5.3 交叉网络

交叉网络(crossing networks)是一种另类交易系统。在将交易订单发送到交易所或其他公开的市场(例如 ECN)之前，先在交叉网络内部进行撮合交易。由于交叉网络只公布交易价格，而不公布交易双方的身份和交易量，因此这使得机构投资者能够更有效地进行大宗交易，而产生较小的市场冲击。但是，由于交叉网络在进行交易的时候不向市场公开信息，所以它必须在市场上以最优买卖价格的中间价成交。一些知名的交叉网络包括：Liquidnet，Pipeline，ITG 的 Posit，以及高盛的 Sigma X。

交叉网络有三种不同的撮合交易的模型：

第一种模型是时间表交叉模型(scheduled crossing model)。Posit、Instinet Crossing 和 NASDAQ Open and Close 使用时间表交叉模型。在时间表交叉模型中，系统中订单对于市场参与者是匿名的，未进行匹配的订单可以取消，等待下次匹配，或者转发到其他实时的交易市场进行匹配。

第二种模型叫连续交叉模型(continuous crossing model)，全天候地提供流动性，以及进行交易协商。连续交叉模型向公众提供了更多信息，也更容易产生信息泄露。

第三种模型称为暗盒模型(Dark Box Model)，是连续交叉模型和时间表交叉模型的一种混合，公司在内部系统隐藏一些流动性，可以在不公开任何信息的情况下为交易双方提供价格的改进。最近几年，交叉网络不断地向市场渗透，市场份额不断上升。这些变化凸显了买方交易员对交易信息隐蔽性的需求的增加。这一类交易系统也被称为流动性暗池(dark pools of liquidity)。

7.5.4 直接入场技术

随着通信能力和技术的提高，交易订单能够通过网络直接到达市场，交易流程发生了极大的改变，这就是直接入场技术（DMA）。它不但使交易员能够通过低的交易成本进行市场操作，还淘汰了交易室里表现不好的交易员。随着 DMA 使用的增加，另类交易系统成为更好的交易选择。

订单智能路由的概念，是由客户基于各自的交易参数所确定的，例如价格、流动性、成本和速度。电子通信网络（ECN）使订单智能路由系统变得透明，并使他们能够执行订单路由，更多地利用网络进行交易操作。从 ECN 出现以后，直接入场技术在美国就成为交易技术中重要的一部分。通过高速的聚集工具，公司可以更快地得到更完整的市场交易信息。聚集技术为投资者创造了一个能在多个市场间交易的便利。聚集工具的发展产生了订单智能路由技术。订单智能路由技术能够分析订单和市场价格数据，然后为交易寻找最高效的市场，从而改进交易。随着更多的投资者和机构利用 ECN，聚集技术的优势也在逐渐增加。

DMA 为投资者提供了一个通过互联网访问电子交易所的直接和高效的方式。个人能够自主地制定交易策略，通过特定的目标市场进行交易（比如做市商、交易所和 ECN）。还有一些交易会继续依赖于人工的联系，但是可以通过即时通信技术或者选择可信赖的对手方来改善交易操作。

总的来说，DMA 技术的优点大致包括：

- 为投资者提供在第三方经纪商情况下得不到的交易速度和更好的价格。
- 尽可能改善流动性。
- 提供多种下单选择，使大宗交易更加隐蔽。
- 访问多种产品和多个市场。

7.6 复杂事件处理

现代的事件处理技术从 20 世纪 90 年代开始发展，加州理工学院的 Mani Chandy，剑桥大学的 John Bates 和斯坦福大学的 David Luckham 分别开始了独立的学术研究。这些研究的目标是开发一种处理事件流数据的新手段，用于识别复杂的事件序列，其中可能包括一些时间或范围的限制，然后根据这些

复杂的事件模式的结果做出响应。

事件是一个很广泛的概念，例如一次交易的完成，一架飞机的降落，或一个参数的输入等等。计算机系统内的事件被定义为状态的改变，例如当一个货物销售出去之后，状态从“待售”变为“已售出”。

简单事件指的是那些不是对其他事件抽象而成的事件。相对应的，复杂事件（或组合事件）就是对两个或更多事件抽象而得到的事件。那些被抽象的事件则被称为成员事件。事件之间存在着相互的联系，例如时间关系、因果关系、组合关系等等。这些关系将事件联系在一起形成了复杂事件，例如，1987年美国股市的大崩盘是由成千上万的简单事件抽象和联系而成，其中包括每个股票的交易情况、每个交易员的抛售等等。

随着越来越多的自动化交易系统的出现，复杂事件处理（complex event process）技术逐渐在算法交易领域取得了很大的成功。通过复杂事件处理技术，投资者可以很容易和清楚地将自己所制定的复杂的交易策略编写到计算机系统当中。计算机系统可以根据投资者的需求，通过分析一系列事件的集合，按照预先设定的规则做出交易决策。另外，金融机构可以将监管的规则以复杂事件的形式编制进入计算机系统，有效地防止违规和欺诈行为。

例如，算法交易中很重要的一个方面就是确定交易的时间和数量，这就需要交易系统不断地观察变化中的市场条件，发现交易机会。下面就是一个价差套利的例子：

```
IF
  MSFT price movesoutside 2% of MSFT-15-minute-VWAP
FOLLOWED-BY(
  S&P500 moving by 0.5%
  AND(
    IBM's price moves up by 5%
    OR
    MSFT's Price moves down by 2%
  )
)
ALL WITHIN
  Any 2 minute time period
THEN
  BUY MSFT
```

SELL IBM

例子中，交易系统对 IBM 和微软（代码：MSFT）股票进行价差套利。我们可以很容易地理解这段规则：如果在两分钟内微软股票价格高于或低于 15 分钟内 VWAP 价格的 2%，然后标准普尔指数移动了 0.5%，同时，IBM 的价格上涨 5%或者微软股票的价格下降了 2%，那么就买进微软股票，卖出 IBM 的股票。

复杂事件处理技术可以为机构投资者在算法交易系统的开发上带来许多的优势和便利。通过复杂事件处理技术，投资者可以很方便和容易地开发算法交易策略。它使交易策略更容易理解，从而提高算法开发的效率，而且方便修改和重新利用。

在金融市场上，监管机构可能会不断地提出新的监管要求，利用复杂事件处理系统会使算法交易系统更方便和灵活地适应变化的环境。利用复杂事件处理技术，算法交易系统的开发者可以更多地关注交易策略的层面，而不是计算机系统本身的开发。这样，不但可以减少系统开发中的错误和风险，而且能够为投资者节省人力，更多地关注能够带来附加值的业务。

第八章 交易成本分析

8.1 交易成本简介

传统的投资组合理论为我们在构建最优投资组合方面给出了很好的指导，但是实际交易操作中的交易成本却没有被考虑在内。一些研究表明，在美国，积极型的股票型基金的年收益率平均低于标准普尔 500 指数 1～2 个百分点，导致这一问题的一个主要原因被认为是交易成本。

当投资者构建投资组合或调整的时候，必然需要买进和卖出的交易操作，而这不可避免地会带来交易成本。哈佛大学的 Perold 在研究中发现，在 1965 年到 1986 年之间，通过股票价值排序可以构建一个能够每年击败市场大约 20%的账面投资组合。但是在实际中，这个投资组合只能够超过市场收益率大约 2.5%，这之间的差异很大程度上就来自交易成本。实际操作中，人们低估了控制交易操作成本的重要性，尤其是当机构投资者经常进行大额交易的时候，会导致相当大的执行落差。当市场收益率不足以弥补交易成本的时候，一个计划中可以盈利的投资组合甚至有可能产生负的收益。由此，不难看出交易成本控制在整个投资过程中的重要性。

以往，机构投资者都有专门的交易员负责交易操作的活动。但是近几年，算法交易实现了迅猛的发展。在欧美市场，算法交易已经成为主要的交易手段之一。算法交易的目标就是通过自动化的交易操作，以最优化的交易策略尽可能地减少交易当中的成本。因此，要了解算法交易，我们首先应该充分地了解交易成本的组成和原因。

8.1.1 什么是交易成本

1. 传统的交易成本

交易成本(transaction costs),又称交易费用,是在一定的社会关系中,人们自愿交往、彼此合作达成交易所需要支付的成本。交易成本是人与人之间关系产生的成本,与一般的生产成本有所不同。从本质上说,只要有人类交往互换活动,就会有交易成本,它是人类社会的交易活动中一个不可分割的组成部分。

交易成本理论是由诺贝尔经济学奖得主科斯(R. H. Coase)所提出,他在《企业的性质》一文中认为,交易成本是"通过价格机制组织生产的、最明显的成本,就是所有发现相对价格的成本"、"市场上发生的每一笔交易的谈判和签约的费用"及利用价格机制存在的其他方面的成本。

科斯提出,为了确定商品或服务的市场公平价格,人们需要了解所有围绕交易的相关问题:市场动态、供求关系、竞争市场上的价格,当然还包括交易成本。没有这些方面的信息,我们很难确定市场上的公允价格。

诺贝尔奖获得者威廉姆森(Oliver Williamson)在其研究中将交易成本简单地分为以下几项:

• 搜寻成本:搜集商品信息和交易对象的过程中产生的成本。

• 信息成本:取得交易对象信息与和交易对象进行信息交换所需的成本。

• 议价成本:针对契约、价格、品质进行讨价还价的成本。

• 决策成本:进行相关决策所需的内部成本。

• 监督成本:监督交易对象是否依照契约内容进行交易的成本,例如追踪产品、监督、验货等。

• 违约成本:违约时所需付出的事后成本。

威廉姆森还指出,在人性因素与交易环境因素交互影响下,有时会产生市场失灵现象,会造成交易困难和交易成本,这些因素大致包括:

• 有限理性:指参与交易的人由于身心、智能、情绪等限制,在追求效益极大化时所产生的限制约束。

• 投机主义:指参与交易的各方为寻求自我利益最大化而采取的欺诈手法,同时会增加彼此不信任与怀疑,进而导致交易过程监督成本的增加。

• 不确定性与复杂性:由于环境因素中充满不可预期性和各种变化,交易双方均将未来的不确定性及复杂性纳入契约中,使得交易过程增加不少制定

契约的议价成本，并使交易难度增加。

• 少数交易：某些交易过程有专属性，或因为异质性信息与资源无法流通，使得交易对象减少，并造成市场被少数人把持，使得市场运作失灵。

• 信息不对称：因为环境的不确定性和自利行为产生的机会主义，交易双方往往握有不同程度的信息，使得市场的先占者拥有较多的有利信息而获益。

• 气氛：指交易双方若互不信任，而且又处于对立立场，无法营造一个令人满意的交易关系，将导致交易过程过于重视形式，增加不必要的交易难度和成本。

上述交易成本的发生原因，进一步追根究底可发现源自交易本身的三项特征。这三项特征影响着交易成本的高低：

• 交易商品或资产的专属性：如果交易的一方投资的资产本身不具有市场流通性，或者契约一旦终止，投资于资产上的成本难以回收或转换使用用途，称之为资产的专属性。

• 交易不确定性：指的是交易过程中各种风险的发生几率。由于人类有限理性的限制使得面对未来的情况时，人们无法完全事先预测。同时，在交易过程中，买卖双方经常发生交易信息不对称的情形下，交易双方会通过契约来保障自身的利益。交易不确定性的升高会伴随着监督成本、议价成本的提升，使交易成本增加。

• 交易的频率：交易的频率越高，相对的管理成本与议价成本也会升高。在交易频率提高的情况下，企业可以将该交易内部化，以节省企业的交易成本。

2. 金融市场的交易成本

在金融市场上，交易成本表示在执行一个投资决策的过程中产生的成本。它包括为了促使交易进行所需要的成本，其中包括佣金、交易税、流动性成本和机会成本等等。例如，佣金是按交易份额支付给经纪商的费用；而流动性成本产生于投资者执行交易的时候，流动性的短缺会导致高于买单决策价格的价格升水，或者低于卖单决策价格的价格贴水。

交易成本出现在执行投资决策的过程中，它一方面使得投资的成本更高，另一方面还降低了投资组合的利润。交易成本使得基金经理和分析员在发现投资机会的过程中所做的许多努力被浪费掉，成为投资组合利润的损耗。许多基金的业绩低于市场水平的原因也来自交易成本的影响。因此，对交易成本进行全面和细致的分析，进而减少交易成本，对于提高投资回报以及改善市场效率的意义不言而喻。

8.1.2 交易成本的组成

交易成本产生于交易决策的执行过程之中。在经济学术语中，交易成本被定义为：由买方支付，但没有被卖方收到，或者是由卖方支付，但没有被买方收到的费用。在金融市场中，交易成本代表比决策价格差的那部分费用。

1. 简单分类

我们先对金融市场上的交易成本进行一些简单的分类，这有助于全面地理解和认识交易成本。交易成本可以分为可见（直接）部分和隐藏（间接）部分：

- 可见（透明）成本是由那些费用结构预先知道的，或可以很容易地通过实际市场数据获得的部分，可见成本只占交易成本的很小一部分。可见成本按要素划分为交易税、佣金等等。
- 隐性成本是费用结构中的那些不容易知道的，或者不容易从实际市场数据观察的部分。例如，市场冲击产生的成本不容易从市场数据观察到，因为市场冲击带来的价格变化只能通过对比下单和不下单的两种情况来得到，但是不可能同时得到这两种状态下的价格。因此，市场冲击成本很难进行估计。投资者必须利用统计推断或其他技术来估计交易成本结构和参数。隐性成本占总交易成本的很大一部分。但是，通过选择交易策略，我们可以改善交易操作的质量，进而提高投资价值。

另外，交易成本还可以分为固定部分和可变部分：

- 固定成本是交易成本中那些固定需要支付的部分。它是不依赖市场价格和交易策略的部分，不能通过调整实施策略来控制。固定成本只占交易成本的很小一部分。
- 可变成本是交易成本中那些由实际市场价格和交易策略所决定的部分。它们随着交易策略实际实施的不同而变化，投资者可以通过执行策略来有效控制可变成本。可变成本占总交易成本的相当大一部分。交易员可以通过控制可变成本来增加投资的价值。

2. 交易成本的详细划分

交易成本的划分不是一个新的概念，有很多学者进行过这一方面的研究。在表 8.1 中，我们把交易成本的九个组成部分按照固定成本和可变成本，以及可见成本和隐性成本进行划分。如表中所示，大部分交易成本是隐性成本和可变成本。这对投资者来说也有好的一面。由于交易成本大部分是可变的，

那么在执行过程中可以对它们进行控制，从而能够降低交易成本，得到更高的回报。负责交易成本管理的交易员在这一过程能够增加投资的价值。但同时，成本结构是未知的，因此投资者需要一些技术来估计这些成本，以便于更好地进行交易操作。

表 8.1

	固定成本	可变成本
可见成本	佣金、交易费用	买卖价差、交易税
隐性成本		延迟成本、价格增长、市场冲击、时间风险、机会成本

(1)佣金

佣金是经纪商为投资者代理买卖证券时按成交金额计算向其收取的费用。在国际上，一般是由证券管理部门或证券交易所确定一个统一的佣金比率，或是上下浮动界限。有时，佣金也会因为交易难度的不同而不同，容易进行的交易佣金比率低，而困难的交易佣金比率高。佣金是一种固定的、可见的交易成本要素。我国证券市场上 A 股交易的佣金规定不超过 0.3%。

(2)交易费用

交易费用是在执行过程中收取的，包括由交易所收取的经手费、清算结算费用、证券交易委员会所收取的会员费、监管费等等。和佣金不同，交易费用是由证券监管部门或证券交易所收取。但是通常情况下，投资者会将这部分费用归入交易经纪商收取的佣金。这些费用是固定的、可见的交易成本要素。

(3)交易税

交易税是对已实现利润或交易操作收取的税，例如资本利得、长期收益、分红、短期利润等等，按不同的税率征收。交易税是可见的和可变的交易成本要素，可见是因为税率是事前知道的，可变是因为执行价格影响交易成本。我国证券交易印花税从普通的印花税发展而来，是税务部门对证券市场上买卖、继承、赠与所书立的转让书据根据市场价格收取的，税率为双边 0.1%。

(4)买卖价差

买卖价差是证券交易中最低买入报价和最高卖出报价的差额。买卖价差用以补偿经纪商为撮合交易而承担的风险。买卖价差是可见的、可变的交易成本要素。可见是因为它们可以在每个时间点上测量，可变是因为买卖价差在一天中是变化的，而且还能在很大程度上被执行策略所影响。

（5）延迟成本

延迟成本表示经理做出投资决策时的价格和交易员下单时市场价格之间的差异。由于经理经常在价格上涨时买入股票，在价格下降时卖出股票，也就是“追涨杀跌”。因此，指令延迟会产生不利的成交价格，导致较高的成本。延迟成本常常产生于交易员下单时的犹豫。延迟成本是一种可变成本组成要素，因为它依赖于投资策略。立即下单的指令将产生相对较小的延迟成本，而由于投资者的犹豫产生的延迟成本会相当的高。通过互联网下单导致的网络延迟可以通过改善硬件设施进行改善。Plexus 集团估计每个决策的平均延迟成本是 62 个基点。昨日收盘价和今日开盘价的价差也会产生延迟成本。在这种情况下，延迟成本是不能被投资者控制的，这是一种不连续交易现象。如果不进行交易成本管理，这些成本会拖累业绩。

【例 8.1】一个基金经理发现一个价值被低估的股票，这只股票的当前市场价格为＄50，要求交易员买进 250 000 股。交易员收到指令后，开始寻找最适合进行交易的经纪商，然而，在交易员挑选经纪商和提交订单的过程中，股价上升到每股＄50.25。在这种情况下，交易员的犹豫让这只基金每股支付了＄0.25或 50 个基点的交易成本。

（6）价格增长

这里所提到的价格增长指的是股价的内在运动趋势。它指的是股价去除不确定因素后的市场表现。它所产生的交易成本表现为在上涨（下跌）的市场中买入股票的成本（盈余），或在下跌（上涨）市场卖出股票的成本（盈余）。价格增长带来的交易成本依赖于对股票走势的预期和执行策略。价格增长是一种不可见和可变的交易成本要素。

【例 8.2】一个经理决定买入 250 000 股股票，当前市场价为每股＄50，期望今年上涨 20%。因此，该股票可能每股上涨＄0.04 或 8 个基点。如果交易员在接下来的 5 天中每天买入 50 000 股，那么他的期望买入的平均价格是每股＄50.08。价格增长带来的交易成本是每股＄0.08 或 16 个基点。

（7）市场冲击

市场冲击指的是由交易引起的股价变动。市场冲击是最高的交易成本之一，而且经常引起不利的价格运动，从而会拖累投资业绩。市场冲击成本来自投资者的流动性需求和股票价格交易信息的泄漏。市场冲击使投资者需要支付升水来实现买单，或提供贴水来实现卖单。市场冲击成本依赖于交易的规模、股票的波动性、交易的方向、市场行情以及执行策略。市场冲击是不可见的和可变的交易成本要素。

【例 8.3】一个交易员接到了一个买入 10 000 股 ABC 股票的指令。然而,市场报价显示给出最低报价的只有 1 000 股,\$50.25 的价位有 2 000 股,\$50.50有 3 000 股、\$50.75 有 4 000 股。交易员只能在 \$50 成交 1 000 股,而其他 9 000 股则需要以更高的价格成交,平均成交价为每股 \$50.50。

为了吸引更多的流动性进入市场,交易员需要支付价格升水,因而导致了市场冲击成本。这个成本是由于交易员的流动性需求改变了市场原有的供给和需求条件。

一个交易员接到一个 250 000 股 XYZ 股票的买单。当这样一个大订单进入了市场,会给市场带来该股票价格被低估的信号。持有股票的投资者将不再愿意按当前的市场价格出售,同时其他投资者会希望买进该股票。这将会导致该股票价格永久的上升。交易信息泄漏引起的市场调整会产生持久的市场冲击。

(8)时间风险

时间风险成本是指交易成本的不确定性。它是由股价波动、市场行情和交易量的不确定等因素造成的。时间风险通常认为是股票价格的波动,这种定义是不完整的。执行成本的不确定性也依赖于实际市场交易量等因素。时间风险导致同样的交易对象和交易策略会产生不同的交易成本。时间风险是一种不可见的和可变的交易成本要素。

【例 8.4】股票的时间风险使得股价的运动具有不可预料性。如果一只股票当前交易价格为 \$50,我们可以有理由相信在接下来的几个小时里股票会在 \$49.50 和 \$50.50 之间进行交易。然而,两天后的股价将不一定在 \$49.50和 \$50.50 之间,更可能的价格区间是 \$48.00 到 \$52.00。当投资者在依次执行交易指令时,股价可能会升高或下降。假设一个交易员接到一个买进 100 000 股 ABC 股票的指令,并且决定消极地将订单在未来几天内执行。如果价格的运动对交易员有利,他将得到一个比预期更好的价格,反之将得到一个更差的价格。

(9)机会成本

机会成本是指交易决策不能够执行所损失的利润。它表示为不能完成订单产生的成本。导致机会成本的原因通常是市场流动性的不足,或价格变化太快。因为基金经理通常要买入上涨的股票或卖出下跌的股票,所以不能按时完成交易将有可能失去获利的机会。这对于投资来说是一种潜在的成本,进而会降低投资的回报。机会成本是一种不可见的和可变的交易成本要素。

【例 8.5】一个基金经理发现了一只被低估的股票，当前交易价格为每股 \$50，他要求交易员在当天买入 250 000 股该股票。为了减少市场冲击，交易员采用分割的方法执行这一订单，但是这导致在当天交易结束时只买入了 200 000 股。收盘时，股票价格为每股 \$51，且此时股票已经到达了公平价格，基金经理取消了剩下的交易。没有在较低价位完成剩下 50 000 股的交易导致了机会成本。

3. 交易成本的另一种划分

下面，我们从另外一个角度将交易成本分为如下三类：投资相关成本、买卖相关的成本、机会成本。

$$\text{总成本}=\underbrace{\left\langle\begin{array}{c}\text{交易税}\\\text{延迟成本}\end{array}\right\rangle}_{\text{投资相关}}+\underbrace{\left\langle\begin{array}{c}\text{佣金}\\\text{交易费用}\\\text{买卖价差}\\\text{价格上升}\\\text{市场冲击}\\\text{时间风险}\end{array}\right\rangle}_{\text{买卖相关}}+\underbrace{\left\langle\text{机会成本}\right\rangle}_{\text{机会成本}}$$

投资相关成本包括在投资决策阶段中可以管理的成本：交易税、延迟成本。交易税依赖于实际执行中的价格和策略，所以是可变的。交易税率是事前知道的，因而交易税又是可见的。但是如果将交易税合并到投资决策过程，会是一个很复杂的问题。延迟成本是不可见的、变化的交易成本。它不仅仅是一种对交易中犹豫不定和等待的惩罚，还包括网络延迟等客观因素。交易员和基金经理在实施某个交易决策的问题上花费的时间越长，价格运动使投资决策的成本更高的可能性就越大。交易员产生犹豫的一个原因是常常没有一个合适的分析法来决定恰当的交易策略，或者无法快速评价适合执行投资决策的经纪商。因此，交易员经常在如何执行、如何选择经纪商或交易地点上花费过多的时间。降低延迟成本最容易的方法是基金经理和交易员之间更多地互相交流，这样交易员可以更好地理解基金经理真正的投资目标。另外，交易员应该配备合适的交易决策执行工具和交易成本度量系统，以便于快速估计成本和决定交易策略。

交易成本的第二类是买卖相关的成本。买卖相关的成本包括了交易成本的大部分因素。这些成本是在执行投资决策的过程中产生的成本，它们可以通过合适的操作方法进行有效的管理。尽管这些成本无法消除，但是我们也可以通过恰当的交易策略大幅度来降低。这些成本中既包含固定成本，也包含可变成本；既包含可见成本，也包含不可见成本。其中的一些，例如佣金、交

易费用和买卖价差，可以被进一步归为交易服务的成本。它们是用来补偿经纪人撮合交易和清算结算成本的。不可见的、可变的交易成本组成要素包括价格增长、市场冲击和时间风险。交易员能够通过实际的交易策略来改善交易操作中的成本和风险。例如，价格增长和时间风险可以通过积极的执行策略来降低，但是积极的执行策略将产生较高的市场冲击成本，而消极的执行策略则产生较低的市场冲击成本。因此，交易员应当在这些要素之间进行权衡和取舍，从而提高投资的回报率。买卖相关的成本依赖于交易清单的安排、订单大小、股票的流动性、波动性、当天的市场环境、价格走势，以及交易执行策略等等许多方面因素。

交易成本的第三类是机会成本。机会成本的原因是交易员不愿在当前的市场价格上交易，或是由于市场流动性较差。降低机会成本的最佳方法是基金经理和交易员一起来决定市场是否容易在基金经理给出的价格区间内吸收订单，而这需要适当的交易前分析和对交易清单的成本估计。如果基金经理断定市场无法吸收足够的股票，他就可以改进交易清单，以便于更好地完成交易，进而降低了机会成本，提高投资回报。

最后，结合以上对于交易成本的划分，我们可以将交易成本表示为：

$$\text{总成本}=\underbrace{\underbrace{\left\langle\text{延迟成本}\right\rangle}_{\text{投资相关}}+\underbrace{\left\langle\begin{matrix}\text{价格评价}\\\text{市场冲击}\\\text{时间风险}\end{matrix}\right\rangle}_{\text{交易相关}}+\underbrace{\left\langle\text{延迟成本}\right\rangle}_{\text{机会成本}}}_{\text{不可见成本}}+\underbrace{\left\langle\begin{matrix}\text{交易税}\\\text{佣金}\\\text{交易费用}\\\text{买卖价差}\end{matrix}\right\rangle}_{\text{可见成本}}$$

这个公式对交易成本进行了细分，是度量和估计交易成本的基础。

8.1.3 交易成本产生的原因

1. 供求法则

隐性交易成本通常是由未来的买家或卖家施加给市场的买卖压力所造成的，这些成本在总交易成本中占的比例最高。如果不能进行正确的量化和控制，它们将会使一个好的交易计划变得无利可图。

隐性交易成本的作用可以由供求法则来进行解释。在经济学中，当需求方需要更多的商品或服务时，就需要抬高价格，在购买价上支付了一个升水来吸引更多的商品或劳务的供给进入市场。当供给方提供更多的商品或劳务时，供给过剩就使得供给方需要折价在市场上出售，以吸引更多的买者。

供求法则在短期内确定证券的市场价格的机制上起到了决定性的作用。当投资者需要买进更多股票的时候，他们就需要支付较高的价格，以吸引足够的卖方。商品或服务的均衡价格是由供给等于需求时的价格决定的(见图8.1)。图中，供给和需求在 E 点达到均衡。在均衡点，需求量和供给量等于 Q^*。当市场参与者希望卖出大量的股票，他们必须打折以吸引更多的买者进入市场，即 $P<P^*$。当市场的参与者想买入大量的股票，他们必须支付较高的价格以吸引更多的卖者进入市场，即 $P>P^*$。这就产生了交易成本。

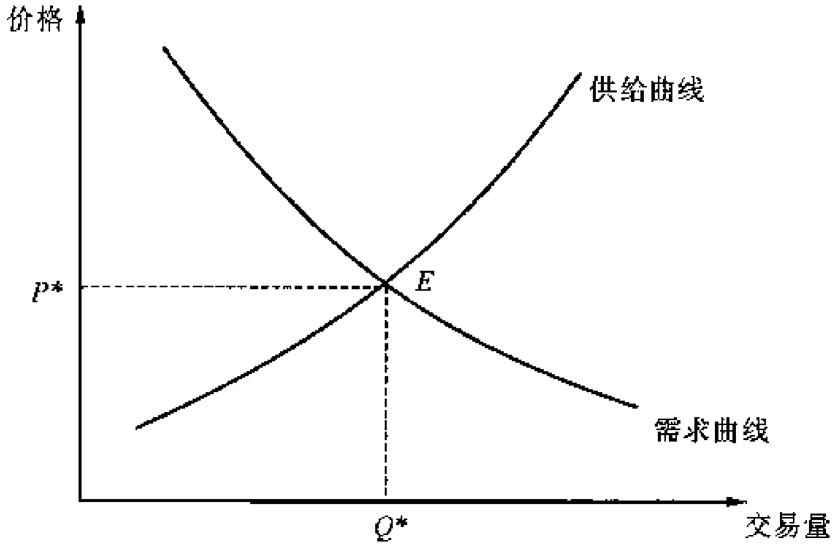


图 8.1 市场的供给和需求

2. 流动性需求者

流动性需求者是积极执行投资决策的投资者，往往是交易的发起者。需求指的是投资者进行交易操作的需求：用现金交换股票，或用股票交换现金。流动性需求者需要新的交易对手方进入市场，这样他们才能够完成交易。在大多数情况下，买方投资者是流动性需求者，因为他们需要执行一个投资决策。但是，流动性需求者的角色并不只限于买方投资者，卖方有时也会扮演流动性需求者的角色来平衡自身所持有的投资组合。例如，一个经纪商在一天内做空了某股票，使自身面临着大量的市场风险和潜在不利价格运动造成的风险敞口。为了扭转这种局面，券商需要买入该股票来抵消空头的风险，平衡投资组合。这样的操作可能使券商需要支付一定的升水。相反，如果券商做多的量太大，他们将需要卖出股票，以平衡投资组合，这样就需要提供价格贴水。在上面的两种情况下，券商都是积极地执行投资决策，并充当着流动性需求者。

3. 流动性提供者

流动性提供者、专营经纪商、做市商、券商和很多对冲基金作为市场的参与者希望以他们指定的价格或更优的价格在市场上进行交易。与流动性需求者有所不同的是，他们不需要制定特别的投资目标，而是通过对市场提供的流动性来收取升水，并获取交易利润，而且只在他们指定的目标价格或更优的价格进行交易。流动性提供者的角色并不局限在卖方，买方投资者也能扮演流动性提供者的角色，但是有着很大的风险。例如，一个有现金流入的投资者，他需要将这些现金投入到一些指定的指数基金。这个投资者只有当价格落在指定的范围内时他才会执行订单。那么在一段时期内，他很有可能扮演的就是流动性提供者的角色。市场价格也可能在这个时期内没有落在指定的价格范围内，那么该投资者可能迫不得已在不利的价格上进行交易。同时，尽管可能会观察不到，流动性提供者也会给市场带来的冲击。

8.1.4 投资周期中的交易成本

交易成本产生于投资决策的各个阶段，并影响着投资目标的实现。实行交易成本控制是基金经理和交易员的职责，并有助于防止投资组合的收益受到损失。较差的交易成本管理将会导致错误资产配置、低效率的股票和投资组合选择、不恰当的执行策略等等，进而产生较高的成本。为了进一步地了解交易成本，下面我们在投资过程的不同阶段来分析交易成本。

1. 资产分配

资产分配是将资金在众多的投资类别中进行分配，有助于分散风险和确定目标回报水平。传统的资产类别主要包括现金、债券和股票。而现在，投资于大宗商品、期货和衍生产品、不动产、私募股权投资基金和对冲基金等也很常见。但是除了资产选择，交易成本在投资组合的收益中起着重要的作用。不同类型的资产和交易方式产生的交易成本也是不同的。如果没有在制订资产分配的计划过程中考虑到交易成本，那么将会影响投资组合的回报和投资目标的实现。

2. 投资组合的建立

投资组合的建立阶段主要包括在每个投资类别中选择需要买卖的投资工具，并制定交易清单。在投资组合的建立过程中，基金经理经常需要在不同类别的股票之间选择，例如不同的市值、不同的板块，或是选择成长型股票和价值型股票。股票的选择基于它们不同的期望回报和相应的风险。如果投资决

策中没有考虑到交易成本，那么，由于股票回报估计的偏差将会降低投资组合的效率。交易成本在实现股票的投资回报的过程中起着非常重要的作用，所以应该在投资组合的建立阶段给以重视。例如，选择消极的大盘股策略和高换手率的小盘股策略为投资组合带来的交易成本是不同的。

3. 交易执行

选定具体的证券之后，投资者需要进行买进和卖出的交易操作，其中包括了做出关于如何、在何时、在何地买入和/或卖出股票的决策。交易员需要评估所有潜在交易选择，并决定交易执行的最佳方法。他们需要选择合适的交易策略、合适的交易方式（委托或代理）、合适的交易地点（传统的经纪人、电子通信网络或交叉盘系统）、执行交易的经纪商。因为这一阶段包含了订单的实际交易，而证券的买进和卖出过程中必然会产生交易成本，所以交易操作必然会对投资回报产生一定的影响。有时，交易成本的因素会使得一些看起来能够获取高回报的投资组合的表现低于预期。

4. 投资组合的归因分析

投资组合的归因分析包括衡量基金业绩表现，以及找出偏离（高于或低于）目标回报的具体原因。这有助于区分回报是由投资组合选择带来的还是纯凭运气。而对于交易操作的阶段来说，归因分析指的是度量交易成本和对交易员的表现的评估，区分交易员的出色表现和运气。对于交易表现的谨慎和准确的评估，有助于发现交易过程中的表现，进而改善交易执行的质量，提高整个投资过程的回报。

8.1.5 衡量交易成本的价格基准

计算交易成本通常需要确定一个基准价格，通过将交易价格和基准价格进行对比，才能够评价交易成本的高低。最常见的价格基准是加权的平均价格，以及交易前后的报价。下面是常用的三类基准价格：

- 交易开始前的某个价格，其中包括前一日收盘价，当天开盘价，交易前最后一个成交价，交易开始时的决策中间价等等。
- 交易当天的加权平均价格，例如 VWAP 和 TWAP。
- 交易完成后的某个价格，例如当天的收盘价。

其中，VWAP（volume weighted averaged price）是指交易量加权的平均价格。VWAP 的计算方法是将交易时段内成交的总的交易量 $\sum P_j v_j$ 除以

总的交易股票数目 $\sum v_j$ 。由于 VWAP 代表着市场上平均的交易价格，所以是应用最广泛的价格基准之一，是算法交易中使用的主要价格基准。如果一个买单价格低于 VWAP，那么交易被认为是好的；如果价格较高，交易就被认为是差的。VWAP 的计算公式如下：

$$\text{VWAP}=\frac{\sum v_j P_j}{\sum\limits_j v_j}$$

TWAP(time weighted averaged price)是时间加权的平均价格，是在不同交易时间点上交易价格的平均：

$$\text{TWAP}=\frac{1}{n}\sum P_j$$

8.2 交易成本分析

8.2.1 执行落差

哈佛大学教授 Perold 的执行落差法是一个普遍使用的度量交易成本的方法。执行落差(implementation shortfall)指的是账面资产组合和实际操作中的收益的差别。在账面资产组合中，投资者关注的是决策价格，而实际的收益来自实际操作当中获得的交易价格。执行落差衡量交易前的决策价格和实际中最终交易价格之间的差距，也就是以决策价格为基准来衡量交易成本。其中，决策价格是指投资者决定用于买进和卖出某只股票的价格。最常见的决策价格例如前一日的收盘价、订单的到达价格(arrival price)等等。

作为投资的目标，保持资产价值需要使用投资决策时的价格 P_d 作为交易的基准价格。因为这个价格是制定投资策略时所选择的价格，也就是说决策价格更加符合投资的目标。P_d 代表基金经理的投资模型(例如投资组合最优化、基础或技术分析系统等等)中所用的价格。使交易成本基准价格与用来做投资决策的价格保持一致是非常重要的，否则，将会导致交易价格和投资的决策理念不匹配。如果要尽可能地保持资产的价值，利用投资决策时的市场价格作为交易价格基准是绝对必要的。

账面组合代表投资决策中投资组合的理想状态。所有证券都以基准价格交易。交易成本、佣金、买卖价差、流动性影响、机会成本、市场趋势和价格偏差都不存在。

实际组合反映现实，所有证券在现实市场中交易得到的投资组合。交易成本、佣金、买卖价差、流动性影响、机会成本、市场趋势和价格偏差都需要考虑在内。

许多机构投资者在交易中非常关注于减少执行落差。下面，我们把这些不可见成本分为投资相关成本、买卖相关成本和机会成本：

$$TC = \underbrace{\underbrace{X(P_d - P_0)}_{\text{投资相关}} + \underbrace{\sum x_j P_j - \sum x_j P_0}_{\text{买卖相关}} + \underbrace{\left(X - \sum x_j\right)(P_n - P_0)}_{\text{机会成本}}}_{\text{不可见成本}} + VC$$

其中，X 表示需要交易股票的总数量，$X>0$ 表示买入，$X<0$ 表示卖出；

x_j 表示在第 j 个交易期中交易的股票数量；

$\sum x_j$ 表示交易股票的总数量；

P_d 表示决策价格；

P_0 表示到达价格，即订单投放到市场时的价格；

P_j 表示第 j 个交易期时的股票价格；

P_n 表示交易结束时的股票价格；

VC 表示可见成本。

其中，我们将 P_d，P_0 和 P_n 取为在相应时间段中买卖价差的中间值。因此，买卖价差包含在买卖相关成本之中。交易价格 P_j 代表实际的交易值。前面对执行落差形式的详细说明有助于我们区分可见成本和隐性成本。这种分类方法是理解和估计交易成本的基础。

投资相关成本是从基金经理做出投资决策到交易员开始交易这段时间资产价值的变化。它是需要交易的股票总数与相应价格变化的乘积，即 $X \cdot (P_d - P_0)$。在大多数情况下，基金经理在上涨时买入，或者在下跌时卖出。因此，与之相关的延迟就会成为基金的成本。延迟产生的成本是相当大的。有研究估计延迟成本会占到每次交易价值的 30 到 50 个基点。然而，如果交易员能够采取正确的交易方式，决策时刻 t_d 与订单投放时刻 t_0 的间隔会变小，那么延迟成本会更接近零，即当 $t_d - t_0 \to 0$ 时，$P_d - P_0 \to 0$。

买卖相关成本是交易的总金额与订单投放时的价格 P_0 产生的价值的差，即 $\sum x_j P_j - \sum x_j P_0$ 。由于订单投放价格 P_0 被估计为买卖价差的中间

值，买卖价差已包含到买卖相关成本的计算中，因而无需再单独列出。买卖相关成本占交易成本的很大一部分，但是如果交易成本管理得当的话，它可以被有效地降低。

机会成本是指由于无法完全实施投资决策所造成的利润损失，表现为未执行的股票数乘以交易结束时的价格 P_n 与订单投放时的市场价格 P_0 的差。这种度量方法指出了基金无法享受到的资产价值变化带来的收益。在整个交易执行的时间范围内（从 t_d 到 t_0），由未执行产生的交易成本分解如下：

$$
\begin{aligned}
OC &= \left(X-\sum x_i\right)(P_n-P_d) \\
&\quad - \underbrace{\left(X-\sum x_i\right)(P_d-P_0)}_{\text{投资过程中产生}} + \underbrace{\left(X-\sum x_i\right)(P_n-P_0)}_{\text{交易中产生}}
\end{aligned}
$$

因此，基金经理将机会成本分为投资决策中产生的成本和交易过程中产生的成本。通过正确的交易成本管理技术将机会成本有效地降低为零也是有可能的。这要求投资者做出决策与订单投放市场的间隔充分的小。交易当中产生的机会成本是因为流动性条件不足，或者是价格向不利的方向大幅的移动。但是，如果基金经理能够和交易员进行充分的沟通，他们就可以较好地评价流动性条件，并且估计执行成本，从而制订出适当的订单规模，以便于能够容易地被市场所吸收。在基金经理给出的价格范围内，未执行的股票数将有效地下降至零，即当 $X-\sum x_j \to 0$ 时，$OC \to 0$。

可见成本包括佣金、买卖价差、交易税和其他所有对投资者收取的费用（例如入场费、清算结算费用）。很多时候，经纪商将这些固定费用捆绑在一起作为由经纪商收取的佣金，投资者看不到这些成本。一般说来，这些可见成本只占交易总成本的最小的一部分，而且无法通过适当的交易成本管理方法进行控制。

在对交易成本根据执行落差分解之后，我们再转向隐性成本的几个组成部分：价格增长、时间风险和市场冲击。通过对这些成本的进一步分析，将有利于我们得出最优的交易执行策略。

8.2.2 价格增长

价格增长是指股票价格的自然变动。它表明在没有不确定性、意外的事情和噪音的情况下，市场中的价格如何变动。价格增长通常也被认为是价格变动趋势，有时也指股票价格自然的上涨。

交易成本依赖于价格变动、交易方式、交易环境、订单种类以及实际执行策略等等因素。如果一只股票的价格有上涨的趋势，那么交易员分多次买进股票会产生较高的成本，因为每次交易会在更高的价格上完成；而分多次卖出可以在较高的价格成交，从而降低成本或节省本金。因此，在这种情况下，买方需要使用积极的交易策略，而卖方则应该采用消极的交易策略。但是，由价格增长而产生的成本只是交易成本的一个组成部分，交易员还需要考虑市场冲击和时间风险的作用。买方大量买入的交易策略会带来很高的市场冲击成本，过度消极的交易策略会使卖方承受很大的时间风险，因为市场仍然可能会产生不利的价格移动。熟练的交易员可以通过交易策略显著地增加交易过程的价值。这些交易策略需要同时考虑价格增长、市场冲击和时间风险三方面的因素。

1. 自然增长

一些研究表明，美国股票市场存在着 7%到 8%的长期增长率，这可以被视为是股票价格自然的增长。股票自然增长的原因可以认为是经济增长的反映，或是企业的自身盈利的增长。通过一些计算，我们可以得到不同期限内股票自然增长的收益率。

如果持有某种资产 t 个时期，那么这 t 个时期内的收益率为：

$$1+R_t=\frac{P_{t+1}}{P_1}=\frac{P_{t+1}}{P_t}\times\frac{P_t}{P_{t-1}}\cdots\times\frac{P_2}{P_1}=\prod_{i=1}^{t}(1+R_i)$$

通过对数化，可以将乘法转换为加法进行运算：

$$R_t=\exp\left[\sum_{i=1}^{t}\ln(1+R_i)\right]-1$$

按一年 250 个交易日计算，若年增长率为 7%，则每个交易日内股票价格的自然增长率为 0.027%。

通过泰勒展开，我们可以得到上面式子的一个近似的计算：

$$R_t=\sum_{i=1}^{t}R_i$$

以这个方式计算，单个交易日内的自然增长率为 0.028%。

2. 基本面分析

除了股票价格的自然增长，交易员也可以通过其他方式来预测交易期限内股票价格的走势，进而更加有效地进行交易成本管理。常见的预测方法包

括基本面分析和技术分析。

基本面分析利用财务分析和经济学上的研究来评估企业价值或预测证券价值的走势。这些被分析的基本资料可以包含一家公司的财务报表和非财务上的信息，如商品需求成长性的预测、企业比较、宏观经济环境的变化。

支撑基本面分析的理论认为，以长期投资来赚钱，一个投资者必须专注于公司本身，而不只关注其股价。基本面分析相信一家公司的真实价值终会因市场力量而反映在股票价格上。但事实上，不论这市场是多么的有效，一些股票短期亦会有高估或低估其价格的现象。

基本面分析通常以三个步骤来分析：

• 总体经济的分析：通常包括国际和国内各经济指标，如 GDP 增长率、通货膨胀、利率、汇率、能源价格等等。

• 产业分析：包括整体产业的发展预期、销售量、价格标准、产品、国内外参与竞争的公司数量等等。

• 公司的财务分析：包括其销售量、价格、新产品、盈利预期等。

尽管基本面分析主要关注长期价格预测，但是当股票价格被高估或者低估时，也可以为短期价格的自然变动提供一个好的指标。

3. 技术分析

技术分析是指只利用历史交易数据（例如价格和交易量）来进行价格增长预测的方法。技术分析师不考虑公司的财务状况，他们相信所有相关信息都体现在这些数据中，并且未来的情况可以从过去中推断出来。技术分析非常依赖直方图、移动平均线、供求不平衡状况，以及一些指标（例如相对强弱指标等）来预测未来，主要是发掘出没有被发现的短期价格趋势。当然，很多技术分析也可用来预测长期价格趋势。

在技术分析所使用的各种方法和工具之中，价格图表是最主要的一种。技术分析师会特别去寻找某些特征的图形，例如著名的头肩反转形态，以及研究此类型的指标，如价格、成交量与价格的移动平均等。另外，许多的技术分析师会根据投资者心理的指标来操作。技术分析有许多个学派，例如 K 线、道氏理论和艾略特波浪理论。许多交易者会同时使用一个学派以上的理论来做分析。技术分析师使用经由经验所得到的判断依据来决定特定股票在一定时间内所显示的图样，以及图样的解释。

技术分析经常会和基本面分析相冲突。技术分析表示，价格在投资者认知到它们之前就已经反映了所有的影响因素，因此只需研究价格本身的作用。有些交易者只使用技术分析或只使用基本面分析，而有些人则同时使用这两

者来决定如何交易。技术分析相信市场存在所谓的价格趋势。技术分析常常预测市场会“上涨”、“下跌”或“盘整”。

技术分析相信,投资者会集体地重复他们之前的投资者的行为。对技术分析来说,市场上的情绪也许是不够理性的,但它们确实存在。因为投资者通常会重复他们的行为,所以技术分析相信可以找到一些可预测的价格图形,并绘制出一个图表来。技术分析有时也不一定局限在图表中,也不仅是只关心价格的趋势而已。例如,许多的技术分析师会监看投资者心理的报告。这些报告会估测出市场参与者的态度,尤其是看空或看多。

8.2.3 时间风险

时间风险是指对于一个特定的交易策略来说,对交易成本估计预测的不确定性,其中包括价格波动和市场交易量等因素的变动。这些不确定性导致对同一只股票的一样的交易会得到不同的价格。

首先,在执行交易清单的过程中,由于市场自身波动、经济层面的变化、新闻公告、价格模式变动、随机噪音等等因素的影响,价格通常会有一定的波动。因此,当前价格常常高于或者低于预期价格,这就是所谓的价格波动或者交易指令的价格风险。其次是市场交易量和当日交易模式的不确定性。市场活动常常高于或者低于预期水平,当投资者在交易量较大的情况下进行交易时,交易指令就会引起较小程度的供需不平衡,从而引起较低的市场冲击。当投资者在交易量较小的条件下交易时,交易指令就会引起很大程度的供需不平衡,从而引起较高的市场冲击。由于市场容量未知引起交易中的不确定性被称为交易指令的流动性风险。最后,来自其他市场参与者的未预料到的买入和卖出压力也会影响交易成本。不过,这种价格变动很大程度会包含在价格波动里,没有必要把它作为时间风险的独立组成部分进行考虑。

股票收益的波动性是股票价格行为的一个重要特征。预测波动率的三个主要目的就是用于风险管理、资产配置,以及针对波动性进行交易。对于交易成本管理来说,预测股票价格的波动同样重要。价格冲击使交易员将大的订单分割,更加缓慢地进行交易,而未来价格的不确定性促使交易员更加积极和快速地进行交易。下面我们介绍一下如何计算一个交易策略的价格波动的风险。

1. 单只股票交易中的价格波动

计算一个交易策略的价格风险就等于计算一个股票头寸从本期到下一期价格变动的风险。如果 h 表示头寸中包含的股票数量,σ^2 表示股票价格方

差，单位是美元/股。头寸 n 期总风险以美元计的 $\sigma(x)$ 的计算如下：

$$\sigma^2(n\text{期})=\underbrace{h^2\sigma^2+h^2\sigma^2+\cdots+h^2\sigma^2}_{n\text{期}}=nh^2\sigma^2$$

所以，

$$\sigma(h)=\sqrt{n}\cdot h\sigma$$

n 是交易时期数目，从上式可以看出随时间变化的波动幅度。现在假设资产组合中股票数量 h_j 因为追加投资或者股票赎回随时间的变化而变化。如果 h_j 表示资产组合中第 j 期所持有某股票的数量，则以美元计资产组合 n 期的总风险如下计算：

$$\sigma^2(h)=h_1^2\sigma^2+h_2^2\sigma^2+\cdots+h_n^2\sigma^2=\sum_{j=1}^{n}h_j^2\sigma^2 \tag{8.1}$$

所以一只股票价格变动的风险为：

$$\sigma(h)=\sqrt{\sum_{j=1}^{n}h_j^2\sigma^2}$$

2. 单只股票交易策略的价格风险

我们可以将交易安排看成随时间变化的头寸变化。第 j 期的头寸 h_j 可以看作是在本期期初未被执行的股票数量，x_j 表示第 j 期交易的股票数量。

$$h_j=\sum_{k=j}^{n}x_k$$

如果 σ^2 表示每期的方差，单位为美元/股，则交易策略的价格风险可以根据(8.1)式计算，如下：

$$\sigma^2(h)=h_1^2\sigma^2+h_2^2\sigma^2+\cdots+h_n^2\sigma^2=\sum_{j=1}^{n}h_j^2\sigma^2$$

$$\sigma(h)=\sqrt{\sum_{j=1}^{N}h_j^2\sigma^2}$$

3. 多只股票交易的价格风险

以上所述的计算方法在股票组合中同样适用。例如，如果 $\boldsymbol{h}$ 是代表股票头寸的列向量，其中 h_i 是第 i 只股票的数量，$\boldsymbol{C}$ 是每期价格风险协方差矩阵，单位为美元/股，那么资产组合的总风险 $\sigma_P(h)$ 计算如下：

$$\sigma_P(\boldsymbol{h})=\sqrt{\boldsymbol{h}^T\boldsymbol{C}\boldsymbol{h}}$$

n 期内资产组合的风险是：

$$\sigma_P^2(\boldsymbol{h})=\underbrace{\boldsymbol{h}^T\boldsymbol{Ch}+\boldsymbol{h}^T\boldsymbol{Ch}+\cdots+\boldsymbol{h}^T\boldsymbol{Ch}}_{n期}=n\cdot\boldsymbol{h}^T\boldsymbol{Ch}$$

$$\sigma_P(\boldsymbol{h})=\sqrt{n\cdot\boldsymbol{h}^T\boldsymbol{Ch}}$$

现在假设每只股票的数量随时间变化而变化。令 $\boldsymbol{h}_k$ 表示列向量，从而 h_{ik} 表示第 k 期第 i 只股票的数量，也就是：

$$\boldsymbol{h}_k=\begin{pmatrix}h_{1k}\\h_{2k}\\\vdots\\h_{mk}\end{pmatrix}$$

则 n 期内，这个交易中的资产组合的总标准差 $\sigma_P(\boldsymbol{h})$ 可以从下面式子得到：

$$\sigma_P^2(\boldsymbol{h})=\boldsymbol{h}_1^T\boldsymbol{Ch}_1+\boldsymbol{h}_2^T\boldsymbol{Ch}_2+\cdots+\boldsymbol{h}_n^T\boldsymbol{Ch}_n=\sum_{j=1}^{n}\boldsymbol{h}_j^T\boldsymbol{Ch}_j$$

4. 交易策略的价格风险

一组股票的交易清单的价格风险可以模仿资产组合交易的价格风险求出。假定一个交易者按照一个特定交易策略执行投资决策。$\boldsymbol{x}_k$ 是一个列向量，其中 x_{ik} 表示时刻 k 交易的第 i 只股票的数量。也就是：

$$\boldsymbol{x}_k=\begin{pmatrix}x_{1k}\\x_{2k}\\\vdots\\x_{mk}\end{pmatrix}$$

这样，交易策略的价格风险可以通过股票交易列表中剩余股票数量求出。令 $\boldsymbol{h}_k$ 是剩余待交易股票数量的列向量，其中 h_{ik} 表示第 k 期第 i 只股票中未交易的股票数。

$$h_{ik}=\sum_{j=k}^{n}x_{ij}$$

那么，与特定交易策略 $\boldsymbol{x}_k$ 相关的总价格风险 $\sigma(x_k)$ 可以如下计算：

$$\sigma^2(x_k)=\boldsymbol{h}_1^T\boldsymbol{Ch}_1+\boldsymbol{h}_2^T\boldsymbol{Ch}_2+\cdots+\boldsymbol{h}_n^T\boldsymbol{Ch}_n=\sum_{j=1}^{n}\boldsymbol{h}_j^T\boldsymbol{Ch}_j$$

$$\sigma(\boldsymbol{x}_k)=\sqrt{\sum_{j=1}^{n}\boldsymbol{h}_j^T\boldsymbol{Ch}_j}$$

一个交易策略 x_k 的价格风险是：

$$R(x_k) = \sqrt{h_j^T C h_j}$$

8.2.4 市场冲击

市场冲击是指由于某一特定交易指令所引起的股票价格变化。它是指有交易指令时的价格和没有交易指令下达市场时的价格之间产生的差异。市场冲击是一个难以估计的事后指标，因为它要求同时关注两种不能同时存在的情形。人们不可能记录下有指令时的价格变动，并记录下没有指令时的价格变动。如果这可能的话，一个指令的市场冲击就可以简单地通过两种情形下的价格差异求出。由于不能同时测量这两种情形，市场冲击就被比作金融的海森堡不可测理论。

市场冲击仍然是一个模糊的概念，通常的做法就是只考虑能引起市场冲击的因素。首先，市场冲击与交易指令的规模呈正向相关，因为大的订单更容易引起市场供需的不平衡。交易指令的规模越大，它所引起的不平衡就越大，从而交易成本就越高。其次，市场冲击和价格的波动性有关。股价波动越大的股票就具有越高的价格弹性和越高的交易成本。再次，市场冲击和交易的速度有关，交易的速度越快则交易的成本越大，反之则交易成本越小。最后，市场冲击成本还取决于交易期内所具有的市场条件。例如，市场流动性越高则交易成本越低，市场流动性越低则交易成本越高。市场冲击可以由下式决定：

$$K = f(\text{订单大小},\text{波动性},\text{交易速率},\text{市场条件})$$

市场冲击通常给投资者带来成本，创造不利的交易环境。为了完成交易指令，买方需要支付额外的费用，而卖方则会损失一部分收益。市场冲击可以被认为是为了得到额外的市场流动性，以使交易指令可以在合适的时候完成而提供的激励措施。没有这种激励措施，投资者就必须等其他指令的自然出现。然而，新的流动性的到来很可能会在更不利的价格上，或是危险的头寸上。所以，这种“附加优惠条件”可以看成是投资者为了激励其他市场参与者在某个时期某个价格交易提供的让步，否则，他们是不会在这个时期和价格上交易的。

1. 引起市场冲击的因素

市场冲击成本来自两个主要的原因：流动性需求和信息泄露。流动性需

求使得市场在原有均衡价格上供求条件不再平衡。这就要求投资者为了完成指令需要支付一个溢价来吸引额外的流动性。信息泄露是将投资计划或者投资者的交易动机泄露到市场上，导致市场价格更高。

(1)供需条件不平衡

在某种程度上有效市场假说是成立的，或者是基本成立的。市场不断地调整价格以促使买方需求等于卖方需求。因此，如果市场处于均衡状态，任何其他指令的出现都会引起在供求平衡状态下的不平衡。如果他们试图卖出股票，则必须提供一个较低价格去吸引市场上的其他买方。买卖指令伴随着股票价格的增长和下降。这两种情况都会给投资者带来较高的成本。这个成本依赖于交易指令的规模或者不平衡程度、股票价格弹性，以及交易方向(买入指令和卖出指令会产生的成本可能是不同的)。供需不平衡带来的市场冲击成本可以通过传统的需求供给曲线来描述。每次投资者带着指令进入市场总是想从市场得到股票或卖出股票。投资者的交易企图就是流动性需求的实现。也就是说，投资者要么是把现金换成股票，要么是把股票兑换成现金。两种情形下，他们都需要对手方为了相反的交易方进入市场。这也是股票市场价格产生的过程。

(2)订单簿

市场冲击成本也可以从投资者的即时交易需求方面解释。投资者的交易指令规模经常会大于市场最佳报价上可提供的数量。因而，为了尽快完成交易，他们经常要跨过限价订单簿的市场价格，使得交易一次比一次昂贵。例如，一个投资者现在有 2 000 股的买入指令，这时最佳买入报价和卖出报价分别是 49.90 美元和 50.10 美元。然而，在最佳的报价上只有 500 股股票。即时交易的需要将可能使投资者在下一个更高的价格上交易，可能是每股50.10 美元买入 500 股股票，每股 50.20 美元买入 1 000 股股票。由于市场价格被定义为买卖的中间价，所以在最优买卖价交易的股票成本等于买卖价差的一半。只要交易指令的规模小于最优报价上交易数量，那么买卖价差的 1/2 可以被认为是冲击成本。然而在很多情况下，指令规模比累计指令簿中最优报价上股票数要多，因而，如果投资者要立即完成交易指令，就需要支付更高的价格。

(3)信息泄露

每当交易指令下达到市场的时候，就相当于传达了关于某些股票价格的信息，从而导致交易变得更加昂贵。人们会相信某些投资者利用私人信息做出买卖交易决策，但是这些信息在市场上是不可得的。具有代表性的

是股票价格低估或者高估的信息。但是，一旦市场得知这些信息并且知道股票被错误估值，股价就会迅速做出调整来更准确地反映股票的市场价值。高明的投资者能够很好地掩饰自己的交易意图，从而降低市场冲击成本。然而，投资者通常并不是基于私人信息做出投资决策，或者他们不是真正认为市场定价错误。那些流动性需求者，例如指数投资者，通常会增持特定数量的特定股票。这些投资者没有向市场传达任何投资信息。但是有时候市场很难分辨流动性交易者和信息交易者。因此，所有交易指令都至少部分被认为是基于信息做出的交易决策，从而引起市场调整股票价格，从而使交易更加昂贵。

另外，关于投资者交易意图的信息泄露也会提高交易成本。一些投资者可以通过研究来估计其他投资者交易序列中的股票种类，利用这些结果来模仿他人的买卖行为，从而成为其他投资者研究成果的搭便车者。他们可以买入或卖出被低估或被高估的股票，而不用花费自己任何的投资研究的资源。这就会制造出更大的买入或者卖出压力，引起股价更快变动，这会降低所发现的好的投资机会的价值。

交易指令还会向市场传达关于投资者交易规模和迫切程度的信息。因而，本来在当前价格上愿意交易的参与者如果相信投资者必须完成这个交易，他们就会要求流动性需求者支付一个溢价来完成本次交易。另外，那些本来可能对这些股票没有兴趣的参与者会很快积累这只股票的头寸，期望日后高价抛出股票卖给该投资者。

2. 暂时冲击和永久冲击

每当交易指令在市场上披露，通常会包含一些信息。大多数情况下，这些信息只涉及交易的即时性和投资者的流动性需求。然而，有些信息会导致对股票内在价值的观点变化。如果传达到市场上的信息只涉及交易的即时性和投资者流动性需求，那么市场冲击带来的价格变动只会在短期内存在。但是，如果交易信息涉及股票未来价值或者宏观经济环境的变化，那么这种偏离会持续很长时间，甚至可能是永远的。这两种价格偏离分别产生了暂时的和永久的市场冲击。就像其名字所蕴含的意义一样，暂时市场冲击是证券价格轨道的短期偏离，长久市场冲击则是股票价格轨道长期，甚至永久的偏离。

(1)暂时市场冲击

每当交易指令在市场上披露，如果没有改变市场长期观点，或者当前价值的基本面信息时，产生的市场冲击就是暂时的。暂时市场冲击通常是短期的。

即时性要求和流动性需求引起短暂的市场冲击效应。在这种情况下，买入指令引起价格的短暂上升，卖出指令引起价格的短暂下降。接着，价格会向初始的价格轨道回归。问题是偏离的时间在微观层面上很难确定，并可能会随着对投资者交易指令规模的认识而变化。

对短暂市场冲击最常见的描述是市场冲击泡沫。投资者的指令引发股票价格上升，使得投资者在更高的价格买入。交易执行后，股票价格将会下跌。投资者在最高价格上，即泡沫的顶端买入股票。

用数学语言描述，短暂市场冲击是指令披露到市场上时价格的短期偏离，紧接着价格会回归到最初的期望价格轨道。令整个偏离过程持续时间 t，准确的时间主要受到所泄露信息的影响。市场上的投资者会推测指令的期望规模和投资者的交易需求。因此，对于买入指令，有交易指令时的价格轨道 $f(x)$（图 8.2 中实线部分）将会高于没有交易指令时的价格轨道 $g(x)$（虚线部分）。一段时间 t 后，两个价格轨道将趋于相同：

$$f(x)=\begin{cases} =g(x) & x<s \\ >g(x) & s\leqslant x\leqslant t \\ =g(x) & x>t \end{cases}$$

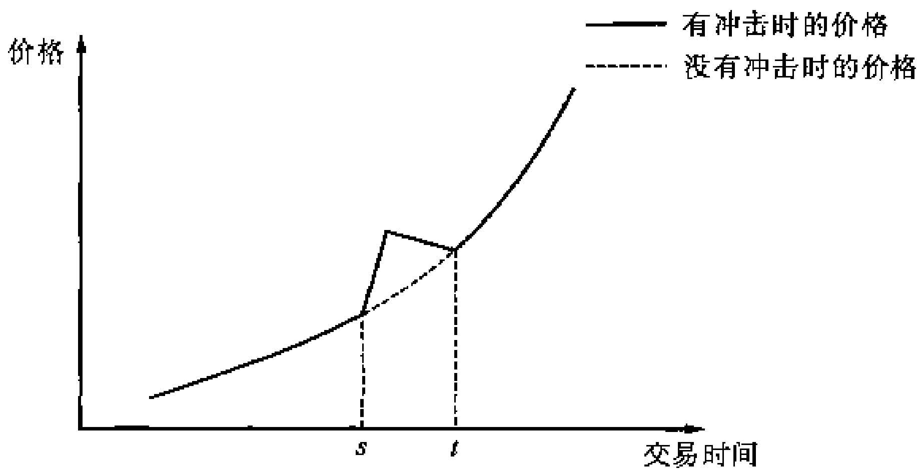


图 8.2 暂时冲击

图 8.2 当中描述了短暂市场冲击的效用。该图说明了买入指令是如何在一个短暂时期内引起价格轨道偏离它的正常轨道达到更高状态的。注意到价格轨道会向正常情况下的预期轨道回归，但是偏离的时间是未知的。价格轨道回归到它原本期望轨道所用的时间被描述成市场冲击的半衰期。它依赖于和交易相关信息的内容，并且随着不同的投资者和市场环境而不同。

例如，一个交易员用一天时间执行一个大的交易指令，相对于一个偏爱即时交易并在一个很短的时间里大量交易的积极型交易员来说，会产生较长时间的暂时市场冲击。即使他们以同样的方式执行相同的指令，也会出现这种情况。市场可能认为消极的基金经理人可能想要购买某种股票，而且他们的意图还没有披露到市场中。所以，一些潜在的卖家都不愿意在当前价格上交易，他们期望消极基金经理人会在更高的价格上交易。匿名交易则可以在很大程度上减小短暂市场冲击成本。

另外，最近很多研究检验了单个交易的市场冲击。其中，研究者发现市场冲击是交易订单规模的凹函数。这也就是说，对于小额的交易来说，冲击成本随着交易规模的变大增加较快；而对于大额交易来说，冲击成本随着交易规模增加较慢。具体的函数形式随市场的不同和时间的不同而不同。Lillo 等人在研究纽约证券交易所的 1 000 只股票的交易和报价数据后，认为市场冲击函数服从指数定律，即如下公式：

$$E(TI)=\frac{\epsilon v^{\alpha}}{\lambda}$$

其中，对于小规模交易，参数 φ 大约为 0.5，而对于大规模交易大约为 0.2。Farmer 认为其中的原因可能是投资者自身根据市场条件的选择。当市场条件不好的时候，投资者会将大订单进行分割，以减少市场冲击；而市场条件好的时候，投资者才有可能交易大的订单。因此，市场冲击随订单规模增长的速度是递减的，也就是上面所提到的凹函数的形式。

(2)永久市场冲击

永久市场冲击是由交易指令引起的长期价格变化。如果交易指令使得市场相信股票的未来价格和最初期望价格不同，或者股票的内在价值发生了变化，就会引起永久冲击。这使得股价迅速做出调整，使得价格快速上升或者快速下降。当指令包含的信息使市场对公司的长期增长潜力认识发生改变时，这种情况就会发生。这就表明股票被错误定价，并且存在套利机会。因此，市场参与者迅速把股票价格调整到新的公平价格。这些自然会影响股票价格的预测，并发现新的价格增长路径。

假设一个交易员开始积累或者卖出一个股票头寸，并多次达到预期效果，并获得巨大的利润。当市场注意到这个交易员买卖股票的行为时，市场就会重新评估股票价值。其他本来对这只股票不感兴趣的参与者就会开始和这个交易员一样建立头寸，并期望获得一部分利润。本来打算在当前价格迅速交易的投资者就会发现市场价格没有合适地反映股票内在价值，他

们就会要求更高的卖出价或更低的买入价。这就会形成一个与股价的自然增长路径相偏离的价格路径，引起一个永久性的价格偏离。换句话说，就是股票价格路径的一个偏离，使价格不会再向原有的路径回归，产生永久性的市场冲击。

如果股票的相对表现，或者作为一种投资工具相比其他可选择的投资机会（例如大宗商品和债券）的吸引力发生变化时，永久市场冲击也可能发生。因而，投资者即使不进行交易也可能面临永久市场冲击。

用数学语言描述，永久市场冲击可以被定义为指令披露时刻 t 之后股票价格类似于图 8.3 的偏离，有指令时的价格轨道 $f(x)$（实线部分）通常高于没有指令时期望价格轨道 $g(x)$（虚线部分）。冲击产生后，股票价格不会再回归到起初的自然轨道上来。也就是：

对于所有的 $x>t$，$f(x)>g(x)$。

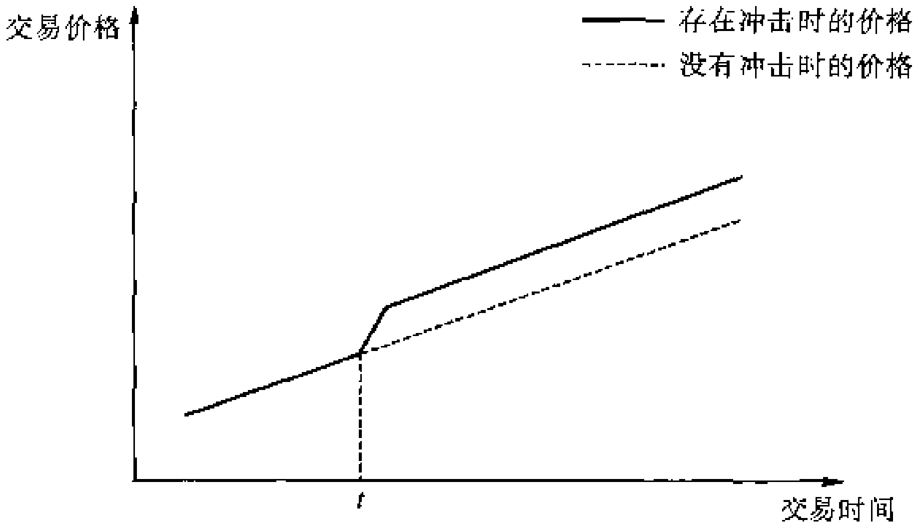


图 8.3　永久市场冲击

图 8.3 中说明永久市场冲击对证券价格轨道的作用。指令披露到市场后价格轨道就偏离了起初的期望轨道，然后价格就沿着新的路径变动，从而达到一个新的长期价格。图中所示的价格轨道和起初的路径平行，这并不是普遍的情形，新的轨道可以和起初的不同。例如，可能以更快的速度增长。

对于单个交易订单来说，如果将永久性冲击成本表示为交易规模的函数，那么对于函数形式的选择，Huberman 和 Stanzl 的研究证明，只有线性的永久冲击函数才能排除操纵价格的套利行为，所以永久性市场冲击通常选用线性函数的形式。

8.2.5 机会成本

机会成本是指由于没有能够完全执行一项投资决定而损失或未能实现的利润。市场流动性的不足导致了机会成本的存在，这种流动性的不足使交易指令不能够顺利完成，或使股价波动超出了目标交易的范围。机会成本等于交易期内股价的变动和未执行(残留)交易股票数量的乘积。机会成本可以通过执行落差法来衡量的，包括与投资相关的机会成本和与交易相关的机会成本两部分，其构成如下：

$$
\begin{aligned}
OC(R) &= \underbrace{(X-\sum X_j)}_{\text{未执行股票数}}\underbrace{(P_n-P_d)}_{\text{股价变动}} \\
&= \underbrace{(X-\sum X_j)(P_0-P_d)}_{\text{投资过程中产生}}+\underbrace{(X-\sum X_j)(P_n-P_0)}_{\text{交易当中产生}}
\end{aligned}
$$

其中，OC 为机会成本

P_d＝作出投资决定时的股价

P_0＝交易开始时的股价

P_n＝交易结束时的股价

投资过程中产生的机会成本是由于交易延迟所造成的。当基金经理向交易员发出指令时犹豫不决，或者交易员未及时向市场下单时，便会产生这类机会成本。通过减少延迟时间和进行适当的交易成本管理，可以很容易地减少这项成本。但是，由于收盘到开盘时产生的价格跳跃所导致的机会成本和延迟成本是无法避免的，这是非连续交易带来的结果。交易当中产生的机会成本是指未能够及时执行交易指令而产生的实际发生的成本。因此，我们将与交易相关的机会成本定义为所有未交易股票数量与价格变动的乘积：

$$(X-\sum X_j)(P_n-P_0)$$

机会成本既可以是成本也可以是收益。首先，我们必须知道每一笔交易都会对市场造成永久的和暂时的冲击。永久的市场冲击所产生的成本会在交易后长期存在。这会导致接下来的交易期内的买价更高，卖价更低。其次，在某种意义上来说，交易者在购买价格上升中的股票，或者卖出价格下降中的股票的过程中会使自己最终的成交价格更加不利。我们知道，机会成本在某种程度上是由于价格的反向变动和投资者试图对不利交易价格的规避带来的结

果。因此，如果买入价格过高或者卖出价格过低，投资者会取消对未交易股票所发出的指令。那么，这就会导致不能够完成交易。当股价更具吸引力的时候交易者很少会取消交易指令，因此，取消交易指令更多的情况下带来的是交易成本，而不是收益。

交易成本管理(TCM)的目的是尽可能地减少与交易相关的机会成本。我们通过对交易成本的估计来决定在一定市场条件下，以及在一定价格范围内指令执行的可能性。交易前的分析工作会提醒交易经理们市场流动性不足的情况，以及与之相对应的指令执行难度。交易成本的估计可以帮助我们预测交易成本和风险，不论指令是否能在一定价格范围内执行。在交易前拥有这些信息可以使公司的经理和交易员之间进行有效的交流，使得一些无法完全执行的指令能及时被修正，并投资于下个更具吸引力的机会中。最后，通过对实时市场信息(以及交易进程)的监控，我们也可以预期到不利的交易条件。

如果一个交易者接到一个购买 100 000 股 ABC 公司股票的交易指令，但是在交易时间结束时只能完成 80 000 股的交易指令。如果公司经理的决策价格是 $P_d=\$20$，交易员在 $P_0=\$20.50$ 的价格上开始交易，当天的收盘价是 $P_n=\$21.50$。我们可以将机会成本分解为与投资相关的机会成本和与交易相关的机会成本两部分：

$$
\begin{aligned}
OC(\text{与投资有关的}) &= (100\ 000-80\ 000)\times(\$20.50-\$20.00)\\
&= 20\ 000\times\$0.50\\
&= \$10\ 000\\
OC(\text{与交易有关的}) &= (100\ 000-80\ 000)\times(\$21.50-\$20.50)\\
&= 20\ 000\times\$1.00\\
&= \$20\ 000\\
\text{总的 } OC &= \$10\ 000+\$20\ 000=\$30\ 000
\end{aligned}
$$

如果一个基金经理发现了两只被低估的股票。第一只股票是 ABC 公司，交易价格是 \$25，并预期它会涨 8%，达到 \$27。第二只股票是 XYZ 公司，交易价格是 \$50，并预期它会涨 7%，达到 \$53。这个经理决定投资 \$1 000 000 来买 40 000 股的 ABC。但是在交易前，交易员通过对这个交易清单的分析之后认为，只有 70% 的可能性能以优于 \$27 的价格执行这一指令，并且提醒了基金经理。基金经理就将前面的交易指令修改为购买 20 000 股的 ABC 和 10 000股的 XYZ，以便使交易能够以 95% 的可能性在既定价格范围内执行每个指令。这个例子显示了，通过用低收益的股票置换高收益的股票可以提高指令执行完成的可能性，从而提高整个投资组合的实际回报率。

8.3 交易成本管理

8.3.1 最优交易操作

在执行交易清单的过程中，交易员需要在互相冲突的目标之间取得平衡。如果交易员执行得太过积极，也就是交易速率过快，将会带来很高的市场冲击成本。相反，如果交易员执行交易过于消极，会将交易指令暴露在很大时间风险中，未来可能出现的不利的价格运动将导致更高的交易成本。交易员应当在风险和成本之间权衡取舍，并决定一个与基金经理投资目标一致的最优点。这种目标的冲突被称为是“交易员的困境”。

最优交易操作对不同的人有着不同的意义。但是，它可以根据价格、时间、规模等不同因素进行归类。例如，价值型和消极的投资者关心价格的运动和资产的保值，成长型和动量型的投资者要求立刻得到回报，而大宗交易员和大型的共同基金则更加需要流动性或扩大规模等等。

最优交易操作是指在投资周期的各个阶段管理交易成本，以保证投资组合实现最高回报的过程。在执行过程中，需要估计成本和评价策略来确定交易策略。然后，它还需要交易员和经纪商不断地根据变化中的市场条件调整交易策略，从而保证使资产保值的可能性最大。

最优执行是在一定的风险水平或价格水平下，尽可能地降低交易成本，以及改进交易价格。指望交易员总是在最佳的市场价格上进行交易，或把所有订单指令执行完成也是不切实际的。即使最优的执行策略也可能会在不利的价格上进行交易，或者在交易结束后还有没执行完的股票。然而，一个最优执行策略和合适的交易方式将确保在大多数情况下能够以恰当的方式进行交易。通常，前面所说的最优交易策略能够保证在公平的市场价格上完成交易，更重要的是确保交易在基金经理指定的成本和价格限度内执行。只有当一个交易策略能以最大的机会实现投资目标，才可以被认为是最优执行策略。

8.3.2 交易执行的目标

人们通常认为投资研究工作的目标是找到能够以最大的可能性实现最大回报率的资产。然而，对于决策的执行却有着不同的观点。一些市场参与者认为交易执行的目标是实现交易量加权平均价格(VWAP)，因为它代表着市场平均的价格；而另一些人则认为执行的目标是实现收盘价，因为它是一些基金制定投资策略的决策价格；还有一些人认为执行目标是实现开盘价或其他的价格基准。这些不同的观点在于选择了不同的价格基准，因而会产生不同的执行策略。但是，所有观点都认为增加资产的价值是非常重要的，也就是说，我们希望新的投资组合的价值和投资组合的固有价值尽可能地接近。

因此我们可以说执行的目标是使平均执行价格 P_{avg} 和做出决策时的决策价格 P_d 的差最小。用数学公式表示如下：

$$\min \varphi = |P_{avg} - P_d|$$

或者用更为常见的离差平方的形式：

$$\min \varphi = (P_{avg} - P_d)^2$$

其中 P_{avg} 代表平均的成交价格。

但是，φ 的值不能描述成一个成本估计值。因为它是随机市场价格 P_t 的函数，所以它自己也是随机变量。因此，最好把 φ 描述成一个有期望成本和风险分布的随机变量。但是，期望成本和风险是往往此消彼长的。当其中一项减少时，另一项就会增加。为了更好地保持资产价值，或实现执行目标，在制订执行策略时分布的期望和方差都要考虑在内。由此，我们可以确定如下几个包括成本和风险项的交易决策标准，可以用来制定合适的执行策略。

目标 1：成本最小化

第一个标准是在可接受的风险水平上使成本最小化。风险水平可能由公司指定，例如最大允许风险敞口，或者与来自投资模型的风险水平一致。数学上这个目标表示为：

$$\min\ Cost$$
$$\text{s.t. } Risk \leqslant R^*$$

其中 R^* 是由公司给定的最大允许风险敞口。

目标 2：平衡成本与风险

第二个标准是平衡成本与风险。对真实的风险敞口的最高水平不确定但

更偏好降低成本所带来的风险的投资者通常更偏爱这个标准。这个目标用公式表示如下：

$$\min: Cost + \lambda \cdot Risk$$

其中 λ 是代表投资者对成本与风险偏好的风险厌恶因子。例如，$\lambda=1$，一般表示投资者对成本和风险偏好程度相同。

目标 3：改进价格

改进价格的目标是指尽可能地提高交易清单的成本小于指定成本的概率。寻找短期回报率最大化的市场参与者，或是寻找使执行优于报价的机会最大化的投资者通常会选择这个标准。它用公式表示如下：

$$\max: \mathrm{Probability}(Cost \leqslant C^*)$$

其中 C^* 是可接受的最大交易成本，即订单平均交易成本的上界。

8.3.2 交易成本和换手率

换手率（turnover ratio）是一定时期内交易的股票总的金额与投资者拥有的资本的比率。它用于说明投资者股票交易的频繁程度。由于股票交易当中有固定成本，所以我们不难看出，换手率对于投资者总的交易成本会产生影响。假定每次交易产生固定比例的交易成本，那么频繁的交易将会导致总交易成本同比例地上升，而经典的买入持有策略将会明显地降低交易的固定成本。

每次交易的成本和换手率是决定总交易成本的两个方面。为了减少交易成本，提高投资的总的回报，需要在制定投资策略的时候将换手率也考虑在内。如果投资者采用的是短期的交易策略，那么交易频率相对较高。这就要求投资策略必须能够产生较高的账面收益率，才能够补偿较高交易成本带来的损失。相反，如果投资策略所确定的持有期限较长，则就不需要投资策略必须产生非常高的账面收益。

通常认为投资风格会影响投资者的换手率，进而影响交易成本。积极型的投资者通常喜欢对股票价格变化做出主动的判断，更多地做出交易决策，导致更高的换手率。相对而言，消极型的投资者则较少做出交易决策，他们更倾向于长期持有指数产品，则会有较低的换手率和交易成本。另外，对于机构投资者而言，较大的交易量可能使他们从经纪商那里得到交易佣金的折扣，在一定程度上可以弥补高换手率所产生的交易成本的损失。

8.3.3 交易成本的管理过程

交易成本管理是实现最优执行的过程。它包括为了制定合适的执行策略所进行的交易成本估计，这是一个需要在投资决策执行前、执行中、执行后实施的三步过程。

第一步：投资决策

交易成本管理应该从投资决策开始。这使得资产分配更加合理，也使基金经理在投资组合阶段能够筛选出低成本高回报的股票。确定投资目标、投资策略和风格的时候应该充分考虑不同资产的交易成本特性，换手率对于交易成本的影响等等因素，以便于交易操作能够更好地达到预期的投资目标。

第二步：交易执行—— 估计成本和评价策略

基金经理和交易员应该在执行过程中密切合作，更好地确定交易清单和不同执行策略的成本和风险估计。

交易前分析能够了解交易清单的特征，例如筛选流动性、波动性、价格趋势和动量，以及按板块、股本、市场进行总结。股票精确的成本和风险估计能够提醒基金经理注意那些不太可能以他们的目标价格被市场吸收的订单。这使得基金经理可以在交易开始之前修改订单的规模，从而降低或消除机会成本。基金经理和交易员通过更好的沟通和协调，共同确定最优交易决策，这样能够更好地实现基金总体的投资目标。

另外，交易前分析为交易员提供了必要的数据和信息。结合第三步，基金经理将得到一份关于他们经纪人过去在类似交易清单中的表现的详细报告。这使得基金经理和交易员可以快速地选择最适合执行特定交易清单的经纪商，进而降低延迟成本。

最后，一个正常的交易成本管理过程会在一天当中连续监视市场情况和成本，交易员可以通过实时调整交易策略来应对变化的市场。交易员为了利用潜在的流动性将积极地交易，或为了避免遭受更多的市场冲击而消极地进行交易。基金经理可以增加订单规模以利用较好的市场价格，或在出现不利的市场价格时减小订单规模。

第三步：进行交易后评价

交易后分析评价投资决策的执行质量。大多数典型的交易后分析包括度量实际的交易成本和评价交易员的表现。此外，还需要评价基金经理的执行决策，以确保他最好地代表了投资者的利益。交易后分析需要评价成本是什

么时间和哪个阶段产生的。这可以让我们了解估计模型的可信度，确保将来有更好的成本估计，并且能够用到投资决策和执行阶段，进而改善交易操作水平。我们需要进行交易后分析，这样基金经理可以评价交易员执行交易的效率。它们测定交易员是否让基金遭受了不必要的成本或风险，是将交易员的技能与风险分离的一种手段，即评估交易员是否真正增加了投资的价值。最后，交易后分析也给出了那个交易员更适合处理什么类型的订单。不断地评价交易员的表现将给基金经理提供充分的信息，从而能在将来的交易中根据股票、列表、板块、市场、股本、市场条件等等更有效地选取合适的交易员。

第九章

算法交易策略介绍

9.1 最优交易问题

9.1.1 最优投资组合问题

风险和收益之间的权衡是金融学的核心问题之一。在马科维茨所确立的传统投资组合理论中，我们可以利用投资组合的收益和方差来得到最优投资组合。投资者可以根据历史数据计算投资组合的历史收益和波动性，并用历史数据作为参考依据，计算最优的投资组合。最优投资组合理论被广泛地应用于金融投资领域，是投资决策的重要理论之一。

最优投资组合理论的问题可以描述为如下两个问题：

1. 在给定的最大风险条件下，求出具有最高收益率的投资组合，其数学描述为：

$$\begin{aligned}&\max:\ E(R)\\&\text{s. t.}\ \mathrm{Var}(R)\leqslant\sigma^2\end{aligned}$$

其中 R 表示投资组合的收益，σ^2 为收益率的方差。

2. 在给定的最小收益条件下，求出具有最小风险的投资组合，其数学描述为：

$$\begin{aligned}&\min:\ \mathrm{Var}(R)\\&\text{s. t.}\ E(R)\geqslant\mu\end{aligned}$$

在不同风险条件下，我们得到具有最大收益的投资组合也不同。根据不同的最优投资组合的收益和风险特性，我们能够得到一个最优投资组合所组成的曲线，被称为有效市场前沿。进一步地，如果我们定义一个风险厌恶系数λ，那么可以将投资组合优化问题写为：

$$\max: E(R)-\lambda\cdot\mathrm{Var}(R)$$

有效市场前沿上不同的最优投资组合对应于投资者不同的风险厌恶度。较大的λ意味着投资者对风险的厌恶度较高，在优化问题中对方差的惩罚度也比较高。较小的λ意味着风险厌恶度比较低。

9.1.2 最优交易问题描述

在前面的章节，我们对投资过程中的交易成本进行了一系列的分析。交易成本包括可见成本和隐性成本，具体被归纳为佣金、交易费用、买卖价差、交易税、延迟成本、价格增长、市场冲击、时间风险、机会成本九个方面。最优交易问题的目的在于尽可能地减少交易成本，获得最优的交易执行策略，改善投资策略的表现。

对于交易操作来说，由于市场流动性的短缺，大额的交易订单通常会带来较大的市场冲击成本。为了减少市场冲击，交易员需要将订单分割成较小的部分，分步、逐渐地进行交易操作。但是，这样会使得交易清单面临着价格增长的成本，股票价格和市场环境波动的时间风险，以及订单不能够完成带来的机会成本。快速的交易操作可以减少这些成本，但是会带来较高的市场冲击成本。这两方面的矛盾给交易员带来了两难的选择。最优交易问题致力于平衡两方面成本因素的大小，以获得更好的交易策略。

另外，由于股票价格的波动、市场交易量的变化等等因素为交易操作带来的风险，使得相似的交易清单和策略有可能产生不同的交易成本和交易表现。因此，交易成本应该被认为是一个随机变量。不同的交易清单和交易策略会有着不同的交易成本和风险的统计分布，最优交易策略的研究通过这些因素的优化，有助于改进交易操作。

本章中，我们将利用均值方差的分析框架来解决最优交易问题，那么，最优交易问题可以写成：

$$\min: E(\mathrm{TC})$$
$$\text{s.t. } \mathrm{Var}(\mathrm{TC})\leqslant k$$

其中 TC 表示交易成本。

类似地，如果我们定义一个风险厌恶系数 λ，那么可以将交易优化问题写为：

$$\min: E(TC)+\lambda \cdot \mathrm{Var}(TC)$$

风险厌恶系数 $\lambda=1$，表示投资者认为风险等于成本的大小。风险厌恶系数 $\lambda\geqslant 1$，表示投资者认为风险比成本更值得注意，那么他更倾向于进行积极的交易策略。风险厌恶系数 $\lambda\leqslant 1$，表示投资者认为成本比风险更值得注意，那么他更倾向于选择消极的交易策略。从经济学意义上讲，λ 表示交易成本和风险之间的替代率。

类似于最优投资组合理论中的有效市场前沿，我们将得到一个有效交易前沿(efficient trading frontier，ETF)。交易员可以根据不同的风险厌恶度在有效交易前沿上选择不同的最优交易策略，在交易成本的期望和方差之间作选择，从而改进交易表现。

9.1.3 交易成本对夏普比率的影响

如果交易员需要交易一个大的订单，通常会将订单分割为许多小的订单，那么在不同交易时间将得到不同的交易价格。根据执行落差法，交易成本可以表示为不同时段内交易平均价格和订单到达价格之差与订单大小的乘积的加总，用数学表示为：

$$TC=\sum x_i(\tilde{p}_i-p_0)$$

其中，TC 表示交易成本，x_i 为在时段 i 内交易订单的规模，$\tilde{p}_i$ 表示时段 i 内交易的平均价格，p_0 为交易开始时的市场价格，即到达价格。这里假定交易完全执行，不存在机会成本。另外，对于买进交易，x_i 为正，如果交易价格高于到达价格，则交易成本为正，反之，则交易成本为负。对于卖出交易，x_i 为负，如果交易价格低于到达价格，则交易成本为正，反之，则交易成本为负。通常情况下，买进交易造成的市场冲击，会使股票价格上涨；卖出交易产生的市场冲击则会导致股票价格下降，产生正的交易成本。但是，如果市场面临和市场冲击相反的价格移动，交易成本也有可能为负值。

在实际投资过程中，交易成本经常被忽略，这将会使投资者很难完成预期的投资目标。下面我们通过夏普比率进行说明，夏普比率是常见的衡量投资

表现的指标之一，其具体的定义为：

$$\text{Sharpe Ratio}=\frac{R_P-R_f}{\sigma_P}$$

其中，R_P 表示投资组合的收益率，σ_P 表示投资组合收益的标准差，R_f 表示无风险利率。

我们用 $\boldsymbol{h}_t$ 表示投资组合中持有的资产数量向量，$\boldsymbol{p}_t$ 表示一段时间内的资产价格向量，其中下标表示时间。在时间$(0,T)$内，投资组合包含交易成本的夏普比率可以表示为：

$$\text{Sharpe Ratio}=\frac{E(\boldsymbol{h}_T'(\boldsymbol{p}_T-\boldsymbol{p}_0))-E(TC)-r_f}{\sqrt{T}\cdot\sqrt{Var(\boldsymbol{h}_T'(\boldsymbol{p}_T-\boldsymbol{p}_0)-TC)}}$$

我们可以明显地看到，分子项包含了交易成本导致预期收益减少，而分母中的方差项中多了交易成本的因素将有可能增加收益的风险，这两方面都将会降低投资的夏普比率。夏普比率是衡量投资表现的重要指标，因此我们可以看到交易成本对投资表现的影响。

9.2 一个最优交易模型

关于最优交易策略有过许多研究。通常的做法是尽可能地减少执行落差，根据到达价格制定最优交易策略。执行落差主要衡量了包括价格增长、时间风险和市场冲击等隐性交易成本。通过对执行落差的优化，我们可以得出最优交易策略。比较著名的研究包括 Bertimas 和 Andrew Lo(1998)，Almgren 和 Chriss(2000)，Obizhaeva 和 Wang(2005)等等。本章所介绍的模型来自 Almgren 和 Chriss 的研究。

9.2.1 模型设定

首先，我们定义一下交易策略。假定在一段时间$(0,T)$内，需要交易股票的数量为 X，将时间 T 分为 n 个区间，那么每个区间长度为 $\tau=T/n$。然后，我们可以将交易列表定义为 $x_1,x_2,\cdots,x_n$，即在每个时间段内交易股票的数量。通过确定交易列表，我们就得到一个交易策略 x。进一步的，定义每个交易时间内未交易的股票数量为 h_t，那么：

$$\sum_{i=1}^{n} x_i = X - h_t \qquad (9.1)$$

另外，定义交易速率为 $v_t=\frac{x_t}{\tau}$，表示在一个交易时段内交易的股票数量。

1. 价格运动

要评估一段时间内交易操作带来的成本，需要对价格运动做出一些假设。我们假设股票价格的变化服从算术布朗运动：

$$p_t = p_{t-1} + \mu\tau + \sigma\tau^{\frac{1}{2}}\varepsilon_t \qquad (9.2)$$

右边第二项表示股票价格的增长，μ 表示其增长率。第三项则表示价格的自然波动，其中 σ 为波动项的标准差，ε_t 表示一个服从标准正态分布的随机变量。在这个假设下，股票价格围绕一个增长的趋势 $\mu\tau$ 进行随机游走。其中既包含了价格增长的趋势，也包含了股票价格波动带来的时间风险。另外，价格增长率 μ 既可以是根据历史数据计算的股票价格的自然增长趋势，也可以是交易员根据基本面分析或者技术分析得出的股票价格增长的预测。

很多研究当中假设股票价格服从几何布朗运动。不过，对于最优交易问题来说，交易期限往往只有一天或几天，在这么短的交易期限内算术布朗运动可以作为一个良好的近似。

2. 市场冲击

市场冲击是指由于某一特定交易指令所引起的股票价格变化。它是指有交易指令时的价格轨道和没有交易指令下达市场时的价格轨道之间的差异。市场冲击成本来自两个主要的原因：流动性需求和信息泄露。流动性需求使得市场在原有均衡价格上供求条件不再平衡。这就要求投资者为了完成指令需要支付一个溢价来吸引额外的流动性。信息泄露是将投资计划或者投资者的交易动机泄露到市场上，使市场价格更高。两者分别导致了永久冲击和暂时冲击。

这里，我们将市场冲击设定为交易速率的线性函数，如下：

$$MI(v)=\gamma v$$

其中，$MI(v)$表示市场冲击，其中包含了暂时冲击和永久冲击的影响。γ 是市场冲击的参数。其中，交易速率 v 的单位包括时间，不符合交易成本的设定。因此，在一个交易时段 τ 内产生的市场冲击为：

$$\tau \cdot MI(v)=\gamma x \qquad (9.3)$$

3. 暂时冲击

当交易指令披露到市场上，如果没有改变市场长期观点，或者当前价值的基本面上的新闻或信息时，产生的市场冲击就是暂时的。暂时市场冲击的产生往往是由于短期内流动性的短缺，并在较短时间内会逐渐消除。

我们这里将暂时冲击 $TI(v)$ 定义为一个交易时段后股票价格的下降，这里，我们假设暂时冲击效应表现为 $t-1$ 时期的交易价格会在 t 时期内有一定的下降：

$$\tilde{p}_t = p_{t-1} - TI \tag{9.4}$$

为了计算的简化，我们将暂时冲击设定为线性函数，也就是说暂时冲击随交易速率呈线性增长：

$$TI(v) = c + \eta v$$

其中，c 和 η 是暂时冲击函数中的参数。暂时冲击分为一个固定部分和一个线性部分，c 可以认为是买卖的半价差。那么交易 x 份股票产生的总的暂时冲击为：

$$x \cdot TI(v) = c \cdot x + \eta \cdot \tau v^2 \tag{9.5}$$

在(9.4)式中，p_{t-1} 是前 $t-1$ 时期价格运动和永久冲击的累积，那么，我们可以进一步计算 t 时期内的交易平均价格为：

$$\begin{aligned}\tilde{p}_t &= p_{t-1} - TI \\ &= p_0 + \mu(t-1) + \sigma\sum_{i=1}^{t-1}\tau^{1/2}\varepsilon_i - \gamma \cdot \sum_{i=1}^{t-1} x_i - c - \eta v_t\end{aligned}$$

结合上面的分析，交易策略 x 产生的交易总投入(对于卖出交易为总收入)可以通过交易期限内的交易数量和成交价格来进行计算。那么，整个交易策略的总投入为：

$$\begin{aligned}X \cdot \bar{p} &= \sum_{t=1}^{n} x_t \tilde{p}_t \\ &= Xp_0 + \mu\sum_{t=1}^{n}\tau h_t + \sigma\sum_{t=1}^{n}\tau^{1/2}h_t\varepsilon_t - \gamma\sum_{t=1}^{n}\tau h_t v_t - cX - \eta\sum_{t=1}^{n}\tau v_t^2\end{aligned} \tag{9.6}$$

上面的式子中，Xp_0 表示股票在交易开始时的账面价值。$\mu\sum_{t=1}^{n}\tau h_t + \sigma\sum_{t=1}^{n}\tau^{1/2}h_t\varepsilon_t$ 表示股票价格的运动，其中包括一个固定漂移项 $\mu\sum_{t=1}^{n}\tau h_t$ 和一个随

机波动项 $\sigma\sum_{t=1}^{n}\tau^{1/2}h_t\varepsilon_t$。$\gamma\sum_{t=1}^{n}\tau h_t v_t$ 表示市场冲击的效应，而 $cX+\eta\sum_{t=1}^{n}\tau v_t^2$ 则表示由于暂时冲击的消除，每个交易时期内股票价格的下降。其中，对于市场冲击项我们可以做如下变化：

$$\sum_{t=1}^{n}\tau h_t v_t=\sum_{t=1}^{n}h_t x_t=\sum_{t=1}^{n}h_t(h_{t-1}-h_t)$$
$$=\frac{1}{2}\sum_{t=1}^{n}[h_{t-1}^2-h_t^2-(h_{t-1}-h_t)^2]=\frac{1}{2}X^2-\frac{1}{2}\sum_{t=1}^{n}\tau^2 v_t^2$$

代入到(9.6)式当中，我们得到交易策略的总投入为：

$$X\cdot\overline{p}=Xp_0+\mu\sum_{t=1}^{n}\tau h_t+\sigma\sum_{t=1}^{n}\tau^{1/2}h_t\varepsilon_t-\frac{1}{2}\gamma X^2-cX-(\eta-\frac{1}{2}\gamma\tau)\sum_{t=1}^{n}\tau v_t^2$$

9.2.2 最优交易策略

1. 执行落差

我们将交易成本定义为执行落差。执行落差是账面资产组合和实际交易中实现的投资组合之间的差异，即 $X\cdot p_0-X\cdot\overline{p}$。根据前面章节的介绍，我们知道执行落差是一个随机变量，它可能在相似的交易条件和策略下得到不同的结果。那么，按照我们前面的分析和设定，对于交易策略 x，即交易列表 $x_1,x_2,\cdots,x_n$，执行落差为：

$$Xp_0-X\cdot\overline{p}=-\mu\sum_{t=1}^{n}\tau h_t-\sigma\sum_{t=1}^{n}\tau^{1/2}h_t\varepsilon_t+\frac{1}{2}\gamma X^2+cX+(\eta-\frac{1}{2}\gamma\tau)\sum_{t=1}^{n}\tau v_t^2 \tag{9.7}$$

执行落差的期望和方差如下：

$$E(x)=-\mu\sum_{t=1}^{n}\tau h_t+\frac{1}{2}\gamma X^2+cX+(\eta-\frac{1}{2}\gamma\tau)\sum_{t=1}^{n}\tau v_t^2$$
$$V(x)=\sigma^2\sum_{t=1}^{n}\tau h_t^2 \tag{9.8}$$

由(9.7)可知，如果 ε_t 服从标准正态分布，那么交易策略 x 的执行落差服从正态分布。那么，根据交易策略的执行落差的期望和方差，我们就可以确定一个交易策略的成本的分布。

2. 有效交易前沿

从前面的分析我们知道，在确定一个交易策略之后，我们可以得到这个交易策略的成本的分布情况。一般情况下，两者之间是此消彼长的。也就是说，降低交易成本的方差会提高交易成本的期望，反之亦然。例如，如果交易员选择较为积极的交易方式，以较快的速率进行交易，那么交易成本的风险就会降低，同时，市场冲击就会使得交易成本的期望上升。相反，如果交易员选择相对消极的交易方式，进行缓慢的交易，交易成本的期望会随着市场冲击的降低而降低，但是时间风险会导致交易成本的风险增加。

因此，为了得到最优交易策略，我们将引进均值方差的分析框架。对于交易员来说，总是希望在特定的风险条件下尽可能地减少交易成本，所以我们可以将最优交易策略定义为：在给定的风险条件下，交易成本最小的交易策略。用数学描述为：

$$\min E(x) \quad \text{s.t.}\ \operatorname{Var}(x)\leqslant k \tag{9.9}$$

也就是说，在给定的风险条件 k 下，我们要找到一个交易成本最低的交易策略 x。在我们前面描述的模型当中，交易成本的均值和方差函数都是二次函数，所以最优解是唯一的。那么对于每一个风险限制 k，我们都能够得到一个最优交易策略 x^*。那么，根据不同的风险限制，我们能够得到一组最优交易策略。我们把这个叫做最优交易前沿，如图 9.1 所示。

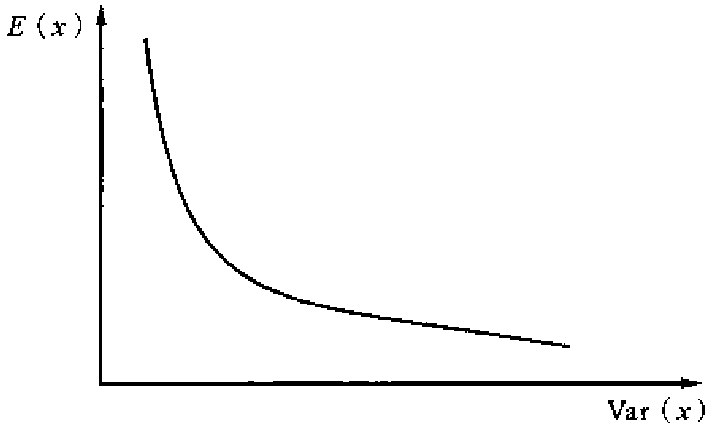


图 9.1　有效交易前沿

为了求解(9.9)式的最优交易问题，我们可以引入拉格朗日乘数 λ，那么最优交易问题变为：

$$\min E(x)+\lambda\cdot\operatorname{Var}(x) \tag{9.10}$$

λ 可以解释为风险厌恶系数。对于给定不同的 λ，我们可以得到不同的最优交易策略。较大的 λ 表示投资者更加重视风险，会更倾向于选择积极的交易策略。较小的 λ 表示投资者相对来说更重视投资的期望收益，因此会更倾向于选择消极的交易策略。

对于(9.8)中描述的交易成本的均值和方差，目标函数 $U=E(x)+\lambda\cdot\mathrm{Var}(x)$ 是一个二次函数，所以我们可以通过 U 对 h_t 的偏导数求 U 的最小值：

$$\frac{\partial U}{\partial h_t}=2\eta\tau\left[-\frac{\mu}{2\eta}+\frac{\lambda\sigma^2}{\eta}h_t-\left(1-\frac{\gamma\tau}{2\eta}\right)\frac{h_{t-1}-2h_t+h_{t-1}}{\tau^2}\right]$$

令 $\frac{\partial U}{\partial h_t}=0$，那么我们得到如下的差分方程：

$$\frac{1}{\tau^2}(h_{t-1}-2h_t+h_{t+1})=\tilde{\kappa}^2(h_t-\bar{h}) \tag{9.11}$$

其中 $\tilde{\kappa}^2=\dfrac{\lambda\sigma^2}{\eta\left(1-\dfrac{\gamma\tau}{2\eta}\right)}$，$\bar{h}=\dfrac{\mu}{2\lambda\sigma^2}$，利用边界条件 $h_0=X$ 和 $h_n=0$，我们可以求得差分方程(9.11)的显式解：

$$h_t=\bar{h}+\frac{\sinh[\kappa(T-t)]}{\sinh(\kappa T)}(X-\bar{h})-\frac{\sinh(\kappa t)}{\sinh(\kappa T)}\bar{h} \tag{9.12}$$

其中 κ 是方程 $\frac{2}{\tau^2}[\cosh(\kappa\tau)-1]=\tilde{\kappa}^2$ 的解，sinh()和 cosh()分别为双曲正弦和双曲余弦函数：

$$\sinh(x)=\frac{e^x-e^{-x}}{2},\cosh(x)=\frac{e^x+e^{-x}}{2}$$

通过(9.12)式，我们可以确定不同时间段持有股票数量的序列 h_t。进而，我们可以得到交易策略 x。

这里，我们利用执行落差的方法建立了一个最优交易的模型，进而得到了具体的交易成本估计和交易策略。由于我们的交易目标是尽可能地减少执行落差，所以，在算法交易当中这一类交易策略被称为 IS(implementation shortfall)策略。在确定交易序列后，交易策略可以由计算机系统自动进行操作，这就是算法交易中所指的 IS 算法。前面介绍的模型相对有些过于简化，通过将不同的市场冲击设定、考虑收益率的自相关、交易量的变化等等更多因素纳入模型可以对交易算法进行改进。由此也可以看出，IS 算法可以有许多不同的版本。不同的交易系统可以使用不同的交易策略。此外，由于该交易策略使用的执行落差以交易的到达价格为基准进行度量，所以得出的交易策

略也可以被称为到达价格(arrival price)策略。

3. 交易成本的 VaR

VaR(value at risk,风险价值)用于描述一个投资组合在一段时间内价值变化的置信区间。它实际上是要回答在概率给定情况下,投资组合价值在下一阶段最多可能损失多少。例如,一个投资组合在未来一年内有 95%的概率损失低于多少美元等等。VaR 完全是基于统计分析基础上的风险度量技术,它有助于投资者对未来投资的风险控制。

风险价值的概念也可以引入到交易操作中来。对于一个交易策略 x,我们定义 $VaR_p(x)$为交易的风险价值,用于表示交易策略 x 所引起的交易成本在 p 概率下不超过某一具体数值。前面的分析中,我们指出特定交易策略的交易成本可以被认为是服从正态分布,那么,根据对交易成本的均值和方差的估计,我们可以计算交易成本分布的置信区间。那么,交易策略 x 的风险价值可以表示为:

$$VaR_p(x)=\alpha_p \cdot \sqrt{Var(x)}+E(x)$$

其中,α_p 表示在概率 p 下交易成本偏离均值的标准差个数,可以通过标准正态分布表获得。用执行落差表达风险价值就是,交易策略的执行落差在 p 的概率下不超过 $VaR_p(x)$。

4. 多只股票组合的交易策略

如果交易员要交易 m 只股票,那么在 t 时刻所持有的头寸向量为 $\boldsymbol{h}_t$ $(h_{1t},h_{2t},\cdots,h_{mt})^T$,其中,如果 $h_{it}<0$,则代表持有股票 i 的空头。交易开始时的头寸为 $\boldsymbol{X}=(X_1,X_2,\cdots,X_m)^T$,$t$ 时刻的交易速率为 $v_t=\dfrac{h_{t-1}-h_t}{\tau}$,如果交易速率 $v_{it}<0$,则代表在 $t-1$ 到 t 时刻之间卖出股票 i。

假定股票价格向量 $\boldsymbol{P}_t$ 服从算术布朗运动:

$$\boldsymbol{P}_t=\boldsymbol{P}_{t-1}+\boldsymbol{\mu}\tau+\boldsymbol{\sigma\varepsilon}_t\tau^{\frac{1}{2}}$$

其中 $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^T$ 表示预期价格增长率向量。$\boldsymbol{\varepsilon}_t=(\varepsilon_{1t},\varepsilon_{2t},\cdots,\varepsilon_{rt})^T$ 表示 r 个独立的布朗运动增量的向量,其中 $r\leqslant m$,$\boldsymbol{\sigma}$ 是一个 $m\times r$ 的波动率矩阵,$\boldsymbol{C}=\boldsymbol{\sigma\sigma}^T$ 是 $m\times m$ 维对称正定的方差协方差矩阵。

市场冲击和暂时冲击也需要用向量来进行表示,我们还是用线性模型:

市场冲击:$\boldsymbol{MI}(\boldsymbol{v})=\boldsymbol{\Gamma v}$

暂时冲击:$\boldsymbol{TI}(\boldsymbol{v})=\boldsymbol{c}+\boldsymbol{Hv}$

其中 $\boldsymbol{\Gamma}$ 和 $\boldsymbol{H}$ 是 $m\times m$ 维矩阵,c 是 $m\times 1$ 维的列向量。$\boldsymbol{\Gamma}$ 和 $\boldsymbol{H}$ 当中的第 ij 个元素表示每单位股票 j 的交易对股票 i 带来的价格冲击。

这样,类似于单只股票的情形,我们可以得到投资组合交易当中产生的交易成本的均值和方差:

$$E(x) = -\sum_{t=1}^{n} \tau\boldsymbol{\mu}^T\boldsymbol{h}_t + \frac{1}{2}\boldsymbol{X}^T\boldsymbol{\Gamma}_S\boldsymbol{X} + \boldsymbol{c}^T\boldsymbol{X} + \sum_{t=1}^{n} \tau\boldsymbol{v}_t^T(\boldsymbol{H}_S - \frac{1}{2}\tau\boldsymbol{\Gamma}_S)\boldsymbol{v}_t + \sum_{t=1}^{n} \tau\boldsymbol{h}_t^T\boldsymbol{\Gamma}_S\boldsymbol{v}_t$$

$$V(x) = \sum_{t=1}^{n} \tau\boldsymbol{h}_t^T\boldsymbol{C}\boldsymbol{h}_t$$

在得到交易成本的均值和方差以后,通过优化技术我们可以得到最优的交易策略。但是,由于问题的复杂性,求解往往需要对假设做出进一步的简化。例如,可以假定股票价格移动只存在一种相关性,即将 $\boldsymbol{\Gamma}$ 和 $\boldsymbol{H}$ 当作一个矩阵,这样可以减少计算的复杂性。

组合的交易策略对于交易操作中的风险控制有着明显的优点。交易过程中的风险控制一般是指尽可能地降低交易清单中未执行部分的风险。如果整体风险能够降低,那么交易员就可以拉长交易的时间,进而可以更好地降低市场冲击成本,以及获取流动性。

类似于市场中性的投资策略,交易员在组合交易过程中也可以在交易清单中同时包含买进和卖出交易,以降低交易的风险。如果两种股票相关度比较高,那么两个方向的头寸就可以互相对冲,买进的一边的价格移动可以被卖出一边的价格移动所抵消。这样就可以降低交易清单总的风险。这种情况下,交易员就有更充分的时间进行交易,而不用顾忌股票价格的变化和波动,可以更好地进行交易,降低市场冲击。

例如,如果一个交易员的交易清单上卖出的指令有 1 000 万美元,买进的指令有 500 万美元,而且买进和卖出的股票的收益相关度很高。那么交易员就可以先交易卖出的指令,使买卖两个方向的头寸逐渐平衡,这样就能够降低交易订单的总风险。

即使是单个方向的交易,也可以利用这个方法。如果股票之间的相关性比较强,交易策略就需要尽可能地减少不同股票或板块之间交易头寸的不平衡,以免一只股票交易产生的冲击引起其他股票价格的变化带来交易成本。

9.3 VWAP 交易策略

VWAP 交易策略是指投资者在进行交易时,使交易的平均成交价格尽可能地接近 VWAP 的交易策略。VWAP 策略在消极型投资者当中很流行。

VWAP 指的是成交量的加权的平均价格。它是对于在一段时间市场上所有的交易活动的平均价格的衡量。很多投资者相信 VWAP 是市场上"平均的"和"公平的"价格。它是一个评估交易员表现好坏的常用手段,代表着投资者的交易表现是否优于市场平均水平。交易员可以通过 VWAP 策略来执行交易清单,这样能够减少市场冲击。但是,这要以时间风险的增加和价格增长的成本为代价。下面,我们将介绍如何得到一个 VWAP 策略。

VWAP 与其他基准价格有所不同,因为它是一个通过计算得到的值,并不是像开盘价、收盘价、最高/最低价等一样的特定时间的市场价格。具体地说,VWAP 等于一段时间里的股票的总的交易额 $\sum P_j v_j$ 除以在这段时间里总的成交量 $\sum v_j$,所以 VWAP 被认为是市场的平均和公平的价格。

$$\text{VWAP}=\frac{\sum v_j P_j}{\sum_j v_j} \tag{9.13}$$

9.3.1 如何实现 VWAP

理论上,如果要在 VWAP 价格上执行交易,交易员需要在每个交易区间内最小化市场冲击,常见的方法是按平均交易量的固定比例进行交易。

【例 9.1】 假定 3 个小时内市场的预期平均成交量如下:

第 1 小时　300 000 股

第 2 小时　400 000 股

第 3 小时　300 000 股

如果一个交易员要交易 100 000 股股票,由于 10 000 股股票占市场总成交量的 10%,那么他可以按如下方式进行交易:

第 1 小时　30 000 股

第 2 小时　40 000 股

第 3 小时　30 000 股

也就是说在每个小时内交易的股票数量都占总交易量的 10%。这种情况下每小时内产生的供需不平衡程度是一样的,产生的市场冲击最小。

但是这样的策略有一个问题。市场成交量的变化比较大,交易员通常无法预测交易期限内的市场成交量有多少。VWAP 交易需要找到一个合适的分割策略,以使得交易员在无法提前知道当天的总交易量的情况下进行交易。另外一个替代方案是按某一时段内占日交易量的比例进行交易。交易员可以

把一天分割成不同的时间段或者交易间隔，并在每个期间计算日交易量的平均百分比组成。许多研究表明，单日内交易量的曲线呈现“U 形”，也就是说交易量变化的模式相对稳定，因此这种 VWAP 策略相对更容易实现。为了得到较好的估计，投资者可以使用最近的观察数据。

【例 9.2】假定 3 个小时内市场的预期平均成交量占总成交量的百分比如下：

第 1 小时　30%

第 2 小时　40%

第 3 小时　30%

这种情况下交易员可以按照交易量变化的相同模式进行交易，也就是在 3 个小时内分别执行交易清单的 30%、40%和 30%。

交易员经常会想要得到一个接近 VWAP 基准的平均执行价格。由于市场上没有人能够保证交易的成交价格，交易员经常会选择实现 VWAP 价格可能性最大的策略来执行交易。在这种情况下，最好的执行策略（这个策略提供了最高可能性获得 VWAP 基准价格）可以表示为：

等式(9.13)可以改写如下：

$$\mathrm{VWAP}=\frac{\sum v_j p_j}{\sum v_j}=\sum\left[\left(\frac{v_j}{\sum v_j}\right)\cdot \bar{p}_j\right]=\sum u_j\,\bar{p}_j \tag{9.14}$$

其中 u_j 是第 j 个交易期间的日交易量的百分比，$\bar{p}_j$ 是第 j 个交易期间的平均价格。

如果 y 是一个订单的分割策略，那么订单的平均交易价格可以表示为 $\sum y_j\,\bar{p}_j$。如果交易员在每个交易期间能够按照平均价格完成交易，则这个策略提供获得 VWAP 的可能性最大。它找到一个使得 VWAP 基准价格和平均执行价格的差异最小的交易策略 y。在数学上，交易目标是使得 VWAP 基准价格和平均执行价格的差异的均方差最小，即：

$$\min:\delta=\left(\sum u_j\,\bar{p}_j-\sum y_j\,\bar{p}_j\right)^2$$

写成向量形式为：

$$\min:\delta=(\boldsymbol{u}'\boldsymbol{p}-\boldsymbol{y}'\boldsymbol{p})^2$$

其中 $\boldsymbol{u},\boldsymbol{y},\boldsymbol{p}$ 为 $n\times1$ 列向量，$\boldsymbol{u}',\boldsymbol{y}',\boldsymbol{p}'$ 为这些向量的转置。

目标函数是一个平方的形式，因此 $\delta\geqslant0$。我们可以看到一个直观的解为

$(y-u)'=0$，即 $y=u$。因此，如果交易员想要达到 VWAP，一个近似的分割策略是交易订单的分割要和这个期间的交易量保持一致。进一步地，这个在任何交易期限内都成立。最后，分割策略 $y=u$ 也是期望价格和 VWAP 之间差异最小的交易策略。

VWAP 策略并不能够保证投资者避免大的价格波动和高成本的风险。例如，投资者用 VWAP 策略在一天时间内执行一个交易，那么肯定会在收盘前执行一部分交易。因此，如果收盘之前价格出现大幅上升，买单将会遭受更高的成本，而卖单就会得到更低的成本。VWAP 并不是成本最小化的策略，而仅仅是最小化暂时市场冲击的策略。

9.3.2 交易量特征的估计

前面我们提到，使期望平均价格和 VWAP 基准价格之间差异最小的策略是 $y_j=u_j$。但是，这里的 u_j 怎样估计出来呢？我们可以通过前 30 天内的移动平均交易量来估计 u_j，选择数据的时间尺度要看交易量曲线的稳定性。例如高流动性的股票需要 20 天的数据，因为这样的股票交易量曲线更稳定；而缺乏流动性的股票，则需要大约 30 天的数据。不过，在任何情况下使用数据越多估计就会越精确。

下面我们对股票的交易量特征进行估计，假设 V_i 是 j 天以前总的交易量，v_{ij} 是第 j 个期间 i 天以前的交易量，那么

$$u_j=\frac{1}{n}\sum_{i=1}^{n}\frac{v_{ij}}{V_i}=\frac{1}{n}\sum_{i=1}^{n}u_{ij}$$

在这里，向量 $u^T=(u_1,u_2,\cdots,u_n)$ 包含了在每个期间内预期的日交易百分比，其中 u_j 为第 j 个期间的日交易百分比。很明显，$\sum u_j=1$。为了用 VWAP 策略进行交易，交易员根据交易量特征 u 来分割他们的订单 X。在任何一段给定的交易期间内，需要交易的股份数量为 $x_j=u_j\cdot X$。

用超过 30 天的交易数据来估计股票的交易量曲线并不一定能够改进估计精度。在大多数情况下，交易量模式是稳定的，而总交易量的波动性则很大，甚至会在两天之间产生很大的变化，日交易量的比例的持续性则相对较好。但是，流动性低的股票出现大额交易的情况是一个例外。

此外，数据的滞后并不像我们想的那样对交易量曲线的估计有很大的损害。交易量曲线的模式是非常稳定的，也许是股票交易中最稳定的一个特征。

用一个月前的数据和用当月的数据并没有太大的不同。因此，分析员可以从以前的数据得到和使用最近的交易数据相接近的估计精度。但是，一般情况下我们建议用最近的数据。因为最近的数据能够更好地反映市场结构的变化。从某种程度上来说，交易员在今天的投资行为与他上个月，或者去年的表现是接近的，所以我们可以认为市场结构是稳定的。

关于交易量曲线的一个有趣的问题是：我们应该对单个股票做出特定的估计（每个股票一个交易量曲线），还是用一些相对一般化的曲线，例如资本市场或者某个行业的曲线？一些研究发现，一般化的交易量曲线和特定的股票的交易量曲线相比总的来说是一致的。然而，为了得到最好的置信区间和误差区间的估计，特定的股票曲线当然更好。

9.3.3 为什么交易员会选择 VWAP 交易策略？

交易员选择用 VWAP 交易策略来执行委托有两个主要原因：首先，VWAP 策略是一种最小化市场冲击成本的交易策略。用 VWAP 策略执行的交易员喜欢根据市场交易量进行交易，这使得在每个交易期中尽可能地保持供需平衡，而导致最小的损失。

其次，对于许多交易员来说，他们的执行业绩经常用 VWAP 作为基准来进行衡量。一个交易员的平均执行价格和 VWAP 相比越好，就表示他的操作业绩越好；相反，如果交易员的平均交易价格和 VWAP 相比差很多，那说明他的操作业绩很差。但是这些并不能保证 VWAP 策略总能够带来好的结果，有时也可能导致基金遭受更高的交易成本。

VWAP 策略确实是一个最小化市场冲击成本的策略。但是，它并不是一个最小化总交易成本的策略。市场冲击成本只是总交易成本的一部分。在执行交易的时候，交易员也面临着价格增长和时间风险的威胁。VWAP 策略最小化了市场冲击成本，但是并不能够避免价格上升带来的影响。例如，交易员用 VWAP 策略买入一只价格上升中的股票，由于 VWAP 策略过于消极，他将会遭受更高的交易成本。类似的，交易员用 VWAP 策略在一个下降的市场上卖出股票，他同样也会招致许多不必要的成本。VWAP 策略并不能使投资者免受价格上升所带来的交易成本。由于 VWAP 并不是实现最小成本的交易策略，所以交易员可以找到在相同风险下的更低成本的策略，或者在相同成本下更低风险的策略。投资者在价格上升时选择 VWAP 策略来执行委托，既不是一个最优的交易方式，有时甚至不是一个合理的交易决定。

如果没有预期的价格上升，VWAP 策略是一个最低的交易成本策略，因为只剩下市场冲击这一主要的交易成本。因此，在这些情形下，VWAP 策略是一个最优的交易策略。然而，即使在这种情形下，VWAP 策略也并没有使投资者避免时间风险。通过和有效交易边界上其他交易策略的对比，我们发现 VWAP 策略是所有最优策略中风险最大的一个。尽管它可能是最低成本的策略，但是它相对较大的方差使得投资者面临着交易成本很大的不确定性，可能出现的价格波动将使投资者面临较大的交易成本，而这是投资者所希望极力避免的。

图 9.2 中，VWAP 是所有最优策略中成本最低的。期望成本为 10 个基点，但是其风险相对较大，为 150 个基点。具有较小的风险厌恶水平的交易员可能会决定选择 VWAP 作为最优交易策略。但是如果交易员风险厌恶度较高，则会选择图形上的策略 A。因为 A 可以带来更小的交易风险，它的期望成本为 12 个基点，风险为 100 个基点。交易员牺牲 2 个基点期望成本就可以使得交易价格的风险下降 50 个基点。我们可以将策略 A 称为 VWAP 对冲策略，也就是说，一个和 VWAP 非常相似的低成本策略。但是相对来说，它却有着较低的交易风险。VWAP 对冲策略是一个按照股票交易量曲线执行委托的策略，但是没有必要整天都执行交易。小的订单可以较快地执行，而大的订单则要在整天执行。最优化的规则是由交易量情况决定的，什么样的订单就用什么样的交易期限来执行。例如，为了保证最小化市场冲击成本，执行 150 000 股股票的订单需要在整天执行。但是，在一个更短的时间段来执行则可以更好地控制交易风险。

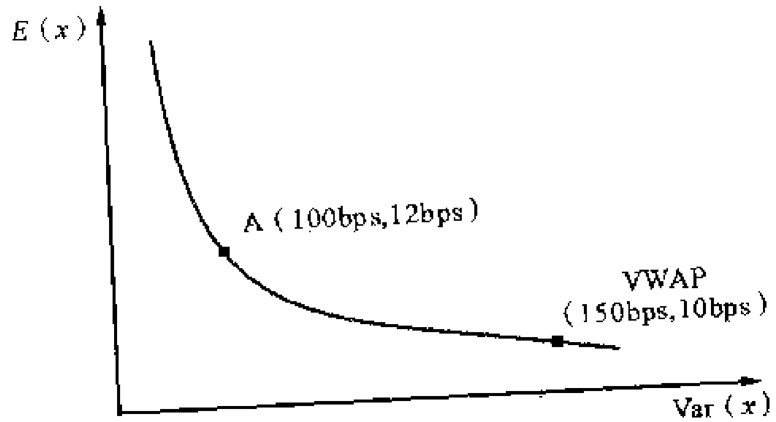


图 9.2 一个有效交易前沿

从前面的分析我们知道，获得 VWAP 价格的交易目标和最优交易操作的目标并不一致。交易员需要在成本和风险上保持平衡，并且需要意识到价格上涨的趋势，才能实现最优的交易操作。然而，VWAP 策略仍然是很多交易员愿意执行的策略。因为许多金融机构经常会用 VWAP 来评价交易员的业绩，而这往往是不恰当的交易方式。

此外，VWAP 策略有许多优点。例如 VWAP 策略最小化市场冲击成本，而且在没有任何关于价格增长的成本的期望的情况下，它还是最低成本的最优化策略。VWAP 策略也经常使得交易员隐藏其委托量大小和他们的交易意图，而这一点很受投资者的重视。在每个时期都以一个和市场交易量相同比例的订单进行交易，这使得市场很难知道他们到底想干什么。因此，许多人声称 VWAP 策略的交易速度刚刚好，既不太快也不太慢。

9.4 算法交易策略简介

9.4.1 价格基准和交易策略

算法交易的目的是通过计算机系统自动地完成交易的执行过程，并尽可能地减少交易成本和风险。计算交易成本通常需要确定一个基准价格，通过将交易价格和基准价格进行对比，才能够评价交易成本的高低。最常见的价格基准是加权的平均价格，以及交易前后的报价。下面是常用的三类基准价格：

- 交易开始前的某个价格，其中包括前一日收盘价、当天开盘价、交易前最后一个成交价、交易开始时的决策中间价等等。
- 交易当天的加权平均价格，例如 VWAP 和 TWAP。
- 交易完成后的某个价格，例如当天的收盘价。

常见的交易算法常常就是通过交易目标的基准价格所命名的，例如最常用的 VWAP 算法、TWAP 算法，以及 Arrival Price 等等。这些算法的目标就是实现接近于基准价格或优于基准价格的平均成交价格。另外，Implementation Shortfall 算法表明这类算法是以执行落差为决策标准的，其交易目标就是尽可能地针对某一特定的基准价格的执行落差，使得交易的平均成交价格尽可能地接近基准价格。因此对于算法交易来说，交易算法常常是和特定的价格基准联系在一起的。不过，不同的金融机构在内部开发的算法会各不相

同。例如，即便同样是 Implementation Shortfall，不同机构提供的算法的具体操作过程也会有所区别，因为他们开发的技术细节是会存在差异的。这也导致不同的金融机构的交易算法之间的目标会有一些细微的差异，有时候不同策略的性能也会有一定的差别。

9.4.2 一些交易算法的介绍

1. VWAP

常见的 VWAP 策略的操作方法是按平均交易量的固定比例进行交易。但是，由于市场成交量波动性比较大，交易员无法预测交易期限内的市场成交量有多少。另外一种实现 VWAP 的策略是按某一时段内占日交易量的比例进行交易。交易员可以把一天分割成不同的时间段或者交易间隔，并在每个期间计算日交易量的平均百分比组成，例如如果一天的某一时间内的交易量占当天 5%，那么交易员就在这个时段交易自己交易清单的 5%。

使用 VWAP 的理由在于：

- VWAP 是最为常见的评估交易员操作表现的价格基准。
- VWAP 产生的市场冲击最小。
- 当交易员没有很强烈的交易意图的时候，会选择 VWAP。因为 VWAP 至少代表着市场平均的价格水平。
- VWAP 的订单分割方式更有利于隐藏交易意图。

同时 VWAP 策略也存在着一些缺陷：

- VWAP 策略和最优交易操作的目标并不一致，没有将交易当中的价格增长和时间风险考虑在内，使得交易更多地面临着价格波动和增长带来的风险。
- VWAP 对于小额交易来说交易速度过于缓慢。小额交易产生的市场冲击较小，没必要设定过长的交易时间，完全可以直接进行交易而不进行分割。这种情况下 VWAP 策略使交易面临的风险较大。
- 由于 VWAP 等于全天内交易成交价格的加权平均，所以里面包含了未来的市场价格。因此，即使交易员造成的市场冲击使市场价格升高，仍有可能实现 VWAP。很明显，如果交易员所执行的交易比例接近于该股票的全部交易量时，无论如何操作都能够实现 VWAP。

2. Implementation Shortfall(IS)

Implementation Shortfall 算法是以执行落差为决策基础的交易策略。交易的目标是尽可能减少交易当中的执行落差。Implementation Shortfall 算法

的思想可以应用于各种价格基准，如到达价格、开盘价格、前一日收盘价等等。因此一些诸如 Arrival Price、Opening、Closing 等算法实际上也可以执行落差为基础来制定。例如，Arrival Price 的决策目标可以描述为：

$$\min:\ \sum x_i(\bar{p}_i - p_0)$$

利用执行落差的方法可以针对这个目标优化，以得到一个具体的 Arrival Price 算法。

Implementation Shortfall 算法的优点在于：

• 执行落差的方法较为全面地分析了交易成本的各个部分，在市场冲击成本和时间风险，以及价格增长等因素之间取得了更好的平衡，更加符合最优交易操作的目标。

Implementation Shortfall 算法根据一些决策价格进行交易过程的优化，更符合投资决策的过程。例如一些投资策略以市价为信号进行交易，那么 Arrival Price 则是一个很自然的选择。

Implementation Shortfall 的组合交易算法能够利用交易清单上股票之间的相关性更好地控制风险。

3. Participate

Participate 算法类似于 VWAP 算法，以市场平均交易量的百分比为决策基础。但是与 VWAP 有所不同的是，Participate 算法以一个固定比例进行交易，这个比例就是所谓的参与率(participation rate)。例如，如果交易员认为可以忍受市场交易量 10%的订单带来的市场冲击，那么就可以一直以 10%的参与率进行交易。因此，对于小订单来说，Participate 算法表现出了比 VWAP 更有利的一面。小订单可能在前两个交易期限就已经完成，使得交易清单更少地面临价格增长和波动带来的风险。

另外，和 Implementation Shortfall 算法一样，Participate 算法也可以和特定的决策价格基准相结合。例如，如果利用开盘价格为价格基准，可以在开盘后的一段时间内以一定的参与率进行交易；如果使用收盘价格为价格基准，可以在收盘前的一段时间内以一定的参与率进行交易。

4. TWAP

TWAP(time weighted averaged price)是时间加权的平均价格，是在不同交易时间点上交易价格的平均，计算方式如下：

$$\mathrm{TWAP} = \frac{1}{n}\sum P_j$$

如果要实现 TWAP，可以在整个交易期限内使用 Implementation Shortfall 算法或者 Participate 算法的思想。这样就能够使交易的平均成交价格尽可能地接近 TWAP。

5. 一些其他算法

算法交易出现以后，大量的金融机构都在开发自己的交易算法。市场上出现了大量不同的算法。有时算法交易的策略会被冠以很奇怪的名字。这里，我们对几种简单的算法进行一些介绍。

Simple Time Slicing，这种算法就是简单地将订单分割为小的部分，在一定的时间间隔内以市价单的形式发送到市场。

有一种交易策略被称为冰山一角(icebergíng)。这个策略的目的就是通过订单分割的方式，可以隐藏或部分隐藏交易的动机，来达到降低成本和发现流动性的目的。策略的名字"冰山一角"形象地表达了其操作方式和目的，就像冰山一样，被发现的永远只是水面上的一小部分。其具体做法就是由基金经理设定交易期限内最大的交易数量，或者每个更短的期限内的具体交易数量，形成一个具体的交易序列，然后交给计算机系统进行交易操作。

游击队(guerrilla)是由瑞士信贷开发的一个交易算法，它实时地不断接收不同交易市场上的公开报价，并进行判断。其目标是发现能够进行交易，而且造成价格移动的可能性较小的报价，以避免影响所交易股票的交易模式。这种算法适合希望尽可能避免市场价格的基金经理。

一些交易所和交叉网络会提供一些隐藏的流动性，并不在传统的公开平台上交易，而只在大的机构投资者和经纪商之间交易。这些流动性通常被称为流动性暗池(dark pools of liquidity)。瑞士信贷开发的一种交易算法叫狙击手(sniper)，就专门用来发现这类隐藏的流动性。在流动性暗池中的交易只公布股票的价格，而不公布交易的投资者和交易量信息，所以能够降低市场冲击。

另外，还有一类交易算法叫做窃听器(sniffer)，专门用来发现其他投资者是否使用算法进行交易操作，以及所用的具体算法，以希望能够利用这类信息获得额外的收益。

这里我们仅仅介绍了一些比较简单的和相对知名的算法。通过进一步的修正和改进，这些算法可以变化出很多的版本，以适应不同的投资者在不同市场环境下的需要。在算法交易领域里，可以说每天都在诞生着新的交易算法。因为市场条件不断在变化，投资者也在不断地进行创新和研究，算法交易也在不断地发展之中。

第十章 交易后分析和算法的选择

10.1 交易后分析

交易后分析是对交易员的表现进行评估的过程。交易后分析需要估计交易执行产生的成本，辨别交易员的表现是靠运气还是技能，以及通过评估投资决策的实施情况来判断是否完成了最优决策。交易后分析不像传统投资组合的绩效分析那样根据一定的时间间隔(例如每年、半年或季度)进行，而是在每次交易完成后进行。

交易后分析由两部分组成:交易成本测量和交易表现评估。交易成本测量用以判断交易成本的大小，以及成本在哪里产生。通过评估，基金经理可以加深对市场结构变化的理解，进而改进资产配置和股票选择的过程。交易表现评估目的是检验所产生的交易成本是否合理。我们判断成本是由交易活动本身和不可避免的价格变动，还是由低质量的交易决策所导致的。这一信息对基金经理的投资决策十分重要，可以帮助他们将更多资源投入能够带来附加值的部分，而将更少资源投入到难以带来改进的地方。

10.1.1 测量交易成本

交易后分析基础就是测量交易成本，把成本看成期望成本和风险的分布来进行估计，进而评估交易员的表现。测量交易成本代表着评价过去的交易表现，而估计交易成本则是对未来交易的预测，其中包括着不确定性。交易表

现的评估是衡量在实际的市场环境中产生的成本是否合理，或者说，交易决策质量的好坏。如果基金经理和交易员不了解这些测量的方法，就容易造成不合理的投资和交易决策。

为什么要测量交易成本？原因很简单。如果基金经理想要降低交易成本，进而改进投资决策，那么首先需要能够测量交易成本。

交易成本是从事经营的成本，需要在决策过程中加以考虑。对于基金经理来说，深入了解经营过程中的成本非常重要。交易成本就是平均执行价格与投资决策的价格之间的差。不过，二者产生差异的原因有很多。

金融行业经常将交易成本定义为执行价格与基准价格（如 VWAP）的差。然而，从基金的投资目标来看，这是没有道理的。即使投资者能够提前知道交易价格会优于或差于 VWAP 5 个基点，有些时候对投资决策也是没有意义的，因为投资者依然不知道他们实际交易的价格。成本测量的应该是执行价格与决策价格的差。

【例 10.1】一个基金经理需要在两种股票 A 和 B 中选择，两只股票的市价都是 50 美元，期望年收益率均为 10%。因此，一年后两只股票的期望价格均为 55 美元。那么，基金经理应该把哪只股票加入投资组合当中？

为了能够选择合适的股票，基金经理需要知道成本是多少。如果 A 的成本为 100 个基点，而 B 的成本为 50 个基点，那么我们可以计算一年后的回报率：

$$\text{平均交易价格}=\text{市价}\cdot(1+\text{成本比率})$$

$$\text{实际回报}=\frac{\text{未来价格}}{\text{平均交易价格}}-1$$

A 以 50.50 美元（50 美元加上 100 个基点，即 1%）买进，则实际回报率为：

$$55/50.50-1=8.91\%$$

B 以 50.25 美元（50 美元加上 50 个基点，即 0.5%）买进，则实际回报率为：

$$55/50.25-1=9.45\%$$

因此，基金经理显然会选择 B，因为它的实际回报率更高。但是，如果没有确切的交易成本信息，经理无法做出最优的交易决定。因此，基金经理需要知道市价和成本才能做出正确的投资决策。

【例 10.2】和上面问题相同的情况，但基金经理不再具有确切的成本信息，而是有每只股票的平均执行价格和 VWAP 价格的历史记录。如果从历史数据看，A 的平均执行价格差于 VWAP 5 个基点，B 的平均执行价格差于 VWAP 20 个基点。那么，基金经理如何利用这一信息决定买入哪只股票？

如果只知道历史信息，经理是不能够做出正确决定的。知道交易差于VWAP的比例不能提供任何交易成本和交易价格的信息。因为，虽然和各自的VWAP相比A表现较好，但是事实上A的VWAP可能要高于B很多。那么在这种情况下，B则是成本更低的股票，也更值得投资。即使基金经理假设两只股票的VWAP价格相等，A比B的成本更低，也仍然无法做出合理决策。交易指令大小和执行策略会对比较执行价格和VWAP产生很大影响。当交易指令的规模越来越大的时候，执行价格就会与VWAP越来越接近。当交易指令的规模占成交量的100%的时候，执行价格就是VWAP。另外，如果以前每只股票的执行目标均不相同，那么与VWAP的对比也会有很大不同。因此，如果在历史交易当中A的交易量比B大，则A比B更容易接近于VWAP。

1. 执行落差

最常用的交易成本评价方法是Perold提出的执行落差法。执行落差是指实际组合回报与账面回报的差异。这一方法考虑了除管理费用以外的所有投资组合交易成本。这一测量方法的目的是测量执行投资决策的能力，已成为交易成本的测量标准。

执行落差将成本分成两部分：执行成本部分和机会成本部分。执行成本由实际市场交易成本构成，如佣金、交易税和价格冲击。在这种情况下，Perold把价格冲击定义为投资决策时的股票价格与实际执行价格的差。正如他所指出的，价格冲击是由很多因素导致的：信息泄露、流动性需求、反向价格运动、市场波动等等。Perold对价格冲击的定义包含了我们的市场冲击、价格上涨、时间风险和延迟成本，而不只把价格变动归咎于交易。因此，这里价格冲击与我们的市场冲击定义不同，市场冲击是特定交易引起的股票价格变动。Perold定义的价格冲击正是我们所定义的交易相关和投资相关成本。执行落差中的机会成本指的是因为交易不能完成而损失的利润。产生机会成本的主要原因是市场缺乏流动性，以及不利的价格变动。

n阶段的账面组合收益定义为n时刻投资组合价值减去投资决策时的价值。账面组合的基本假设是在市价条件下可以无限地交易股票，而不产生价格移动。另外，账面组合没有佣金、买卖价差、交易费用和税收的成本。

账面组合所使用的市价定义为买卖报价的中间值。值得注意的是，账面回报的定义要求在买卖价差中间值执行所有的交易指令，否则投资组合会产生平均为买卖价差一半的成本。假设投资者同时买进和卖出一只股票。如果账面组合以买入报价买入，以卖出报价卖出，那么投资者因为买卖价差而会造

成损失。但是,Perold 的执行落差法认为账面组合不存在交易成本。另一方面,n 阶段投资组合的实际回报定义为 n 时期组合的价值与执行时的组合价值(去除佣金、税和费用与机会成本)的差。计算如下:

X_i=要执行的股票 i 的股数

x_{ij}=时期 j 执行的股票 i 的股数

$X_i, x_{ij}>0$ 表示买;$X_i, x_{ij}<0$ 表示卖

P_{id}=投资决策时股票 i 的价格

P_{ij}=第 j 次交易时股票 i 的执行价格

P_{in}=交易结束时股票 i 的价格

$$\text{账面回报} = \sum_i X_i P_{in} - \sum_i X_i P_{id}$$

$$\text{实际回报} = \sum_i X_i P_{in} - \sum_i \sum_j x_{ij} P_{ij} + \text{可见成本}$$

然后,我们计算执行落差成本如下:

(1)情况 1

首先,我们考虑 t_n 时刻交易完全执行的情况:

$$\begin{aligned} IS &= \text{账面回报} - \text{实际回报} \\ &= \left(\sum_i X_i P_{in} - \sum_i X_i P_{id}\right) - \left(\sum_i X_i P_{in} - \sum_i \sum_j x_{ij} p_{ij}\right) + \text{可见成本} \\ &= \underbrace{\sum_i \sum_j x_{ij} p_{ij}}_{\text{执行价格}} - \underbrace{\sum_i X_i P_{id}}_{\text{决策价格}} + \text{可见成本} \end{aligned}$$

(2)情况 2

考虑在交易结束时有没能执行的交易的情况。与情况 1 相同,执行落差是账面回报与实际回报的差,即使每个组合执行的股数不同也是如此。这直接引出 Perold 的机会成本定义。

y_i=未执行的股票 i 的股数

$\sum_j x_j$ = 股票 i 的执行股数

$$\begin{aligned} IS &= \text{账面回报} - \text{实际回报} \\ &= \left(\sum_i X_i P_{in} - \sum_i X_i P_{id}\right) - \left(\sum_i \left(\sum_j x_{ij}\right) P_{in} - \sum_i \sum_j x_{ij} p_{ij}\right) + \text{可见成本} \end{aligned}$$

因为交易列表中有 y_i 股未被执行,那么:

$$y_i = X_i - \left(\sum_j x_{ij}\right)$$

$$\sum_i X_i P_{in} = \sum_i \left(\sum_j (x_{ij}) + y_i\right) P_{in} = \sum_i \sum_j x_{ij} P_{in} + \sum_i y_i P_{in}$$

$$\sum_i X_i P_{id} = \sum_i \left(\sum_j (x_{ij}) + y_i\right) P_{id} = \sum_i \sum_j x_{ij} P_{id} + \sum_i y_i P_{id}$$

经过代换有

$$\begin{aligned}
IS = & \left[\left(\sum_i \sum_j x_{ij} P_{in} + \sum_i y_i P_{in}\right) - \left(\sum_i \sum_j x_{ij} P_{id} + \sum_i y_i P_{id}\right)\right] - \\
& \left(\sum_i \sum_j x_{ij} P_{in} - \sum_i \sum_j x_{ij} p_{ij}\right) + \text{可见成本} \\
= & \sum_i \sum_j x_{ij} P_{ij} - \sum_i \sum_j x_{ij} P_{id} + \sum_i y_i P_{in} - \sum_i y_i P_{id} + \text{可见成本} \\
= & \left(\sum_i \sum_j x_{ij} P_{ij} - \sum_i \sum_j x_{ij} P_{id}\right) + \\
& \left[\sum_i \left(X_i - \sum_j x_{ij}\right) P_{in} - \sum_i \left(X_i - \sum_j x_{ij}\right)\right] P_{id} + \text{可见成本} \\
= & \underbrace{\left(\sum_i \sum_j x_{ij} p_{ij} - \sum_i \sum_j x_{ij} P_{id}\right)}_{\text{执行成本}} + \underbrace{\sum_i \left(X_i - \sum_j x_{ij}\right)(P_{in} - P_{id})}_{\text{机会成本}} + \\
& \text{可见成本} \qquad (10.1)
\end{aligned}$$

因此，执行落差被划分为执行成本、机会成本和可见成本。值得注意的是，只要用于分析的时间区间等于或长于交易期限，分析的期限对于结果就没有影响。

(3)情况 3

考虑基金经理确定每只股票交易的规模而不采用股票数量，是货币值 D_i。$D_i>0$ 表示买进，$D_i<0$ 表示卖出，这样隐含地确定了经理打算买的股数 X_i：

$$X_i = \frac{D_i}{P_{id}}$$

这样，计算执行落差的方法可以参考第 1 种情况。

2. 扩展执行落差方法

Perold 的执行落差法正是基金经理所需要的测量执行成本的方法。然而，对于交易的归因分析，执行落差法却不是那么的有效。Wagner 的研究指出，指令的交易成本可以分成延迟部分和交易部分。延迟成本是从经理做出决定到交易员执行交易之间因延迟而损失的利润。交易成本代表交易活动本身的成本。拥有这些信息可以使基金经理区分成本是基金管理成本，还是交易本身的成本。重要的是，通过合理的交易成本管理可以减少大量的延迟成本，进而改进组合表现。将 Wagner 的方法扩展到执行落差法当中，我们可以得到进一步分辨交易成本出处的公式：

t_0＝指令到达市场的时间

t_d＝交易开始的时间

t_n＝交易结束的时间。其中，$t_d < t_0 < t_n$

P_{i0}＝股票 i 的交易指令下达时的价格

我们有：

$$t_n - t_d = (t_n - t_0) + (t_0 - t_d)$$
$$P_{in} - P_{id} = (P_{in} - P_{i0}) + (P_{i0} - P_{id})$$

于是我们计算执行成本和机会成本如下：

$$\begin{aligned}\text{执行成本} &= \sum_i \sum_j x_{ij} P_{ij} - \sum_i \sum_j x_{ij} P_{id} \\ &= \left(\sum_i \sum_j x_{ij} P_{ij} - \sum_i \sum_j x_{ij} P_{i0}\right) + \left(\sum_i \sum_j x_{ij} P_{i0} - \right. \\ &\quad \left. \sum_i \sum_j x_{ij} P_{id}\right)\end{aligned}$$

$$\begin{aligned}\text{机会成本} &= \sum_i \left(X_i - \sum_j x_{ij}\right)(P_{in} - P_{id}) \\ &= \sum_i \left(X_i - \sum_j x_{ij}\right)(P_{in} - P_{i0}) + \sum_i \left(X_i - \sum_j x_{ij}\right)(P_{i0} - P_{id})\end{aligned}$$

最后，展开机会成本项并带人(10.1)的 IS 式：

$$\begin{aligned}IS = {} & \underbrace{\sum_i \sum_j x_{ij}(P_{i0} - P_{id})}_{\text{投资成本}} + \underbrace{\left(\sum_i \sum_j x_{ij} P_{ij} - \sum_i \sum_j x_{ij} P_{i0}\right)}_{\text{执行成本}} + \\ & \underbrace{\sum_i \left(X_i - \sum_j x_{ij}\right)(P_{in} - P_{id})}_{\text{机会成本}} + \text{可见成本}\end{aligned} \tag{10.2}$$

$$\begin{aligned}OC = {} & \sum_i \left(X_i - \sum_j x_{ij}\right)(P_{in} - P_{id}) \\ & - \underbrace{\sum_i \left(X_i - \sum_j x_{ij}\right)(P_{id} - P_{i0})}_{\text{投资过程产生}} + \underbrace{\sum_i \left(X_i - \sum_j x_{ij}\right)(P_{in} - P_{i0})}_{\text{交易当中产生}}\end{aligned}$$

IS 的扩展公式使得基金经理能更好地辨别成本在何时何地发生。通过合理的交易成本管理，交易员和基金经理能更好地合作，以减少延迟成本和机会成本。另外，通过应用恰当的交易策略，基金经理和交易员能够更多地减少成本。我们的扩展公式也能区分隐性成本（延迟、价格上涨、市场冲击、时间风险和机会成本）和可见成本（佣金、交易税、交易费用和买卖价差）。

于是，交易引起的成本为：

$$\varphi = \underbrace{\sum_i \sum_j x_{ij} P_{ij} - \sum_i \sum_j x_{ij} P_{i0}}_{\text{执行成本}} + \underbrace{\sum_i \left(X_i - \sum_j x_{ij}\right)(P_{in} - P_{i0})}_{\text{机会成本}}$$

虽然 IS 的扩展公式能辨别成本何处产生，但是仍有一些限制。这是由订单管理和交易系统的缺陷造成的。例如，这种方法需要基金经理和交易员每次交易都要记录投资决策时的市场价格、指令进入时间和实际执行时间。

当经理获得决策价格，交易员获得每次交易的价格时，以及在指令进入市场时，这个价格已经存在延迟了。在这些时候，相关价格（如开盘价）经常被用来作为代替，使得投资和交易相关的成本精度下降，但它不影响总体成本的精度。此外，为了考虑长时期的损失，这种方法需要基金经理管理两只组合（一只名义组合、一只实际组合），那么就需要更高的管理水平。

从方程(10.2)中可以很容易看出，基金经理和交易员能通过相互合作对交易成本进行控制，以减少延迟和机会成本。

第一，延迟成本的产生往往是由于交易员需要检查交易清单，以了解其特征。然后，交易员需要选择经纪商和最好的交易市场。这经常需要交易员在得到交易单后进行研究。因此，应该持续地监控交易执行情况，并提前了解每个经纪商最适合执行哪一类指令操作，这样能够减少延迟时间。另一个导致延迟成本的原因是交易员常常没有理解基金经理的交易意图，因此，交易员需要更多的时间研究交易清单。例如，如果基金经理在刚开盘时将交易单交给交易员，潜在成本会变得很大。然而，基金经理与交易员更好的交流能够使交易员可以在得到交易单时就开始执行。

第二，如果经理与交易员相互合作，就可以避免机会成本。机会成本代表错失投资机会而损失的利润。然而，交易指令不能完全执行是因为缺乏流动性，或者是反向价格运动。如果交易员和基金经理合作，则很可能在交易前预测到交易指令潜在的困难，以便基金经理将资金投入到下一个更有吸引力的投资机会。这种情况下，用组合的账面回报去评估机会成本是不合适的。因为基金经理交易的已经不是相同的股票。最好的解决方案是采用经济学对机会成本的定义，就是错过最有吸引力的投资机会的成本。利用这种思想，很有可能将经济机会成本引入执行缺陷测度，如下：

$$IS = \underbrace{\sum_i \sum_j x_{ij} (P_{i0} - P_{id})}_{\text{延迟成本}} + \underbrace{\sum_i \sum_j x_{ij} P_{ij} - \sum_i \sum_j x_{ij} P_{i0}}_{\text{执行成本}} + \underbrace{\sum_k \alpha_k \left(\sum_i X_i P_{id}\right) R_k^*}_{\text{经济机会成本}} + \text{可见成本}$$

其中，R_k^* 代表下一个最有吸引力的投资的实际回报，α_k 是投资到每个投资方向 k 的初始资金比例，$\sum \alpha_k = 1$。

3. 评估交易表现

表现评估的主要目的是评价交易员的能力和表现。前面我们介绍了如何测量交易成本,以及辨别交易成本是在何处发生。可是,它不能够判断成本是否合理。合理的表现评估可以判断交易员是在交易过程中降低了成本还是增加了成本。另外,表现评估可以区分交易员是靠能力还是运气。不管产生多少成本,靠运气的交易员从长期看对基金是有害的,因为运气不可能总伴随一个人。

一个最常见的,但不正确的表现评估方法是利用基准进行比较。这种情况下,常常把平均执行价格与基准价格(如开盘价、VWAP、收盘价)进行比较。如果执行价格比基准价格更有利,则认为表现好,反之则认为是表现差。但事实上,基准分析的效果有限。它很难比较不同交易日和不同股票的交易之间的表现。此外,它不能真正地评价表现的好坏。例如,仅因为平均执行价格差于基准就认为交易表现差是不合理的。如果交易员在下跌的市场买入股票,收盘价可能是当日最低价。这种情况下,所有的交易价格都差于收盘价,但差异的产生是因为市场价格的移动,而不是因为交易员做错了什么。传统的基准比较方法会导致不恰当的结论,因为它没考虑市场条件、价格趋势和执行策略。一个更好的评估表现的方法是相对表现测量法(RPM),它给出执行的百分比等级。这是一个更直观的方法,它能比较不同股票和不同交易时间的表现。同时,它考虑了市场条件、价格趋势和特定的执行策略。后面我们将对其展开讨论。

交易后分析的表现评估阶段判断交易员在执行交易时是否表现出色。我们想找到测量参与投资决策执行的各个部门的表现的方法,以确定成本是否合理,以及产生这些结果的原因。差的执行能力也完全有可能出现低成本交易的情况,反之亦然。

4. 基准比较

常用的评价交易表现的方法是比较实际执行的头寸与基准之间价值的差异。对一只股票来说是:

$$\text{表现}_i = \sum_j x_{ij} P_{ij} - X_i P_b$$

对于一个交易清单来说:

$$\text{表现} = \sum_i \sum_j x_{ij} P_{ij} - \sum_i X_i P_b$$

其中,$x_{ij}>0$ 表示买进,$x_{ij}<0$ 表示卖出,P_b 是基准价格。

这里,因为表现度量的实际上是交易成本,所以正值表示交易执行差于基

准，负值表示执行优于基准。

这里得到的货币价值可以通过除以股数转变为每股的交易成本，除以指令执行时的市场价值转变为百分比所表示成本。如果平均执行价格好于基准，认为是好的表现，反之则为差的表现。例如，如果执行价格好于基准价格10个基点，就可以认为是表现好。但是这样的推理效果并不好，而且事实上也并不全面。它导致依赖于这一标准的从业者做出不合理的执行决策，不符合我们的最优交易的标准。

交易后分析所使用的价格基准可以根据时间分为交易前、日内基准和交易后基准。交易前基准包括前一日的收盘价、当天开盘价、决策价格和指令的到达价格。它们用作测量交易成本的参照。日内基准包括VWAP，以及开盘价、收盘价、最高价和最低价四个的平均价格(OHLC)。它们用于比较执行价格与日内平均价格。交易后的基准包括当天收盘价和未来某时的收盘价(例如1天或5天后的收盘价)。

交易前基准 交易前的基准衡量的是交易成本，而不是表现。例如，基金经理经常在收盘后做投资决定，因此决策价格往往是收盘价。交易单是交易员在第二天早上开始执行的。因此，这些基准可以提供延迟成本和执行成本，而不是表现。即使在决策价格和指令进入价格已知的情况下，他们也只能更好地衡量成本，而不是表现。前一日收盘价和当天开盘价只是分别代表决策价和市场进入价。交易前基准可被用来计算延迟和执行成本，但不能计算机会成本，除非与交易后价格结合起来。因此，交易前基准不能衡量表现。

日内基准 日内基准与表现测量最为一致。VWAP是交易当天市场公平价格的一个代表，而OHLC是对平均价格的一个很好的估计。在没有时间和交易量信息的时候，OHLC是一个很好的测度，而VWAP很难进行计算。在美国，普遍应用的是VWAP，而在世界上一些其他的市场OHLC依然流行。因为有时时间和交易信息很难获得。对于日内基准，普遍认为如果执行价格好于基准，就表现出色。但是，它不能用于所有的交易策略。例如，如果执行的目标是最小化风险敞口，与VWAP的比较就是毫无意义的。因为VWAP测度的目标是成本期望值的最小化，而没有充分考虑风险。

交易后基准 交易后基准用来测量交易指令产生的市场冲击。通过将执行价格与未来价格比较，我们可以发现是否发生均值反转现象。如果发生均值反转现象，说明交易指令容易产生市场冲击。然而我们知道，交易订单会产生市场冲击，价格会被永久和暂时市场冲击所影响，所以这一测量方法只是证明了我们已知的结论。有人认为，这种比较可以提供价格反转的大小，但是仍

然不能区分均值反转是合理的还是由差的执行能力引起的。市场价格的变化受很多因素影响。可是,即使价格的波动仅受指令影响,交易后比较也只能衡量暂时冲击,而不能测量全部的市场冲击,因为永久市场冲击与未来价格有关。另外,当基金经理在价格上升时买入,以及在下跌时卖出的时候,交易后分析看起来会很好。事实上,Perold 强调了这一缺点。单独使用时,交易后分析不能够提供任何成本的计算。然而,交易后分析能够用于衡量指数基金与某个基准指数的跟踪误差,但低估了交易的真实成本。

混合基准 许多从业者坚持采用混合基准,例如 30-40-30 基准,即开盘价的 30%、OHLC 的 40%加上收盘价的 30%。但是,这只是采用不同权重的 OHLC:开盘价的 40%,最高价的 10%,最低价的 10%和收盘价的 40%。因此,混合策略只有在表现出比传统基准更好的时候才会被采用。

5. 基准比较的局限

基准比较存在一些不足。它不能与具体的执行策略结合,不能用于不同股票的表现,有时还会被交易员所影响。即使表现用每股成本或基点表示,也仍是这样。

第一,如果基金经理和交易员采用委托监督,他们总会采用最优策略执行交易。但是,没有一个基准能够判断策略是否是最优的。有时,一个最优策略可能会得到较低的基准分数。同时,交易员会制定最小化成本、平衡成本与风险、改进价格的策略。VWAP 测度可以很好地估计公平市场价格,可用于成本最小化策略。基准评价不适用于通常会比单纯 VWAP 策略更有效的 VWAP 对冲策略。此外,没有基准能很好地估计价格改进和风险厌恶策略的公平价格。尽管这些策略为各自的目标提供了最好的选择,但平均执行价格和基准之间差异的变化非常大,它们无法评估操作表现的好坏。

第二,基准比较不能够进行不同股票比较,以及同一只股票的不同时期比较。例如,如果与 VWAP 基准比较,A 差于 VWAP30 个基点,而 B 是差 50 个基点。表面上看 A 的表现比 B 好,因为它有更低的差异数字。但这可能是由于指令规模大小、波动率不同,交易模式不稳定等等。此外,考虑 A 和基准的差异昨天是 30 个基点,而今天是 50 个基点。即使订单的大小一样,仍不能对其进行比较,因为它们都依赖于实际市场条件、交易模式和价格范围。例如,非常可能昨天的价格波动范围小于 50 个基点,那么 30 个基点成本的指令很可能是以当天最差的价格执行的;而如果今天 A 的价格波动范围是 5%,和基准 50 个基点的差异相比应该算是表现更好。不幸的是,这些单纯通过基准进行比较是不能发现的。

第三，交易的基准可能会被交易员所影响。例如，如果交易员知道开盘或收盘价是评估的基准，他们会集中在这些时间交易。这可能会使基金产生不必要的高市场冲击成本。这些情况下，交易员看起来表现很好，但基金的投资表现会很差。如果交易订单很大，并需要分散到多天来进行交易，交易员可以根据自己的表现加速或减速交易，以获得更好的基准分数。例如，如果交易员与某一基准比较，而下午的市场价格可能会使整个交易表现变差，交易员可能会停止当天的交易，而在第二天继续交易，因而增加了时间风险。相反，如果当天价格使交易员的表现更好，他们就会增加当日交易，而造成更高的市场冲击成本。最后，对于大订单和 VWAP 基准，交易员只要在当天交易占市场交易量较大比例的指令，就可以使他们的表现趋于 VWAP。因为随着订单规模的增大，执行价格就会接近于 VWAP。

10.1.2 相对表现测量

相对表现测量(RPM)的目标是计算价格差于某次交易的平均价格的交易数量。与 VWAP 相同的是，它能与实际市场的价格变动和交易活动相结合。但是它相比于 VWAP 有一些优点，因为它可以针对不同交易时间和股票进行比较。RPM 采用了学校测验学生成绩的百分比排名模型，并做了一些有意义的改善。

例如，考虑一个学生在一个测验中得 52 分，另一个学生在另一个测验中得 117 分。在没有其他信息的时候，很难说只凭借考试分数来评价学生成绩的好坏。但如果我们知道第一个学生的成绩处在所有学生的前 5%，第二个学生处于最后的 10%，我们就可以说第一个学生的成绩是非常好的，而第二个学生成绩很差。我们不需要其他信息来做比较。

RPM 的计算过程为把平均执行价格与交易期内所有市场活动相比较，计算差于平均执行价格的市场活动的百分比。例如，对于买入，RPM 计算高于执行价格的交易的百分比；而对于卖出，计算低于执行价格的交易的百分比。在计算 RPM 时要同时根据市场成交量和交易次数两个指标，以避免由于极端价格时大单交易造成的潜在偏差。具体计算如下：

$$RPM_{(\text{成交量})}=\frac{\text{价格差于执行价格的总成交量}}{\text{总市场成交量}}$$

$$RPM_{(\text{交易数量})}=\frac{\text{比执行价格差的交易数}}{\text{总交易数}}$$

更具体的，如果 P^* 是指令的平均执行价格，则买入的 RPM 的计算是：

$$RPM_{(成交量)}=\frac{\text{sum}_{(成交量)}\text{ for } P_i\geqslant P^*}{\text{sum}_{(成交量)}}$$

$$RPM_{(交易数量)}=\frac{\text{sum}_{(交易数量)}\text{ for } P_i\geqslant P^*}{\text{sum}_{(交易数量)}}$$

卖出的计算与之相同，除了我们评估的是 $P_i\leqslant P^*$ 的活动。因此，平均 RPM 是：

$$RPM=\frac{1}{2}(RPM_{(交易数量)}+RPM_{(成交量)})$$

RPM 比基准分析要好，因为它可以进行不同时间和不同股票的交易之间的比较。百分比可以用于横截面的比较，而基准不能。这使得 RPM 成为一种更有效的表现评估方法。如果交易员的表现用基准来比较，A、B 两只股票的交易分别差于基准 30 个和 50 个基点。在没有其他信息的情况下，我们不能说两个交易员的表现是好还是差。然而，如果股票 A 的 RPM 是 10%，而 B 的 RPM 是 95%，我们可以说 A 表现差，而 B 表现好。这表示交易员在交易 B 时比交易 A 时的表现好。我们只需要计算 RPM 就可以了。

1. 计划交易策略的 RPM

经理经常给交易员具体的交易指令。在这种情况下，用 RPM 来衡量交易员的表现是不公平的，因为经理的指令确定了交易员执行大多数交易的方式和时间。例如，积极的策略会在市场下跌时买进，而产生低的 RPM，在市场上涨时卖出，而产生高的 RPM。消极的策略则正好相反。因此，为了区分交易员的表现和基金经理的指令，必须计算基于具体交易策略的 RPM：

$$RPM_i^*{}_{(策略)}=\sum_j\frac{x_{ij}^*}{X_i}RPM_j$$

RPM_j 是 j 时期的测度，x_{ij}^* 代表实际执行策略。加权 RPM 提供了交易员在被基金经理要求交易时表现的分析。

2. 实际交易策略的 RPM

假设投资者给交易员下达了买入指令，要求在中午之前完成交易。结果交易员偏离了交易策略，在整个交易日完成了交易。虽然交易员最小化了市场冲击成本，决策依然增加了基金的交易成本，因为他在收盘前以较高价格买入了一些股票。这样交易员的 RPM 很可能在 50%附近，但是却没有揭示出交易员偏离策略指令而引发的高成本。

为了考虑这一点，并将交易员的决策能力量化，我们定义一个新的 RPM 来确定产生价值增加的交易数量：

$$价值增加=\frac{RPM_{(x_i^*)}-RPM_{(x_i)}}{RPM_{(x_i^*)}}=\frac{\sum_{j=1}^{n}(x_{ij}^*-x_{ij})RPM_{ij}}{RPM_{(x_i^*)}}$$

价值增加测量了 RPM 中由于偏离决策而引起的价值增加百分比。如果交易员偏离了策略，并得到较好的价格，他们能提高表现，增加的价值大于零。但是，如果偏离策略导致了更差表现，增加的价值则小于零。容易看出，如果交易员服从策略，增加的价值为 0，因为 $x_{ij}^*-x_{ij}=0$。在 $x_{ij}=0$ 时，我们假设交易员在每一期都获得平均价格，并计算各期的 RPM_{ij}。对于 15 分钟、30 分钟的交易时间间隔来说，这一假设是合理的。这一测量方法可以判断交易员的执行决策是否增加了价值。

很多人会说，当可以获得更好的价格时，为什么要以平均价格成交呢？当然，每个人都想获得好于平均的表现，但是这样会增加额外的风险。数量化的 RPM 测度可以很容易辨别交易员是否在承担不必要的风险。那些获得高测量值和低测量值一样多的交易员就在赌博，而能够一直都有好的表现的交易员就是好的交易员。

3.定性分析

到目前为止，我们的表现分析都是基于数量的测度。但是，很可能两个数量相近的测量值的表现的性质差别不大。例如，RPM 为 72%和 77%的两个交易的实际表现可能非常接近，甚至 72%和 67%的也如此。由于市场噪音，这些比较并不精确。为了简化，我们可以将 RPM 的测量值概括为几个大类（例如好、一般、差）。这样，通过计算成交量的 RPM 和交易数量的 RPM 的平均值，我们可以将定量 RPM 转换为定性 RPM。

具体步骤如下：

第一步：计算 RPM 的平均值

$$RPM=\frac{RPM_{(成交量)}+RPM_{(交易数量)}}{2}$$

第二步：转换为定性指标。

$$定性指标=\begin{cases}极好 & 80\%<RPM\leqslant 100\% \\ 好 & 60\%<RPM\leqslant 80\% \\ 一般 & 40\%\leqslant RPM\leqslant 60\% \\ 差 & 20\%\leqslant RPM<40\% \\ 最差 & 0\%\leqslant RPM<20\%\end{cases}$$

这样，基金经理就可以很容易地评价交易员的表现了。

4. RPM 与 VWAP 的比较

在前面所有评价交易表现的方法中，VWAP 与 RPM 最接近。它们都利用一个交易员的平均执行价格与其他所有市场活动相比较。VWAP 的平均值是 RPM 的 50%分位数。

VWAP 与 RPM 也有着相同的缺陷。首先，当交易量占日总交易量的比例很高的时候，交易的平均价格就会接近于 VWAP，而 RPM 会接近于 50%。因此，从这些数据里很难得到有用信息。然而，如果我们能够将 RPM 与成本的期望相结合，就可以比较精确地判断交易表现。其次，如果订单交易在很短的时间内完成了，那么与整个交易日活动相关的测量方法都可能包含了交易期限之外的价格移动，而这些价格移动与交易操作质量无关。

RPM 相比于 VWAP 也存在着一些优点。首先，正如前面提到的，VWAP 测量的是交易价格和 VWAP 的差异，但是它没有提供任何其他有关的信息。它很难比较不同股票和不同时间的交易表现。有时，它甚至很难比较同一只股票在同一天的交易表现。例如，考虑一个高于 VWAP 价格 10 个基点的交易操作和一个高于 VWAP 价格 15 个基点的操作相比，我们知道第一个好些，但好多少？由于市场波动，10 个基点的交易很可能是比 15 个基点好一点。但是，10 个基点在今天是好的表现并不意味着在明天也是好的表现。RPM 对于不同股票和交易期是一致的。我们知道 65%的 RPM 是好的表现，这一点不会因为不同交易日而改变。

另一个 RPM 优于 VWAP 的地方是，VWAP 的分布是不对称的。如果 VWAP 是对称的，则 50%的交易价格和成交量高于 VWAP，而 50%低于 VWAP。但是一般情况下，大部分交易处在 VWAP 的一边。因此，差于 VWAP 基准并不一定意味着表现不好。有时，虽然交易员差于 VWAP，但仍好于大多数其他交易员，而 VWAP 无法对此区分。

考虑以下情况。一只股票以相对小的数量在三个价格增量交易。但是在收盘时，一个大单以更高价格进行交易，结果，除了那笔大额交易，所有市场价格都低于 VWAP。最终导致 75%的交易处在 VWAP 一边，而 25%的交易处在 VWAP 的另一边，分布很不对称。如果一个交易员以平均执行价格高于市场 75%完成了买入交易，但仍低于 VWAP。很差的表现却优于基准，利用 RPM 则不会出现这种情况。

10.1.3 交易后分析过程

交易后分析过程不仅仅是应用基准进行交易评价。首先，必须理解经理和交易员的目标，然后我们可以测量成本和评价执行。合理和良好的交易后分析有助于投资者在未来交易改善交易质量，提高投资回报。具体地说，可以通过如下七步进行：

第一步：评估执行决策

交易后分析的第一步是判断执行决策是否是最优的，这里我们要回答两个问题：

- 策略是否在有效交易边界上？
- 策略是否满足我们可接受的执行目标？

要判断策略是否最优，我们只需要判断它是否在一定的风险条件下成本最小，或者是在一定成本期望下风险最小，即是否处于 ETF 上。因为经理和交易员不知道市场条件未来是什么样子的，他们需要利用他们的期望来做出最好的判断。

如果 $E(\Omega)$是交易时段内的期望市场条件（包括日交易量期望、买卖压力、交易模式、价格趋势等等），$\theta_k=(\hat{\varphi}_k,\hat{\Re}_k)$是策略 x_k 的期望成本和风险的估计。于是，我们需要判断 $\theta_k=(\hat{\varphi}_k,\hat{\Re}_k)$是否满足两个条件：

- 是否 $\hat{\varphi}_k=\min\ \varphi(x_k\,|\,E(\Omega),\Re(x_k)=\hat{\Re}_k)$
- 是否 $\hat{\Re}_k=\min\ \Re(x_k\,|\,E(\Omega),\varphi(x_k)=\hat{\varphi}_k)$

这一分析可以用图 10.1 表示，使用决策时的期望市场条件得到 ETF。图 10.1 中我们绘出了 ETF 和两种策略 A 和 B。A 是最优策略，而 B 不是。投资者选择策略 B，不管执行结果如何，都不是最优执行策略。

很多时候，负责交易后分析的分析师不知道基金经理的具体意图。但是，只需要知道策略 x_k 是否是最优，并在 ETF 上就够了。Almgren 和 Chriss 的有效交易边界不但可以作为制定最优交易策略的工具，也可以作为交易后分析的重要工具。

第二步：使用扩展 IS 方法测量成本

交易后分析的下一步需要测量实际交易成本，并辨别成本从何而来。这里可以通过扩展的执行落差法来进行计算：

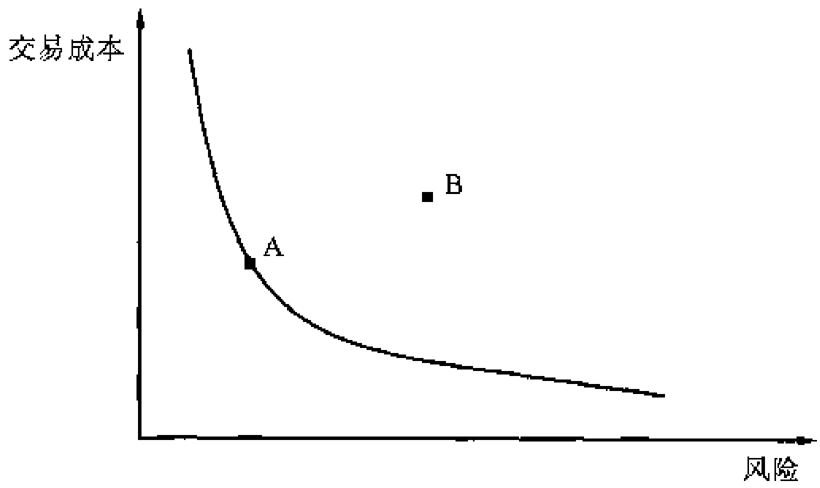


图 10.1　有效交易前沿

$$IS = \sum_i \sum_j x_{ij} (P_{i0} - P_{id}) - \sum_i \sum_j x_{ij} P_{ij} - \sum_i \sum_j x_{ij} P_{i0} + \sum_i \Big(X_i - \sum_j x_{ij}\Big)(P_{in} - P_{id}) + \text{可见成本}$$

以及，

$$\varphi = \underbrace{\sum_i \sum_j x_{ij} P_{ij} - \sum_i \sum_j x_{ij} P_{i0}}_{\text{执行成本}} + \underbrace{\sum_i \Big(X_i - \sum_j x_{ij}\Big)(P_{in} - P_{id})}_{\text{机会成本}}$$

第三步：执行后成本评估

在制定执行策略的时候，投资者并不知道市场条件将会如何变化，所以他们用预期中最可能出现的情形。但是，我们知道实际条件会和预期产生差异。因此，具体的策略在实际市场条件和期望市场条件下的成本就会不同。所以在实际评估执行表现时，我们需要比较实际市场条件下的交易成本与之前估计的成本，否则无法做出公正的评价。例如，考虑最优执行策略的预期的成本为 5 个基点，交易员服从了基金经理的指示。可是，因为实际市场条件缺乏流动性，实际成本为 25 个基点。用实际成本和正常交易量下的预期成本比较，并认为成本高出 20 个基点是不公平的，因为预期中的成本是不可实现的。只有在实际市场条件下计算成本才是公平的。这一过程不再需要最优化，只需要计算成本和风险。我们把这一成本估计视为执行后成本估计。这一估计如下：

$$E[\theta^*(x_k) \mid \Omega^*] = (\varphi^*, \Re^*)$$

这里 Ω^* 是交易期内的实际市场价格。

第四步：实际交易成本与执行后估计成本的比较

在给定市场条件下，我们可以比较实际交易成本与预期的交易成本。令

φ为实际的执行成本和机会成本，交易后测量的交易成本为：

$$E[\theta^*(x_k)|\Omega^*]=(\varphi^*,\Re^*)$$

于是，实际与期望的差别为：

$\Delta\varphi$=实际成本－预期成本

即

$$\Delta\varphi=\varphi-\varphi^*$$

在上面的几个方程中，负值表明成本低于估计值，说明交易中节省了成本。因为成本和风险的估计值随股票变动很大，很难比较不同股票间的差异。因此，可以根据期望成本和风险的参数将成本正规化：

$$\delta_i=\frac{\varphi_i-\varphi_i^*}{\Re_i^*}$$

这样，就得到了正规化的差异。为简洁起见，我们假设该参数服从标准正态分布。然而在这一步我们仍然不能确定表现是由于价格波动，还是来自交易员的能力。

第五步：使用 RPM 测量执行表现

接下来是评估交易员的表现。正如我们前面提到的，基准比较方法是不足以满足我们的需要的，因为它不是一个一致的方法。一个 10 个基点的测量值对于一只股票可能是好的表现，对于另一只股票则是差的表现。出于这些原因，RPM 比基准比较更为有效。例如，如果 RPM 大于 80%，我们可以确定交易员表现非常好。如果 RPM 小于 20%，我们可以确定交易员表现差。这里没有什么值得辩论的。我们需要确定交易员是否获得市场的公平价格，是增加还是减少了基金价值。这个目标可以用加权的 RPM 完成。

$$RPM_i(\text{策略})=\sum_j\frac{x_{ij}^*}{X_i}RPM_j$$

其中 x_{ij}^* 表示实际执行策略。

基金经理可以确定，当交易员的 RPM 值为 40%～60%时，他们获得了公平市场价格，高于 60%的可以认为获得了好于市场的价格，低于 40%的获得差于市场的价格。

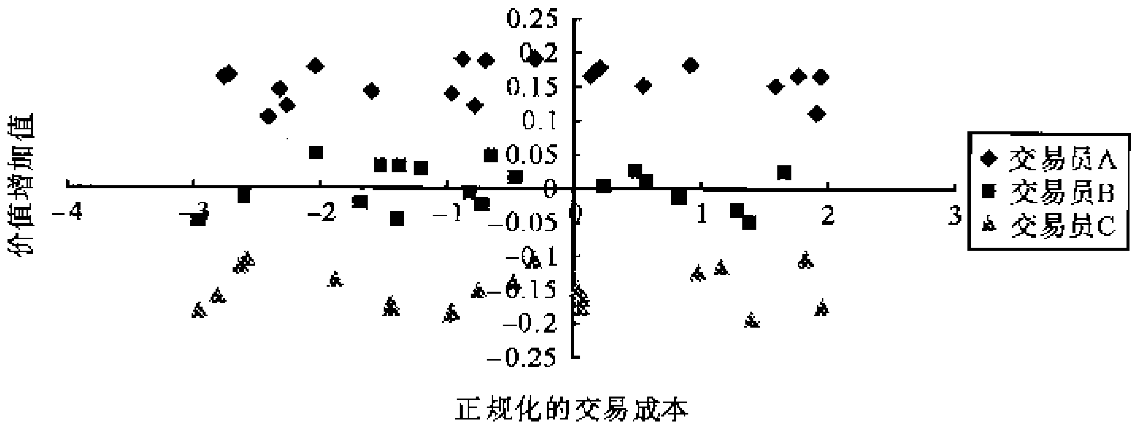
$$\text{价值增加}=\frac{\sum_{j=1}^{n}(x_{ij}^*-x_{ij})RPM_{ij}}{RPM(x_i^*)}$$

其中 x_{ij}^* 表示实际执行策略，x_{ij} 表示具体执行策略。

价值增加的测度衡量了交易员的实际决策能力。基金经理给交易员制定具体的策略进行交易。当交易员偏离原有交易策略的时候，可能是因为他们觉得当前市场价格更好。因此，通过比较与策略的偏离，我们可以评估交易员的决策能力所带来的价值增加。正的价值增加值表示好的决策能力，负的价值增加值表示差的决策能力。需要注意的是，分析交易员的决策能力时，需要充足的观察数据，这样才能过滤掉匆忙决定所产生的结果的影响。如果交易员没有偏离策略，也就是说，他们的偏离策略的价值增加值为 0。这显示了他们既没有增加成本，也没有减少成本。

第六步：比较价值增值的成本差异

只要画出正规化的交易成本和交易员的价值增加值的图（如图 10.2），投资者就能精确评估由交易员的交易决策导致的成本。在图 10.2 中，上半边的点是交易员增加交易价值的点，下半边的点是交易员减少交易价值的点。右



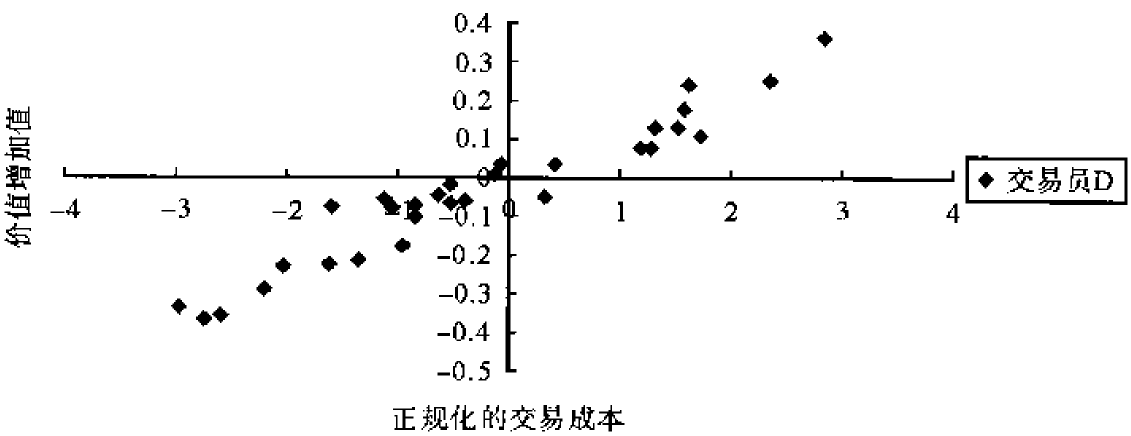


图 10.2 交易成本归因分析

半边的点表示成本低于预期，左半边的点表示成本高于预期。这使投资者很容易辨别交易超过或低于预期成本的原因。也就是说，它是由执行能力造成的，还是价格变动造成的。有高增加值的点是由于好的执行能力，低增加值的点是由于差的执行能力。

图 10.2 中我们给出两个例子。第一张图描绘了三个交易员的价值增加值。A 偏离了指定决策并做了正确决定，表现很出色。B 偏离了指定决策，但做了错误决策，因此表现很差。C 遵守指定决策，因此得到了平均表现。第二张图展示了 D 的交易执行结果。这个交易员做出的有降低成本的好决策，也有增加成本的差决策。但是这并不代表他得到了平均的表现，因为如果不能够改进交易成本，那么我们更希望交易员能够按照指定交易策略执行交易。所以，为了更好地分析结果，不仅应该分析增加值的平均值，还要分析标准差。

第七步：记录交易员的表现

保存交易员的实际交易成本和表现的历史数据是很重要的。这样，基金经理和交易员就能够得到真实、精确的交易成本和表现信息。同时，这些信息能够被投资者用来减少延迟成本，改进交易操作，进而提高投资组合的回报。

10.2 如何选择和应用算法交易

10.2.1 Kissell 的决策框架

在算法交易出现以后，各式各样的算法不断地涌现出来。因此，当投资者选择使用算法交易的时候，往往会面临着很多的选择。那么，如何选择合适的交易算法呢？Kissell 和 Malamut(2005)给出了一个分为 3 步的决策方法：

- 选择一个合适的价格基准；
- 确定交易风格；
- 确定如何根据市场条件的变化进行调整。

选择交易算法的第一步是选择一个适合的价格基准。选择基准的目标应该和投资策略结合起来。例如，如果投资者是一个消极型的投资者，那么他对于决策价格敏感度就比较低，这样投资者往往会选择 VWAP 作为交易的基准，因为 VWAP 代表着市场上平均的交易价格。如果投资者做投资研究时用的决策价格是每日收盘价，那么交易的价格基准则会是前一日收盘价或当

天收盘价。

选定价格基准之后，投资者可以根据所选的基准得到有效交易前沿。最优交易策略是给定风险下成本最低的交易策略，或者给定成本条件下风险最小的交易策略。有效交易前沿就是最优交易策略的交易成本和风险的组合。很显然，不同的基准价格意味着所度量出来的交易成本和风险特征将会不同。因此，对于不同的基准价格，我们将会得到不同的有效交易前沿。例如，如果将开盘价和前一日收盘价相比较，开盘价包含着隔夜风险，所以用开盘价衡量的交易成本将会比用前一日收盘价衡量的成本有着更低的风险。

通过确定价格基准，投资者可以确定有效交易前沿，而有效交易前沿代表着最优交易策略的组合。接下来，投资者要从有效交易前沿上选择适合自身交易目标的策略，那么他就需要确定风险厌恶度，或者叫交易紧迫度、交易风格。如果投资者不急于完成交易，那么他就会选择较为消极的交易策略。这样通常会拉长交易的时间，减少市场冲击成本，但是要承担更多的价格增长和波动的风险。相反，如果投资者急于完成交易，那么他就会选择相对积极的交易策略，进而减少时间风险，但是增加了市场冲击成本。我们知道，最优交易策略的目标函数为：

$$\min: E(TC)+\lambda \cdot Var(TC)$$

其中的 λ 就是对交易成本风险的惩罚因子，表示风险厌恶度，TC 即交易成本。较大的 λ 代表着较高的风险厌恶度，进而产生较为消极的交易策略。在算法交易中，投资者可以通过选择自己的风险厌恶度来确定合适的交易风格。

交易算法决策的第三步是确定根据市场条件如何进行调整。在交易之前，人们不可能预见到市场环境所有的变化，例如不利的价格移动、流动性的高低、交易对象的公司的新闻等等。因此，投资者应该确定根据市场变化的调整方式，以便更好地适应新的市场条件。例如，如果市场流动性很高，那么投资者可以选择更为积极的交易方式，从而可以降低时间风险。相反，如果市场流动性很差，投资者则可以选择更为消极的交易方式，减少市场冲击。此外，投资者也可以根据价格变化进行调整。Kissell 给出了两种比较典型的方式：AIM(aggressive in money)和 PIM(passive in money)。AIM 的做法是在市场出现有利的价格移动的时候加快交易速度，在出现不利的价格移动时则减缓交易速度。这样的做法可以理解为在价格有利时注重把握交易机会，而在价格不利时选择等待。PIM 则与 AIM 完全相反，在市场出现有利的价格移动的时候减缓交易速度，而在出现不利的价格移动时，则加快交易速度。这样

做的目的是在价格不利时限制交易损失，在价格有利时则等待价格能够继续向有利的方向变动。这两种方式本身不存在优劣之分，交易员可以根据自身的喜好进行选择。

10.2.2 针对不同类型的股票进行操作

算法交易可以理解为通过计算机实现特定的交易策略，或者说是订单分割策略。通过算法交易进行操作的时候，计算机系统会按照预先设定的方式向市场下单。但是，当市场流动性不好的时候，算法交易未必能够降低交易成本。因为在某些情况下，即使是小的订单，如果不能被市场所消化，仍会产生相当大的市场冲击，所以等待流动性出现比分割订单更有利于减小市场冲击。

通常情况下，一个交易员的买进交易会在如下几种方式中进行选择：

- 下单；
- 等待对手方的叫价；
- 不急于下单，等待更小的买卖价差或者流动性机会。

很显然，后两种选择相比第一种会增加交易的风险。同时，承受这些风险的好处是降低交易的成交价差和冲击成本。在这里，我们需要考虑的变量是电子竞价系统的买卖价差、波动性、流动性。这里，我们将股票的交易行为根据两个简单的比率做出划分。它们就是最小价格增量(tick size)对波动性的比率，以及买卖价差和波动性的比。我们利用这两个比率将股票分为三类，见表10.1。

表10.1 股票交易模式的一个分类

最小价格增量/波动性	买卖价差/波动性
高	高
低	高
低	低

对于第一类股票，最小价格增量对波动性的比率高，而且买卖价差和波动性的比也比较高。因此在短时间范围内，波动性对交易成本的影响相对并不明显。这种股票的特点是在流动性好的交易日里，大量的报价都处在最好的买进价和卖出价附近。对于这种股票来说，最好的交易策略就是耐心地等待系统更好的报价出现。有时候为了减小买卖价差需要等待几个小时。算法交易对这类股票的交易可以带来很高的附加价值，因为计算机能够持续地监视

订单簿上的流动性状况，用供给和需求之间的相对关系来判断未来的价格走向。如果算法交易判断价格会有一个反向的移动，那么就应该马上下单，否则就应该等待更好的报价。而对于订单分割策略，小的分割对这种股票并不适合，因为订单在订单簿上排队的时间一般较长。交易员应该等待机会，并做出正确判断，尽快下单。由于大量的交易在最优的买进价和卖出价附近成交，所以加入更多的流动性不会影响到供需之间的不平衡，订单不会产生很大的市场冲击。

第二类股票一般是中盘股，缺乏流动性，同时有着较大的买卖价差。很多人建议不要用算法引擎来交易这类的股票，因为很难达到好的表现。而另外一种观点则认为，对于这种股票来说，内部消息泄露会对交易产生很大的影响，而使用电子化交易至少能够保证交易的隐蔽性。对于这种股票，正确的策略是投机性操作，在极短时间内把握住小买卖价差和高流动性的机会。如果这类股票有着较小的买卖价差和较好的流动性，就说明该股票处于公平价格。因此，对于这种股票，等待更好的报价所承受的风险并不会得到充分的回报。随着时间的流逝，如果保持一定的交易速率，价差会逐渐变大。发现交易机会的时候可以加快交易速率，但是必须注意内部消息泄露会导致供需的失衡，从而导致价格的反向移动。

最后一类股票流动性更高，低报价最小增幅与波动率比和低成交价差与波动率比的股票一般是蓝筹股，定价比较有效。这类股票很容易通过人工和算法进行交易，可以被称作是“低接触交易”或商品化的股票。这类订单需要很好的分割策略，以防止较高的风险。因为把握成交价差所得到的回报相对较小，不足以抵消等待价格波动所带来的风险，所以订单在订单簿上等待的时间往往比较短。对这类股票进行算法交易能够增加交易的价值，因为它能够节省人力，使交易员有更多精力关注难度更高的交易。

10.3 算法交易的影响

10.3.1 算法交易能够节省成本吗?

由于算法交易属于各个金融机构私营的业务，所以公开的数据很少。2005年，ITG公司的Domowitz和Yegerman对算法交易的交易成本进行了

研究。他们利用2004年超过40家机构的250多万个交易订单进行了分析，样本中的交易总量达到了约100亿股。其中，相当大一部分交易使用的是VWAP算法，达到了订单总数的40%。

在考虑了交易难度、不同市场、交易方向和市场波动等因素之后，他们认为，以执行落差为交易成本的衡量标准的话，算法交易比其他交易方式的成本要低。但是，算法交易的优势只存在于平均日交易量10%以下的交易。这也说明了算法交易在大额交易的情况下并不优越。另外，VWAP算法的表现也相当地好，和基准的平均差异只有2个基点。

相比于VWAP，Implementation Shortfall的表现随订单规模的增长下降得更快。这一方面是由于VWAP算法随着订单规模的增长更容易接近基准，另一方面，对于大订单，Implementation Shortfall算法也难以避免市场冲击对交易成本的影响。

Domowitz和Yegerman还认为，要表现出充分的优势，交易员必须坚持使用算法进行交易。他们从数据中发现，算法交易产生的交易成本的分布是对称和发散的，表现出了肥尾的特性。这说明算法交易的表现并不稳定，而且好的结果和差的结果基本一样多。因此，他们认为依靠小样本交易数据对特定算法进行判断很容易出现错误，交易员要依靠大数定律来实现算法交易的优势。

针对于不同的供应商，Domowitz和Yegerman发现对于小规模的交易来说，他们之间的差异并不大。但是，随着订单规模的增加，不同供应商间会出现一些差异。差异表现在交易成本的标准差，也就是稳定性上。

10.3.2 算法交易的问题和影响

随着市场电子交易能力的提高，利用算法交易和其他电子下单的方式在不断地增加。市场参与者认识到这种变化并不是来自市场结构本身，而是为了市场竞争的需要。不同的市场主体都需要使用更多电子化的交易技术方式来隐藏和寻找流动性，以及减少交易成本等。自动化的交易系统当中的下单、修改和订单撤销的操作都会比人工方式要多得多，因此订单流会导致网络当中的通信量猛增。随着算法交易的迅猛发展，这会对电子化交易的网络以及股票交易市场的处理能力带来很大的负担。有数据显示，近几年美国股票市场的数据量呈现了指数增长。Aitc Group预计到2011年美国股票市场每天的通信消息数量会达到每天12亿条。此外，投资者通过算法交易可以任意选

择经纪商和交易市场进行交易。通过订单的分割和对交易市场、经纪商的选择和分配，投资者很容易隐藏自己的交易意图，而ECN和交叉网络当中都是进行匿名交易。这样的匿名性和交易隐蔽性给证券交易的监管也带来了严峻的挑战。

算法交易的发展过程中，经纪商开发交易算法的能力逐渐会成为其竞争的主要手段之一。为了赢得更多的客户，经纪商不得不投资于算法交易系统的开发当中去。此外，各种OMS、技术、硬件和数据供应商也会参与到竞争当中，这些使得算法交易市场的竞争愈演愈烈。机构投资者也会开发一些私有的交易算法，以适应特殊的需要或者获得更高的投资回报。为了竞争的需要，算法交易开发商也需要逐渐地适应投资者的各种需求，使算法变得越来越灵活和具备定制化。但是，各个金融机构在投资于算法交易的软硬件的同时也需要考虑投入和收益，应该量力而行，选择合适的解决方案。

可以预见，算法交易在未来几年会逐渐地进入中国。它不但有利于投资者更好地进行投资交易，改善证券市场的效率，同时也是我国证券市场跟国际先进的证券市场接轨的需要。国内的券商之间会在算法交易领域进行激烈的竞争，以获得更多的经纪业务的市场份额。算法交易会成为我国证券市场的新亮点。

第十一章 配对交易

11.1 历史

配对交易起源于20世纪80年代中期。摩根斯坦利的宽客(quant)Nunzio Tartaglia召集了一批物理学家、数学家和计算机科学家，利用数量方法去发现股票市场的套利机会。他们采用复杂的统计方法建立了自动化交易系统，以数量化指标代替交易员的经验判断，发展了一系列的数量化交易方法，其中之一就是配对交易。1987年，他们利用配对交易为摩根斯坦利创造了5 000万美元的利润。但不幸的是，在经历了两年损失之后，该团队最终解散了，但配对交易的方法却在华尔街流传开来。实际上，根据Thorp(2003)的记载，摩根斯坦利的Garry Bamberger才是配对交易的真正发现者。但由于感觉公司没有给予相应的报酬，他于1985年离开了摩根斯坦利，加入了Thorp的公司。

11.2 数学基础

11.2.1 基本统计知识

平均值是一组数据的平均状态，分为算数平均值和几何平均值。算数平均值是用数据的和除以数据的个数。计算公式如下，后面有关的数学描述均

以 x_t 表示 t 时刻数据的值：

$$\overline{X}=\frac{\sum x}{n}=\frac{x_1+x_2+\cdots+x_n}{n}$$

几何平均值是数据乘积的数据个数次方，计算公式如下：

$$\overline{X}=\sqrt[n]{x_1\cdot x_2\cdot\cdots\cdot x_n}$$

方差是用来衡量数据离散程度的数据，计算公式如下：

$$s^2=\frac{\sum(x-\overline{x})^2}{n-1}$$

标准差是方差的平方根，计算公式如下：

$$s=\sqrt{\frac{\sum(x-\overline{x})^2}{n-1}}$$

正态分布是所有分布中最常见、使用最广泛的分布。在工业和商业中，许多变量都服从正态分布。由机器生产的绝大多数产品也服从正态分布。误差正态曲线的发现，通常归功于数学家高斯，他认为对物体反复测量的误差通常呈正态分布。因此，正态分布有时也称作高斯分布或误差正态曲线。正态分布具有以下性质：

1. 它是一种连续型概率分布。
2. 它关于均值对称。
3. 它以横轴为渐近线。
4. 它为单峰分布。
5. 它是一簇曲线。
6. 曲线下的面积为 1。

正态分布是对称的。以均值为中心，左右两边相互对称。由于对称性，两侧的概率值相等。有时正态曲线也被称作钟形曲线。它是单峰曲线，因为所有的值只在曲线中心聚集。正态分布实际上是一簇曲线。每组不同的均值和标准差对应着一条不同的正态曲线。而且任一正态曲线下的总面积都为 1。曲线下的面积表示概率，所以一个正态分布所有概率和等于 1。此外，由于正态分布是对称的，所以均值每一侧的面积都等于 0.5。

正态分布可由两个参数予以描述：均值和标准差。每组均值和标准差确定了一个正态分布。正态分布的概率密度函数为：

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

对于正态分布，偏离均值小于一个标准差的概率为68%，偏离均值小于两个标准差的概率为95%，偏离均值小于三个标准差的概率为99.7%。

11.2.2 平滑技术

利用一些技术可以对平稳的或不包含明显趋势、周期性或季节性的时间序列数据进行预测。这些技术被称为平滑技术，原因在于这些技术得到预测值的基本原理是将时间序列数据中的不规则波动效应“平滑”掉。这包括平均模型和指数平滑模型。

1.平均模型。平均模型由简到繁分为三种：简单平均、移动平均和加权移动平均。

(1)简单平均。最基本的平均模型是简单平均模型。在这个模型中，第t期的预测值等于给定的前若干期数值的平均值，公式如下：

$$F_t=\frac{x_{t-1}+x_{t-2}+\cdots+x_{t-n}}{n}$$

(2)移动平均(MA)。移动平均数是一个对每一个所考察的新时期都进行更新或重新计算的平均数。每个移动平均数都利用了最新的信息。公式与简单平均相同，只是数据是更新后的数据。但移动平均有以下缺点：①难以选择计算移动平均数的最优步长。②移动平均数并不总是能够根据诸如趋势、周期或季节性等时间序列效应做出调整。为了确定计算移动平均数的最优步长，需要根据几种不同的步长进行预测，然后比较相应的误差。

(3)加权移动平均(WMA)。预测者可能希望对某些时期的数据赋予比其他数值更高的权数。如果某些时期得到的权数不同于其他时期的权数，则移动平均称为加权移动平均。计算公式如下：

$$F_t=\frac{w_{t-1}x_{t-1}+w_{t-2}x_{t-2}+\cdots+w_{t-n}x_{t-n}}{\sum\limits_{i=t-n}^{i=t-1}w_i}$$

其中w_i为赋予该时期数值的权重。

2.指数平滑模型(EWMA)。采用指数平滑方法进行预测时，对以前时期数据的加权方法是重要性依指数形式递减。指数平滑方法是：将当前时期的实际值x_t乘以一个介于0～1之间的数值，该数值称为α，将当前时期的预测

值 F_t 乘以$(1-\alpha)$，然后把两个乘积相加。计算公式如下：

$$F_{t+1}=\alpha X_t+(1-\alpha)F_t$$

α 的数值由预测者自己确定。这个程序的实质在于新的预测值是当前预测值和当前实际值的组合。如果选择的 α 小于 0.5，则对实际值赋予的权重小于对预测值的权重。如果选择的 α 大于 0.5，则对实际值赋予的权重大于对预测值的权重。

11.2.3 平稳时间序列

平稳过程：对于随机过程$\{x_t:t=1,2,\cdots\}$，如果对于每一个时间指标集 $1\leqslant t_1<t_2<\cdots<t_m$ 和任意 $h\geqslant1$ 的整数，$(x_{t_1},x_{t_2},\cdots,x_{t_m})$的联合分布都与$(x_{t_1+h},x_{t_2+h},\cdots,x_{t_m+h})$的联合分布相同，那么这个随机过程就是平稳的。

协整理论：如果$\{y_t:t=0,1,\cdots\}$和$\{x_t:t=0,1,\cdots\}$是两个 I(1)过程，则一般来说，对于任意 β，$y_t-\beta x_t$ 都是 I(1)过程。但在某些情况下，存在不为 0 的 β，使得 $y_t-\beta x_t$ 有可能为 I(0)。也就是说，这一差值有固定均值、固定方差和仅取决于序列中任意两个变量之间时间间隔的自相关。如果存在该 β 值，则称 y_t 和 x_t 是协整的。在实际使用时，β 往往不能提前知道。所以先采用回归方法对 β 进行估计。回归采用式：

$$y_t=\hat{\alpha}+\hat{\beta}x_t$$

$\hat{\beta}$ 是 β 的估计。进行上述回归后，可得到残差 $\hat{u}_t=y_t-\hat{\alpha}-\hat{\beta}x_t$。对残差进行 DF 和增广 DF 检验，如果能够拒绝存在单位根的假设，就可以支持 I(0)的假设，只是这时的检验要考虑到 β 是估计值。如果得到支持 I(0)的假设，就可认为 y_t 和 x_t 是协整的。

11.3 配对交易方法

11.3.1 基本概念

在存在做空机制的情况下，投资股票市场从本质上讲，就是买入低估的股票，卖出高估的股票。但在无法准确估计股票价值的时候，很难分清哪只股票

是低估的，哪只股票是高估的。配对交易选择规避这一问题，这种交易方法采用相对估值的方法。具体说来，选择两只类似的股票，则两只股票的价差应维持一个稳定数值。当价差偏离稳定值较远时，价差倾向于回归到稳定值。对于这一思想背后的理论基础存在两种观点：一是遵从 APT 理论。根据 APT 理论，当两只股票有几乎相同的风险因素时，则这两支股票的预期回报应该大致相同，所以由噪音或个别事件引起的回报偏差最终会消失；另一种观点认为，这一现象是基于资金的轮动。这一观点认为，资金在相似的股票间是轮动的，当一只股票相对于一只相似的股票价格上涨过大，则资金持有者会认为上涨的股票相对被高估，因此会抛售上涨的股票并购买相对价格较低的相似股票，使得价差回归稳定数值。

利用这一思想，就可以在价差偏离稳定数值较远时，卖出价格相对高的股票，买入价格相对低的股票，以在价差回复到稳定数值时平仓获利。下面的部分将详细讲解配对交易的整个操作过程。概括说来，配对交易大致需要三个步骤：

1. 选择合适的股票对。这一步骤可以采用两种方法：一是股票的基本面分析，或者采用纯粹的统计分析方法。多数情况下是两种方法综合使用。

2. 在选择了合适的股票对之后，采用后面将要介绍的不同的统计方法，分析历史数据，得到不同方法需要的参数。

3. 得到所需参数后决定交易的起始和中止点、止损位置和交易时间。

后面的章节将按这一顺序逐一介绍每一步骤的详细内容。

11.3.2 基本术语

做空：指借入股票并在市场中将其卖出，平仓时从市场中买回相同数量的该股票，并将其归还借出者。做空与做多完全相反，它是在所做空股票的下跌过程中获利的。例如，股票 A 现价 30 元，甲借入该只股票 100 股，并将其在市场中卖出，获得资金 3 000 元。一段时间后，A 下跌到 25 元，甲平仓，从市场中以 25 元买入股票 A 100 股归还给 A 的借出者，花费 2 500 元。这样，甲从市场下跌中获利 3 000－2 500＝500 元。同样，如果平仓前 A 涨到 35 元，则甲需要花费 3 500 元买回 A 归还借出者，因此甲在股票上涨时亏损 3 500－3 000＝500 元。借出股票往往有时间或流动性的限制，所以做空往往不能像做多那样长期持有，并且常常有保证金的限制。因此做空与做多并不完全对等。

价差：是指配对的两只股票同一时刻的价格差值，其值可正可负。

市值：一个公司所有股票的市场价值。也就是用该公司股价乘以该公司在外发行的股票数目。

流动性：流动性是指交易活动的水平，它决定了买卖的容易程度和市场冲击成本的大小。具有流动性的证券，交易活动很活跃，执行交易的买卖价差很低。而非流动性的证券，交易活动很少，买卖价差很高，市场冲击成本较大。

市净率：用一个公司的证券市场市值除以其账面价值所得到的比率。也可以用每股市值除以每股账面价值而得到。该指标是一个测量证券价格便宜度和证券价值评估的比率。低市净率的股票一般认为是便宜的、有价值的股票。

市盈率：用公司的证券市场市值除以公司的利润得到的比率。

无风险利率：指一个无风险资产带来的收益率，一般指国债的利率。它是一个所有投资者都期望可以得到的基本收益率。根据现代资产组合理论和资本资产定价模型，超过无风险利率的收益率只能通过承担市场风险来获得。

有效市场理论：认为证券市场已经将所有的信息置于价格之中，因此对公开所得的信息进行分析，不能产生额外的收益。

随机游走：金融理论的经典假设，认为股票的收益率之间彼此独立，并符合正态分布。也是有效市场理论的必然结果。

11.3.3 股票对的选择

单纯采用统计分析方法很难确定所需的股票对，这并不是因为统计方法没有效果，主要是因为所要分析的股票对太多，计算量过大，难以全部采用统计分析进行覆盖。举例来说，假设股市中有 1 500 只股票，则采用纯统计分析方法，需要分析的股票对将有 1 124 250 个，这样将耗费大量的时间。所以，股票对的选择一般先采用基本面的分析。

经验地看，处于同一板块的股票，往往有较大的相关性。所以，在选择股票对时，优先选择同一板块的股票。在此基础上，在板块股票之间进行统计分析，这样可以大大减少所要分析的股票对数。

统计分析的主要方法包括相关系数的计算或协整分析。相关系数的计算前面的数学基础部分已经介绍过了，这里不做赘述。对所选择的板块内的股票，每两只股票计算其相关系数。设定一个门槛值，相关系数高于此门槛值时，将此股票对加入进一步分析的范围。协整分析前面也已介绍过，这里谈一下针对股票对的使用方法。

理论上，股票价格的收益率符合随机游走。这一假定已被广泛用于期权定价理论。但随机游走是非稳序列，所以，我们的目的是要找到一对合适的股票对，如果这对股票价格的对数序列是协整的，那么它们的价差就是平稳序列，就可以进一步采用配对交易的策略。因此，对于同一板块的股票，可以对每一对股票价格的对数采用协整分析，如果价差通过平稳性检验，则将此股票对加入进一步分析的范围。

实际上，相关系数分析和协整分析适用于不同的交易分析方法。相关系数分析适用于较为简单的交易分析方法 1，协整分析适用于交易分析方法 2。后面将会详细介绍不同的交易分析方法。

11.3.4 交易分析方法和参数计算

目前，配对交易大致存在两种主要的交易分析方法。按数学工具的从易到难列述如下：

1. 简单交易法

简单交易法只需很少的数学工具。简单交易法认为，一对匹配较好的股票对的每日价差符合正态分布。选择股票对时只计算相关系数，不需协整检验。判断可用股票对时经验的作用大于数学工具的使用。一些完整的操作例子可参见 Whistler(2004)。简单交易法的思路很简单，既然假设股票对的每日价差符合正态分布，那么就可以通过历史数据计算该股票对的均值和方差。有了均值和方差，就可以用来设计交易了。如图 11.1 所示，这是一个平稳序列，序列中的每一点都符合均值为 3，标准差为 1 的正态分布。图中明显标出了一个标准差和两个标准差的位置。从图中可以看出，如果在高于均值两个标准差的位置采用价差卖出的方法，则当价差回归均值时，可以获得两个标准差的利润。同样，在低于均值两个标准差的位置采用价差买入的方法，当价差回归均值时，同样可以获得两个标准差的利润。

一个合理的问题是：为什么采用两个标准差？原因在于事件发生的概率。根据正态分布的理论，偏离均值小于一个标准差的概率为 68%，偏离均值小于两个标准差的概率为 95%。也就是说，偏离均值超过两个标准差的事件发生的概率只有 5%。这样，存在两个好处：一是在建立头寸后，发生亏损的概率只有 5%，有效地控制了风险；另一个好处是保证了交易的频度，如果把交易的门槛定得高于两个标准差，虽然亏损的概率更小，但交易的次数也会大大下降，这样做的结果就是，大多数时间没有交易，不足以弥补资金的机会成本。

当然两个标准差只是一个经验的数值，没有较好的理论证明其是最好的交易门槛。更优的交易门槛的选择会用到更为复杂的方法。对于简单交易法，往往采用两个标准差加上交易员的经验判断，经验往往占主要地位。

简单概括简单交易法的交易思想就是：当价差偏离均值达到两个标准差时买入或卖出价差（依偏离方向而定），当价差回归均值时平仓。

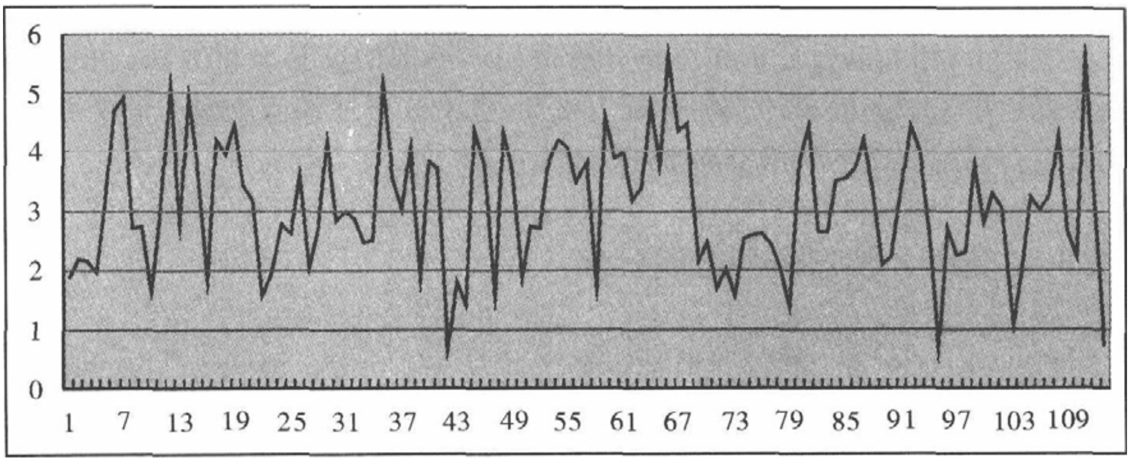


图 11.1　符合正态分布的平稳序列

2. 基于协整的交易法

正如前面所述，协整交易法基于随机游走的假设。这符合有效市场理论（EMH），认为股票价格的收益率符合布朗运动。如果能够找到协整的股票对则它们的对数价差是一个平稳序列，基于此，就可以设计交易方法。

假设有两只股票 A、B，其价格为 P_A、P_B，价格对数为 $\ln P_A$、$\ln P_B$，则其协整回归(cointegrating regression)为：

$$\ln(P_t^A)-\gamma\ln(P_t^B)=\mu+\varepsilon_t$$

其中 γ 为协整系数，μ 类似于价差均值，ε_t 为残差，是通过平稳性检验的平稳序列。这样，股票 A 和 B 就可以采用基于协整的交易方法。其思路是：如果 A 和 B 是协整的，那么买入一股 A 同时卖出 γ 股 B，在 t_1 和 t_2 时的情况为：

$$\ln(P_{t_1}^A)-\gamma\ln(P_{t_1}^B)=\mu+\varepsilon_{t_1}$$
$$\ln(P_{t_2}^A)-\gamma\ln(P_{t_2}^B)=\mu+\varepsilon_{t_2}$$

两式相减得：

$$\ln\left(\frac{P_{t_2}^A}{P_{t_1}^A}\right)-\gamma\ln\left(\frac{P_{t_2}^B}{P_{t_1}^B}\right)=\varepsilon_{t_2}-\varepsilon_{t_1}$$

由上式可以看出，左面恰为头寸在 t_1 和 t_2 间的获利或亏损情况，右边为两个残差的差值。因为作为交易规则，残差值 ε_{t_1} 和 ε_{t_2} 的值是由投资者选择的，所以通过选择适当的门槛残差值，可以使得收益不为 0（如果收益为负，就采用反向头寸，即卖出一手 A，买入 γ 手 B）。这样类似于简单交易法，价差偏离均值超过门槛时，进行价差买入或卖出（与偏离方向有关），当价差回归均值或达到其他门槛值时平仓。之所以可以采用这样的交易方法是因为，通过协整检验的股票对 A 和 B，其按照上面方法建立的头寸是平稳序列，拥有均值回归的性质。其值在均值附近的一个区间内震动，所以存在着从门槛值向均值的回归。所以关键是获得 γ 和 μ 并确定门槛。

实际操作时，采用股票 A 和 B 的历史数据进行回归，即可得出 γ 和 μ 数值。因此，该方法的难点在于确定门槛值。一般认为平稳序列 ε_t 符合正态分布，所以门槛往往采用偏离均值一倍或两倍标准差。方差可以通过 ε_t 的历史数据计算。当然，也不一定采用整倍方差，也可对历史数据进行检测。采用不同的门槛，根据历史数据计算不同的总利润值，然后比较不同门槛所获利润的大小，得到能获得利润最大的门槛值。可采用此门槛值进行交易。但这样使得交易过于依赖历史数据，如果现实数据和历史数据的性质有较大差别，则这样选择的门槛值可能大大影响交易系统的获利能力。

11.3.5 具体交易细节的设计

不论采用哪种交易分析方法，交易细节的设计都是类似的。首先，对于选择好的股票对，根据历史数据计算出价差的均值和方差（简单交易法），或通过回归计算 γ 和 μ（基于协整的交易法）。然后确定交易门槛值，相关的方法已经介绍过了。确定了门槛值之后，根据门槛值决定何时开始交易。但对于何时结束交易，观点不一。大致分为三种观点：

1. 在门槛值处开始交易，在回归均值处停止交易。

2. 在门槛值处开始交易，当达到反向的门槛值时停止交易。同时开始另一次交易。

3. 在门槛值处开始交易，当回归均值时停止交易。同时开始另一次交易，在价差达到门槛值时停止交易，并开始新的交易。

三种方式很难说哪一种更好，这必须在交易频率和单笔获利之间做出权衡。从单笔获利上看，第二种方法的单笔获利最高，第一和第三种方法相同。但第二种方法开始交易后，停止交易的时间明显长于另外两种方法，因为必须

先回归均值，才有可能达到反向的门槛。但这样就增加了资金使用的机会成本，其总盈利能力未必胜过其余两种方法。第三种方法的交易频率高于第一种方法，其是一种连续交易的方法。但该种方法亦存在机会成本问题。因为达到门槛的频率不能过高，否则门槛会被设置得过小。这样一方面会使单笔获利过小，未必能弥补交易费用；另一方面也使得价差向不利于头寸的方向移动的可能性增加，增加了交易的风险。因此设置的门槛值一般需要经历较长时间才能达到，而根据第三种方法，这段时间资金被用于配对交易，获利未必会弥补资金的机会成本。

除了上面三种结束交易的方法，配对交易还必须有止损条件。这是因为：第一，无法保证交易结束的目标一定会达到，即无法保证价差会回归均值或达到反向门槛值。比如两只汽车制造公司的股票，某天其中一家公司的股票价格大幅下跌，使得价差偏离均值达到门槛值，根据交易方法的设定，交易开始。但最终股价下跌的公司被破产摘牌，这样价差将永远无法回复到均值，交易会面临巨大损失。

第二，即使价差最终会回到均值，如果回到均值前，价差变动方向不利于交易头寸程度过深，交易也会被迫中止。例如，公司 A 股价为 10 元，公司 B 股价为 5 元，价差均值为 3 元，标准差为 1 元。这样，价差偏离均值 2 元，达到两倍标准差，交易被触发，卖出一股 A，并买入一股 B。但如果一段时间后，A 由于被收购，股价涨到 20 元，B 的股价保持不变，则 A 股的损失将达 100%，会被追加保证金。如果交易者没有足够的资金追加保证金，A 股的头寸将被迫平仓。这样即使最终价差回归均值，交易还是会有很大损失。实际上，后面第十五章将要提到的长期资本管理公司就是因为不利头寸损失过大，无力追加保证金才面临危机的，最终，没有撑到头寸获利。

第三，交易的时间长度也是必须要考虑的。即使没有发生前面说的头寸不利情况，但如果交易开始时间过长而没有盈利，资金的使用成本压力增大，也会迫使交易决策者被迫中止交易。

所以，基于上面三个原因，在上面提及的三种交易判定方法之外，必须设定止损条件。止损条件一般分为不利方向止损和交易时间止损。由于各种投资者或交易员的资金成本和抗压能力不同，并没有最优的止损方法。这里只是提出止损的思想。

首先是不利方向止损，由于前面提到的原因，价差向不利方向移动导致损失是非常可能的，所以要根据投资者能够承受资金损失的能力，设定不利方向的止损线。例如，以一倍标准差为门槛，以两倍标准差为止损线。这样，当价

差偏离均值一倍标准差时触发交易，但当价差偏离均值达到两倍标准差时触发止损线，不管将来价差是否会回归均值，交易都立刻被终止。

然后是时间止损。这取决于投资者的资金使用成本和投资机会。如果投资者的其他投资机会较少且资本使用成本较低，配对交易的单次持续时间就可以长些，反之则持续时间较短。因此，根据投资者的资金使用时间限制对配对交易设置时间止损线。例如，可以设定交易开始后，如果一个星期交易未被中止，就直接中止交易，以更有效率地使用资金。

除了上面两种止损的判定方法，配对交易还需要判定价差的结构是否发生了改变。所谓价差结构的改变，是指由于股票对中的股票对应的公司的基本面发生了变化。例如，某个公司被并购，或注入了优质资产等，这样该公司的股价可能比原来股票对中的另一只股票的价格有非暂时性提升。价差将可能无法回到原来的均值水平，也就是说，价差的均值提高了。因此，应用原来计算的均值和方差，将无法与改变后的价差结构相匹配，交易模型的模型风险大大提升。所以，需要判断价差的结构是否发生了改变。

这里采用 Cuscore 方法来帮助判定结构的改变。Cuscore 计算以下指标：

$$Q = \sum (y_t - \beta t)t$$

其中 y_t 是 t 时刻的观测值，t 代表时间，β 是斜率系数。为了说明 Cuscore 的应用，我们举下面的例子。有一条由两条斜率不同的直线组成的折线，一条斜率为 0.7，另一条为 1.3，如图 11.2 所示。从图中可以看出，虽然斜率从0.7

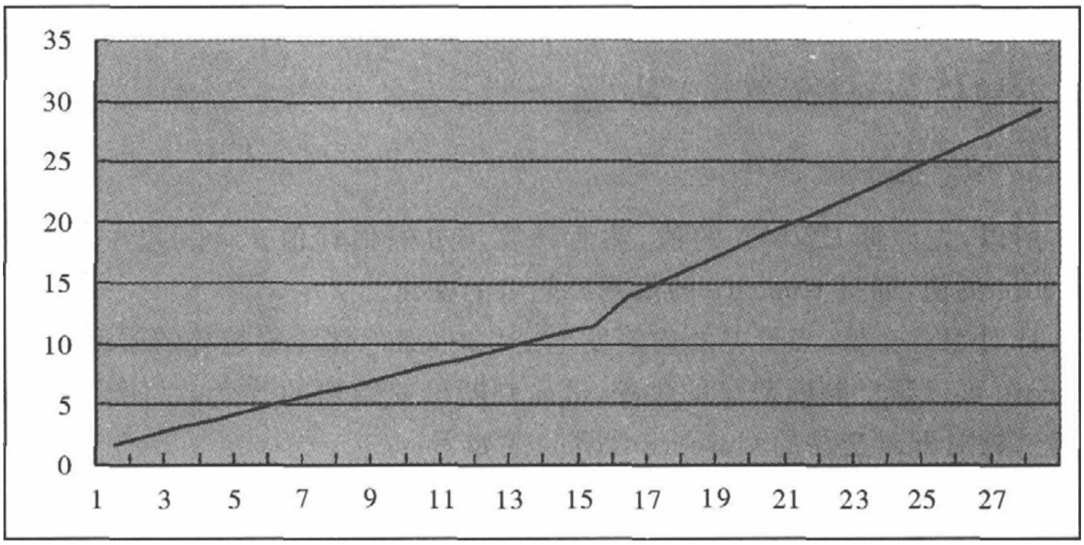


图 11.2　斜率的改变

变到 1.3，增加近 50%，但直接观测却并不明显。不容易辨别趋势的改变。图 11.3 是同样数据的 Cuscore 的图，从图中可以看出，在趋势变化时，Cuscore 的变化比数据本身的变化明显得多。因此，可以用 Cuscore 来帮助检验趋势的变化，以对变化作出更快的反应。

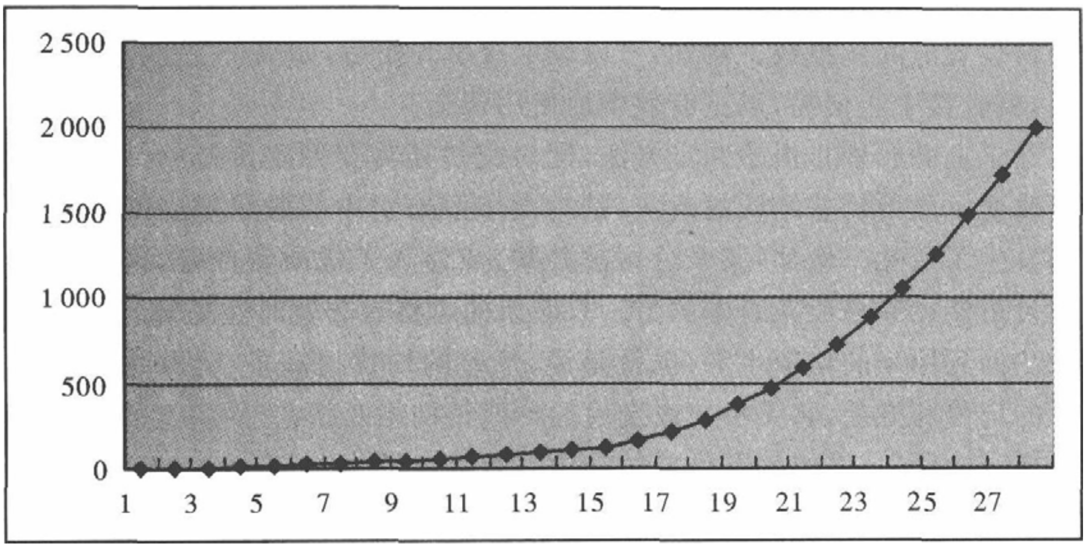


图 11.3 Cuscore 图

实际应用中，由于斜率系数 β 并不已知，需要使用历史数据进行估计。由于股票价格数据的连续性比较差，经常使用股票价格的 EWMA 数值估算斜率系数 β。EWMA 的算法前面已经提及，这里不再赘述。β 的常用估计式为：

$$\hat{\beta}=0.25(EWMA_t-EWMA_{t-4})$$

这样 Cuscore 的计算式变为：

$$Q=\sum(y_t-\hat{\beta}_t t)t$$

所以，先计算 EWMA 的值，然后计算出 β 的估计值 $\hat{\beta}$，利用上式计算出 Cuscore 的值，通过 Cuscore 的值判断趋势的变化。

针对配对交易，当价差的均值发生变化时，必然伴随着趋势的改变。利用 Cuscore 可以帮助判断趋势的改变，进而判断价差结构的改变，以便设计止损位置。针对配对交易的 Cuscore 计算公式如下：

$$Q=\sum(\text{spread}_t-\hat{\beta}_t t)t$$

最后，为了保证参数符合数据的变化，采用历史数据计算的均值、方差、γ 和 μ 等需要定期或随时根据新的数据进行调整。例如，参数是通过近两个月

的历史数据计算得出的，则可以把最新的数据加入其中，将最早的数据剔除，这样使数据随时间向前滚动。这在价差结构发生变化时尤其重要，因为结构变化了，意味着均值、方差等参数发生了较大变化，历史数据无法表征这一变化，新的数据就显得格外重要。

为避免混乱，这里将这些细节总结如下：

1. 交易开始和结束的条件设计，包括三种方式。
2. 不利方向止损和交易时间止损。
3. 利用 Cuscore 来判断价差结构的变化。
4. 数据滚动计算参数。

11.4 思路的扩展

上面只是配对交易最基础的方法和思路，实际上，配对交易可以进行多方面的扩展。本节介绍一些思路上的扩展。

11.4.1 交易内容的扩展

上面介绍的都是针对股票交易的配对交易，其实，配对交易并不局限于股票交易。商品期货、外汇、指数都可进行配对交易。只要两个交易对象存在较好的相关性或符合协整条件，都可进行配对交易。实际上，期货和外汇市场的做空较股票市场更加容易，更便于进行配对交易。例如，在外汇市场上，澳大利亚元兑美元的汇率和新西兰元兑美元的汇率的相关性就非常高。同时澳大利亚元兑美元的汇率和黄金期货的价格的相关系数高达 0.8，这些都可能是配对交易的目标交易对。

11.4.2 交易数量的扩展

配对交易也不限于只交易两只股票。实际上，从本质上讲，配对交易就是相信，相对于被卖出的股票，买入的股票表现要相对地好一些，体现的是相对价值的思想。因此，只要是存在相对价值的强弱差别，都可以采用配对交易的方法。Whistler(2004)就描述了三只股票的配对交易，卖出两只相对弱势的股票，买入一只相对强势的股票，当然，三只股票有很好的相关性。更深一步

讲，配对交易只是市场中性策略的一个特例。市场中性策略涉及买入投资者认为相对强势的股票，卖出相对弱势的股票。这样，相较于只是采用买入策略的开放基金，市场中性策略利用相对价值的概念，可能可以降低组合的风险。实际上对冲基金最早出现时，就是采用市场中性策略的，这远早于配对交易的产生。所不同的是，配对交易更容易数量化，从而使交易更加系统化，而较少依赖于交易员的经验判断。股指期货出现后，市场中性策略又有了新的发展。投资者可以买入自己认为相对于股指更加强势的股票组合，同时卖出相应的股指期货；或者卖出自己认为相对于股指更加弱势的股票组合，同时买入相应的股指期货。关于股指期货的详细讨论见第十四章。

11.4.3 分析技术的发展

配对交易方法从发现到现在，经历了从简单到复杂的一系列技术上的发展。从简单的相关系数的计算，到协整的应用，再到随机过程，再到更复杂的神经网络、遗传算法、小波分析等。由于篇幅所限，这里不能详细探讨这些更新的技术，希望可以在后续的书中进行探讨。这里只想谈谈配对交易的改进方向。

1. 更有效地选择股票对。前面只是提到了选择股票对的最简单的方法。选择股票对是配对交易的前提，更有效地选择股票对可以大大提高交易效率。当面对全球市场时，庞大的股票数量使得对每只股票做基本分析变得极为困难。因此如何利用先进的算法更有效地选择股票对，是十分重要的问题。

2. 配对比例。协整方法比简单交易法改进的地方在于，协整方法能够算出配对的比例 γ，而简单交易法往往采用一比一的比例。更高级的技术可以发现更好的配对比例，以使配对后的结果更加便于操作。

3. 门槛值的选择。门槛值对于交易的盈利至关重要。门槛值过低，可能无法弥补交易成本；过高则减少交易次数，增加资金使用成本。这一点协整方法并没有好的解决方案。对于价差更加细致的建模包括 ARMA、GARCH、随机过程等等。

最后，复杂的数学技术不能保证一定可以改进交易结果。在设计交易系统时不能盲目相信复杂的数学工具用得越多，盈利能力就越强。正如摩根斯坦利研究部创始人巴顿·比格斯所说："如果在华尔街有人告诉你，他发现了一个复杂的数学模型可以赚到大钱，最好离他远一点。"

第十二章

可转换债券套利

12.1 可转换债券特征

所谓可转换债券，就是这样一种债券，其息票支付和归还本金的设计与一般债券一样，但它赋予债券购买人在一定时期，按一定比例，以一定价格将债券转换为股票的权利。有些可转换债券为了保证发行公司的利益，赋予发行公司在一定时间，以一定的价格赎回债券的权利。同时，也有些可转换债券为了保证购买人的利益，赋予债券购买人在一定时间，以一定价格将债券回售给该债券发行公司的权利。

最早的可转换债券出现于19世纪，是美国铁路公司为了吸引投资者而发行的。在那之后，可转换债券经历了一段空白的岁月，直至20世纪初，同样是由铁路公司重新开始发行。例如1901年，巴尔的摩到俄亥俄的铁路发行了1 500万美元的可转换债券。同时，一些非铁路公司也开始发行可转换债券，例如，1906年AT&T发行了1.5亿美元的可转换债券。现代可转换债券市场形成于20世纪50年代的美国。那时，很多航空公司采用可转换债券来改善资本结构。因为可转换债券可以支付较少的利息，同时又避免了股权的稀释。到20世纪60年代，很多中小型企业开始发行可转换债券。这导致了人们普遍认为，可转换债券的评级不高。直到1984年，IBM发行了AAA级的可转换债券，才改变了人们的这一观念。在20世纪80年代末至90年代初，银行业由于监管需求的原因，开始发行可转换债券募集资本。到了90年代，高科技企业开始采用可转换债券。

最早量化可转换债券的并非 Black、Scholes 或 Merton，而是数学家 Thorp 和经济学家 Kassouf。在他们 1967 年的书籍《击败市场》(*Beat the Market*)中，Thorp 和 Kassouf 第一次采用数学的观点分析可转换债券(当时还没有 Black-Scholes 公式)。两位作者首次将可转换债券分拆成债券和权证(当时尚没有期权)，然后分别分析债券部分和权证部分，以辨别是否有套利机会。这奠定了现代可转换债券的复杂数学定价和套利方法的基础(不知为何没给两人诺贝尔奖)。同时，两人创立的权证和可转换债券基金收益也很好。

12.2 基本术语

可转换债券有一些较为复杂的条款，为了后面的叙述方便，这里将一些术语简单进行介绍。

到期日：到期日是可转换债券的发行者以票面值向债券购买者支付的时间。债券的到期日的范围很广，最长可达 20 年。最近，5 到 10 年期的可转换债券逐渐开始流行。

息票：息票是可转换债券的利息，通常是固定利率和按固定日期支付的。该利率一般低于发行公司发行不可转换债券的利率。这主要是因为，可转换债券赋予了债券购买者将债券转换为标的股票的权利，这实际上是卖给了债券购买者一份看涨期权(通常为美式)。这样，债券购买者就需要减少获得的利息，以抵补期权费用。

转换比率：转换比率是一张可转换债券转换为标的股票的数目。

转换价值：转换价值是任何时候，可转换债券可以转换成的股票的市场价值。它代表了可转换债券的最低价值，可转换债券的市场价格不能低于转换价值。否则，套利者就可以以低于转换价值的价格买入可转换债券，将其转换为标的股票，然后卖出股票获得转换价值。实际上仍有一些可转换债券的价格低于转换价值。这是因为，在购买可转换债券到转换为标的股票并卖出之间存在时滞。也就是说，在一些市场，买入可转换债券进行转换后，不能立刻获得标的股票，从而在市场上卖出。这样，从买入可转换债券到卖出标的股票之间，股票价格可能发生变化，使得股票价值可能低于买入时的转换价值，套利存在风险。因此，在这些市场，可能可转换债券的价格相对转换价值存在一个折价，以弥补套利者的持有风险。标的股票的波动率越大，折价就越高。在一些新兴市场，这一折价可能达到 10%～30%。

溢价：溢价是可转换债券的市场价格高于转换价值部分，以相对于转换价值的百分比衡量。投资者愿意支付一个高于转换价值的溢价，主要是因为可转换债券为投资者提供了下跌风险的保护，并且一般可以获得高于股利的利息收入。

债券下限(bond floor)：债券下限是指，将可转换债券看成普通的公司债券，按照折现的方法，将息票和本金按照LIBOR加信用风险利差折成现值所得的值。这一价值代表了可转换债券的最低价值，它衡量了没有其他特征的债券部分的价值。要获得债券下限，需要知道利率和发行公司的信用利差。信用利差是指发行公司发行普通债券要支付的利率与无风险利率的差值。利差越大，公司的违约风险越高。

回购条款：回购条款是指发行人有权利在可转换债券到期日前按照约定的赎回价格在约定的赎回期将可转换债券提前赎回。由于，提前赎回权利的存在，约定的赎回价格一般要高于可转换债券的票面价值。

回售条款：回售条款是指投资人有权利在可转换债券到期日前按照约定的价格在约定的回售期将可转换债券提前卖回给发行人。约定的回售价格一般等于或略低于可转换债券的票面价值。

12.3 可转债定价

可转换债券作为一种复杂的衍生品，其定价根据发行条款的不同，会有着不同的复杂程度。最简单的可转换债券，可以看成一份看涨期权加一份普通的债券。而复杂的情况时，则无法给出解析的解，只能通过数值解法，有时甚至数值解法都很难得到有效的价格。本节先简单介绍一般期权定价的基础Black-Scholes公式，然后由易到难介绍可转换债券的一些定价方法。由于不是单纯讲可转换债券的书，所以我们在此没有考虑更复杂的信用风险，所介绍的方法也不是最前沿、最复杂的方法，因为那样将涉及很复杂的衍生品定价知识，导致偏离主题。

12.3.1 Black-Scholes 公式

由于本书主要探讨套利方法，所以不打算详细介绍Black-Scholes公式。为了后面可转换债券公式的推导，这里只给出Black-Scholes的简单推导。

考虑一个期权与标的股票的组合，设其价值为$\prod$，期权价格为 V(S，t)，标的股票价格为 S，时间为 t。组合构成为买入一手期权，卖出 Δ 手标的股票，组合的总价值为：

$$\prod = V(S,t) - \Delta S$$

假设股票价格服从对数正态随机游走：

$$dS=\mu S dt+\sigma S dX$$

则经过 dt 的时间，组合价值改变：

$$d\prod = dV-\Delta dS$$

根据伊藤引理，

$$dV=\frac{\partial V}{\partial t}dt+\frac{\partial V}{\partial S}dS+\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}dt$$

带入组合价值改变的公式得：

$$d\prod =\frac{\partial V}{\partial t}dt+\frac{\partial V}{\partial S}dS+\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}dt-\Delta dS$$

如果令 $\Delta=\frac{\partial V}{\partial S}$，则上式变为：

$$d\prod =\frac{\partial V}{\partial t}dt+\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}dt$$

由上式可见，$d\prod$ 中不存在随机项，根据无套利假设，此时组合价值的改变只能按照无风险利率变化，即 $d\prod = r\prod dt$，其中 r 为无风险利率。代入上式并整理得：

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}+rS\frac{\partial V}{\partial S}-rV=0$$

此为 Black-Scholes 方程。根据不同的边界条件，可以计算不同种类期权的值。

为了方便后面讨论，简单介绍一下希腊字母(Greeks)。希腊字母是一系列的微分，包括 delta、gamma、vega、theta 等。本节只简单介绍前两个希腊字母 delta 和 gamma。

Delta 是期权或可转换债券的价值对标的股票价格的微分，衡量的是期权

或可转换债券的价格对标的股票价格变化的敏感性。

$$\Delta=\frac{\partial V}{\partial S}$$

Delta 不仅仅是一个微分，而是有着非常实际的意义。从前面的 Black-Scholes 公式的推导可以看出，delta 的选择使得整个组合没有风险。换句话说，通过卖空 delta 手标的股票，在所考虑的极短的时期，整个组合等价于无风险资产。实际中，交易员卖出一手期权后，往往会买入 delta 手标的股票来对冲风险，使得整个组合头寸没有风险。这通常也叫作 delta 中性策略。当然，要想完全消除组合的风险，按照 Black-Scholes 公式的要求，交易员就需要连续地进行对冲，也就是根据 delta 的不断变化，不停地买卖标的股票。但实际中这是不可能做到的，买卖不可能连续。如果考虑到交易成本，连续的买卖造成的交易成本几乎是无限大的。所以，实际中交易员是间断对冲的。

Gamma 是期权或可转换债券的价格对标的股票价格的二阶微分，衡量的是 delta 对标的股票价格变化的敏感性。

$$\Gamma=\frac{\partial^2 V}{\partial S^2}$$

Gamma 决定的是，当标的股票价格发生改变时，为了维持 delta 中性，需要买入或卖出的标的股票的数量。由于实际中无法连续对冲，gamma 的大小影响对冲的频率。如果 gamma 较大，那随着标的股票价格的变化，delta 的变化就很大，就需要频繁买卖标的股票。如果 gamma 很小，甚至接近于 0，那么 delta 就几乎和标的股票价格的变化无关，这样就不需要频繁地买卖标的股票。实际中也有采用买卖其他类型的期权以达到使组合的 gamma 为 0 的策略，称为 gamma 中性策略。

12.3.2 利率期限结构

由于可转换债券的存续时间一般较长，所以相关的无风险利率一般不能视为常量。因此往往将利率也认为是随机的。一般的，设利率服从：

$$dr=u(r,t)dt+w(r,t)dX$$

设想有两种到期日不同的债券，价格为 V_1 和 V_2。组合为买入一手债券 1，同时卖空 Δ 手债券 2，于是组合的价值为：

$$\prod = V_1 - \Delta V_2$$

则组合的变化量为：

$$\mathrm{d}\prod = \frac{\partial V_1}{\partial t}\mathrm{d}t + \frac{\partial V_1}{\partial r}\mathrm{d}r + \frac{1}{2}w^2\frac{\partial^2 V_1}{\partial r^2}\mathrm{d}t - \Delta\left(\frac{\partial V_2}{\partial t}\mathrm{d}t + \frac{\partial V_2}{\partial r}\mathrm{d}r + \frac{1}{2}w^2\frac{\partial^2 V_2}{\partial r^2}\mathrm{d}t\right)$$

为消除随机项，取

$$\Delta = \frac{\partial V_1}{\partial r}\Big/\frac{\partial V_2}{\partial r}$$

于是，组合的价值应按无风险利率变化，即：

$$r\prod\mathrm{d}t = r\left(V_1 - \left(\frac{\partial V_1}{\partial r}\Big/\frac{\partial V_2}{\partial r}\right)V_2\right)\mathrm{d}t$$
$$= \left[\frac{\partial V_1}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_1}{\partial r^2} - \left(\frac{\partial V_1}{\partial r}\Big/\frac{\partial V_2}{\partial r}\right)\left(\frac{\partial V_2}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_2}{\partial r^2}\right)\right]\mathrm{d}t$$

将 V_1 和 V_2 的项放到等号的两边，可得：

$$\frac{\frac{\partial V_1}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_1}{\partial r^2} - rV_1}{\frac{\partial V_1}{\partial r}} = \frac{\frac{\partial V_2}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_2}{\partial r^2} - rV_2}{\frac{\partial V_2}{\partial r}}$$

由上式可以看到，等式左边只与债券 1 有关，等式右边只与债券 2 有关，所以，等式的两边一定都是等于一个与债券到期日无关的函数，于是：

$$\frac{\frac{\partial V}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} - rV}{\frac{\partial V}{\partial r}} = a(r,t)$$

为方便起见，设 $a(r,t) = w(r,t)\lambda(r,t) - u(r,t)$，则方程变为：

$$\frac{\partial V}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} + (u - \lambda w)\frac{\partial V}{\partial r} - rV = 0$$

则债券的价格可由 u 和 w 的具体函数形式确定。

下面解释一下 λ 的含义。单一债券的变化量为

$$\mathrm{d}V = w\frac{\partial V}{\partial r}\mathrm{d}X + \left(\frac{\partial V}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} + u\frac{\partial V}{\partial r}\right)\mathrm{d}t$$

由债券定价方程得：

$$\mathrm{d}V = w\frac{\partial V}{\partial r}\mathrm{d}X + \left(w\lambda\frac{\partial V}{\partial r} + rV\right)\mathrm{d}t$$

整理可得：

$$dV - rVdt = w\frac{\partial V}{\partial r}(dX + \lambda dt)$$

由上式可见，债券不是无风险资产，否则右面应该为零。由此可见，承担了 dX 的风险，应该获得 λdt 的额外收益，所以 λ 被称为风险价格。

根据上面的介绍，采用不同的 $u(r, t)$ 和 $w(r, t)$，可以得到不同的期限结构，常用的假设包括：

Vasicek 假设：$dr = k(\theta - r)dt + \sigma dX$

Dothan 假设：$dr = ardt + \sigma rdX$

CIR 假设：$dr = k(\theta - r)dt + \sigma\sqrt{r}dw$

12.3.3 可转换债券定价

可转换债券根据时间长短，分为两种形式。对于短期的可转换债券，可以视无风险利率为固定的。对于长期的可转换债券，则要考虑随机的期限结构。首先介绍短期可转换债券定价。

1. 短期可转换债券定价

设无风险利率 r 为常量，可转换债券的价格为 $V(S, t)$。与前面一样，考虑一个组合，包括买入一手可转换债券，卖出 Δ 手标的股票，则组合价值的变化量为：

$$d\Pi = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}dt - \Delta dS$$

取 $\Delta = \frac{\partial V}{\partial S}$，则可得方程为：

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

2. 长期可转换债券定价

计算长期可转换债券价格时，无风险利率不能视为常量，而应符合随机期限结构。同样设可转换债券价格为 $V(S, r, t)$，注意此时可转换债券的价格也是无风险利率的函数。标的股票价格服从对数正态模型：

$$dS = \mu Sdt + \sigma SdX_1$$

无风险利率服从前面利率期限结构的假设，即：

$$\mathrm{d}r=u(r,t)\mathrm{d}t+w(r,t)\mathrm{d}X_2$$

两个随机过程项的相关系数为 $\rho(r,S,t)$，即：

$$E[\mathrm{d}X_1\mathrm{d}X_2]=\rho\mathrm{d}t$$

其中 $-1\leqslant\rho(r,S,t)\leqslant1$。

根据伊藤引理，$\mathrm{d}V$ 为：

$$\mathrm{d}V=\frac{\partial V}{\partial t}\mathrm{d}t+\frac{\partial V}{\partial S}\mathrm{d}S+\frac{\partial V}{\partial r}\mathrm{d}r+\frac{1}{2}\left(\sigma^2S^2\frac{\partial^2V}{\partial S^2}+2\rho\sigma Sw\frac{\partial^2V}{\partial S\partial r}+w^2\frac{\partial^2V}{\partial r^2}\right)\mathrm{d}t$$

考虑一个组合，包括买入一手可转换债券，卖空 Δ_2 手债券，卖空 Δ_1 手标的股票。所以组合价值为：

$$\prod=V-\Delta_2Z-\Delta_1S$$

计算和上面一样，取

$$\Delta_2=\frac{\partial V}{\partial r}\bigg/\frac{\partial Z}{\partial r}$$

$$\Delta_1=\frac{\partial V}{\partial S}$$

消去风险项得方程为：

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2V}{\partial S^2}+\rho\sigma Sw\frac{\partial^2V}{\partial S\partial r}+\frac{1}{2}w^2\frac{\partial^2V}{\partial r^2}+rS\frac{\partial V}{\partial S}+(u-\lambda w)\frac{\partial V}{\partial r}-rV=0$$

得到了方程之后，根据可转换债券的条款，加上相应的边界条件，就得到了可转换债券的定价框架。

12.3.4 数量算法

除了最简单的情况，通常可转换债券的价格没有解析的形式，所以只能采用数值算法来计算。通常使用的数值算法有两种，分别是差分法和 Monte Carlo 算法。差分法的优势在于计算的效率较高，并且对希腊字母的计算较为方便，计算速度较快。缺点在于编程较之 Monte Carlo 方法复杂，而且只能应对变量较少的情况(一般只能计算 3～4 个变量)，对于条款较为复杂的可转换债券的计算能力差。Monte Carlo 算法的优势在于编程较为方便，同时能够应对多变量情况，便于计算条款复杂的可转换债券。缺点在于计算效率较之于差分算法要慢，并且计算希腊字母的效率较低。下面分别介绍这两种方法。

1. 差分法

差分法的基本思想是，将变量分成等距的点。如对于期权定价来说，将时间和股价分成等距的点，间距为 δS 和 δt。则第 i 时的股票价格为 $S=i\delta S$，第 k 时的时间为 $t=k\delta t$，为了计算方便，往往将第 k 时的时间计算为 $t=T-k\delta t$，其中 T 为期权的到期日。从图 12.1 可以看出，这一思想实际上是将时间和股价平面分成等距的格子。格子上的每一个点代表一个股价和时间的组合 $(i\delta S, T-k\delta t)$。对于期权定价，时间为 0 到 T，所以 k 取 $0\leqslant k\leqslant K$，$k=K$ 时，t 为 0。为了计算，需要给股价 S 一个上限，一个较大的股票值 S_m，则 i 取 $0\leqslant i\leqslant I$，$i=I$ 时，$S=S_m$。

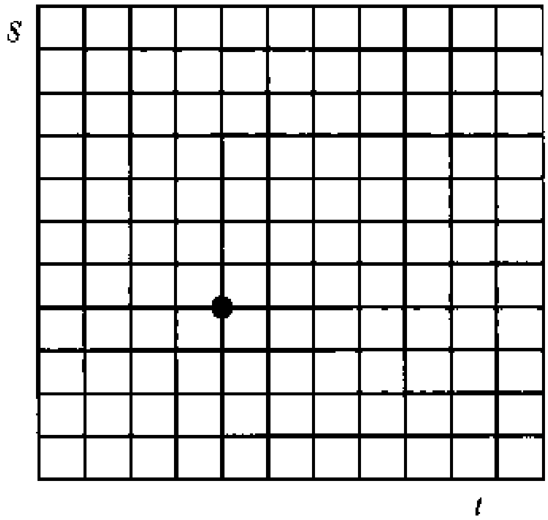


图 12.1 差分法

在每一个格点，期权值为 $V_i^k=V_{(i\delta S, T-k\delta t)}$。在此基础上，我们介绍 Black-Scholes 方程的差分解法。

首先，微分的定义是

$$\frac{\partial V}{\partial t}=\lim_{h\to 0}\frac{V(S,t+h)-V(S,t)}{h}$$

差分的方法是用较小的间距 h 来近似代替 h 趋于 0 的情况，即：

$$\frac{\partial V}{\partial t}=\frac{V_i^k-V_i^{k+1}}{\delta t}$$

于是对于 S 的微分，可以近似为：

$$\frac{\partial V}{\partial S}=\frac{V_{i+1}^k-V_i^k}{\delta S}$$

从上式可以看出，右侧差分是采用 $i+1$ 时的值减去 i 时的值，该式称为向前差分。还存在两种差分方式，即向后差分和中间差分：

$$\frac{\partial V}{\partial S}=\frac{V_i^k-V_{i-1}^k}{\delta S},\ \frac{\partial V}{\partial S}=\frac{V_{i+1}^k-V_{i-1}^k}{2\delta S}$$

二阶微分采用类似的方式近似，采用下式：

$$\frac{\partial^2 V}{\partial S^2}=\frac{V_{i+1}^k-2V_i^k+V_{i-1}^k}{\delta S^2}$$

由此可得 Black-Scholes 方程的差分形式为：

$$\frac{V_i^k-V_i^{k+1}}{\delta t}+\frac{1}{2}\sigma^2\ (i\delta S)^2\ \frac{V_{i+1}^k-2V_i^k+V_{i-1}^k}{\delta S^2}+ri\delta S\ \frac{V_{i+1}^k-V_{i-1}^k}{2\delta S}-rV_i^k=0$$

这里采用了中间差分法，因为中间差分法的精度更高。具体的推导和精度分析，这里不做详细分析。读者可参考相关的书籍。

从上式可以看出，对于时间，仅有 k 和 $k+1$ 的项，因此如果已知第 k 项的所有 i、$i-1$、$i+1$ 的位置的值，即可算出第 $k+1$ 处的值。而第 k 项的所有 i、$i-1$、$i+1$ 的位置的值就要看初始条件和边界条件了。对于期权定价，初始条件实际上是到期时的期权价值，因为我们假设 $k=0$ 时 $t=T$。由于期权规定了到期日的支付情况，所以在 $k=0$ 时的所有 i 的值都是已知的。另外，期权定价还会存在所谓边界条件。边界条件的情况多种多样，在这里就不一一列举了。对于看涨期权，当股价为 0 时，期权价格显然为 0。对于可转换债券，边界条件就是相应的条款。利用初始条件和边值条件就可逐层递进，先采用 0 时刻的值计算 1 时刻的值，以此类推，当通过计算得到 k 时刻的值后，利用 k 时刻的值计算 $k+1$ 时刻的值。

当然，这里只是采用了差分方法的显式方法。所谓显式，就是 $k+1$ 时的值可以由 k 时的值直接计算出。该方法的优点在于，计算和编程都非常方便。但缺点在于，显式方法的收敛速度较慢，需要采用较小的步长，计算次数较多。而隐式方法的收敛速度较快，可以放大步长，减少计算次数。但缺点在于，该方法需要矩阵运算，编程较之显式方法要难。一个隐式方法的例子是：

$$\frac{V_i^k-V_i^{k+1}}{\delta t}+\frac{1}{2}\sigma^2\ (i\delta S)^2\ \frac{V_{i+1}^{k+1}-2V_i^{k+1}+V_{i-1}^{k+1}}{\delta S^2}+ri\delta S\ \frac{V_{i+1}^{k+1}-V_{i-1}^{k+1}}{2\delta S}-rV_i^{k+1}=0$$

从该例可以看到，无法从 k 时的值直接计算 $k+1$ 时的值，这就需要进行矩阵运算来求解。由于本书不是讲解数值计算方法的书，详细的求解方法，就不在这里赘述了。

通过差分法，可以得出相应于各种初始标的股票价格的可转换债券的价格。这样，按照定价时的标的股票价格即可查得相应的可转换债券的价格。Delta 的计算也很方便，根据

$$\frac{\partial V}{\partial S}=\frac{V_{i+1}^{t}-V_{i}^{t}}{\delta S}$$

只要把初始时的可转换债券价格和其临近的一个价格带入上式，就可得到 delta。

2. Monte Carlo 方法

Monte Carlo 方法源自物理，最早是为了计算核裂变而产生的。该方法最大的优点就是简单，能够应对非常复杂的情况。目前对于那些条款复杂的金融产品，往往只能采用 Monte Carlo 方法。特别是现在 CPU 计算速度的提高，正在弥补 Monte Carlo 计算效率低的缺点。

Monte Carlo 方法的思想很简单。因为衍生产品定价的基本思想是，计算衍生产品在到期时的期望值，再将这一期望值折成现值。按照这一思想，Monte Carlo 方法模拟标的资产的路径，按照这一路径计算相关衍生品的到期价值。按此方法计算很多条不同的路径，然后计算这些路径对应的衍生产品的价值。将所得到的这些衍生产品的价值求平均值，以此来近似期望值，再将此值折现成现值。可以证明，当模拟次数逐渐增大时，该方法计算出的值将趋近理论值。

具体说来，对于标的资产为股票的衍生品，采用 Monte Carlo 法模拟股票的轨迹。因为假设股票服从对数正态，则可将股票价格写成下式：

$$\delta S=rS\delta t+\sigma S\sqrt{\delta t}\varphi$$

其中，φ 是标准正态随机数，r 为无风险利率，σ 为波动率，S 为股价，δt 为时间的步长，δS 为股价增量。操作时先生成一个标准正态的随机数，然后代入上式得到股价的增量，在原估价上加上增量，得到下一步的股价。如果波动率 σ 为常量，可以得到 S 的表达式为：

$$S_{(t+\delta t)}=S_{(t)}\exp\left[\left(r-\frac{1}{2}\sigma^{2}\right)\delta t+\sigma\sqrt{\delta t}\varphi\right]$$

也可由上式得到股票价格轨迹。

对于长期可转换债券，需要考虑股价和利率两个随机数值，且两个随机数有一定的相关性。这时就需要采用 Cholesky 方法。

Cholesky 方法的目的是产生符合某种相关性的随机数。假设要产生的

正态分布的随机数列向量为 $\boldsymbol{\varphi}$，随机数之间符合相关系数矩阵 $\boldsymbol{P}$，即：

$$E[\boldsymbol{\varphi}\boldsymbol{\varphi}^T]=\boldsymbol{P}$$

则 Cholesky 方法是，先产生同样数量的一列独立的正态分布的随机数，列向量为 $\boldsymbol{\varphi}$，则所要得到的随机数列向量 $\boldsymbol{\varphi}$ 为：

$$\boldsymbol{\varphi}=\boldsymbol{M}\boldsymbol{\varphi}$$

其中 $\boldsymbol{M}$ 是一特殊矩阵，它服从：

$$\boldsymbol{M}\boldsymbol{M}^T=\boldsymbol{P}$$

所以，先由相关系数矩阵求得 $\boldsymbol{M}$，然后产生随机数列向量 $\boldsymbol{\varphi}$，再由公式计算出要求得的随机数列向量 $\boldsymbol{\varphi}$。

以长期可转换债券为例，需要考虑股价和无风险利率的变动，股价服从前面公式，而利率服从

$$\delta r=[u(r,t)-\lambda(r,t)w(r,t)]\delta t+w(r,t)\sqrt{\delta t}x$$

其中 x 也是标准正态分布的随机数。φ 和 x 的相关系数为 ρ，即：

$$E[\varphi x]=\rho$$

则可得相关系数矩阵为

$$\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$$

由此可计算出矩阵 $\boldsymbol{M}$

$$\begin{pmatrix}\sqrt{1-\rho^2} & \rho\\ 0 & 1\end{pmatrix}$$

从而产生两个独立的标准正态分布随机数：ε_1、ε_2，则

$$\begin{pmatrix}\varphi\\ x\end{pmatrix}=\begin{pmatrix}\sqrt{1-\rho^2} & \rho\\ 0 & 1\end{pmatrix}\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\end{pmatrix}$$

再利用所得到的 φ 和 x，计算股价和利率的变动。

得到了标的股票和期限结构的轨迹后，可以利用其计算可转换债券的价格，并以得到的期限结构折现回初始时刻。然后多次重复这一过程，得到多个可转换债券的值，求平均得到理论价格。

利用 Monte Carlo 法计算 delta 较为麻烦。采用的方法是将初始的标的

股票价格改变一个很小的数值，重新计算可转换债券的理论价格，再根据

$$\frac{\partial V}{\partial S}=\frac{V_{i+1}^{k}-V_{i}^{k}}{\delta S}$$

计算出 delta。其中 δS 是初始股价改变的数值。

最后谈标准正态随机数的生成。很多软件都自带均匀 0-1 分布随机数，这一随机数的生成不在本书讨论范畴。以下介绍两种通过 0-1 分布随机数计算标准正态分布随机数的方法。第一种方法较为简单。假设 φ_i 为服从 0-1 分布的随机数，则利用下式可得标准正态的随机数：

$$\left(\sum_{i=1}^{12}\varphi_i\right)-6$$

实际上就是 12 个 0-1 分布的随机数的和减去 6

另一种是 Box-Muller 方法。该方法首先生成两个独立的 0-1 分布随机数 x_1、x_2，按照下式生成 y_1 和 y_2，这两个数都是标准正态分布。

$$y_1=\sqrt{-2\ln x_1}\cos(2\pi x_2),\quad y_2=\sqrt{-2\ln x_1}\sin(2\pi x_2)$$

附录中给出了 Monte Carlo 法和差分法的可转换债券的定价程序。

12.4 可转换债券套利

12.4.1 套利技术

可转换债券的套利技术有很多种，但多数需要复杂的衍生产品才能实现，如股票互换、CDS 等。由于这些衍生产品在我国短期内不太可能出现，本书不倾向于介绍这些复杂的套利方法。这里只介绍较为简单但使用较多的 delta 套利，也是较早被使用的套利技术。考虑到融券业务正开始试点，这一套利技术将更有实际意义。

简单的可转换债券套利往往涉及买入可转换债券，卖空标的股票。Delta 套利正是如此。对于 delta 套利来说，可转换债券的定价模型至关重要。回顾前面可转换债券定价方程的推导可以发现，如果该理论模型是准确的，那么，买入可转换债券，同时卖空 delta 手标的股票，可以使得组合短期等同于

无风险资产，从而获得无风险利率的收人。如果可转换债券价格被低估，也就是可转换债券的价格低于其理论价格，那么如果投资者采用同样的方法，即买入一手可转换债券，卖空 delta 手标的股票，投资者就以比理论低的成本获得了无风险组合。这样，理论上看，短期投资者获得了无风险利率，但成本却低于理论的无风险组合的成本，所以，投资者的实际收益将高于无风险利率。也就是说，投资者持有无风险资产，获得了超过无风险利率的收益，套利由此产生。这也就是 delta 套利的基本思想。

由这一思想可以看出，准确的理论值对套利十分重要。如果理论模型较差，就无法准确分辨可转换债券是否被低估，同时也无法计算准确的 delta 值，套利就很难达成。但实际操作时，很难确定可转换债券的定价模型的准确性，涉及条款复杂的可转换债券，就更难确定准确的价值。所以本章并未介绍更为复杂的定价模型，因为很难确定这些新的模型是否能使套利者获利更多。对于可转换债券的定价，至今仍没有公认的最优模型，所以，本章只是假设获得了一个投资者自己认为较好的理论值，进而探讨套利中的一些问题。至于理论值如何得到，就由投资者自己决定了。

其实，可转换债券的套利与一般期权或权证的套利类似，所涉及的方法也非常相像，所不同的是，可转换债券可以获得利息，而且这一利率一般高于标的股票的股利支付率，而这一特征使得可转换债券的套利较一般期权或权证的套利更具吸引力。

当投资者根据理论模型发现被低估的可转换债券时，投资者可以建立初始头寸，即，买入一手可转换债券，卖空 delta（由理论模型确定）手标的股票。但交易并不由此结束。因为从理论模型可以看出，该组合只在很短的时间可视为无风险组合（应该是无限短的时间），而且 delta 的值随时在发生变动。按照 Black-Scholes 模型的假定，只有连续地根据 delta 的改变而改变所持有的标的股票的头寸，才能维持组合是无风险的。但实际中这是根本不可能做到的。这时就存在着跟踪误差和交易成本的平衡问题。如果投资者根据 delta 的改变较为频繁地调整标的股票的头寸，则组合可以更好地跟踪无风险组合，交易的风险就越小，但这样会产生很高的交易成本，使得套利利润下降，甚至可能造成亏损。如果投资者较少更改标的股票的头寸，则交易成本很低，但对于无风险资产的跟踪效果就可能会很差，导致组合的风险提高。对于跟踪误差和交易成本的平衡，有一些研究文章，但依然没有定论。同时，这一平衡还与投资者的风险承受能力和所承担的交易成本有关，由于涉及的文献较多，观点各异，这里就不介绍了，有兴趣的读者可以参看相关的参考文献。按照

delta 的改变和投资者自己的偏好，投资者在建立初始头寸之后，逐渐调整标的股票的头寸，直到：(1)可转换债券的价格达到其理论值；(2)可转换债券到期；(3)获利达到投资者的预期标准。这时或通过转换平仓，或通过卖出可转换债券，同时买进标的股票平仓。

12.4.2 相关风险

可转换债券套利的相关风险有以下几种：

1. 股票市场风险。如果对冲并非市场中性的，套利交易就会有股票市场波动的风险。

2. 利率风险。可转换债券的价格和普通债券一样，与利率的变动方向相反。可转换债券越接近其债券下限，其对利率的变动就越敏感。不过，因为股票的价格往往会与利率呈一定的负相关的关系，所以卖空的标的股票头寸的盈利与利率存在一定的正相关关系。这样，标的股票的卖空头寸对可转换债券的利率风险起到了一定的保护作用。如果在欧美那样衍生品发达的市场，可以采用利率互换来分散利率风险。

3. 信用风险。因为可转换债券的债券性质，使得其对于发行公司的信用等级非常敏感。如果公司债券与国债的信用利差扩大，债券的价值就会下降。不过，如果信用利差扩大，也有可能引起股票价格的下跌。这使得标的股票的卖空头寸又对可转换债券的信用风险提供了一定的保护。

4. 杠杆风险。由于套利的利差往往比较小，投资者常会采用杠杆来扩大收益。但这在扩大利润的同时，扩大了投资者承担的风险。过高的杠杆，可能会使得对投资者不利的变动所造成的损失超过投资者的资本，直接导致投资者陷入财务困境。后面详细分析的长期资本管理公司，就是因为采用过量的杠杆才引来破产之祸的。

5. 股利支付风险。由于套利者持有空头标的股票头寸，如果在套利期间，公司支付股利，则套利者必须向股票借出者支付股利。这减少了套利的收益。

6. 流动性风险。如果可转换债券的流动性差，那么买卖价差就会很大。而且如果标的股票的流动性差，则套利者很难借入股票进行卖空，也就无法建立套利组合头寸。这类风险对于信用等级较低公司发行的可转换债券尤其重要。

7. 逼空风险。因为套利组合需要卖空标的股票，该组合就可能面临逼空风险。所谓逼空，源于卖空的机制。因为卖空需要先借入股票，但最终要归还

借入的股票。如果在投资者打算平仓时，市场的流动性下降，则投资者可能面临无法买回借入股票的风险。这时投资者就需要出更高的价钱。如果有逼空者事先买入大量该股票，然后抬升股价，卖空者就会遭受非常大的损失。这就是逼空风险。

8. 汇率风险。如果可转换债券涉及两种或以上的货币，套利者就可能面对货币之间的汇率风险。该风险可以通过汇率远期来规避。

除了使用相关的衍生品，为了减少以上这些风险，就需要采用投资组合的方法，即分散投资多个可转债套利，每个套利头寸只占全部资金的一小部分。

附　录

1. 长期可转换债券的 Monte Carlo 定价程序，5 年期，一年支付一次利息，欧式。C 语言程序。

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
#define PI 3.1415926
long in;
double random()
{
float xn;
long m=2147483647,q=127773,r=2836,a=16807;

  if((a*(in%q) -r*(in/q))>=0)
  in=a*(in%q)-r*(in/q);
  else
  in=a*(in%q)-r*(in/q)+m;

  xn=((float)in)/((float)m);
    return(xn);
}

double normrand(double u,double v)
```

```
{
  double x;
  x=sqrt(-2 * log(u)) * sin(2 * PI * v);
  return x;
}

void corand(double a,double b,double *c,double *d,double cor)
{
  *c=sqrt(1-cor * cor) * a+cor * b;
  *d=b;
}
double convertible (double s0, double strikeprice, double vol, double rate,double k,double theta,double vegar,double correlation,double coupon)
{
double intert,rv,s1,maxs=0,optionp,sumop=0,randa,randb,r[501],r1,r2,r3,r4,sumr;
  int i,j;
  intert=5/500;
  rv=rate-vol * vol/2;
  r[0]=rate;
  for(i=0;i<1000;i++)
  {
    sumr=0;
    s1=s0;
    maxs=s0;
    optionp=0;
    for(j=0;j<500;j++)
    {
      sumr+=r[j];
corand(normrand(random(),random()),normrand(random(),random()),&randa,&randb,correlation);
      s1=s1 * exp(rv * intert+vol * sqrt(intert) * randa);
      r[j+1]=k * (theta-r[j]) * intert+vegar * sqrt(intert) * randb
```

```
            +r[j];
            rv=r[j+1]-vol*vol/2;
            if(j==99)
              r1=sumr/100;
            if(j==199)
              r2=sumr/200;
            if(j==299)
              r3=sumr/300;
            if(j==399)
              r4=sumr/400;
          }
          if(100/strikeprice*s1<100+100*coupon)
            optionp=100+100*coupon;
          else
            optionp=100/strikeprice*s1;
          sumr=sumr/500;
          optionp=optionp*exp(-sumr*5);
          r1=100*coupon*exp(-r1*1);
          r2=100*coupon*exp(-r2*2);
          r3=100*coupon*exp(-r3*3);
          r4=100*coupon*exp(-r4*4);
          optionp+=r1+r2+r3+r4;
          sumop+=optionp;
          rv=rate-vol*vol/2;
          r[0]=rate;
        }
        sumop=sumop/1000;
        return sumop;
      }
      void main()
      {
        double
vol,rate,s0,strikeprice,delta,vega,optionp,k,theta,vegar,coupon,correla-
```

```
tion;
        in=rand();
        printf("please input the volatility\n");
        scanf("%lf",&vol);
        while(vol<0)
        {
          printf("the volatility is less than zero,please input again\n");
          scanf("%lf",&vol);
        }
        printf("please input the interest rate\n");
        scanf("%lf",&rate);
        while(rate<0)
        {
          printf("the interest rate is less than zero,please input again\n");
          scanf("%lf",&rate);
        }
        printf("please input the stock price\n");
        scanf("%lf",&s0);
        while(s0<0)
        {
          printf("the stock price is less than zero,please input again\n");
          scanf("%lf",&s0);
        }
        printf("please input the strike price\n");
        scanf("%lf",&strikeprice);
        while(strikeprice<0)
        {
          printf("the strike price is less than zero,please input again\n");
          scanf("%lf",&strikeprice);
        }
        printf("please input the k of interest rate\n");
        scanf("%lf",&k);
        while(k<0)
```

```
    {
      printf("the k of interest rate is less than zero,please input again\
n");
      scanf("%lf",&k);
    }
    printf("please input the theta of the interest rate\n");
    scanf("%lf",&theta);
    while(theta<0)
    {
      printf("the theta of interest rate is less than zero,please input again
\n");
      scanf("%lf",&theta);
    }
    printf("please input the vega of interest rate\n");
    scanf("%lf",&vegar);
    while(vegar<0)
    {
      printf("the vega of interest rate is less than zero,please input again
\n");
      scanf("%lf",&vegar);
    }
    printf("please input the coupon\n");
    scanf("%lf",&coupon);
    while(coupon<0)
    {
      printf("the coupon is less than zero,please input again\n");
      scanf("%lf",&coupon);
    }
    printf("please input the correlation\n");
    scanf("%lf",&correlation);
    while(correlation<0)
    {
      printf("the correlation is less than zero,please input again\n");
```

```
        scanf("%lf",&correlation);
    }
    optionp=convertible(s0,strikeprice,vol,rate,k,theta,vegar,correlation,coupon);
    printf("the price of convertible is %lf\n",optionp);
    delta=(convertible(1.01*s0,strikeprice,vol,rate,k,theta,vegar,correlation,coupon)-optionp)/0.01;
    printf("the delta of this option is %lf\n",delta);
    vega=(convertible(s0,strikeprice,1.01*vol,rate,k,theta,vegar,correlation,coupon)-optionp)/0.01;
    printf("the vega of this option is %lf\n",vega);
}
```

长期可转换债券的 Monte Carlo 定价程序，5 年期，一年支付一次利息，欧式。Matlab 程序：

normrand.m 文件：

```
function x=normrand()
x=sqrt(-2*log(rand(1)))*sin(2*3.14*rand(1));
```

corand.m 文件：

```
function [c,d]=corand(a,b,cor)
c=sqrt(1-cor*cor)*a+cor*b;
d=b;
```

convertible.m 文件：

```
function optionp=convertible(s0,strikeprice,vol,rate,k,theta,vegar,correlation,coupon)
maxs=0;
sumop=0;
intert=5/500;
rv=rate-vol*vol/2;
r(1)=rate;
for i=1:1000
```

```
    sumr=0;
    s1=s0;
    maxs=s0;
    optionp=0;
    for j=1:500
      sumr=sumr+r(j);
      [c,d]=corand(normrand(),normrand(),correlation);
      s1=s1 * exp(rv * intert+vol * sqrt(intert) * c);
      r(j+1)=k * (theta-r(j)) * intert+vegar * sqrt(intert) * d+r(j);
      rv=r(j+1)-vol * vol/2;
      if j==100
        r1=sumr/100;
      end
      if j==200
        r2=sumr/200;
      end
      if j==300
        r3=sumr/300;
      end
      if j==400
        r4=sumr/400;
      end
  end
  if 100/strikeprice * s1<100+100 * coupon
    optionp=100+100 * coupon;
  else
    optionp=100/strikeprice * s1;
  end
  sumr=sumr/500;
  optionp=optionp * exp(-sumr * 5);
  r1=100 * coupon * exp(-r1 * 1);
  r2=100 * coupon * exp(-r2 * 2);
  r3=100 * coupon * exp(-r3 * 3);
```

```
        r4=100 * coupon * exp(-r4 * 4);
        optionp=optionp+r1+r2+r3+r4;
        sumop=sumop+optionp;
        rv=rate-vol * vol/2;
        r(1)=rate;
    end
    sumop=sumop/1000;
    optionp=sumop;
```

主函数 main. m 文件：

```
vol=input('please input the volatility,vol=');
while vol<0
    vol=input('the volatility is less than zero,please input again,vol=');
end
rate=input('please input the interest rate,rate=');
while   rate<0
    rate=input('the interest rate is less than zero,please input again,
rate=');
end
s0=input('please input the stock price,stock=');
while   s0<0
    s0=input('the stock price is less than zero,please input again,stock=');
end
strikeprice=input('please input the strike price,strikeprice=');
while   strikeprice<0
    strikeprice=input('the strike price is less than zero,please input a-
gain,strikeprice=');
end
k=input('please input the k of interest rate,k=');
while   k<0
    k=input('the k of interest rate is less than zero,please input again,
k=');
end
```

```
theta=input('please input the theta of interest rate,theta=');
while  theta<0
    theta=input('the theta of interest rate is less than zero,please input again,theta=');
end
vegar=input('please input the vega of interest rate,vega=');
while  vegar<0
    vegar=input('the vega of interest rate is less than zero,please input again,vega=');
end
coupon=input('please input the coupon,coupon=');
while  coupon<0
    coupon=input('the coupon is less than zero,please input again,coupon=');
end
correlation=input('please input the correlation,correlation=');
while  correlation<0
    correlation=input('the correlation is less than zero,please input again,correlation=');
end
optionp=convertible(s0,strikeprice,vol,rate,k,theta,vegar,correlation,coupon);
optionp
```

2.短期可转换债券定价差分法。一年期,美式。C语言程序。

```
#include<stdio.h>
#include<math.h>
void main()
{
  double ds,dt,vol,rate,s0,coupon,V[100][500],pal,delta,gamma,strikeprice;
  int i,j,intert;
  printf("please input the volatility\n");
```

```
scanf("%lf",&vol);
while(vol<0)
{
  printf("the volatility is less than zero,please input again\n");
  scanf("%lf",&vol);
}
printf("please input the interest rate\n");
scanf("%lf",&rate);
while(rate<0)
{
  printf("the interest rate is less than zero,please input again\n");
  scanf("%lf",&rate);
}
printf("please input the stock price\n");
scanf("%lf",&s0);
while(s0<0)
{
  printf("the stock price is less than zero,please input again\n");
  scanf("%lf",&s0);
}
printf("please input the strike price\n");
scanf("%lf",&strikeprice);
while(strikeprice<0)
{
  printf("the strike price is less than zero,please input again\n");
  scanf("%lf",&strikeprice);
}
printf("please input the coupon\n");
scanf("%lf",&coupon);
while(coupon<0)
{
  printf("the coupon is less than zero,please input again\n");
  scanf("%lf",&coupon);
```

```
        }
    ds=2 * strikeprice/40.0;
        dt=0.9/(vol * vol)/(40.0 * 40.0);
        intert=(int)(1/dt)+1;
        dt=1/intert;
        pal=100+100 * coupon;
        for(j=0;j<40;j++)
        {
          if(100/strikeprice * j * ds>pal)
            V[0][j]=100/strikeprice * j * ds;
          else
            V[0][j]=pal;
        }
        for(j=0;j<intert;j++)
          V[j][0]=pal * exp(-rate * dt * j);
        for(i=1;i<intert;i++)
        {
          for(j=1;j<39;j++)
          {
            delta=rate * j/2 * (V[i-1][j+1]-V[i-1][j-1]);
gamma=vol * vol * (j * ds) * (j * ds)/2/ds/ds * (V[i-1][j+1]-2 * V[i-
1][j]+V[i-1][j-1]);
            V[i][j]=V[i-1][j]+(delta+gamma+rate * V[i-1][j]) * dt;
          }
          V[i][39]=2 * V[i][38]-V[i][37];
          for(j=0;j<40;j++)
          {
            if(100/strikeprice * j * ds>V[i][j])
              V[i][j]=100/strikeprice * j * ds;
          }
        }
        printf("%lf\n",V[intert-1][(int)(s0/ds)]);
    }
```

3. 短期可转换债券定价差分法。一年期，美式。Matlab 程序。

```
V=zeros(300);
vol=input('please input the volatility,vol=');
while vol<0
    vol=input('the volatility is less than zero,please input again,vol=');
end
rate=input('please input the interest rate,rate=');
while  rate<0
    rate=input('the interest rate is less than zero,please input again,
rate=');
end
s0=input('please input the stock price,stock=');
while  s0<0
    s0=input('the stock price is less than zero,please input again,stock=');
end
strikeprice=input('please input the strikeprice,strikeprice=');
while  strikeprice<0
    strikeprice=input('the strike price is less than zero,please input a-
gain,strikeprice=');
end
coupon=input('please input the coupon,coupon=');
while coupon<0
    coupon=input('the coupon is less than zero,please input again,
coupon=');
end
ds=2*strikeprice/40.0
dt=0.9/(vol*vol)/(40.0*40.0);
intert=fix(1/dt)+1;;
dt=1/intert;
pal=100+100*coupon;
for j=1:40
    if 100/strikeprice*j*ds>pal
        V(1,j)=100/strikeprice*j*ds;
```

```
        else
            V(1,j)=pal;
        end
    end
    for j=1:intert
        V(j,1)=pal * exp(-rate * dt * j);
    end
    for i=2:intert
        for j=2:39
            delta=rate * j/2 * (V(i-1,j+1)-V(i-1,j-1));
gamma=vol * vol * (j * ds) * (j * ds)/2/ds/ds * (V(i-1,j+1)-2 * V(i-
1,j)+V(i-1,j-1));
            V(i,j)=V(i-1,j)+(delta+gamma+rate * V(i-1,j)) * dt;
        end
        V(i,40)=2 * V(i,39)-V(i,38);
        for j=1:40
            if 100/strikeprice * j * ds>V(i,j)
                V(i,j)=100/strikeprice * j * ds;
            end
        end
    end
    intert
    fix(s0/ds)
    V(intert,fix(s0/ds))
```

4. 短期可转换债券定价的 Monte Carlo 程序。欧式，1 年期。C 语言程序。

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
#define PI 3.1415926
long in;
double random()
{
```

```
    float xn;
  long m=2147483647,q=127773,r=2836,a=16807;

   if((a*(in%q)-r*(in/q))>=0)
   in=a*(in%q)-r*(in/q);
   else
   in=a*(in%q)-r*(in/q)+m;

   xn=((float)in)/((float)m);
     return(xn);
  }

  double normrand(double u,double v)
  {
  double x;
  x=sqrt(-2*log(u))*sin(2*PI*v);
  return x;
  }

  double convertible(double s0, double strikeprice, double vol, double
rate,double coupon)
  {
   double intert,rv,s1,maxs=0,optionp,sumop=0;
   int i,j;
   intert=1/100;
   rv=rate-vol*vol/2;
   for(i=0;i<1000;i++)
   {
     s1=s0;
     maxs=s0;
     optionp=0;
     for(j=0;j<100;j++)
     {
```

```
    s1=s1 * exp(rv * intert+vol * sqrt(intert) * normrand(random(),ran-
dom()));
        }
        if(100/strikeprice * s1>=100 * (1+coupon))
          optionp=100/strikeprice * s1;
        else
          optionp=100 * (1+coupon);
        sumop+=optionp;
      }
      sumop=sumop/1000;
      optionp=sumop * exp(-rate * 1);
      return optionp;
    }
    void main()
    {
      double vol,rate,s0,strikeprice,delta,vega,optionp,coupon;
      in=rand();
      printf("please input the volatility\n");
      scanf("%lf",&vol);
      while(vol<0)
      {
        printf("the volatility is less than zero,please input again\n");
        scanf("%lf",&vol);
      }
      printf("please input the interest rate\n");
      scanf("%lf",&rate);
      while(rate<0)
      {
        printf("the interest rate is less than zero,please input again\n");
        scanf("%lf",&rate);
      }
      printf("please input the stock price\n");
      scanf("%lf",&s0);
```

```
    while(s0<0)
    {
      printf("the stock price is less than zero,please input again\n");
      scanf("%lf",&s0);
    }
    printf("please input the strike price\n");
    scanf("%lf",&strikeprice);
    while(strikeprice<0)
    {
      printf("the strike price is less than zero,please input again\n");
      scanf("%lf",&strikeprice);
    }
    printf("please input the coupon\n");
    scanf("%lf",&coupon);
    while(coupon<0)
    {
      printf("the coupon is less than zero,please input again\n");
      scanf("%lf",&coupon);
    }
    optionp=convertible(s0,strikeprice,vol,rate,coupon);
    printf("the price of coupon is %lf\n",optionp);
    delta=(convertible(1.01*s0,strikeprice,vol,rate,coupon)-optionp)/0.01;
    printf("the delta of this option is %lf\n",delta);
    vega=(convertible(s0,strikeprice,1.01*vol,rate,coupon)-optionp)/0.01;
    printf("the vega of this option is %lf\n",vega);
}
```

5. 短期可转换债券定价的 Monte Carlo 程序。欧式，1 年期。Matlab 程序。

normrand. m 文件：

```
function x=normrand()
x=sqrt(-2*log(rand(1)))*sin(2*pi*rand(1));
```

convertible. m 文件：

```
function optionp=convertible1(s0,strikeprice,vol,rate,coupon)
maxs=0;
sumop=0;
intert=1/100;
rv=rate-vol * vol/2;
for i=0:999
   s1=s0;
   maxs=s0;
   optionp=0;
   for j=0:99
      s1=s1 * exp(rv * intert+vol * sqrt(intert) * normrand());
   end
   if 100/strikeprice * s1>=100 * (1+coupon)
      optionp=100/strikeprice * s1;
   else
      optionp=100 * (1+coupon);
   end
   sumop=sumop+optionp;
end
sumop=sumop/1000;
optionp=sumop * exp(-rate * 1);
```

main. m 文件：

```
vol=input('please input the volatility,vol=');
while vol<0
      vol=input('the volatility is less than zero,please input again,vol=');
end
rate=input('please input the interest rate,rate=');
while  rate<0
      rate=input('the interest rate is less than zero,please input again,
rate=');
end
```

```
s0=input('please input the stock price,stock=');
while  s0<0
    s0=input('the stock price is less than zero,please input again,stock=');
end
strikeprice=input('please input the strike price,strikeprice=');
while  strikeprice<0
    strikeprice=input('the strike price is less than zero,please input again,strikeprice=');
end
coupon=input('please input the coupon,coupon=');
while coupon<0
    coupon=input('the coupon is less than zero,please input again,coupon=');
end
optionp=convertible1(s0,strikeprice,vol,rate,coupon)
```

第十三章

风险套利

13.1 什么是风险套利

风险套利一般来说是指与改变公司资本结构的事件相关的交易。先介绍一下什么是资本结构。当一家公司募集初始资本时，它可以采用两种方法。一是采用债务融资，即从贷款者处借入资金，到期时偿还本金和利息。或者它可以采用股权融资，即分出一部分公司所有权给投资人，投资者根据其资金所占的比例参与分红，并且根据比例承担风险。公司可能会选择以一种方式筹集资金，或者同时采用两种融资方式，即一部分采用债务融资，一部分采用股权融资。债务和股权在公司资本中所占的比例称作公司的资本结构。再介绍一下事件，这里的事件特指与公司经营相关的事件，它表示公司影响股东和债权人的经营活动。典型的事件包括支付股利、股票拆细、并购、发行新股等等。在本书中，我们仅介绍与并购相关的风险套利方法。

并购套利始于半个多世纪之前，这一方法最初是由高盛的合伙人 Gustave Levy 于 20 世纪 40 年代中期发现的。他的后继者就是 Robert Rubin，同样是高盛合伙人，后来成为美国财政部部长。整个 20 世纪 80 年代，高盛的风险套利部都是最盈利的。

13.2 并购(M&A)

1.定义

并购包括收购(acquisition)和兼并(merger)。

所谓收购,就是一个公司(收购方)购买另一家公司(目标方)。收购可以是友好的,也可以是具有敌意的。在友好收购时,交易双方相互协商,希望可以达成一致意见。在敌意收购时,目标方不愿被收购,或者目标方董事会对于要被收购毫不知情。一般情况下,收购是大公司购买小公司。不过有时也会出现小公司买下大公司甚至是有着悠久历史的大公司的情况,这种情况被称作反向接管(reverse takeover)。另一种收购的形式是反向兼并(reverse merger),这是指未上市公司买下已上市的壳公司。通常这类已上市公司没有实际的经营项目,并且只有不多的资产。未上市公司之所以这样做,是因为对自身前景看好,想上市融资,同时打算节约上市成本。收购被证明很难取得成功,研究发现,只有50%的收购最终取得成功。

所谓兼并就是两家公司合并成为一家更大的公司。这种交易通常是双方自愿的,并通过股票互换、现金支付等形式完成。股票互换是最常用的方法,因为它使得并购双方可以共担交易的风险。兼并与收购很相似,只是常会导致拥有新名字和新品牌的企业出现。有些时候,使用兼并而不是收购只是出于政治或营销的原因。

2.兼并的分类

(1)水平兼并:指这样两个企业的合并:拥有相似生产线和市场的企业,彼此之间存在竞争。

(2)垂直兼并:指企业和其供货商的兼并或企业与其客户或销售商的兼并。

(3)市场扩展兼并:指在不同市场销售不同产品的两家公司的合并。

(4)产品扩展兼并:指在同一市场销售不同但相关产品的两家公司的合并。

(5)聚合(conglomeration)兼并:指两家没有任何经营共同点的公司的合并。

3.并购的动机

(1)规模经济:因为公司合并后可以减少重复的部门,增加和供货商、销售

商的议价能力，扩大市场份额，从而增加利润率。

（2）减少竞争，增强定价能力：公司收购竞争者后，可以扩大自己的市场份额，减少竞争压力，同时也可以增强自己在市场上的定价能力。

（3）互补销售：企业可以并购生产和自己产品相关的产品的企业，这样可以同时增加两种产品的销量。

（4）税收因素：一家盈利很好的企业可以收购一家亏损的企业，以使用这家企业的亏损减少自己的应缴税收。

（5）资源共享：资源在公司间不是平均分配的，并购可以使收购方和目标方交换信息、矿产、客户等稀缺资源。

（6）开拓市场：在经济全球化的今天，开拓海外市场是十分必要的。通过并购海外企业，往往可以绕过政策壁垒，更有效地打开一国的市场。

本章将不区分兼并和收购，统一称为并购。

4. 并购风险

当一个并购交易被公布时，目标公司的股票价格一般会上涨 20%以上。但是，尽管股价会有大幅上涨，目标公司的股票价格相对于收购方的收购价格还是会有 1%到 3%的折扣。出现这种情况的原因是总会存在并购失败的可能性。导致并购失败的原因有很多。可能来自政府的监管部门，因为政府倾向于阻止垄断。所以，当并购有可能会产生垄断时，政府就会出面阻止并购，从而导致并购交易失败。经营环境的改变，经济条件的变化，也会使收购方或目标方中止并购。如果并购会影响股东的利益，则股东也可能集体投票阻止交易。无论交易失败的原因是什么，对目标方股票价格的影响趋势都是一样的。只要交易最终没有完成，股价通常会下跌 25%左右。目标公司的股东可以有两个选择。要么他们坚持持有目标方的股票，承担并购可能失败的风险，希望并购最终完成，获得收购方支付的收购价格；或者，他们可以在收购最终完成前卖出持有的股票，避免并购失败的风险。那些为避免并购失败风险而将股票卖出的人，一般是把股票卖给了并购风险套利者。

并购风险套利者擅长评估交易成功的可能性，他们承担交易取消从而股价大跌的风险，以换取交易成功完成时 1%到 3%的股价升值收益。交易失败的可能，给了并购风险套利者从自己的策略中获利的机会。如果公布的并购交易总是能够成功完成，那么目标方的股价和收购价格的差值就只能反映无风险利率，同时也就没有并购风险套利者的投资机会了。并购风险套利者需要足够的收益来补偿其承担的风险，这一收益就是套利价差，也就是购买时股票价格与最终收购价格的差值。正像投资组合一样，并购风险套利者会采用

组合的方式来分散风险。他们会投资多个并购交易以分散风险。

5.美国并购潮

Brealy 和 Myers 发现，美国的并购存在着波动，主要的五次波动都是始于新商业结构的改变，终于萧条时期。这五个时期包括：

(1)从 1898 年到 1903 年的美国工业的托拉斯化。

(2)从 1923 年到 1929 年组装线的引入。

(3)20 世纪 60 年代的集团化热潮。

(4)20 世纪 80 年代的第二次重组浪潮。

(5)20 世纪 90 年代的由于监管的解除造成的并购。

第一波并购潮可以回溯到 1898 到 1903 年间，许多中小型石油、钢铁和制造业企业谋求合并，以形成在这些行业的垄断。这主要是为了应对 19 世纪末的通货紧缩。第一波并购潮止于罗斯福总统的反托拉斯计划，同时也由于 1902 年开始的经济萧条。

第二波并购潮发生在 1923 年到 1929 年，特征是在电力、燃气和制造业寡头垄断出现后，很多行业出现的并购现象。那一时期，高科技和高增长的制造业是汽车、飞机制造、电视和电力行业。第二波止于 1929 年开始的大萧条。

第三波开始于 20 世纪 60 年代，这一时期，不同行业的企业互相兼并组成大的集团。同时，许多投资银行开始成立并购套利部，赚取了很高的利润。第三波止于 20 世纪 70 年代开始的经济萧条。

第四波始于 20 世纪 80 年代，起因是遏制通货膨胀和国际竞争。这一时期，由于较低的利率和管制的放松，敌意收购频繁发生。并购主要出现在石油、天然气、银行和食品产业。随着一系列金融丑闻的出现，特别是一些并购套利者内幕交易的曝光，并购的黄金年代在 1989 年突然结束。这一时期，很多杠杆化的公司纷纷倒闭，包括垃圾债券之王 Drexel Burnham Lambert。其间，最出名的两个人，莫过于 Michael Milken 和 Ivan Boesky。Milken 的最大功绩在于其大力发展了垃圾债券或高收益债券市场，为这一时期，公司间敌意收购采用垃圾债券融资提供了有效渠道。但最终他因为内幕交易而被监管机构起诉，被罚款 2 亿美元并服刑 22 个月。Boesky 是 Hudson 基金的经理，被起诉之前，曾是美国年轻人的偶像，著名的并购套利专家。1986 年，他因内幕交易被捕，被罚款 1 亿美金。

第五波开始于 20 世纪 90 年代中期。起因是技术变迁和全球金融市场自由化。遍及金融、电信和公用事业的管制放松，刺激了为了创新和降低成本的并购。这一波止于 2001 年新经济泡沫的破裂和随之而来的萧条。

13.3 并购的风险套利方法

并购的风险套利的方式与收购方采用的收购方式有关。最直接的方式是采用现金收购，更复杂的是采用股票收购。对于不同的收购方式，采用不同的风险套利方法，下面将一一介绍。

13.3.1 采用现金收购的风险套利方法

收购最直接的方法就是采用现金直接购买目标公司的股票，而这又可分成两种方法，包括投标竞价(tender offer)和现金并购(cash merger)。在投标竞价的情况下，收购方直接从目标方的股东手中买入股票。在现金并购的情况下，收购方向目标方公司支付现金，目标方公司再将现金支付给股东用以购买在外发行的股票。尽管投标竞价和现金并购有着法律和税收上的不同，对于风险套利者来说，这两者最重要的不同点是现金并购比投标竞价完成的时间要长。

不论是投标竞价还是现金并购，风险套利者所采用的方法是一样的。所用的方法是：在交易的消息宣布后买入目标公司的股票，持有直至并购完成。在交易完成时，套利者获得并购的收购价格与买入目标公司股票时的价格的差价。

下面举一个美国的例子来说明这种方法。

2001年10月22日，有消息称可口可乐有意收购Odwalla公司。当日，Odwalla的股票大涨48%，从前一日的收盘价6.80美元涨到22日的10.05美元。接下来的五天，由于投机者认为协议最终会达成，Odwalla的股票上涨至11.83美元。

10月30日，可口可乐宣布已经与Odwalla的董事会达成一致意见，可口可乐将以15.25美元一股收购Odwalla的全部流通股，收购采用投标竞价的方式。15.25美元意味着相对于并购宣布前一天的股价有29%的溢价。如果是与消息发出当日的股价比较，则溢价为144%。并购被宣布的当天，Odwalla的股票收盘于15.13美元，相比于收购的价格折价0.79%。

套利者可以在消息散布时立即买入Odwalla的股票，买入价为10.05美元一股，希望并购最终以一个更高的价格完成。如果交易最终没有完成，

Odwalla 的股票价格会大幅下跌，导致套利者出现重大损失。最终，交易顺利完成，并且成交价远高于 10.05 美元一股，所以套利者在 6 天内收益率达到 51%。

这个例子说明，根据消息投资回报是丰厚的，同时也可能产生重大的损失。要估计并购交易顺利完成的概率和交易的实际价格是非常困难的（如果不是不可能的话）。所以许多套利者选择知道明确的协议公布时才建立头寸。

如果一个套利者等到可口可乐宣布以 15.25 美元一股收购 Odwalla 以后才购入 Odwalla 的股票，则其买入价为 15.13 美元一股。这样，套利者实际是希望可口可乐能够按照协议收购 Odwalla 的股票，使得其可以获得 0.79% 的收益。实际上，在 12 月 11 日，并购宣布的 30 天之后，可口可乐按照约定完成收购。套利者一个月的收益率为 0.79%，年化收益率为 6.8%。

如果可口可乐没有按照协议完成对 Odwalla 的收购，则 Odwalla 的股票价格会下跌几块钱。由此可见，并购套利的收益和损失是非常不平衡的，收益是 0.12 美元，损失却是好几块。因此，只有当并购成功的概率远远大于失败的概率时，套利的预期收益才能超过无风险利率。

虽然当公司采用更复杂的技术完成收购时套利的方法更加复杂，但基本原理与采用现金收购时采用的套利方法的原理是一样的。一旦价差（执行价与买入价的差）能够对损失风险提供足够的补偿，并购风险套利者就会锁定价差，获取利润。下面会介绍非现金收购时的套利方法。

13.3.2 固定股票交换率并购的套利方法

本节通过一个具体的例子来介绍固定股票交换率并购的套利方法。

2001 年 9 月 3 日，惠普和康柏公司宣布，两家公司已经达成协议，惠普将以股票交换的方式收购康柏。交易合约表明，到并购完成时，一股康柏公司的股票将交换 0.6325 股惠普公司的股票。因为 0.6325 的交换比率在并购合约中已经写明，所以将来不会发生改变，这样的收购称作固定股票交换率收购。

在固定股票交换率收购中采用的套利方法比在现金收购中采用的套利方法要复杂。除了买入目标公司的股票，套利者还要卖空收购方的股票。在惠普康柏的例子中，套利者每买入一股康柏公司的股票，就要卖出 0.6325 股惠普公司的股票。

在 2001 年 9 月 4 日，也就是并购被宣布的第二天，康柏的股价接近 11.08 美元而惠普的股价接近 18.87 美元。套利者会卖空 0.6325 股惠普的股票，获

得11.94(0.6325×18.87),同时买入1股康柏的股票,花费11.08美元。0.86美元(11.94－11.08)就是套利的价差。如果交易顺利完成,套利者的康柏股票就会被换为0.6325股惠普的股票。这样,套利者在惠普的头寸恰好相互抵消,使得套利者获得等于价差的利润。

以上的例子忽略了三个影响最终套利收益率的现金流。第一,套利者买入了康柏公司的股票,因此有权利获得康柏公司的股利。第二,套利者同样卖出了0.6325股惠普的股票,所以有义务支付相当于0.6325股惠普股票的股利。第三,套利者可以获得由卖空惠普公司股票所获现金产生的利息。

如果在9月4日套利者建立头寸,希望交易能够完成,那么如果交易如期完成,他将获得7.8%的收益率。考虑到一般这类交易会持续3.5个月,年化收益率为29.4%。这样看来,交易的收益率是非常高的。然而,如果交易受到阻碍,套利者受到的损失将远远大于预期的7.8%的收益。而实际上不幸的事也确实发生了,11月6日,惠普董事会成员Walter Hewlett公开声明,他将投票反对并购。对于套利者来说,Hewlett的决定是未预料到的和痛苦的。在他声明他的决定的那天,惠普的股价上升2.92美元而康柏的股票下跌0.44美元。套利价差从7.8%扩大至47.4%。如果套利者投入100美元,则现在只剩72美元。

最终,2002年5月3日,经过一系列的努力,惠普最终完成了对康柏的收购。包括卖空惠普股票支付的股利,买入康柏股票收到的股利和卖出惠普股票所获现金产生的利息,套利者获利8.9%,年化收益率为14%。

13.3.3 可变股票交换率并购的套利方法

在固定股票交换率并购过程中,收购方用来交换目标方的股票数量比率是提前确定的。在可变股票交换率并购中,收购方用来交换目标方的股票数量,取决于一段特定时期收购方股票的平均价格,这一时期一般接近于并购完成日。这段日期一般被称作定价时期。

可变股票交换率并购有很多种方式,受篇幅限制,这里只介绍最简单的一种,更复杂的情况将在后续的书籍中分析。最简单的方式是,每一股目标公司的股票交换确定价值的收购方公司的股票。实际最终交换的收购方的股票数量,使用已确定的收购方的股票价值除以定价时期收购方股票的平均价格来确定。如果平均价格低,则目标方收到的收购方的股票就相对较多。相反如果平均价格较高,目标方收到的股票就相对较少。股票数量的变动使得目标方收到的资本值维持一个不变的数字。

在定价时期开始之前，可变股票交换率并购和现金并购类似，因为收购方向每一股目标方股票支付的金额与收购方的股票价格无关。所以，这时套利者只要买入目标公司的股票，但不需要卖空收购方的股票。定价时期结束后，收购方与目标方的股票交换比率就确定了，可变股票交换率并购就与固定股票交换率并购相同了。这样，为了获得套利价差，套利者就需要按照交换比率买入目标方股票同时卖空收购方股票。因此，在定价时期，套利者卖空收购方股票，从现金并购套利向股票并购套利转换。

13.4 影响因素

详细的套利分析涉及对历史并购数据的分析、估计分布、建立数学模型等。但是，涉及并购的风险套利实际上很难用数学模型加以预测，因为每次并购的情况都不相同。数学模型的指导意义并不像在配对交易中那么明显。所以本书没有介绍这方面的数学模型，如果有需要，将会在后续的书籍中介绍。相比之于数学模型，一些涉及法律、政策和经济财务基础的因素对并购风险套利可能更加重要。所以，我们把一些需要考虑的因素列在下面。

1. 合约的明确执行。在很多情况下，在并购双方对所有条件都达成一致意见之前，并购的消息就已被公布出来。例如，一种常见的情况是，确定交易价格之后，并购的消息就公布了，但并购双方对于许多具体的问题并未决定。只有当所有的并购条款都达成一致意见，并购才会上报证券与交易委员会。许多交易即使确定了交易价格，还是无法在其他条款上意见一致，最终导致交易无法达成。

2. 收购方的融资能力。收购方能否获得必要的收购资金，在现金交易中是最关键的问题。如果收购方并没有准备好足够的现金完成收购，那么它就必须通过股权或债权融资。一旦收购方的财务状况使得债务或股权融资成本太高，则收购方有很大的可能无法完成收购。

3. 监管部门的批准。并购协议起草和签订之后，该协议会被上交监管部门批准。监管部门审查并确保合约的条款完备，并且与并购相关的信息没有误导投资者。如果监管部门认为合约不完善，它会把合约退回企业修改。虽然监管部门不一定会导致并购的失败，但它会拖延并购的时间，降低年化收益率，并增加阻碍并购实现事件发生的概率。

4. 政府部门的批准。反垄断是评估交易失败概率时需要考虑的非常重要

的问题。这一点在欧美国家非常重要，严格的反垄断法使得大企业的合并受到严格的控制，常常会导致合并的失败。在国际并购的情况下，政府部门的批准也非常重要，例如华为收购3COM的失败主要是因为没有得到美国政府的批准。中国已颁布反垄断法，所以对相关企业的并购活动会产生一定的约束。

5. 相关人员的限制。并购会影响许多相关人员的利益，包括雇员、债务人、竞争者等。这些人可能通过法庭或其他途径向并购双方施压，可能增加交易失败的可能性。

6. 并购对宏观经济环境波动的敏感性。1998年的亚洲金融危机提供了关于这一点的一个非常好的例子。在某些行业，危机从根本上改变了某些并购企业的经济状况，从而使得一些并购交易最终失败。

7. 收购方或目标方经营表现与预期表现的偏离。收购方和目标方的经济健康状况对于并购的最终完成至关重要。如果在并购过程中目标方的现金流和前景恶化，收购方就会倾向于重新就收购价格进行谈判，以压低收购价格。如果目标方的恶化程度较为严重，则收购方可能会停止收购。同样，如果收购方的前景严重恶化，收购方的管理者可能会停止收购，转而关注自己公司出现的问题。

13.5 并购套利基金

存在两种并购套利基金。一种采用集中投资的方法，因为他们相信自己比其他人能够更好地发现和解释与并购交易相关的信息。为了能够有优异的表现，这类基金必须有丰富的信息来源，甚至是一些市场上未曾出现的信息。这使得这类基金的风险较大。另一种方式是采用分散的投资组合的方法。在这种基金中，单一交易的最大头寸一般不超过资金总量的5%。获得和分析市场信息是十分重要的，因为这可以帮助基金评估各单一交易之间的相关性，从而有效地分散风险。

两种基金同样需要完善的运营结构，其必须能够完成财务分析、金融建模、法律分析，并具有很强的交易能力。例如，较扎实的公司财务和会计知识有助于分析收购方和目标方的财务状况，评估收购方的财务健康程度，以分析收购完成的可能性。金融建模则可以帮助分析复杂的收购工具，如优先股、期权等。虽然交易能力一般被认为不如分析能力重要，但是对于并购套利来说，交易能力十分重要。这是因为套利的利差非常小，只有降低交易成本才能提

高利润。最后，因为并购涉及一系列的法律问题，基金必须熟悉相关的法律法规和监管制度。另外，除了基金管理者的能力，另一个非常重要的因素是基金管理的资本量。分析和收集信息是基金运营的重要成本，而且一般是固定的。大的套利基金可以将这些成本分散于更多的投资资金之中，从而达到规模经济。同时，大的基金拥有更强的卖空能力。要卖空一只股票，首先要将这只股票借入，而大的基金公司的借入能力更强。从而，对于某些难以借入的股票，大的基金公司可以借入卖空从而盈利，但小的基金公司却无法做到。

从国外并购套利基金的业绩来看，并购套利基金的收益还是很好的。Hedge Fund Research 的研究表明，如果在 1990 年 1 月将 1 美元采用并购风险套利方式投资，则到 2001 年，该方法年化收益率为 12.2%，标准差为 4.5%。相比之下，同期的股票市场年化收益率为 12.1%，标准差为 14.9%。可见，在收益相似的情况下，并购套利的波动更小。

第十四章

股指期货套利

14.1 股票指数

股票指数是个股的加权组合。股票指数的值是根据组成指数的个股的股价按一定的计算方法得出的，其随着个股股价的变动而变动，代表了某种股票组合的整体变动。

股票指数一般由交易所或商业机构设计，设计包括以下几个方面：

1.加权方法。加权方法用来建立成分股与股票指数的关系，一般包括以下三种加权方法：

(1)根据资本值加权。这种方法要求每只成分股的权重与其市场价值成比例。市场价值采用股价乘以在外发行股数计算。这种设计认为，股票的市值决定了其对股指的影响。只有在成分股的市场价格和在外发行股数发生改变时，权重才需要调整。大多数股票指数都是按资本值加权的。

(2)价格加权。这种方法要求各只股票的权重与其价格成比例。这种设计认为，高价格的股票对指数有更大的影响。只有当成分股的在外发行股数发生变动时，按照该种方案设计的股票指数才需要调整。很少有指数是价格加权的。价格加权的例子如，道琼斯工业平均指数、日经 225 指数。

(3)相等权重。这种方法要求所有成分股有相同的权重。当成分股价格改变的比率不相同时，按照该种方案设计的股票指数需要调整。相等权重的指数需要频繁地进行调整。很少有指数采用相等权重。

2.平均值方法。股票指数采用以下两种平均值方法：

(1)算数平均值。指数值按下式计算：

$$\text{Index value}=\omega_1 P_1+\omega_2 P_2+\cdots+\omega_i P_i$$

其中 ω_i 为第 i 只股票的权重，P_i 为第 i 只股票的价格。

(2)几何平均值。指数值按下式计算：

$$\text{Index value}=(P_1^{\omega_1})\times(P_2^{\omega_2})\times\cdots\times(P_i^{\omega_i})$$

其中 ω_i 为第 i 只股票的权重，P_i 为第 i 只股票的价格。

大多数指数采用算数平均值。

3. 股利的处理。这是关于现金股利是否被包含进指数。多数股票指数只与股价相关，为便于计算而忽视了股利。某些指数在计算指数值时将股利加入股票价格，以使指数中包括股利。例如德国 DAX 指数。

4. 股票的选择。什么股票会被包含进指数？这既可以是任意的，也可以(多数时候)服从一定的规则。一个相关的问题是：是否应该包含那些很大比例股份不能自由交易的股票？对此尚没有结论。

表 14.1 给出了全球主要股票指数的计算方法说明。

表 14.1 全球主要股票市场指数

国 家	指 数	描 述
澳大利亚	S&P/ASX200	资本值加权，包含 200 只股票
加拿大	CAD60	资本值加权，包含 60 只股票
德 国	DAX	资本值加权，包含 30 只蓝筹股
英 国	FTSE—100	资本值加权，包含 100 只最大公司的股票
法 国	CAC—40	资本值加权，包含 40 只股票
香 港	恒生指数	资本值加权，包含 33 只股票
美 国	S&P500	包含 500 只广泛持有的股票
日 本	日经 225 指数	价格加权，包含 225 只日本蓝筹股

14.2 期货

期货合约就是交易的双方决定在将来某一特定时刻，以一个签订合约时确定的价格，交易某种商品的合约。期货合约一般在交易所交易。交易所规

定每份合约交易的商品数量，合约都是标准化的。为了防止违约，期货交易采用保证金交易机制。保证金是期货合约所代表的商品价值的一个比率。期货一般在交易所内交易，世界上最大的两家期货交易所是芝加哥交易所(CBOT)和芝加哥商品交易所(CME)。

1. 期货交易术语

期货交易包括一些术语，现简单介绍如下：

平仓：平仓是持有一个与初始交易头寸相反的头寸。例如，某投资者在7月3日买入一份9月到期的大豆期货合约，到8月5日，他可以通过卖出一份9月到期的大豆期货合约来平仓。这样，投资者就不用实际交割合约标的物。其实，绝大部分期货合约都是在到期前就平仓了，只有很少的一部分期货合约进行实物交割。这主要是因为按照期货合约的规定进行交割较为不便，而且通常成本较高。

交割月份：交割月份是期货合约中规定的合约进行实物交割的月份。交割月份由交易所规定，一般是季度月，即3、6、9、12月。由交易所指定特定月份合约开始交易的时刻，交易所同时也对给定的合约的最后交易日做了规定。

保证金：保证金之所以存在，主要是因为交易者之间的履约风险。如果两个交易者直接签订一份期货合约而不通过交易所，这样，到期时有损失的一方就很可能违约，同时，这一方可能也没有足够的资金来履约。因此，为了防止这一类风险，期货交易合约由交易所统一规定，并要求交易者支付保证金。为了后面分析方便，这里介绍一下保证金的操作方法。假设投资者买入一份期货合约，合约的标的资产价值为10 000元。如果交易所规定，该种合约的保证金比例为20%，那么该投资者就需要存入初始保证金2 000元。每天交易结束后，保证金账户都要根据期货价格进行调整，以反映投资者的盈亏，这就是盯市。例如，投资者买入合约的第二天，期货价格下跌到9 800元，那么就要从2 000元的保证金中扣掉200元，以反映投资者的亏损。同样，如果期货价格上涨到10 200元，则投资者的保证金账户中就会增加200元到2 200元。当保证金下降到一定数值时，即下降到维持保证金时，投资者就会收到追加保证金通知。维持保证金是初始保证金的一个比率，是交易所为了避免保证金账户为负而设立的。保证金的数额不能低于维持保证金的数值，如果达到维持保证金的值，就会被追加保证金。如果投资者不能按时追加保证金，合约将被强行平仓。

2. 期货价格的理论值

考虑投资者采用这样一个组合：t时期，投资者以借入资金，按现价S买

入标的资产，借款利率为 r；同时，投资者以期货价格 F 卖出一份期货合约，到期时间为 T。假设标的资产不支付任何收益。到 T 时，投资者将 t 时买入的资产用于期货合约的交割，则投资者获得收益为 F。同时，投资者需要支付贷款本金和利息。假设连续复利，则需支付本金加利息为 $Se^{r(T-t)}$。由于所有的现金流都可以在 t 时刻确定，没有风险，且投资者没有投入，则投资者的收益应该等于贷款的支付。所以：

$$F=Se^{r(T-t)}$$

这就是期货的理论价格。如果期货的价格与该值不等，则套利者就可以按照上面的方法进行套利，使得期货价格趋近于理论值。

如果标的资产支付收益（如股息），那么上面的组合到 T 时的收益为 F，但所支付的贷款变为 $(S-D)e^{r(T-t)}$，其中 D 为收益的现值。这样，期货定价公式变为：

$$F=(S-D)e^{r(T-t)}$$

这是收益离散支付的情况，如果收益按照比例连续支付，则 T 时的收益为 F，到期需支付贷款为 $Se^{(r-d)(T-t)}$，d 为收益支付率。期货定价公式变为：

$$F=Se^{(r-d)(T-t)}$$

14.3 股票指数期货

股票指数期货是重要的交易所交易的衍生产品之一。它产生于 20 世纪 80 年代早期，现在在发达国家和发展中国家非常普遍。

1. 股票指数期货的特征

(1)标的股指。股票指数期货的标的股指既可以是含有广泛股票的指数，如标准普尔 500 指数；也可以是分行业的指数，如高科技类公司组成的指数；或者根据资本量的指数，如小盘股指数；甚至跨国的指数，如 MSCI 指数等。

(2)交割机制。股票指数期货采用现金交割机制。这主要是受到成分股交割难以执行的限制。

2. 股指期货定价存在的问题

(1)可使用的无风险利率。理论上，期货定价采用的利率是无风险利率。交易员可以以同样的无风险利率贷出和借入资金。实际上，存在一系列可用

的利率，包括国债利率、互换利率、回购利率等。不同的利率会计算出不同的价格。

（2）股利估计。理论上，需要知道期货从签订到到期时实际支付的股利值。但实际上，股利的支付量只能估计。因为期货的到期时间一般比较短，一般少于三个月，估计股利的风险相对较小。

（3）套利限制。股票指数期货需要套利来保证有效定价。这就需要借入股指成分股并卖空。实际中这是非常困难的。当指数包含大量成分股时，套利需要交易指数所包含的所有成分股。这会产生很高的交易成本，并且受制于流动性，做空不一定能够完成。这使得定价的有效性受到影响。

3. 错误定价的原因

实际中，股指期货以接近公平价值的价格交易。然而，时常也会发生价格大幅偏离公平价值的情况。错误定价发生的原因包括：

（1）市场的非有效性。由于卖空受限、复制指数精度不足、低流动性和高交易成本等原因，导致套利受到限制，可能导致错误定价。

（2）税收因素。对于股利的税收处置会影响实际股票与股票指数的相对价值。这会扭曲合约的价值。

（3）市场波动性因素。当市场波动性很高时，现货期货市场的关系可能会被扭曲，导致期货被错误定价。

4. 其他类型标的指数

股指期货除了以包含广泛成分股的指数（如标准普尔 500 指数）为标的指数外，还存在其他的情况，这里简单介绍如下：

（1）新部门指数。这包括一系列新技术公司股票组成的指数，还有包含部门内容更窄的新技术股票组成的指数，例如半导体公司指数。

（2）周期部门指数。这种指数使得交易员可以在行业之间轮动，以利用商业周期的影响。相关的创新还包括基于建筑等强周期股票和零售、食品制造等防御性很强的股票组成的指数。

（3）商品部门指数。一些股票指数产品是基于商品类股票的。这包括基于商品类公司（如石油、金属类）的股票的指数。

（4）利率敏感类指数。这包括基于银行类和房地产类公司股票的指数。

（5）根据资本值大小不同的指数。这类指数集中于按照资本值的大小将市场分类。一个典型的例子是基于小盘股的股票指数。

（6）投资风格指数。这包括价值型股票指数和增长型股票价格指数。

（7）跨国指数。这类指数着眼于全球市场，如 MSCI 指数。

14.4 套利方法

股指期货的套利方法包括：跨期套利、期现套利、跨市场套利、跨品种套利。下面就对其一一进行介绍。

14.2.1 跨期套利

首先谈跨期套利。跨期套利是指利用两个不同交割月份的股指期货合约之间的价差进行的套利交易。一般来说，相同标的指数的股指期货在市场上会有不同交割月的若干合约同时交易。例如美国通常采用季度月作为交割月，即交割月是3、6、9、12月。在CBOT上市交易的道指期货合约在任何时间都有3、6、9、12四个交割月的合约在交易，而CME的标准普尔500指数期货合约则有从当前季度算起的八个季度月的合约在交易。由于同时交易的不同交割月合约均是基于同一标的指数，所以，在市场预期相对稳定的情况下，不同交割日期合约间的价差应该是稳定的，一旦价差发生了变化，则会产生跨期套利机会。严格来讲，跨期套利不是无风险套利，它实际上类似于前面所讲的配对交易，投资者需要对不同到期月的期货合约的价差做出预测，具有投机性，但因为交易行为是建立在价差的基础上，所以，风险要远远小于纯粹的投机交易(即单方向做多，或单方向做空)。跨期套利的操作重点在于判断不同到期月合约的价差将来是扩大还是缩小，而不是整个市场的未来走势。

跨期套利一般包括以下三种模式：牛市跨期套利、熊市跨期套利和蝶式套利。

牛市跨期套利即执行卖近买远策略。如果投资者看多股市，认为较远交割期的股指期货合约涨幅将大于近期合约的涨幅，或者说较远期的股指期货合约跌幅将小于近期合约的跌幅，就可以采用此策略。换言之，牛市套利即是认为较远交割期合约与较近交割期合约的价差将变大。从价值判断的角度看，牛市套利认为远期的股指期货的价格应高于当前远期股指期货的交易价格，当前远期股指期货的价格被低估。我们就可以买入远月合约，同时卖出近月合约。

例如，假定某一指数期货某天3月合约的价格为3 000，6月合约的价格为3 040，则两者价差为40。如果该指数期货3月和6月合约的价差在80左

右，则可能现在6月合约被低估。投资者可以买入6月合约，卖出3月合约。一段时间后，3月合约上涨到3 050，6月合约上涨到3 125，两合约价差为75，接近80。此时投资者平仓，即卖出同样数量的6月合约，买入同样数量的3月合约。假设股指期货每个点为300元，则该投资者的总盈亏见表14.2。

表14.2 牛市跨期套利例1

	3月合约	6月合约	价差
交易开始	3 000	3 040	40
交易结束	3 050	3 125	75
盈亏	−50	85	35
总盈亏＝35×300＝10 500(元)			

牛市跨期套利并非没有风险，如果一段时间后，3月合约上涨到3 050，但6月合约上涨到3 070，价差为20，即两份合约的价差反而缩小，那么该投资者的总盈亏见表14.3。

表14.3 牛市跨期套利例2

	3月合约	6月合约	价差
交易开始	3 000	3 040	40
交易结束	3 050	3 070	20
盈亏	−50	30	−20
总盈亏＝−20×300＝−6 000(元)			

所以，牛市跨期套利能否盈利，取决于投资者对于价差方向预测准确与否，其实际上就是前面所讲配对交易在股指期货的应用，只不过是由不同品种变为同一品种不同时期的配对。

熊市跨期套利与牛市跨期套利相反，即执行买近卖远策略。认为较远交割期合约的跌幅将大于近期合约，或者说远期的股指期货合约涨幅将小于近期合约涨幅，换言之，熊市套利即是认为较远交割期合约与较近交割期合约的价差将变小。在这种情况下，远期的股指期货合约当前的交易价格被高估。我们就可以买入近月合约，同时卖出远月合约。

例如，假定某一指数期货某天3月合约的价格为3 000，6月合约的价格为3 040，则两者价差为40。如果该指数期货3月和6月合约的价差在10左

右，则可能现在6月合约被高估。投资者可以卖出6月合约，买入3月合约。一段时间后，3月合约下跌到2 980，6月合约下跌到2 985，两合约价差为5，接近10。此时投资者平仓，即买入同样数量的6月合约，卖出同样数量的3月合约。假设股指期货每个点为300元，则该投资者的总盈亏见表14.4。

表14.4 熊市跨期套利例1

	3月合约	6月合约	价 差
交易开始	3 000	3 040	40
交易结束	2 980	2 985	5
盈 亏	−20	55	35
总盈亏=35×300=10 500(元)			

熊市跨期套利并非没有风险，如果一段时间后，3月合约下跌到2 980，但6月合约下跌到3 030，价差为50，即两份合约的价差反而扩大，那么该投资者的总盈亏见表14.5。

表14.5 熊市跨期套利例2

	3月合约	6月合约	价 差
交易开始	3 000	3 040	40
交易结束	2 980	3 030	50
盈 亏	−20	10	−10
总盈亏=−10×300=−3 000(元)			

所以，与牛市跨期套利类似，熊市跨期套利能否盈利，同样取决于投资者对于价差方向预测准确与否，其实际上也同样是前面所讲配对交易在股指期货的应用。

基于上面的分析，既然牛市和熊市跨期套利本质上都是配对交易，那么，第十一章所介绍的配对交易的方法，同样都可以用在股指期货跨期套利中。这包括协整方法的使用，此时P_A、P_B分别代表不同交割月的合约的价格。止损分析和Cuscore也是类似的。

蝶式套利实际上是上述买近卖远策略和卖近买远策略的组合，即两个方向相反、共享中间交割月份的跨期套利的组合，参与交易的期货合约为同一标的指数的三个不同交割月合约。当投资者认为中间交割月份的股指期货合约

与两边交割月份合约价格之间的价差将发生变化时，会选择采用蝶式套利。采用蝶式套利时，当近月合约到期后，可以将近月与中月合约进行相同数量的反向操作平仓了结。剩余的另外一半（中月合约与远月合约）可以继续进行套利活动，也可以全部平仓了结。

例如，假定某一指数期货某天 3 月份合约的价格为 3 000，6 月份合约价格为 3 040，9 月份合约价格为 3 060。如果投资者认为 6 月份合约价格被低估，3 月和 9 月合约价格被高估。那么，他可以买入两份 6 月份合约，分别卖出一份 3 月份合约和 9 月份合约。如果一段时间后，3 月份合约价格为3 020，6 月份合约价格为 3 090，9 月份合约价格为 3 100。假设股指期货每个点为 300 元，则该投资者的总盈亏见表 14.6。

表 14.6　蝶式套利例 1

	3 月合约	6 月合约	9 月合约
交易开始	3 000	3 040	3 060
交易结束	3 020	3 090	3 100
盈　　亏	−20	2×50−100	−40
总盈亏＝(−20＋100−40)×300＝12 000(元)			

当然，与牛熊市跨期套利相同，蝶式套利也会出现亏损，例如，如果一段时间后，3 月份合约价格为 3 030，6 月份合约价格为 3 070，9 月份合约价格为 3 110，则该投资者总盈亏见表 14.7。

表 14.7　蝶式套利例 2

	3 月合约	6 月合约	9 月合约
交易开始	3 000	3 040	3 060
交易结束	3 030	3 070	3 110
盈　　亏	−30	2×30＝60	−50
总盈亏＝(−30＋60−50)×300＝−6 000(元)			

实际上，由于套利操作有利于减小股指期货价格的波动和维持股指期货的合理定价，交易所为鼓励投资者套利而给予套利操作更低的交易成本。绝大多数交易所都允许价差交易，即套利者无须同时买入和卖出股指期货合约，而可直接买卖两份合约的价差，从而大大节省占用的保证金。

14.2.2 期现套利

1. 期现套利原理

期现套利是指股指期货与股指现货之间的套利，是利用股指期货合约与其对应的现货指数之间的定价偏差进行的套利交易。即在买入(卖出)某个月份的股指期货合约的同时卖出(买入)相同价值的标的指数的现货股票组合，并在未来某个时间对两笔头寸同时进行平仓的一种套利交易方式。

由于股指期货合约在到期时是按照现货指数的价格来进行现金交割，即期货合约价格在到期时会强制收敛于现货指数，这也就会使得在正常交易期间内，期指与现指会维持一定的动态联系。在各种因素影响下，由于期指相对于现指对信息的反应速度更快，因此，其波动性会大于现指，经常会与现指产生偏离，但当这种偏离超出一定的范围时，就会产生套利机会。期现套利属于无风险套利。只要定价偏差的收益能覆盖掉交易成本，就可以进行期现套利操作，而不用关心市场的未来走势。

在套利过程中涉及买入或卖出股指现货，但实际上不存在股指现货。常用的方法是采用指数基金、ETF 或股票现货组合来模拟指数。然而，无论采取哪种方式，构建出来的投资组合均会与标的指数产生或大或小的偏差，这就是跟踪误差，它衡量的是在一定时期内，投资组合收益率与标的指数收益率的偏离程度。

由于存在交易成本和跟踪误差，股指期货期现套利并不是当期货价格回到理论值时平仓，而是存在一个无套利区间。当股指期货价格与理论价格的差值超过无套利区间时，才开始交易；当股指期货价格回到无套利区间时终止交易。

期现套利的方式可分为正向套利和反向套利两种。正向套利即买入现货，卖出期货。若股指期货与股指现货的价格比高于无套利区间上限，套利者可以卖出股指期货，同时买入相同价值的指数现货，当期现价格比回落到无套利区间之后，对期货和现货同时进行平仓，获取无风险套利收益。反向套利即卖出现货，买入期货。是指当股指期货与股指现货的价格比低于无套利区间下限时，套利者可以买入股指期货，同时卖出相同价值的指数现货，在期现价格比回落到无套利区间内时，对期货和现货同时进行平仓，获取无风险套利收益。

2. 无套利区间的计算

股指期货的理论公式为：

$$F_t = S_t e^{(r-d)(T-t)}$$

其中，F_t 为股指期货的理论价格，S_t 为股票指数现货的价格，r 是无风险利率，d 是股利支付率，t 为现在的时间，T 为合约到期时间。

设 F_t^0 表示 t 时刻股指期货的实际价格；C 为期现套利的交易成本，以头寸的百分比表示；$K_{t,T}$ 为套利期间以现货模拟指数的跟踪误差，以套利期间现货模拟组合与标的指数的收益率的误差占现货的百分比表示。整个期现套利的总成本为(包括交易成本和跟踪误差)：

$$(F_t+S_t)C+S_tK_{t,T}$$

当股指期货的实际价格与其理论价格的差超过总成本时，可以正向套利，即卖出股指期货，买入股指现货组合。此时：

$$F_t^0-F_t>(F_t^0+S_t)C+S_tK_{t,T}$$

将 F_t 的理论表达式带入，经整理得：

$$\frac{F_t^0}{S_t}>\frac{C+\mathrm{e}^{(r-d)(T-t)}+K_{t,T}}{1-C}$$

当股指期货的理论价格与其实际价格的差超过总成本时，可以反向套利，即买入股指期货，卖出股指现货组合。此时：

$$F_t-F_t^0>(F_t^0+S_t)C+S_tK_{t,T}$$

将 F_t 的理论表达式带入，经整理得：

$$\frac{F_t^0}{S_t}<\frac{\mathrm{e}^{(r-d)(T-t)}-C-K_{t,T}}{1+C}$$

因此，无套利区间为：

$$\frac{\mathrm{e}^{(r-d)(T-t)}-C-K_{t,T}}{1+C}<\frac{F_t^0}{S_t}<\frac{\mathrm{e}^{(r-d)(T-t)}+C+K_{t,T}}{1-C}$$

因此当股指期货与现货价格比超过无套利区间时，按正向或反向套利的方法建立头寸。当这一比值回到无套利区间时，结束交易。

3. 现货指数的模拟

期现套利的一个关键问题是现货指数的模拟。股指期货标的指数基金是最简单的现货替代品种。它可使投资者避免对指数成分股组合直接进行买卖交易，而将现货的模拟环节交由基金管理公司来执行，同时，该现货模拟方法具有交易成本低、拟合度高的特点，较为简单易行。但股指期货标的指数基金存在规模和份额限制，该现货模拟方法不适于进行短期或大规模资金的期现

套利活动。

另一种模拟指数的方法是ETF。与普通的指数基金相比，ETF具有以下优点：首先，ETF基金具有交易成本低、交易方便，交易效率高等特点。其次，ETF采用完全被动的指数化投资策略，管理费较低，操作透明度较高，可以让投资者以较低的成本投资于一篮子标的指数成分股。但套利期间收到现金分红的可能性非常小，而且ETF不太可能获取丰厚的超额收益。

第三种方法是使用指数的成分股来模拟指数，一般有三种方法：

(1)购买标的指数中的所有成分股票，并且按照每种成分股票在标的指数中的权重确定购买的比例以构建指数组合，从而达到复制指数的目的。这是最原始的方法，也是最繁杂的方法。它的弱点很明显：①需要对指数成分股以及成分股权重的变动做精确的跟踪。②由于股票指数的成分股数量比较多，即使选取的都是流动性较高的股票，在同一时间买进或卖出几十上百只股票，其难度相对较高。因此以完全复制法构建指数现货的建仓成本和维护成本都较高。

(2)分层抽样法。采用两阶段优选法。第一阶段是抽样，根据一定标准选出样本股票；第二阶段是权重优化配置，通过最优化算法，得到跟踪误差最小的最优权重，同时保证较小的调整频率和跟踪成本。

(3)优化抽样法。其与分层抽样法的本质区别在于不需要进行独立的抽样，采用单阶段优化法，直接通过最优化算法确定成分股品种和权重。优化抽样法和分层抽样法减少了成分股的数量，降低了建仓成本和维护成本，但是需要注意以下两点：①注意所选样本必须是在所有成分股中所占权重较大的股票；②必须是经营状况良好，且是股票价格不易受到操控的股票。

总的来看，抽样复制易于操作、成本较完全复制低，但抽样拟合度很难达到理想水平。完全复制虽然拟合度较高，但交易和冲击成本较高，可行性较小。

14.2.3 跨市场套利

如果同一标的指数在两个以上的市场上均有股指期货，那么各个市场的同一股指期货的价值应该接近。如果出现了较大的偏离，就可以进行跨市套利。投资者可以在不同市场上，买入价值低的卖出价值高的同一股指期货形成套利头寸。例如日经225指数期货同时在新加坡和东京进行交易，如果两个期货的价格偏差较大，投资者就可以进行套利。实际上，搞垮巴林银行的里森，最开始就是采用日经225指数期货的跨市场套利获利较好，才获得上司赏识得以有权进行大额头寸交易的。尤其在接近到期日时，两个指数期货的价格由于要趋向

标的指数，其价格应该非常接近，如果此时价差较大，进行套利风险较低。

为了说明跨市场套利的方法，先假设在同一个市场存在着两种标的指数为同一指数的指数期货合约，每种合约的指数点价格不同，分别为 S_A 和 S_B。两种指数期货在 t 时的点数分别是 X_A^t 和 X_B^t，由于标的是同一指数，所以在合约到期日 T，两种指数期货的点数都是 X^T。现在假设一个投资者买入 n_A 手股指期货 A，卖出 n_B 手股指期货 B，则到期时，该投资者的收益为：

$$(X_B^t - X^T)S_B n_B + (X^T - X_A^t)S_A n_A$$

整理可得：

$$X_B^t S_B n_B - X_A^t S_A n_A + (S_A n_A - S_B n_B)X^T$$

由上式可以看出，前两项在 t 时刻已经确定，投资者只要选择 n_A 和 n_B，使得其满足 $S_A n_A = S_B n_B$，则投资者 T 时刻的收益将完全由 t 时刻确定，不存在风险，即存在套利机会。只要 $X_B^t S_B n_B - X_A^t S_A n_A$ 大于交易成本，就可以进行套利。

现在来看跨市场的情况，基本假设与上面相同，只是现在两种指数期货位于不同的市场，因而采用不同的货币，设一单位 A 国货币可兑换 π_{AB} 单位 B 国货币，则投资者到期的收益为：

$$(X_B^t - X^T)S_B n_B + (X^T - X_A^t)S_A n_A \pi_{AB}^T$$

整理可得：

$$X_B^t S_B n_B - X_A^t S_A n_A \pi_{AB}^T + (S_A n_A \pi_{AB}^T - S_B n_B)X^T$$

从上式可以看出，不可能通过选择 n_A 和 n_B 来消除 T 时刻的影响，因为 T 时刻的汇率无法预知。这种情况下，不存在同一市场下的套利机会。

由上面的分析可以看出，汇率风险的存在，使得同一市场情况下的套利机会消失了。如果可以确定汇率，则可能重新获得套利机会。因此可以考虑使用汇率远期来规避汇率风险。在 t 时投资者签订远期汇率协议，在 T 时，采用 F 的汇率进行交易。F 服从远期定价公式，在 t 时刻是确定的。这样投资者到期收益为：

$$(X_B^t - X^T)S_B n_B + (X^T - X_A^t)S_A n_A F$$

整理可得：

$$X_B^t S_B n_B - X_A^t S_A n_A F + (S_A n_A F - S_B n_B)X^T$$

从上式可以看出只要选择 n_A 和 n_B，使得 $S_A n_A F = S_B n_B$，就可以消去 X^T 项，使得收益在 t 时就可确定，收益为 $(X_B^t - X_A^t) S_A n_A F$，只要该收益大于交易成本，就可进行套利。

现在考虑交易成本以推算无套利区间。设 A 国的交易成本为 C_A，B 国的交易成本为 C_B，均为交易量的某个比例，则总交易成本为：

$$X_A^t S_A n_A F C_A + X_B^t S_B n_B C_B$$

根据 $S_A n_A F = S_B n_B$ 则上式可化为：

$$(X_A^t C_A + X_B^t C_B) S_A n_A F$$

收益大于成本要求：

$$(X_B^t - X_A^t) S_A n_A F > (X_A^t C_A + X_B^t C_B) S_A n_A F$$

整理得：

$$X_B^t - X_A^t > X_A^t C_A + X_B^t C_B$$

由于交易是对称的，所以也存在买入 B，卖出 A 的情况。所以套利存在条件为：

$$|X_B^t - X_A^t| > X_A^t C_A + X_B^t C_B$$

14.2.4 跨品种套利

跨品种套利与跨期套利的基本思想是类似的，都是遵从配对交易的方式，即两种存在某种相关性的资产之间的价差稳定在某个值附近，如果偏离该值过远，就建立套利头寸以在价差缩小的时候获利。跨品种套利中，有相关性的资产是相同到期日的不同指数期货。如果两个不同的股指期货合约具有相互替代性和一定的关联度，或者它们会受到同一市场因素的影响，那么它们之间的价差就可能在一个相对稳定的值附近波动。如果价差偏离该值过远，就存在套利机会。

例如标准普尔 500 指数期货和标准普尔 100 指数期货。由于同是代表美国股市的指数，彼此之间存在一定的相关性。当两种指数的价格差值偏离正常值较远时，就可以建立价差头寸进行套利。由于其基本思想与跨期套利一致，所以相关的套利方法是类似的，这里就不赘述了。并且，前面介绍的配对交易的协整方法，也可以用在这里。

第十五章

套利失败的教训—长期资本管理公司的教训

前几章列举了一些国外常用的套利方法。实际上一些对冲基金一直在采用这些方法。这些方法多数都有辉煌的历史战绩,但同时也有着惨痛的历史教训。所以,本书所谈的套利不是学术上所定义的套利,因为学术上的套利是没有损失风险的。并且,由于一般认为套利是低风险的交易方法,对其进行的风险监控往往较为宽松,且套利方法的模型依赖性非常强,这样就使得一旦出现模型未分析到的情况,套利交易的损失甚至会超过一般交易的损失。历史上最著名的套利基金恐怕要数长期资本管理公司了(LTCM)。两位诺奖得主,所罗门的明星套利团队,再加上前政府官员组成的黄金组合,使得长期资本管理公司一度成为华尔街耀眼的明星,创造了惊人的收益,控制了惊人的资本,但当其陷入危机时,同样也为投资人制造了惊人的损失。长期资本管理公司采用了各种套利技术,先进的模型,复杂的数学工具。在当时的华尔街,该公司在陷入危机前一直披着神秘的面纱,一群金融巫师挥动着魔法棒,不停地发现金融市场的无效点,就像从空气中挤出钱来一样。无数机构想方设法地去获得长期资本管理公司的交易模型,或不惜一切代价成为该公司的服务商。但最终,长期资本管理公司却差一点将整个华尔街拖入深渊。本章将简单回顾长期资本管理公司的发展和衰落史,简单分析他们采用的交易方法,借以对套利交易做一个总结。

长期资本管理公司的多数核心交易人员来自所罗门公司的套利部,是约翰·麦利维瑟一手创立的。首先需要介绍一下约翰·麦利维瑟,因为他是整个长期资本管理公司的核心。约翰·麦利维瑟生于1947年,毕业于西北大学,后获得芝加哥大学MBA,与后来的高盛总裁乔·柯兹纳是同班同学。27岁时加盟所罗门兄弟公司,凭借其在债券交易上的天赋,成为所罗门兄弟公司

的明星员工。1977年，他负责创立套利部。这个部门可以说是后来长期资本管理公司的原型。套利部的一项创举就是利用所罗门公司的自有资金，进行风险很大的债券交易，而不是仅替客户买卖债券。这在当时是全新的业务领域，几乎没有竞争对手，所以该部门的利润是非常丰厚的。麦利维瑟的另一项创举是将学术界的研究人员引入交易市场。此前华尔街也招募这些学术派的人员，但仅限于让他们做研究工作，从不会让他们接触交易。而麦利维瑟认为，只有那些将市场视为一门严谨科学的人，才适合做交易。所以，他开始大量招募学术人员，包括：哈佛商学院助理教授艾里克·罗森菲尔德，伦敦经济学院金融硕士维克多·哈格哈尼，MIT金融博士格里高利·霍金斯，MIT博士威廉·卡拉斯科，MIT双博士劳伦斯·希利布兰德。

但约翰·麦利维瑟却阴沟翻船，正当套利部开疆扩土之时，一件丑闻迫使他离开了所罗门兄弟公司。事情起因于一名叫作保罗·莫泽尔交易员。1991年，保罗·莫泽尔向约翰·麦利维瑟承认，他向美国财政部提交了一份假标书，并在政府债券拍卖时多买进了一部分未得到批准的政府债券。得知此事，约翰·麦利维瑟立刻向董事会上报此事。董事会一致认为事态严重，但并没有采取任何补救行动。1991年8月份，公司进一步发现，保罗·莫泽尔还有其他许多犯罪事实没有交代。无奈之下，公司立即将情况向美国财政部和联邦储备委员会做了通报，两大机构极其愤怒。这一丑闻引起了华尔街的一场震动，媒体大肆批判。在此压力之下，管理层不得不辞掉了麦利维瑟。当然，约翰·麦利维瑟并没有就此停止。他开始筹划建立一个崭新的套利基金，并开始拉拢在所罗门的旧部下。

1993年，约翰·麦利维瑟开始筹建他的基金——长期资本管理公司，目标规模为25亿美元。这在当时是相当大的一个数字，多数基金是从2 500万美元开始的。而且麦利维瑟的基金的收费标准也与众不同。首先是要收取每年2%的资产管理费，同时，麦利维瑟及其合伙人还要向其投资者收取全部利润的25%作为分成。而一般的对冲基金只收取1%的资产管理费和10%的利润分成。并且，长期资本管理公司还要求所有的投资者承诺三年之内不得撤资。

1993年初麦利维瑟的旧部开始离开所罗门公司。艾里克·罗森菲尔德、维克多·哈格哈尼、格里高利·霍金斯、劳伦斯·希利布兰德纷纷投效长期资本管理公司。这使得长期资本管理公司拥有原来所罗门兄弟公司套利部的实力。但麦利维瑟并没有止步于此，要想成立一只有全球影响力的基金，只有原来的团队是不够的。麦利维瑟想塑造更良好的品牌形象，使得整个团队显得不仅是一群债券交易员，而且是一群经验丰富的金融高手。所以，麦利维瑟这

次请到了诺贝尔奖得主，期权定价公式的发现者罗伯特·莫顿和马尔隆·斯科尔斯。尽管后来这两位诺奖得主并没有参与长期资本管理公司的交易模型设计，但两位泰斗级的人物的加盟大大增加了公司的声望。

开始时，基金的募集并不顺利，主要的原因是：(1)麦利维瑟拒绝向投资者透露他的投资战略，这使得投资人根本搞不清楚自己的钱将被怎样使用。(2)很多投资者对长期资本管理公司如此高的财务杠杆表示十分震惊(关于杠杆，后文将会介绍)。(3)大的机构投资者无法忍受长期资本管理公司收取的高额管理费。加之莫顿教授和几位合伙人的宣传能力较差，导致投资者在路演时昏昏欲睡。麦利维瑟还亲往奥马哈，想拉巴菲特投资长期资本管理公司，但巴菲特只是对其进行一番鼓励，没投一分钱。好在美林证券的创新设计有了效果，它设计了一系列针对不同税负和法律地位的投资者的子基金。例如针对普通美国投资者的基金，专门为免税养老基金设计的基金，为日本投资者设计的将利润与日元进行对冲的基金，甚至为欧洲客户设计了在交易所上市的基金。但所有这些基金最后都将汇集到长期资本管理公司。同时，所罗门公司公布了以前5年套利部的利润情况，这使得麦利维瑟可以用其以往的辉煌业绩作为宣传的筹码。

1994年初，另一位重量级人物——戴维·马林斯，美国联邦储备委员会副主席——加盟长期资本管理公司。这使得长期资本管理公司获得了全世界许多国家和地区的官方的资金，马林斯加盟后不久，长期资本管理公司就获得了新加坡政府投资公司、台湾银行、科威特政府养老基金和中国香港土地发展署的资金。随之而来的是更多的信任和更多的资金。意大利中央银行外汇管理局投资1亿美元，日本住友银行投资1亿美元等。同时，还有一大批名人投资该公司，如好莱坞著名经纪人迈克尔·欧维兹、耐克CEO菲尔·奈特、纽约石油巨子罗伯特·贝尔夫等。最终，1994年2月底，长期资本管理公司正式开业，开业时资金总量为12.5亿美元，虽然只有目标的一半，但已是历史纪录了。

在长期资本管理公司开始经营时，格林斯潘送给了该公司一份大礼。1994年2月，格林斯潘宣布提高银行短期贷款利率。这大出市场意料。债券市场暴跌，投资者纷纷撤离。由于部分对冲基金采用高杠杆交易，利率的提高，流动性的缺失，使得这些对冲基金蒙受了巨大损失。斯坦因哈特4天损失8亿美元，索罗斯2天损失6.5亿美元。但幸运的是，长期资本管理公司刚刚募集完资金，债券市场的暴跌不仅没有为该公司造成损失，反而为该公司创造了投资机会。按照麦利维瑟的观点，市场疯狂的抛售只会使利差放大到不合理的程度，而这正为长期资本管理公司创造了赌利差会回归正常状态的机会。

刚刚募集的巨额资金和良好的市场机会，使得长期资本管理公司从开始经营时就赚钱了，而且是大笔的钱。

其实长期资本管理公司所用的方法很类似于前面几章所讲的套利方法，只是将其运用于债券和衍生品市场。由于后面详细分析长期资本管理公司的交易方式（现在已不是什么秘密了），这里仅举一个例子。这个例子就是30年期美国政府债券套利。30年期美国政府债券因流动性存在着一个价差。具体说来，每次30年期政府发行6个月或更长时间后，其流动性都会大大下降而成为非当期(off the run)的债券。这主要是因为债券信用等级较高，投资者买入之后大多长期持有，很少出手，所以交易次数少。由于缺少流动性，通常该种债券相比于发行时间少于半年的同样的30年期政府债券会有折扣，也就是收益率之间会有一个利差。1994年，长期资本管理公司发现，刚发行的30年期政府债券的收益率是7.24%，而发行时间超过半年的同类型债券的收益率7.36%。也就是说由于流动性较差，非当期债券要多支付12个基点的利息。而长期资本管理公司的合伙人认为这是不正常的，毕竟29.5年以后到期和30年以后到期没有太大差别，不应有如此大的利差。所以他们买入非当期的债券，同时卖空发行时间少于半年的债券，希望利差会大大缩小。但是，12个基点的利差很小，即使利差最后减小到2个基点，该交易的收益也只有1%，而且还要等上几个月的时间。这样收益率并不高。但如果使用杠杆，情况就不同了。采用10倍杠杆就可将1%的收益率变为10%，如果采用20、30倍的杠杆率，收益就会被放得更大。当然，杠杆的影响是双向的，也就是说，在放大收益的同时，也会放大亏损。最终，长期资本管理公司买进了10亿美元的非当期债券，卖空了10亿美元的当期债券。这里有一点值得注意，前面介绍过，长期资本管理公司控制的资金总共有12.5亿美元，而他们在这笔交易的头寸却是20亿美元，他们如何做到的呢？实际上，长期资本管理公司几乎没有使用自己的钱就完成了交易。因为虽然买入债券要花费10亿美元，但卖空债券可以获得10亿美元。当然，这里忽略了保证金和利息的问题。但实际上，由于长期资本管理公司的名气和资本量，加之麦利维瑟的讨价还价的能力，长期资本管理公司所做的所有交易都不交保证金，这几乎给了该公司创造无限杠杆的能力。

交易最终被证明是成功的，利差向预料的那样缩小了。该笔交易使长期资本管理公司获利1 500万美元，而所动用的自有资金几乎是0。第一年长期资本管理公司的收益为28%，扣除管理费和提成后，投资者的收益为20%。第二年的收益为59%，扣除管理费和提成后的投资者收益为43%。随后长期

资本管理公司又成功地募集了10亿美元的资金，1995年采用的杠杆为28倍，如果没有杠杆，1995年的收益率将仅有2.45%。至1996年春天，该公司被30倍杠杆放大之后的资本总量为1 400亿美元，已成为全美最大的对冲基金，比第二位的对冲基金大了4倍。而到了1997年，长期资本管理公司未经杠杆放大的资产已达到50亿美元，一切看起来都非常完美。

但随着资金量的扩大，长期资本管理公司面临任何一个资金扩大的基金都会面对的问题——如何为日益增加的资金找寻投资方向。当你只是一家规模较小的基金时，你可以努力去赚市场的钱。但当你达到成为市场的一个非常大的组成部分时，就是另一回事了，利润就会大大减少，因为你能赚谁的钱呢？自己的么？长期资本管理公司，面对的正是这一情况。因为采用的是相对价值交易，按公司合伙人的理解，就是赚确定的小钱，并通过杠杆放大，所以能够交易的机会并不算多，单笔交易能容纳的最大资金量有限。同时，随着公司名气的扩大，华尔街各大机构纷纷成立套利部门，采用非常类似的方法交易。竞争的加剧使得交易机会更少了。因此，要想维持前几年的收益率，长期资本管理公司必须寻找新的投资方向。

长期资本管理公司的合伙人将目光投向了股票市场，开始采用配对交易和风险套利的方法(这在前面都介绍过)。但他们对这一市场并没有研究，只是相信自己在任何一个领域都能做得比别人好。不仅如此，长期资本管理公司的合伙人，尤其是哈格哈尼和希利布兰德，还在自己的未知领域出手大方，完全不顾风险。例如，哈格哈尼在英国壳牌和荷兰皇家石油的配对交易中头寸竟达23亿美元。而希利布兰德则热衷风险套利交易，在西屋电器公司并购哥伦比亚广播公司的交易中，他不断买进哥伦比亚公司的股票，将该公司股票价格推高至只比西屋电器公司的收购价低62美分，并且投资杠杆比率达20倍。

同时，长期资本管理公司的多数合伙人还将自己的多数资产投入自己的公司，甚至个人向银行贷款再投入基金。希利布兰德向法国里昂信贷银行借了2 400多万美元投入了长期资本管理公司，汉斯·赫舒米德向该行借了1 500万美元，其他合伙人也或多或少借了一些。就这样，长期资本管理公司不仅作为企业采用极大的杠杆，它的合伙人同样采用了杠杆，杠杆的大小达到惊人的地步。

但利润终究还是下降了，1997年上半年，长期资本管理公司的收益率只有13%，表明投资的机会减少了。所以，当长期资本管理公司的合伙人结束了与美林高层的高尔夫聚会后，公司就宣布了减资决定。公司决定，公司所有的外部投资者必须收回其一半的资金。到1997年底，公司将1994年开业之

前投入公司的原始投资在该年的全部利润，以及以后投入的全部资金及应分配利润，一次性退还给投资者。返还对象不包括长期资本管理公司的合伙人和员工，同时也不包括比较大的战略投资者。

虽然向投资者退回了资金，但长期资本管理公司的杠杆后资本规模却并没有减少，也就是说，他们使用了更大的杠杆。同时，长期资本管理公司的合伙人在此利用自己的杠杆，增加了自己在公司基金中的份额，当然也欠下了惊人的债务。就这样，在投资机会越来越少的时候，长期资本管理公司的合伙人却通过杠杆，将自己推到了悬崖的边缘。

长期资本管理公司的失败主要可以归于两个原因。第一是对利差过于自信。正如本书前面几章所写，很多套利的方法，实际上是基于某种差值，即相对价值策略。套利者相信，当这一差值偏离正常值或合理值太远时，最终会回到正常值。当然，也有套利者会赌差值会扩大，但相对较少。长期资本管理公司正是这一差值回归思想的坚信者。差值越是扩大，他们就越是认为差值回归的机会增大，就越是会扩大头寸。但正像配对交易那章所写，差值不一定会回归合理值，因为合理值是根据历史数据算出的，但没人能够保证，历史的条件会一直保持下去。如果某些条件改变了，合理值就有可能发生改变，这样差值有可能永远无法回归以前的合理值。另外，如果考虑到资金的机会成本，特别是考虑使用杠杆时的借贷成本，交易的时间长度就非常关键。如果差值长时间没有回归合理值，则即使最终回归合理值，套利者所获得的收益可能也无法弥补资金成本。也就是说，套利者可能是对的，但却挣不到钱。当存在盯市时，问题就更加严重。因为如果在交易过程中，差值进一步扩大，则套利者会被追加保证金。如果套利者的杠杆很大，则可能因为资金不足，无法满足追加保证金的要求。一旦如此，清算商就会强行平仓，在价格对套利者不利时终止套利者的头寸，使得套利者蒙受巨大损失。如果再考虑到套利者是采用高杠杆的，那么这样的平仓甚至可能造成套利者的破产。而这正是长期资本管理公司发生的情况。

另一个失败的原因是投资组合理论的缺陷。投资组合理论的基本观点通俗说来，就是不要把鸡蛋放在一个篮子里。分散地持有，可以把风险有效地分散。因为资产之间不是完全相关的，一种资产的损失会被另一种资产的盈利弥补。但这一理论的关键点是资产之间的相关性不能很高，如果所有资产的价格都向一个方向运动，那么无论怎么分散，也无法减少风险。而长期资本管理公司正是认为自己将资金分散于很多交易之中，交易彼此之间相关性低，所以基金总体的风险很小。但是，东南亚经济危机以及随后的俄罗斯债务危机，

却使得全球股市同时大跳水，流动性差的债券价格也出现暴跌。换句话说，几乎所有长期资本管理公司认为会回归合理值的差值，都同时扩大了，分散的投资并没有减少风险。

正是在以上两个原因作用之下，长期资本管理公司的合伙人一步一步坠入了深渊。

1998年初，随着IMF救助了韩国，整个亚洲金融危机似乎过去了，股票市场一再创下新高，一切看起来都很好。同一时期，为了寻求新的投资机会，长期资本管理公司的合伙人正在离他们擅长的套利策略越来越远。他们开始进入外汇市场、公司股票期权市场等，而且头寸很大。希利布兰德甚至试图卖空巴菲特的伯克希尔·哈撒韦公司的股票。而且很多投资头寸并未像以前那样，利用相关资产的反向头寸进行对冲，而是纯粹的投机，逐渐背离其相对价值投资的理念。同时，公司对于风险的控制也流于形式，几乎没有人可以对希利布兰德和哈格哈尼进行约束。而他们两个人早已脱离了原来保守的交易策略，利用公司的资金为所欲为了。

5月，危机的征兆开始出现，利差开始向长期资本管理公司预计方向的相反方向扩大，水平不断提高。于是，不仅是长期资本管理公司，其他在债券市场进行套利的机构也都出现了亏损。而这又使得一些亏损的企业(如所罗门美邦)出售资产以弥补资金的减少。同时，抵押担保债券出现了大幅下挫。资金的需求使得各机构不得不进一步抛售资产。新兴市场债券成为抛售的对象之一。于是，东南亚刚刚被救助的各国(如新加坡、泰国、印尼等)再次面临大量资金外流，本币大幅贬值。随后，抛售的目标转向了俄罗斯。俄罗斯央行不得不将卢布利率提高三倍以阻止资金外逃。此时，投资者开始将资金转向相对安全的美国国债，使得国债收益率下降。同时，其他债券如公司债券由于被抛售，收益率开始上升。因此，这类债券与美国国债的利差进一步扩大，导致长期资本管理公司的损失继续增加。1998年5月，长期资本管理公司的资产损失了6.7%，但这还只是刚刚开始。

进入6月，几乎所有长期资本管理公司表现活跃的市场，利差都在扩大。至6月中旬，30年期美国国债收益率跌到了5.58%，这是首次发行30年期国债以来的最低点。表明投资者避险情绪高涨，纷纷买入美国国债，抬高了国债的价格，从而减少了收益率。但同一时期，长期资本管理公司却在大量卖空美国国债，以对冲他们买入的其他债券品种，依然认为利差会回归合理值。但实际上，不动产抵押债券与国债利差，由96个基点上升至113个基点；垃圾债券与国债利差，由224个基点上升至266个基点；公司债券与国债利差，由99个

基点上升至105个基点；非当期国债与当期国债的利差，也由6个基点上升至8个基点。总之，长期资本管理公司的每个市场的利差都扩大了，出现了巨额亏损。整个1998年上半年，长期资本管理公司的资本亏损14%。虽然为了资金的需要，公司卖出了一些资产，但卖出的是与价格直接涨跌相关的交易品种，但坚持保留与利差缩放有关的品种，甚至对这些品种增加头寸，财务杠杆竟达31倍。

7月，俄罗斯市场动荡不安，短期债券收益率上升到120%！整个市场都在关心俄罗斯政府是否会违约。部分投资者已开始抛售俄罗斯的国债。但哈格哈尼接受了经济学家大力宣传的说法，即俄罗斯作为核大国，是不会违约的，未对其俄罗斯的国债进行减持。8月初，俄罗斯债券价值大跌，各大机构都在大量抛售俄罗斯债券。同时，全球市场都陷入了卖压之中。8月第二周，俄罗斯央行终于无法抵挡市场对卢布的抛压，外汇储备枯竭，于13日宣布对卢布进行管制。同时，整个银行系统瘫痪，股市被迫休市，短期利率升至200%。两个月的时间，长期资本管理公司的俄罗斯债券价值跌去50%。但合伙人依旧认为，利差已经如此之大，所以不可能进一步扩大，所以很快就会缩小。因此，他们不仅没有减少俄罗斯国债的头寸，反而还增加了持仓量，仿佛将公司的未来都赌在了俄罗斯。

8月17日，俄罗斯政府宣布将延期偿还所有债务，放任卢布贬值。这无异于是一枚重磅炸弹。俄罗斯的违约动摇了投资者对于新兴市场的信心，资金纷纷从新兴市场大举出逃。几乎所有稍有投资风险的资产都遭到了抛售，投资者只想要安全。于是，大量资金涌入美国30年国债。这直接导致利差越来越大。长期资本管理公司的合伙人的乐观，并没有使他们摆脱困境，亏损越来越大。每时每刻都在亏损，每分钟以百万美元计。一天之内，基金亏损5.53亿美元，占资金总额的15%。

亏损还在逐日的扩大。长期资本管理公司要想继续维持它的头寸，就需要资金的注入，否则就可能被强行平仓。合伙人到处奔走，希望能获得宝贵的资金，但效果并不好。资金的注入速度根本赶不上资产的损失速度。到8月底，银行由于担心对冲基金的风险，纷纷抽紧银根，迫使对冲基金普遍斩仓。抛售热潮的加剧使利差继续扩大，而收紧的银根使得长期资本管理公司几乎筹措不到资金。双面打击之下，长期资本管理公司的合伙人近乎绝望了。麦利维瑟甚至把一处地产归到了自己妻子名下。并且，由于几乎所有人都在抛售，所以市场几乎没有买家。就算长期资本管理公司想要卖出资产套现以弥补损失，实际上也根本找不到买家。整个8月，长期资本管理公司亏损19亿

美元,资金余额只剩22.8亿美元。但公司的风险并没有减少,杠杆竟然上升到55倍。

9月初,麦利维瑟不得不向投资者公布了长期资本管理公司的损失。市场哗然。随之而来的是更加惨痛的事实,华尔街在落井下石。长期资本管理公司所持有的资产几乎都比市场的一般水平跌得要惨,表明其余机构都在想方设法逃离长期资本管理公司持有的资产,以免长期资本管理公司被迫减仓时的抛压加剧自己的损失。但这种情况使得长期资本管理公司更加惨痛,而且所持有头寸过大,根本无法抛售,因为一旦抛售就会引起市场的剧烈震荡,加剧抛压,损失可能更大。

此时,长期资本管理公司的终结者出现了。负责公司清算业务的贝尔斯登,要求长期资本管理公司在该公司的账户资金不得低于5亿美元,否则将停止清算服务。随后,各大机构也都开始了瓜分长期资本管理公司的行动,不仅抛售,甚至卖空长期资本管理公司所持有的投资项目。因为几乎所有人都预料,长期资本管理公司迟早会因为资金需求抛售资产,从而使得价格下跌。9月中旬,长期资本管理公司的资本仅剩15亿美元了,仅仅几天后,资产净值更是只剩5.55亿美元。

最终,9月底,为了防止长期资本管理公司破产导致的金融市场动荡,美联储出面调和,美国15家金融机构注资37.25亿美元购买长期资本管理公司90%的股权,从而接管了长期资本管理公司。长期资本管理公司的大部分合伙人都损失了90%的个人财富。由于本书以套利为主,对于救助的细节,这里就不介绍了,有兴趣的读者可以看《营救华尔街》一书。值得一提的是,麦利维瑟并未就此放弃。15个月后,麦利维瑟和他的部分合伙人成功筹集了1.5亿美元,重新开始了他们的投资之旅。

最后,我们分析一下长期资本管理公司的投资方法。作为历史上最著名的套利基金,长期资本管理公司使用了很多种套利方法,包括:

1. 当期债券与非当期债券套利,这一方法已在前面讨论过,这里不再赘述。

2. 房屋抵押证券交易

首先介绍一下IO和PO。IO和PO是房屋抵押证券的衍生产品。所谓房地产抵押证券,简单说来就是将购房者的抵押贷款放入一个资产池。然后把这一资产池的现金流均等分割为数份债券,出售给投资者。而IO和PO是将这一资产池中的对于利息的还款支付给IO的购买者,将本金的还款支付给PO的购买者。

这样，如果贷款利率下跌，则购房者倾向于转贷款，即使用新的贷款来偿还原来的贷款，以减少利息支付。这样，由于原来的贷款被提前偿还，基于原来贷款的资产池的可得利息就大大减少了，所以IO的价格就会大幅下降。而由于本金被提前支付，PO的价格会大幅上涨。反之，如果贷款利率上升，买房者就不倾向于提前还款，而且本金的偿还比例还会因为利率的提高而下降。所以，此时IO的价格就会上涨，而PO的价格就会下跌。

在1993年，美国房屋抵押贷款利率在越战后第一次降至7%以下。利率的下降使得贷款购房者纷纷开始转贷款，那一年，美国大约有40%的人进行了转贷款，IO价格大幅下降。其实IO价格下降的幅度远远超过其应该下降的位置。长期资本管理公司的模型显示，这一价格的下降，只有认为全体美国人都进行了转贷款才能解释，而这显然是不现实的。因此，市场的恐慌使得IO的价格被低估了。所以，长期资本管理公司借助杠杆，大约买入了20亿美元的IO。为了对冲风险，他们同时买入了美国国债，因为国债的价格变动与IO相反。到1994年春天，利率上涨，IO价格猛升。所以，长期资本管理公司虽然在国债上有所损失，却在IO上获利颇多。

3.欧洲国债

长期资本管理公司成立之初，欧洲正在开始货币一体化的进程，但最终是否会采用统一的货币并没有一致的看法。所以欧洲各国的外汇和债券市场非常热闹。长期资本管理公司注意到，欧洲各国的利差较大，如果最终各国采用同样的货币，则利差不会这么大，应该趋向一致。所以，长期资本管理公司买入当时利率比较高的西班牙和意大利的国债，同时卖空利率比较低的德国和法国的国债，希望国家间的利差会缩小。

4.股票配对交易。该方法前面已经介绍过，不再赘述。

5.并购风险套利。该方法前面已经介绍过，不再赘述。

以上只是长期资本管理公司的几种主要的套利方法，并没有涵盖其所有的套利方法。从以上几种方法可以看出，长期资本管理公司都是采用相对价值的投资方法。即买入一项资产的同时卖空一项与其相关的资产，这样实际上并不需要对某一种资产的价格变动进行预测，只需要分析两种相关资产的相对价格变动。表面上看来，这样的投资风险小于对单一资产的投资。但从长期资本管理公司的最终失败可以看出，相对价值投资的风险并不一定就小于绝对价值投资，特别是当价差较小而采用大的杠杆的时候。所以，不应对相对价值套利策略掉以轻心。总之，本书所介绍的套利技术，不可能保证投资者一定获益而没有风险。

参考文献

1. R. Almgren, N. Chriss, "Value under Liquidation", *Journal of Risk*, 1999

2. R. Almgren, N. Chriss, "Optimal execution of portfolio transactions", *Journal of Risk*, 2000

3. R. Almgren, C. Thum, E. Hauptmann, and H. Li, "Equity Market Impact", *Journal of Risk*, 2005

4. D,Bertsimas, AW. Lo, "Optimal control of execution costs", Journal of Financial Markets, 1998

5. Michael A. Berry, Edwin Burmeister, and Majorie B. McElroy, "Sorting Out Risks Using Known APT Factor", *The Journal of Finance*, April 1988

6. Fischer Black,"Toward a Fully Automated Stock Exchange", *Financial Analysts Journal*, 1971

7. Gary Brinson, L. Randolph Hood, and Gil Beebower, "Determinants of Portfolio Performance", *Financial Analyst Journal*, 1986

8. G. Brinson, B. Singer, and G. Beebower, "Determinants of Portfolio Performance II:An Update", *Financial Analysts Journal*, 1991

9. Louis Chan, Josef Lakonishok, "The Behavior of Stock Proices around Institutional Trades", *Journal of Finance*, 1995

10. Clarke, R. H. de Silva, S. Thorley, "Portfolio Constraints and the Fundamental Law of Active Management." *Financial Analysts Journal*, 2002

11. Clarke, R. H. de Silva, S. Thorley, "The Fundamental Law of Active PortfolioManagement", *Journal of Investment Management*, 2006

12. Ronald H. Coase, "The Nature of the Firm", *Economica*, 1937

13. K. Cohen, J. Pogue, "An Empirical Evaluation of Alternative Portfolio Selection Models", *Journal of Business*, April 1967

14. J. B. Copas, "Regression Prediction and Shrinkage", *Journal of the Royal Statistical Society*, Vol. 45, No. 3, 1983

15. D. K. Dey, C. Srinivasan, "Estimation of a Covariance Matrix Under Stein's Loss", *Annals of Statistics*, 1985

16. Ian Domowitz, "A Taxonomy of Automated Trade Execution Systems", IMF Working Papers, 1992

17. Ian Domowitz, Henry Yegerman, "The Cost of Algorithmic Trading: A First Look at Comparative Performance", 2005

18. B. Efron, "Bayesians, Frequentists, and Scientists", *Journal of the American Statistical Association*, 2005

19. E. Elton, M. Gruber, "Estimating the Dependence Structure of Share Prices: Implications for Portfolio Selection", *Journal of Finance*, December 1973

20. R. Engle and R. Ferstenberg, "Execution risk", *Journal of Portfolio Management*, 2007

21. Paul Erlich, "Cash Optimization", Acadian Research, Boston, 1997

22. John Evans, Stephen Archer, "Diversification and the Reduction of Dispersion: An Empirical Analysis", *Journal of Finance*, Dec. 1968

23. F M. rank, P. Wolfe, "An Algorithm for Quadratic Programming", *Naval Research Logistics Quarterly*, 1956

24. *Financial Times*, "Trading with the help of 'guerrillas' and 'snipers'", March 19, 2007

25. G. Frankfurter, H. Phillips, J. Seagle, "Performance of the Sharpe Portfolio Selection Model: A Comparison", *Journal of Finance and Quantitative Analysis*, June 1976

26. P. Frost, J. Savarino, "An Empirical Bayes Approach to Efficient Portfolio Selection", Journal of Financial and Quantitative Analysis, 1986. 9

27. R. Grinold, R. Kahn, "*Active Portfolio Management*", McGraw-

Hill, 1999

28. Gur Huberman, Werner Stanzl, "Price Manipulation and Quasi-Arbitrage", *Econometrica*, 2004

29. John Hull, "*Options, Futures and Other Derivatives*" (6th edition), Prentice Hall, 2005

30. Britten Jones, "Sampling Error in Mean-Variance Efficient Portfolio Weights", *Journal of Finance*, Dec 2002

31. Bruce I. Jacobs, Kenneth N. Levy, "*Market Neutral Strategies*", John Wiley & Sons, 2005

32. Kim Kendall, "*Electronic and Algorithmic Trading Technology*", Elsevier, 2007

33. Amir Khandni, Andrew Lo, "What Happened to the Quants in August 2007?", web. mit. edu/alo/www/papers/august07. pdf

34. Robert Kissell, Merton Glantz, "*Optimal Trading Strategies: Quantitative Approaches for Managing Market Impact and Trading Risk*", AMACOM, 2003

35. Robert Kissell, Roberto Malamut, "Algorithmic Decision Making Framework", *Journal of Trading*, Winter 2006

36. Robert Kissell, Roberto Malamut, "Understanding the Profits and Loss Distribution of Trading Algorithms", *Journal of Trading*, Spring 2005

37. F. Lillo, J. D. Farmer, and R. N. Mantegna, "Master curve for price-impact function", *Nature*, 2003

38. C. Morris, "Stein's Paradox in Statistics", *Scientific American*, May 1977

39. Anna Obizhaeva, Wang Jiang, "Optimal trading strategy and supply/demand dynamics", MIT working paper, 2005

40. Andre Perold, "The Implementation Shortfall: Paper Versus Reality", *Journal of Portfolio Management*, Spring 1988

41. Richard Roll, "A Critique of the Asset Pricing Theory's Tests Part: On Past and Potential Testability of the Theory", *The Journal of Finance*, March 1997

42. Richard Roll, Stephen Ross, "An Empirical Investigation of the Arbitrage Pricing Theory", *The Journal of Finance*, December 1980

43. William Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", *The Journal of Finance*, 1964. 9

44. William Sharpe, "The Arithmetic of Active Management", *Financial Analysts Journal*, Vol. 47, 1991

45. Filippo Stefanini, "*Investment Strategies of Hedge Funds*", John Wiley & Sons, 2006

46. Nicholas Stein, "Small World, After All", *Fortune*, July 22 2002

47. Hans R. Stoll, "Electronic Trading in Stock Markets", *Journal of Economic Perspectives*, 2006 Volume 20

48. Ed Thorp, "A Perspective on Quantitative Finance: Models for Beating the Market", *The Best of Wilmott*, 2003 vol. 1

49. Robert Tibshirani, "*An Introduction to the Bootstrap*", New York: Chapman and Hall, 1993

50. The Trade, "*Algorithmic Trading, a buy-side handbook*", 2005

51. Ganaphathy Vidyamurthy, "*Pairs Trading: Quantitative Methods and Analysis*", John Wiley & Sons, 2004

52. W. Wagner, M. Edwards, "Best execution", *Financial Analyst Journal*, Vol. 49, 1993

53. Mark Whistler, "*Trading Pairs: Capturing Profits and Hedging Risk with Statistical Arbitrage Strategies*", Wiley, 2004

54. OE Williamson, "*Markets and Hierarchies: Analysis and Antitrust Implications*", New York: Free Press, 1975

55. 查尔斯 P. 琼斯(Charles P. Jones)著，李月平、陈宏伟译，《投资学分析与管理(第十版)》，机械工业出版社，2008

56. EdwinJ. Elton, Martin J. Gruber, Stephen J. Brown, William N. Goetzmann 著，余维彬(译)，《现代投资组合理论与投资分析》，机械工业出版社，2008

57. Chi-fu Huang, Robert H. Litzenberger 著，宋逢明译，《金融经济学基础》，清华大学出版社，2006

58. Roger Lowenstein 著，孟立慧译，《营救华尔街》，上海远东出版社，2000

59. 国信证券，数量化投资技术综述，2008 年 12 月

60. 何杰，"交易系统自动化程度的国际比较"，证券市场导报，2001 年 2 月

61. 上海证券报,"我国证券交易技术规范发展的里程碑——《证券交易数据交换协议》介绍", 2005.6

62. 张昕,《基于复杂时间处理的金融软件系统实现和改进》,浙江大学硕士论文, 2007

63. 中国证券监督管理委员会,《中国证券期货统计年鉴(2007)》

64. FIX 协议组织 http://www.fixprotocol.org/

65. 维基百科 http://en.wikipedia.org/