

## § 12 金融数据中的非线性

在本书中，我们讨论的经济方法几乎都是设计查找金融数据中的线性结构。例如在第 2 章中，我们发展时间系列理论来预测资产收益，使用有利的资产自相关联合---线性预测理论是主要的观点。第 4 章的事件理论和第 5 章的 CAPM 和 APT 都是建立在期望收益的线性模型上。甚至在以后的章节中包括其他的经济变量如消费，红利和利率，这些模型依然是线性的。将重点放在线性上并不奇怪，因为许多从金融数据中得到的经济模型就是线性的。

然而，许多经济行为并不一定是线性的。实验证据和偶然的自省发现，被调查者对风险和期望收益的态度是非线性的。许多金融合同中的术语如期权和其他衍生证券是非线性的。市场参与者的战略相互影响，信息融入证券价格的过程，经济波动理论等都是自然的非线性，因此，非线性结构模型自然成为金融经济的前沿。

这是一个相当大的挑战，因为非线性模型的集合远大于线性模型的集合。毕竟任何不是线性的就是非线性。而且，非线性模型一般比线性的更难分析，很少能得到一个闭式(close form)表达式。能被轻松地操纵和以经验为主地执行，在一些例子中，唯一的分析模型是计算型的。这对习惯于解析、直观和线性思维的我们而言是相当不习惯的。但新一代的经济学家创造了新的模型和工具，来描绘经济现象中的非线性。其中一些是本章的重点。在动力学系统理论，非线性时间系列分析，随机波动模型，非参数统计和人工神经网络等方面的令人鼓舞的进步激起了人们最近对金融数据中的非线性的兴趣，我们在下面的章节中将逐个讨论。

§ 12.1 中，我们将重温 § 2 中提及的主题。但是从线性到非线性，我们介绍一种模型的分类法，区别线性(偏离鞅假设)和非线性违反独立原则而非鞅假设。

§ 12.2 探讨了大量非线性模型，包括单变量和多变量的自回归有条件 *GARCH* 模型和随机波动率模型。

在 § 12.3 和 12.4 我们跳出参数的时间系列模型，通过固定变量间的非线性关系来探讨非参数理论。包括光滑技术和人工神经网络。尽管这些技术可以揭示非线性的变化但他们严重地依赖数据和计算。为说明技术的功能。我们介绍一种衍生证券的定价和套期保值以及估计状态价格密度。

我们在 § 12.5 讨论了这些技术的局限性。最重要的局限是一对孪生的问题，过拟合和数据窥察，这个问题虽然没同一程度，但也对线性模型产生麻

烦。很遗憾，我们对如何处理这些问题，除了一些特定的例子，几乎一无所知。因此，这是一个具有许多尚待解决问题的领域。

### § 12.1 单变量的时间系列的非线性结构

一个典型的时间系列模型使得一个观测的时间系列  $X_t$  和一系列的扰动  $e_t$  相关。在线性时间系列的分析中，扰动被假设为不相关的但不必要设定为 IID。世界代表理论认为任何时间系列能写成一个这些扰动的无限阶的线性移动平均。并且这个线性移动平均陈述总结了系列的无条件方差和自协方差。

在非线性的时间系列解析中，这些基础的扰动项典型地设为 IID，但我们找到一个可能的非线性函数，使系列  $X$  和历史的扰动项相关，基本表示如下：

$$X_t = f(e_t, e_{t-1}, e_{t-2} \dots) \quad (12.1.1)$$

其中，扰动假设为均值为 0 或单位方差。 $f(\cdot)$  一个未知函数。这种表示的一般性使得它很难处理。在实际应用中，大部分模型被更严格地限制。

$$X_t = g(e_{t-1}, e_{t-2} \dots) + e_t h(e_{t-1}, e_{t-2} \dots) \quad (12.1.2)$$

$E_{t-1}(X_t) = g(e_{t-1}, e_{t-2} \dots)$ ，函数  $g(\cdot)$  代表以过去信息为条件的  $X_t$  的均值，在  $X_t$  中的创新与  $e_t$  是按比例的，比例系数为  $h(\cdot)$ 。由于  $E_{t-1}(X_t - E_{t-1}[X_t])^2 = h(e_{t-1}, e_{t-2} \dots)^2$ ，因此，这个函数的平方代表以过去信息为条件的  $X_t$  的方差。非线性  $g(\cdot)$  模型被称为均值非线性，而非线性  $h(\cdot)^2$  模型被称为方差非线性。

为理解式(12.1.2)对(12.1.1)的约束，我们将(12.1.1)对给定的  $e_{t-1}, e_{t-2} \dots$  围绕  $e_t=0$  泰勒展开，得

$$X_t = f(0, e_{t-1}, e_{t-2} \dots) + e_t f_1(0, e_{t-1}, e_{t-2} \dots) + \frac{1}{2} e_t^2 f_{11}(0, e_{t-1}, e_{t-2} \dots) + \dots \quad (12.1.3)$$

其中  $f_1$  是  $f$  对  $e_t$  的一阶导数， $f_{11}$  是二阶导数，以此类推。为得到(12.1.2)，我们将泰勒展开式中的高阶去掉，并令

$g(e_{t-1}, e_{t-2} \dots) = f(0, e_{t-1}, e_{t-2} \dots)$ ， $h(e_{t-1}, e_{t-2} \dots) = f_1(0, e_{t-1}, e_{t-2} \dots)$ 。通过去掉高阶项，我们将  $X_t$  的条件高阶矩的时间方差固定地与  $X_t$  的条件二阶矩的时间方差联系起来，由于对所有指数  $P \geq 2$ ， $E_{t-1}[(X_t - E_{t-1}[X_t])^P] = h(\cdot)^P E[e_t^P]$ 。对那些主要关心  $X_t$  前两项条件矩的人认为这些约束作为(12.1.2)的约束更为温顺。

等式(12.1.2)成了非线性时间系列里条件均值  $g(\cdot)$  模型与条件方差  $h(\cdot)^2$  模型的自然分割。大部分的时间系列模型都只关注非线性的一种或另一种。比如，一种简单的非线性移动平均值模型表示如下

$$X_t = e_t + a e_{t-1}^2 \quad (12.1.4)$$

这里  $g(\cdot) = a e_{t-1}^2$ ， $h(\cdot) = 1$ 。这个模型是均值非线性并非方差非线性。另一方面，Engle(1982)的一阶(ARCH)模型表示如下：

$$X_t = e_t \sqrt{a e_{t-1}^2} \quad (12.1.5)$$

这里  $g(\cdot)=0$  ,  $h(\cdot)=\sqrt{ae_{t-1}^2}$  。这个模型是方差非线性而不是均值非线性。

均值非线性与方差非线性的一个区别是  $X_t$  的阶矩。我们已经强调过非线性能够被推定,其二阶(自协方差) $E(x_t x_{t-i})$ 对所有  $i>0$  都等于 0。在上面所举的两个例子中,我们很容易知道  $e_t$  是对称分布的,也就是说它的三阶为 0。例如,对于非线性移动平均值(12.1.4),当  $E[e_{t-1}^3]=0$ ,则有  $E[X_t X_{t-1}] = E[(e_t + ae_{t-1}^2)(e_{t-1} + ae_{t-2}^2)] = aE[e_{t-1}^3] = 0$ 。

现在讨论  $E[x_t x_{t-i} x_{t-j} x_{t-k}]$  的高阶矩。当  $i, j, k \dots > 0$  时,均值非线性模型允许高阶非 0。而方差非线性模型满足鞅过程有  $E[x_t / x_{t-1}, \dots] = 0$ , 所以当  $i, j, k \dots > 0$  时,其高阶为 0。如果至少一个滞后项指数  $i, j, k \dots$  为 0 时,这些模型只能有非 0 高阶。在非线性移动平均值的例子(12.1.4)中,三阶并且  $i=j=1$ ,

$$\begin{aligned} E[X_t X_{t-1}^2] &= E[(e_t + ae_{t-1}^2)(e_{t-1} + ae_{t-2}^2)^2] \\ &= aE[e_{t-1}^4] + 2a^2 E[e_{t-2}^2] E[e_{t-1}^3] \neq 0 \end{aligned}$$

在一阶 ARCH 例子, (12.1.5), 同样第三阶矩  $E[x_t x_{t-1}] = E[(e_t \sqrt{ae_{t-1}^2}) e_{t-1}^2 ae_{t-2}^2] = 0$ 。但当  $i=0, j=k=1$  时,这个模型的第四阶矩,  $E[x_t^2 x_{t-1}^2] = E[e_t^2 a^2 e_{t-1}^4 e_{t-2}^2] \neq 0$ 。

我们将第 12.2 节中讨论 ARCH 和其它变化方差的模型。在本节的剩余部分,我们关注条件均值的非线性模型。在 12.1.1, 我们探讨几种将非线性模型参数化的方法,在 12.1.2 中我们用参数模型来促进和解释一些在检验单变量时间系列中的非线性的方法,包括 Brock, Dechert, 和 Scheinkman(1987)。

### 12.1.1 一些参数模型

即使在参数模型的子集中,我们也不可能将所有非线性规范都列举出来, Priestley(1988), T?ersvirta, Tj?stheim 和 Granger(1994), 和 Tong(1990) 提出了许多涵盖最流行的非线性时间系列模型,包括一些具有迷人名字的专门模型,如, self-exciting threshold autoregression(SETAR), amplitude-dependent exponential autoregression(EXPAR) 和 state-dependent models(SDM)。为规定此领域的范围,我们在本节将讨论四个例子。多项式模型(Polynomial models),分段线性模型(折线法 piecewise-linear models),马尔可夫开关模型(Markov-switching models)和确定性的混沌模型(deterministic chaotic models)。

#### 多项式模型

函数  $g(\cdot)$  的一种表示法是将其关于  $e_{t-1} = e_{t-2} = \dots = 0$  泰勒展开，产生了离散时间 Volterra 系列。(见 Volterra[1959])：

$$g(e_{t-1}, e_{t-2}, \dots) = \sum_{i=1}^{\infty} a_i e_{t-i} + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} b_{ij} e_{t-i} e_{t-j} + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \sum_{k=j}^{\infty} c_{ijk} e_{t-i} e_{t-j} e_{t-k} + \dots \quad (12.1.6)$$

式(12.1.6)中单个求和项是标准的线性移动平均值，连续两次求和表示两个创新的滞后项的交叉乘积的影响，连续三次求和表示三个创新的滞后项的交叉乘积的影响，其他以此类推。为避免重复计算给定的创新的交叉乘积，求和指数  $j$  从  $i$  开始， $k$  从  $j$  开始等。这个想法以过去创新的真实非线性函数描绘成创新的多项式函数的加权和。等式(12.1.4)是这种形式的一个简单例子。Robinson(1979)和 Priestley(1988)对此作了详尽的阐述。

多项式模型也可以写成自回归形式。函数  $g(e_{t-1}, e_{t-2}, \dots)$  将条件均值与过去扰动相关，也可表示为另外一种函数  $g^*(x_{t-1}, x_{t-2}, \dots)$ ，其将条件均值与  $x_t$  的滞后项相关联。则(12.1.6)的自回归形式表示为：

$$g^*(x_{t-1}, x_{t-2}, \dots) = \sum_{i=1}^{\infty} a_i^* x_{t-i} + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} b_{ij}^* x_{t-i} x_{t-j} + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \sum_{k=j}^{\infty} c_{ijk}^* x_{t-i} x_{t-j} x_{t-k} + \dots \quad (12.1.7)$$

它也可能得到混合的自回归/移动平均值的形式，ARMA 模型的非线性等价物。比如，在双线性模型中，使用  $x_t$  的滞后值， $e_t$  的滞后值以及它们的交叉乘积：

$$g^{**}(e_{t-1}, x_{t-1}, \dots) = \sum_{i=1}^{\infty} a_i e_{t-i} + \sum_{i=1}^{\infty} b_i x_{t-i} + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} g_{ij} x_{t-i} e_{t-j} \quad (12.1.8)$$

在纯非线性移动均值模型或非线性自回归模型不能描述非线性时，这个模型可以极度节省地表示(通过有限的短的时滞长度)。Granger 和 Andersen(1978)与 Subba Rao 和 Gabr(1984)详细地研究了双线性模型。

### 分段线性模型

另一种分析非线性结构的常用方法是使用分段线性函数，比如在一阶的阈自回归(threshold autoregression)(TAR)：

$$x_t = \begin{cases} a_1 + b_1 x_{t-1} + e_t \\ a_2 + b_2 x_{t-1} + e_t \end{cases} \quad (12.1.9)$$

这里  $x_t$  对它的滞后项  $x_{t-1}$  的回归的截距和斜率系数取决于  $x_{t-1}$  与阈  $k$  的大小。这个模型可以推广到高阶和多个阈，在 Tong(1983, 1990)中有详细解释。

分段线性模型也包括变化点模型和，经常在经济学文章中出现的具有结构断点的模型。在这些模型中，参数都假设在固定的取样期间里移动---曾经很典型地，目标是估计这两套参数和变化点或结构断点。Perron(1989)将此方法应用到宏观经济时间系列，Brodsky(1993)和 Carlstein, Muller 和

Siegmund(1994)使用了更现代化的方法来处理变化点问题，包括非参数估计和贝叶斯推理(Bayesian inference)。

变化点理论在统计与运算的研究文章中得到了很好的建立，但在经济学模型中的应用并不是没有争论。众所周知，在典型的工程学中，结构断点存在于给定的数据表中，但我们无法确切地说断点存在于经济时间系列中。如果由于一些重大的经济事件如股市崩溃，我们认为结构断点存在，这种数据驱动的设置搜索将误导我们戏剧性地推论发现事实上并不存在的断点(见 Leamer[1978]与 Lo 和 MacKinlay[1990])。

### 马尔可夫转换模型(Markov-Switching Models)

Hamilton(1989,1990,1993)和 Sclove(1983a,1983b)关于马尔可夫转换模型紧密地与 TAR 相关。他们的主要区别是变迁的变化不是由过程的水平决定，而是由没有观测到的状态变量，其变量典型地由马尔可夫链构成。比如：

$$x_t = \begin{cases} \mathbf{a}_1 + \mathbf{b}_1 x_{t-1} + \mathbf{e}_{1t} & s_t = 1 \\ \mathbf{a}_2 + \mathbf{b}_2 x_{t-1} + \mathbf{e}_{2t} & s_t = 0 \end{cases} \quad (12.1.10)$$

$s_t$  是一个没有观测到的两态马尔可夫链，由某个概率矩阵  $P$  转换而来。注意到 (12.1.10) 中不同的时间惯例： $s_t$  决定了在  $t$  时刻的变迁而不是  $s_{t-1}$ 。在两个变迁项中， $s_t$  是一个  $AR(1)$  (autoregressive 一阶自回归)，但参数(包括误差项的方差)在整个过程中都不一样。它们的变化是随机的，与概率系列相关的。

这些模型明显地有对经济前景的要求。制度变迁是由因素引起的，而不是由我们构造的系列引起的( $s_t$  决定结构，而不是  $x_t$ )。的确我们不知我们处于什么制度中，但经过这件事后，我们在有一定的把握知道我们处于哪种制度中。( $s_t$  通过 Hamilton[1989] 过滤程序可以估计)。并且，马尔可夫转换模型(Markov-Switching Models)不受一些统计偏差的影响，而结构断点模型则会。通过数据与马尔可夫链之间的相互作用，结构的移动能够被“识别”，而不是通过数据推出结果的。Hamilton[1989]在经济周期上的应用就是这项技术的影响和范围的极好例子。

### 确定性的非线性动力学体系

在构造确定性的非线性动力学体系模型中取得了许多令人鼓舞的进步，这也激发了许多技术应用到非线性的关系的估计中。相对简单的普通的微分和差分方程已经展示了相当复杂的动力学。对它们的复杂性常用的术语是“蝴蝶作用”。它的定义是“巴西蝴蝶的拍打引起得克萨斯的龙卷风”。这种说法，仅仅是半开玩笑，由 Lorenz(1963)为构造天气规律的模型而提出的一般确定性差分方程的简单系统：

$$\dot{x} = 10(y - x) \quad (12.1.11)$$

$$\dot{y} = xz + 28x - y \quad (12.1.12)$$

$$\dot{z} = xy + \frac{8}{3}z \quad (12.1.13)$$

Lorenz(1963)观测到，对这个系统的起始值作很小的变动---比如在小数点后四位---甚至仅仅进行很短的时间，也会戏剧性地产生不同的试样路径。这种对开始条件的灵敏性是混沌理论的一个特点。

一个简单的平滑的混沌系统的例子是出名的帐篷图形：

$$x_t = \begin{cases} 2x_{t-1} & x_{t-1} < \frac{1}{2} \\ 2(1-x_{t-1}) & x_{t-1} \geq \frac{1}{2} \end{cases} \quad x_0 \in (0, 1) \quad (12.1.14)$$

这个帐篷图可以看作是一个一阶的没有振动  $e_t$  并且  $a_1 = 0, b_1 = 2, a_2 = 2, b_2 = -2$  的阈自回归。如果  $x_{t-1}$  在 0 与 1 之间，那么  $x_t$  也在这个区间；这样，就如图 12.1 描绘的一样，这个帐篷图描绘了单位间隔回归自己的过程。由(12.1.14)产生的数据显示它们随机的，它们均匀地分布在单位区间，连续的不相关的。而且，数据也显示对起始条件的依赖敏感度，这将在问题 12.1 中检验。Hsieh(1991)举了一些其他的例子，而 Brock(1986)，Holden(1986)，Thompson 和 Stewart(1986)则提供了更多的关于混沌系统的正式的数学讨论。

尽管在非线性的动力学体系中的重大的突破都有对物理、生物和其他的“硬”科学的直接的含意，但对经济和金融的影响就没有那么生动了。许多经济学上的应用都已经被接受了，但没有一个特别有吸引力的，特别是从实验式得到的观点。

在将经济现象构造成确定性的非线性动力学系统模型中存在两个严重的问题。首先，不象许多自然科学的理论，经济学理论并不特别在意函数形式。因此，经济学家很少有理论的原因在期望发现非线性的某种形式。第二，经济学家很少有能引导受约束的实验，这样几乎不可能推导出支配经济现象的确定性动力学系统的参数，甚至于这个系统是存在的和低维的。当受约束的实验是切实可行的，例如，在粒子物理学中，相当精确地恢复动力学是可能的，通过在紧密时间间隔上采集许多系统的“快照”。这项技术，赋予了甚至最抽象的非线性动力学的定义以实验的内容，但很遗憾，它不能应用到非实验的数据。

通过一系列简单的非线性确定性等式产生我们在金融市场中所看见的复杂性的可能性是可望而不可及的，但如果我们不能以任何程度的精确恢复这些等式，这就没有什么意义。而且，对一个系统的起始条件的敏感程度的统计的抽样误差使得动力学系统更少具实践性。当然，随着在这个领域的快速发展，也许几年后这些疑惑就没什么了。

### 12.1.2 单变量的非线性结构检验

尽管先前章节的警告，混沌理论的数学原理激发了新的对独立的非线性结构的统计检验。现在我们来讨论这些检验。

### 对高阶的检验

我们先前对高阶的非线性模型的讨论，可以作为对统计学的非线性的检验的基础。例如，Hsich(1989)定义了一个三阶：

$$j(i, j) \equiv \frac{E[x_t x_{t-i} x_{t-j}]}{E[x_t^2]^{3/2}} \quad (12.1.15)$$

并发现在 IID 下或对鞅模型仅在方差非线性，对所有  $i, j > 0$  有  $j(i, j) = 0$ 。他建议通过以下方式估计  $j(i, j)$  的值：

$$\hat{j}(i, j) \equiv \frac{\frac{1}{T} \sum_t x_t x_{t-i} x_{t-j}}{[\frac{1}{T} \sum_t x_t^2]^3} \quad (12.1.16)$$

在这种零假设即  $j(i, j) = 0$  下以及对  $x_t$  加以足够的常规条件使得高阶存在。

$\sqrt{T}\hat{j}(i, j)$  是常规的渐近的，它的方差可通过以下式子测算：

$$\hat{v} \equiv \frac{\frac{1}{T} \sum_t x_t^2 x_{t-i}^2 x_{t-j}^2}{[\frac{1}{T} \sum_t x_t^2]^3} \quad (12.1.17)$$

Hsich 检验使用了一种特殊的数据三阶，但它同时也可能几阶。如，自回归多项式模型(12.1.7)提出了均值的非线性的一个简单的检验，将  $x_t$  对自己的滞后项回归和与滞后项的叉积，来检验非线性中的连点的重要性。Tsay(1986)提出一种使用二阶的检验方法和  $M$  个滞后项代替  $M(M+1)/2$  个非线性的回归量。我们可以计算出异方差性相容的标准方差，因此，对方差非线性的检验就很牢固了。

### 积分相关和维数相关

为将确定性的混沌过程从真实的随机过程中区别开来，有必要将数据放在足够的高阶形式中检验。比如，在帐篷图中，由于  $x_t$  均匀分布如果  $x_t$  在单位区间中则数据呈随机分布。但如果  $x_t$  和  $x_{t-1}$  在单位平方，则如图 12.1 所示，数据都会落在帐篷线上。这种直接的渐进会产生奇怪的现象，就象在第 3 章中所分析股票价格的不连续性一样。但如果要计算高阶或更复杂的非线性就会很困难。Grassberger and Procaccia(1983)已经提出了一种常规的渐进方法来实现这个基础的想法。通过组织数据(prefiltered, if desired, to remove linear structure) $n$  次  $x_t^n$ ，定义

$$x_t^n = \{x_{t-n+1}, \dots, x_t\} \quad (12.1.18)$$

参数  $n$  是嵌入的维数。

下一步接着计算部分 (of  $n$ -histories that are “close” to one another)。为测量紧密度，我们选择一个数  $k$ ，并取一对相近的  $n$  次方  $x_s^n$  和  $x_t^n$  其相应的差的最大的绝对值小于  $k$ ，即  $\max_{i=0,\dots,n-1} |x_{s-i} - x_{t-i}| < k$ 。我们定义紧密指标系数  $K_{st}$ ，当两个  $n$  次数相近时为 1 否则为 0：

$$K_{st} = \begin{cases} 1 & \max_{i=0,\dots,n-1} |x_{s-i} - x_{t-i}| < k \\ 0 & \text{其他} \end{cases} \quad (12.1.19)$$

取  $C_{n,T}(k)$  为在  $T$

$$C_{n,T}(k) \equiv \frac{\sum_{s=1}^T \sum_{t=s}^T K_{st}}{T(T-1)/2} \quad (12.1.20)$$

相关积分  $C_n(k)$  等于取样长度提高时分数的极限：

$$C_n(k) \equiv \lim_{T \rightarrow \infty} C_{n,T}(k) \quad (12.1.21)$$

同样地，随便地选择一对  $n$  次数是紧密的具有概率性。

很明显，相关积分取决于嵌入维(embedding dimension)  $n$  和参数  $k$ 。为研究  $k$  如何影响，我们取嵌入维  $n=1$ 。N-histories 由单一的数据连结构成，我们假设这种情况，数据满足 IID 并均一地分布在单位区间  $(0,1)$  间。在这种条件下，数据点的小数部分在经过基准数据点的距离  $k$  为  $2k$ ，当基准数据点处于单位区间的中间时(在  $k$  与  $1-k$  之间)，但是，如果基准数据点处于单位区间的边缘，它会更小。在极端的情况下，当基准数据点为 0 或 1 时，only a fraction  $k$  of the other data points are within  $k$  of the benchmark。对紧邻基准点  $b$  的数据点的小数公式一般为  $\min(k+b, 2k, k+1-b)$ 。然而，随着  $k$  的收缩，由“边缘问题”引起的复杂化变得可忽略的，相关积分接近  $2k$ 。

Grassberger and Procaccia(1983)研究了当距离测度  $k$  收缩时相关积分的变化过程。他们计算  $\log C_n(k)$  与  $\log k$  对  $k$  的比值：

$$v_n = \lim_{k \rightarrow 0} \frac{\log C_n k}{\log k} \quad (12.1.22)$$

当我们减少表示接近度的参数时，用以测量相近基准点的分数部分的成比例减少程度。在满足 IID 条件下并且  $n=1$ ，当  $k$  收缩时， $\log C_n(k) / \log k$  比值接近  $\log 2k / \log k = (\log 2 + \log k) / \log k = 1$ 。这样，对  $v_1 = 1$  并且满足 IID 条件的数据，对足够小  $k$ ，相近的数据点的分数以  $k$  收缩一样的速度收缩。

现在讨论更高维数  $n$  情况下相关积分的变化过程。当  $n=2$  时，我们在二维图表上作数据的 2-histories，就如图 12.1 and asking what fraction of the 2-histories lie within a square whose center is a benchmark 2-history and whose sides are of length  $2k$ 。在均匀分布的 IID 数据下，数据点的 a fraction  $4k^2$  在一正方形内，当基准的 2-history 足够远离单位平方

的边缘。同样，我们通过让  $k$  收缩来处理边际问题，我们发现当  $k$  收缩时， $\log C_2(k) / \log k$  接近  $\log 4k^2 / \log k = (\log 4 + 2\log k) / \log k = 2$ 。所以，当  $v_2 = 2$ ，对满足 IID 的数据，对足够小的  $k$ ，相邻数据点的分数(fraction)收缩的速度是  $k$  收缩的两倍。同样可证明，当  $v_n = n$ ，对满足 IID 条件的数据，对足够小  $k$ ，相互紧靠的数据点的  $n$  维的分数( the fraction of pairs  $n$ -histories that are close to one another)收缩的速度是  $k$  收缩的  $n$  倍。

当数据是由非线性确定性过程产生时，相关积分的变化十分不同。为说明这个问题，假设数据由帐篷图产生，在一维中，所有数据均匀地分布在区间的直线上。我们得到  $v_1 = 1$ ；但在 2 维中，所有数据均匀地分布在帐篷线上，如图 12.4 所示。对足够小  $k$ ，一对对相互紧密的点的分数(the fraction of pairs of points that are close to one another)收缩的速度同  $k$  收缩的一样。所以， $v_2 = 1$ 。在高维中，以此类推，当数据由帐篷图产生时，对所有的  $n$ ，得  $v_n = 1$ 。

当  $v_n$  的极限存在时，我们定义相关维数为  $v_n$  的极限当  $n$  趋向无穷大时。

$$V = \lim_{n \rightarrow \infty} v_n \quad (12.1.23)$$

非线性确定性的过程以有限的  $v$  为特征。

非线性确定性数据与 IID 均衡数据的对比归纳为具有其他分布的 IID 数据，由于不管如何分布，对于 IID 数据  $v_n = n$ 。由于我们在计算相关积分时，逐个取  $n$ -history 为基准  $n$ -history，因此分布的效果最终达到平衡。由此，grassberger and procaccia(1983)提出，通过取不同的  $n$ ，来计算  $v_n$ ，看看其随着  $n$  增长还是保持在一个固定的值。我们可以将 IID 随机数据与非线性确定性数据区分开来。这种要求大量的数据，因为必须取非常小的  $k$  来计算  $v_n$ ，没有什么分布理论适合  $v_n$ 。

The brock-dechert-scheinkman test

Brock, Dechert, Scheinkman(1987)发展了交替逼近法，这种方法更适合在经济和金融中典型出现的有限的数据。他们发现，即使  $k$  为有限，只要数据满足 IID，则对任何  $n$  有：

$$C_n(k) = C_1(k)^n \quad (12.1.24)$$

为理解这个结果，可以将  $C_{n+1}(k) / C_n(k)$  的比值理解为有条件的概率：

$$\begin{aligned} \frac{C_{n+1}(k)}{C_n(k)} &= \Pr\left(\max_{i=0, \dots, n} |x_{s-i} - x_{t-j}| < k \mid \max_{i=1, \dots, n} |x_{s-i} - x_{t-i}| < k\right) \\ &= \Pr\left(|x_s - x_t| < k \mid \max_{i=1, \dots, n} |x_{s-i} - x_{t-i}| < k\right) \end{aligned} \quad (12.1.25)$$

也就是说， $C_{n+1}(k)/C_n(k)$  表示在前  $n$  个数据点是接近的情况下， $n+1$  点与  $n$  点接近的概率。如果数据满足 IID，这等同于两个数据接近的无条件概率  $C_1(k)$ 。对所有正数  $n$ ，令  $C_{n+1}(k)/C_n(k) = C_1(k)$ ，得到 (12.1.24)。

Brock, Dechert, Scheinkman(1987) 建议 BDS 检验统计量，

$$J_{n,T}(k) = \sqrt{T} \frac{C_{n,T}(k) - C_{1,T}(k)^n}{\hat{s}_{n,T}(k)} \quad (12.1.26)$$

其中， $C_{n,T}(k)$  和  $C_{1,T}(k)$  为式 (12.1.20) 所定义的相关积分的取样， $\hat{s}_{n,T}(k)$  为  $C_{n,T}(k) - C_{1,T}(k)^n$  的标准偏差的渐进值。BDS 统计量在 IID 零假设下，是渐进的标准常规，它在 Hsieh(1989) 和 Scheinkman 和 LeBaron(1989) 中得到应用和解释，他们给  $\hat{s}_{n,T}(k)$  直接的表达式。Hsieh(1989) 和 Hsieh(1991) 报告了在有限的例子中对 BDS 统计量的 size and power 的 Monte Carlo 结果。

虽然，对 IID 数据，满足  $C_n(k) = C_1(k)^n$ ，存在一些不规则的非线性模型，但 BDS 统计量相对其他大部分非线性模型方法而言还是具有优越性的。当然，还要注意的，是相对方差的非线性模型而非平均值的非线性模型。这样，BDS 拒绝并不一定要求时间系列有随时间变化条件的平均值，仅仅需要具有明显的随时间变化的条件方差。例如，Hsieh(1991) 就强烈拒绝对一般股票收益进行 BDS 检验要求满足 IID 条件。后来他构造了收益的随时间变化的条件方差的模型，更难拒绝从这类模型中得到的残差服从 IID 的零假设。

## 12.2 变化波动率的模型

在本节中，我们将讨论其他的方法来构造时间系列  $h_{t+1}$  的变化波动率的模型。§ 12.2.1 介绍了单变量的 ARCH 模型和随机波动模型；§ 12.2.2 证明如何将它们推广到多变量的情况；§ 12.2.3 延伸到条件平均值的时间变差与条件方差的时间变差相联系的模型，这些模型的平均值和方差都是非线性。

为集中在波动率上，我们假设  $h_{t+1}$  是一项创新，也就是说，以  $t$  时刻的信息作预测，它的条件均值为 0。在金融应用中， $h_{t+1}$  可能表示在资产收益中的创新。我们定义  $s_t^2$  为  $t$  时刻  $h_{t+1}$  的条件方差或等价地，为  $h_{t+1}^2$  的条件期望值。我们假设以  $t$  时刻的信息为条件，创新服从正态分布：

$$h_{t+1} \sim N(0, s_t^2) \quad (12.2.1)$$

创新的无条件方差  $s^2$  就是  $s_t^2$  的绝对期望值：

$$s^2 \equiv E[h_{t+1}^2] = E[E_t[h_{t+1}^2]] = E[s_t^2]$$

因此， $s_t^2$  围绕均值的变化并不影响无条件方差  $s^2$ 。

然而， $s_t^2$  的变动却影响  $h_{t+1}$  的绝对分布的高阶。特别，伴随时间变差  $s_t^2$ ， $h_{t+1}$  的绝对分布较正态分布产生更严重的肥尾现象。为说明问题，我们首先令：

$$h_{t+1} = s_t e_{t+1} \quad (12.2.2)$$

$e_{t+1}$  是满足平均值为 0 和单位方差(如同以前章节介绍的) IID 随机变量, 满足正态分布(在以前章节中我们没有假定)。

如同我们在第 1 章讨论那样, 测量一个随机变量  $y$  的分布的尾部浓度的有效的方法是标准的 4 阶或峭度  $K(y) \equiv E[y^4] / E[y^2]^2$ 。众所周知, 正态随机变量的峭度是 3; 因此,  $K(e_{t+1})=3$ 。但对创新  $h_{t+1}$ , 我们有:

$$\begin{aligned} K(h_{t+1}) &= \frac{E[s_t^4]E[e_{t+1}^4]}{(E[s_t^2])^2} \\ &= \frac{3E[s_t^4]}{(E[s_t^2])^2} \\ &\geq \frac{3(E[s_t^2])^2}{(E[s_t^2])^2} = 3 \end{aligned} \quad (12.2.3)$$

第一个等式从  $s_t$  和  $e_{t+1}$  的独立性而来, 不等式包含在 Jensen 不等式。直觉地, 绝对分布是正态分布的混合, 方差小的说明大量的点分布在均值附近, 方差大的则大量的点分布在尾部。因此, 混合分布比正态分布有更肥的尾部。

我们现在考虑新的方法构造  $s_t^2$  的模型和估计它的值。有关这方面的文章数量极大, 因此, 我们的评论毫无疑问是有选择性的。Bollerslev, Chou, and Kroner (1992), Bollerslev, Engle, and Nelson (1994), Hamilton (1994) 提供了更全面的调查。

### 12.2.1 单变量模型

在介绍波动率的参数时间系列模型之前, 有关时间变差的波动率的研究将波动率估计从资产收益数据中得出。例如, Officer (1973) 使用了一种流动标准偏差----这种标准偏差通过随时间变化的二次抽样得出----来及时估计每点的波动率。其他研究人员应用给定日期的高低价之间的差额来估计那天的波动率 (Garman and Klass [1980], Parkinson [1980])。如果目标仅仅是测量一点的波动率, 这些方法还是相当精确的。就如 Merton (1980) 所说, 如果资产定价以一稳定的波动率传播, 如, 几何布朗运动, 如果测量价格足够频繁的话, 取任意短的区间则波动率估计相当精确。Nelson (1992) 已经证明甚至波动率随时间变化, 类似的结论也可以得出, 只要收益的条件分布不要太肥尾和波动率的变化是渐进的。

然而, 建立在结果随着时间运动上, 存在一段时间的波动率为常数的假设基础上的波动率测量导致了逻辑矛盾和统计上的低效率。为处理这些, 最近的研究首先确定波动率的参数模型, 然后应用这个模型从收益的数据中估计波动率。

### ARCH 模型

一项有关收益数据的研究表明，大收益(of either side)似乎紧跟着更大的收益(of either sign)。换句话说，资产收益的波动率显得系列相关。这从图 12.2 很容易看出，此图是从 1926 到 1994 年间相对 CRSP 加权股指的每月超额收益。个别月份的收益变化非常大，但他们都在一个随时间变化相当慢的范围运动。在 30 年代，收益的范围变化很广，而 50 年代、60 年代则窄得多。

理解这种现象的另一种方法是，计算超额收益的平方或绝对值的系列相关系数。分别为 0.23 和 0.21，这些系列的一阶系列相关系数是收益本身的一阶系列相关系数 0.11 的 2 倍，这具有很重要的统计学上的意义，因为在没有序列相关的零假设条件下的标准误差是  $1/\sqrt{T}=0.036$ 。在前 12 个自相关系数的平均数的差别更具戏剧性。0.2 为超额收益率的平方，0.21 为超额收益率的绝对值，而 0.02 为超额收益率本身。这反映了收益率的平方和绝对值的自相关消失得很慢。为描述波动率的系列相关，Engle(1982)提出了自回归有条件的异方差或 ARCH 的模型。这里将条件方差作为过去创新的平方的已分摊差滞：

$$s_t^2 = w + a(L)h_t^2 \quad (12.2.4)$$

其中  $a(L)$  是滞后算子的多项式。为保证条件方差为正， $w$  和  $a(L)$  中的系数必须为非负。

作为构造波动率持久运动的模型，不用估计高阶多项式  $a(L)$  中非常大量的系数。Bollerslev(1986)提出了 GARCH 模型：

$$s_t^2 = w + b(L)s_{t-1}^2 + a(L)h_t^2 \quad (12.2.5)$$

其中  $b(L)$  也是滞后算子的多项式。与 ARMA 模型类似，多项式  $b(L)$  的指数为  $p$ ， $a(L)$  的指数为  $q$  则模型称为 GARCH( $p, q$ ) 模型。最常用的 GARCH 模型是 GARCH(1,1)，表示如下：

$$\begin{aligned} s_t^2 &= w + bs_{t-1}^2 + ah_t^2 \\ &= w + (a+b)s_{t-1}^2 + a(h_t^2 - s_{t-1}^2) \\ &= w + (a+b)s_{t-1}^2 + as_{t-1}^2(e_t^2 - 1) \end{aligned} \quad (12.2.6)$$

在(12.2.6)的第二个等式中， $(h_t^2 - s_{t-1}^2)$  项以  $t-1$  时刻的信息为条件，均值为 0，可以认为是波动率的扰动。系数  $a$  表示现在波动率扰动连通下一期波动率的幅度，而  $(a+b)$  表示比率。(12.2.6)中第三个等式又将波动率扰动记为  $s_{t-1}^2(e_t^2 - 1)$ ，标准正态变量的平方减去它的均值----即，服从  $c^2(1)$  (自由度为 1 的卡方分布)的随机变量减去均值----乘以滞后一阶的波动率  $s_{t-1}^2$ 。

GARCH(1,1) 模型也可以表示为创新的平方  $h_{t+1}^2$ 。有：

$$h_{t+1}^2 = w + (a+b)h_t^2 + (h_{t+1}^2 - s_t^2) - b(h_t^2 - s_{t-1}^2) \quad (12.2.7)$$

这个表达式清楚地表明  $GARCH(1,1)$  模型是对创新平方的  $ARMA(1,1)$  模型。但一个标准的  $ARMA(1,1)$  模型有同方差扰动，而这里扰动  $(h_{t+1}^2 - s_t^2)$  则是异方差的。

### 持久和平稳

在  $GARCH(1,1)$  模型中很容易构造波动率的多期预测。当  $a + b < 1$  时， $h_{t+1}$  的无条件方差或等价的  $s_t^2$  的无条件期望值是  $w/(1-a-b)$ 。对 (12.2.6) 进行迭代，应用迭代期望定理，前期波动率  $j$  的条件期望值为：

$$E_t[s_{t+j}^2] = (a+b)^j \left( s_t^2 - \frac{w}{1-a-b} \right) + \frac{w}{1-a-b} \quad (12.2.8)$$

多期波动率预测以速度  $(a+b)$  回归无条件均值。这种在单期与多期预测之间的关系与具有自回归系数的线性  $ARMA(1,1)$  模型相同。对高阶  $GARCH$  模型，多期预测可以被构造成相同的形式。

当  $a + b = 1$ ，前期波动率  $j$  的条件期望为：

$$E_t[s_{t+j}^2] = s_t^2 + jw \quad (12.2.9)$$

$a + b = 1$  时的  $GARCH(1,1)$  模型有单位自回归根，所以现在的波动率影响不确定将来的波动率的预测。这就是完整的  $GARCH$  也叫  $IGARCH(1,1)$  模型。

对  $s_t^2$  的  $IGARCH(1,1)$  过程很象带漂移  $w$  的线性随机漫步。然而，Nelson (1990) 表示这种类推还有待商讨。线性随机漫步在两方面来说是不稳定的。第一，它没有稳定的分布，因此这个过程不是严格的稳定。第二，它没有无条件的一阶或二阶，因此，它不是协方差稳定。另一方面，在  $IGARCH(1,1)$  中， $s_t^2$  是严格稳定的，即使它的稳定分布缺少无条件阶。因此， $IGARCH(1,1)$  模型是严格稳定但不是一般协方差稳定。

很明显，当  $w = 0$  时， $IGARCH(1,1)$  有稳定分布。则 (12.2.9) 可简化为  $E_t[s_{t+j}^2] = s_t^2$ ，则波动率满足鞅条件。同时，波动率保持有界，因为它不可能为负数。但鞅一致理论要求有界的鞅必须收敛。在这种程度来说，它只能收敛于 0。则对  $s_t^2$  的稳定分布是退化的分布，大部分点集中在 0 附近。这也说明对  $h_{t+1}$  的稳定分布也收敛于 0。因此，对  $s_t^2$  和  $h_{t+1}$  的稳定分布也是有阶的，但是为 0。

当  $w > 0$  时，Nelson (1990) 表明存在非退化的  $s_t^2$  稳定分布。但这些分布没有有限的均值和高阶。创新  $h_{t+1}$  有稳定分布并且平均值为 0，但由于尾部太浓，不存在二阶或更高阶。

### 代替函数形式

在标准 GARCH 模型里，未来方差的预测是现在与过去方差的线性关系，收益的平方校正预测。名为绝对值 GARCH 模型的代替模型，使得未来标准偏差与现在和过去标准偏差为线性关系，并用收益的绝对值来校正预测值。例如，绝对值 GARCH(1,1) 模型可表示为：

$$s_t = w + bs_{t-1} + as_{t-1}|e_t| \quad (12.2.10)$$

Schwert(1989) and Taylor(1986) 估计绝对值 ARCH 模型，而 Nelson and Foster (1994) 讨论绝对值 GARCH(1,1) 模型。

至今为止，我们讨论的模型中负的和正的扰动  $e_{t+1}$  的作用是均衡的，对波动率的影响是一样的。然而，Black(1976) 和许多其他学者指出，在股票市场中存在着不对称：同样数量的负的创新对股票收益产生的波动比正的大。在第 12.2.3 节我们将对这种不对称作出可能的解释。为处理这些情况，我们推广绝对值 GARCH 模型：

$$s_t = w + bs_{t-1} + as_{t-1}f(e_t) \quad (12.2.11)$$

$$\text{其中 } f(e_t) = |e_t - b| - c(e_t - b) \quad (12.2.12)$$

这里，移动参数  $b$  和倾斜参数  $c$  测量两种不同类型的不对称。 $b$  是不受限制的但  $|c| \leq 1$  来保证  $f(e_t) \geq 0$ 。当  $c=0$  和  $b \neq 0$  时，扰动对波动率的影响依赖于  $b$ 。所以，没有扰动时波动率的增长比有扰动  $b$  时增长得更快。当  $b=0$  和  $c \neq 0$  时，零扰动对波动率的影响最小，但负扰动与正扰动是有区别的。对给定大小的扰动，对波动率的影响负的比正的大，反之亦然。根据 Hentschel(1995)，一种理解(12.2.12)的很好的方法是用  $f(e_t)$  代替  $e_t$ ，就象图 12.3 所示。图(a)表示绝对值函数( $b=0$  和  $c=0$ )，在其它图上用虚线表示。图(b)表示移动绝对值函数( $b=0.5$  和  $c=0$ )；图(c)倾斜绝对值函数( $b=0$  和  $c=0.25$ )，图(d)表示移动和倾斜绝对值函数( $b=0.5$  和  $c=-0.25$ )。

Hentschel(1995)推广(12.2.11)允许  $f(e_t)$  的指数而不是  $f(e_t)$  本身来影响波动率；允许  $s_t$  的指数而不是自己本身作为线性差分方程的变量。得：

$$\frac{s_t^I}{I} = w + b\left(\frac{s_{t-1}^I - 1}{I}\right) + as_{t-1}^I [f(e_t)]^v \quad (12.2.13)$$

等式(12.2.13)定义了包括了文章出现的大部分常用的 GARCH 模型。令  $I = v = 2$ ， $b=c=0$  得标准 GARCH 模型。Glosten, Jagannathan, and Runkle(1993) 将标准 GARCH 模型推广到非零  $c$ 。Engle and Ng(1993) 则允许非零  $b$ 。令  $I = v = 1$ ， $b, c$  取任意数，得绝对值 GARCH 模型。(12.2.12)家族的其他特殊的重要成员还有 Nelson(1990)的指数 GARCH 或 EGARCH 模型，令  $I = 0$ ， $v = 1$ ， $b=0$  得：

$$\log(s_t) = w + b \log(s_{t-1}) + a[|e_t| - ce_t] \quad (12.2.14)$$

这个模型吸引人之处在于它不对任何参数作要求来保证收益的条件方差是正的。同样，当  $a+b=1$  时，它也是不稳定和协方差不稳定，所以它就没有  $IGARCH(1,1)$  模型所具有的不一般的平稳性质。换句话说， $EGARCH$  模型的未来自方差的多期预测比较难计算。象 (12.2.8) 那样无紧密形式的表达式就可以利用了。

## 评价

我们介绍了令人困惑的波动率模型的种类。为研究这些模型的哪些性质在金融数据中的有重要作用，我们必须估计模型的参数。很高兴的是，对  $GARCH$  模型和其它由 (12.2.13) 所定义的类型来说，这是很直观的。以模型参数和起始方差估计为条件，正态分布的数据，我们可以构造相当准确的似然函数。记模型参数的向量为  $q$ ，定义  $s_t(q)$  为  $t$  时刻的条件标准偏差，包括参数和收益的历史，定义  $e_{t+1}(q) \equiv h_{t+1} / s_t(q)$ 。当包括了模型的真实的参数， $e_{t+1}(q)$  是 IID 且有满足正态分布的密度函数  $g(e_{t+1}(q))$ ：

$$g(e_{t+1}(q)) = \frac{1}{\sqrt{2p}} e^{-e_{t+1}(q)^2 / 2} \quad (12.2.15)$$

因此， $h_{t+1}$  的条件似然对数为：

$$\begin{aligned} \ell_t(h_{t+1}; q) &= \log(g(h_{t+1} / s_t(q))) - \log(s_t^2(q)) / 2 \\ &= -\log(\sqrt{2p}) - h_{t+1}^2 / 2s_t^2(q) - \log(s_t^2(q)) / 2 \end{aligned} \quad (12.2.16)$$

最后项是 Jacobian 项，我们观察是  $h_{t+1}$  而不是  $h_{t+1} / s_t(q)$ 。整个数据集  $h_1, \dots, h_T$  的似然对数为：

$$l(h_1, \dots, h_T) = \sum_{t=1}^T \ell_t(h_{t+1}; q) \quad (12.2.17)$$

极大似然估计就是选择参数  $q$  使得 (12.2.17) 最大。

尽管很容易证明极大似然估计是相容的，但要证明它是渐进正态的就很难了。困难在于它要求匀称条件，但在  $GARCH$  过程中很难验证。Lee and Hansen (1994) 给出了一些  $GARCH(1,1)$  的结果，但其他的结果很少。实证研究人员典型地忽视这个问题，假设常规的匀称条件存在。一些模拟的证据 (bollerslev and Wooldridge [1992] and lumsdaine [1995]) 支持这个观点。

Hentschel (1995) 运用从 1926 到 1990 年期间的每天和每个月的股票收益数据对 (12.2.13) 家族中的许多模型进行了极大似然估计。要得到参数  $1$ ， $v$ ，Hentschel 发现需要非常大量的日常数据的观察值。结果显示  $1$  接近 1 (就如在绝对值  $GARCH$  模型)，而  $v$  大于 1，事实上接近 1.5。在每天和每月的数据中，Hentschel 发现通过移动参数  $b$  构造模型比倾斜参数  $c$  的好。这样由于挠动的移动绝对值的提高升到三阶同分数，美国的股票收益可以通过  $GARCH$  模型很

好地表示条件标准差，。尽管在二次大战期间，持续的程度是敏感的，但在所有的估计模型中，波动率过程是稳固的。

### 附加解释变量

直到这点，我们只用了收益的过去历史来构造波动率的模型。增加其他的解释变量是很直接的：例如，我们可以将扩张的  $GARCH(1,1)$  模型表示如下：

$$s_t^2 = w + gX_t + bs_{t-1}^2 + ah_t^2 \quad (12.2.18)$$

其中， $X_t$  是  $t$  时刻的任一变量。只要  $X_t \geq 0, g \geq 0$ ，则这个模型依然保证波动率为正。此外，我们可以将  $EGARCH$  模型增加解释变量而不用任何限制。Glosten, Jagannathan, and Runkle(1993)在不同的  $GARCH$  模型增加了短期名义利率并发现这对股市波动率有显著的正的效果。

### 条件非正态

我们讨论的  $GARCH$  模型隐含了以过去的收益历史为条件的收益分布是正态分布的。相同地，这些模型的标准残数， $e_{t+1}(q) = h_{t+1} / s_t(q)$  是正态分布的。很遗憾的是，在实践中  $GARCH$  模型的标准残数存在余数峭度，尽管比自然收益小。(见，例如，Bollerslev[1987]和 Nelson[1991])。

解决这个问题的一种方法就是继续运用由(12.2.16)和(12.2.17)所定义的条件正态似然函数，但是被认为是准极大似然估计(White[1982])。如在 Bollerslev 和 Wooldridge(1992)所讨论的那样，通过应用一加强协方差矩阵估计就可以计算参数估计的标准误差。

同样，我们可以直接地构造遵循  $GARCH$  过程的挠动的肥尾分布的模型。例如，Bollerslev(1987)，在具有自由度  $k$  的  $t$  分布：

$$g(e_{t+1}(q)) = \Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{k}{2}\right)^{-1}(k-2)^{-1/2}\left(1 + \frac{e_{t+1}^2(q)}{k-2}\right)^{-(k+1)/2} \quad (12.2.19)$$

其中， $\Gamma(\cdot)$  是伽玛函数。当  $k$  增加时， $t$  分布收敛于正态分布，但有过分峭度。事实上，当  $k \leq 4$  时，它的 4 阶就无穷大了。类似地，Nelson(1991)应用了一般误差分布，而 Engle 和 Gonzalez-Rivera(1991)非参数化地估计了误差密度。

$GARCH$  模型也可以由 GMM(广义矩 Generalized Method of moments)估计。当条件波动率  $s_t^2$  能够写成观测到的过去变量(过去收益的平方和附加变量如利率)的简单函数形式时，是相当有吸引人的。模型还暗示，收益的平方减去观测变量的适当函数，与观测变量是正交的。GMM 估计所具有的一般优势是它不需要对收益的挠动密度作详细说明。

### 随机波动率模型

对以过去的收益为条件的收益的非正态的反应是，假设存在以正态分布的收益为条件的随机变量，但这个变量----我们称为随机波动率----并不能直接观测到。在连续时间理论模型中，经常存在这种假设，在这些模型中，资产定价通常跟随波动率参数的扩散，这些参数也跟随扩散。Melino and Turnbull(1990)和 Wiggins(1987)争论离散时间随机波动率模型自然地接近这个程序。如果我们将离散时间程序作为随机波动率的参数，则产生了过滤问题：处理观测到的数据来测量随机波动率的参数和及时地估计每点的波动率水平。一个简单的随机波动率模型如下：

$$h_t = e_t e^{a_t/2}, \quad a_t = \beta a_{t-1} + x_t \quad (12.2.20)$$

其中， $e_t \sim N(0, s_e^2)$ ， $x_t \sim N(0, s_x^2)$ ，并且，我们假设  $e_t$  和  $x_t$  是连续不相关和彼此独立。其中， $a_t$  表示收益的条件对数标准偏差和它的均值之间的差额。它遵循均值为零的 AR(1)过程。

我们将收益方程式平方后取对数得：

$$\log(h_t^2) = a_t + \log(e_t^2), \quad a_t = \beta a_{t-1} + x_t \quad (12.2.21)$$

这是一个线性状态空间矢量，除了(12.2.21)中第一个等式是  $\log x^2$  分布而不是正态分布存在误差。为考虑非正态的重要性，我们只需考虑一个事实，那就是当  $e_t$  非常接近 0 时(“内层”)， $\log(e_t^2)$  就是一个非常大的负的外层。

这个系统能被各种方法所估计。Melino and Turnbull(1990)和 Wiggins(1987)用 GMM 估计。尽管这种方法很直接，但效率不高。Harvey, Ruiz, and Shephard(1994)提出了准极大似然估计，忽略了  $\log(e_t^2)$  的非正态并继续下去，似乎(12.2.21)中两个等式都有误差项。最近，Jacquier, Polson, and Kim (1994)也提出了 simulation-based 精确的极大似然估计。

### 12.2.2 多变量模型

到现在，我们只研究了单个资产收益的波动率。更一般的情况是，我们可能遇到资产收益的向量，它的条件协方差矩阵随时间而变化。假设我们有资产  $N$ ，其收益创新为  $h_{i,t+1}$ ， $i=1\dots N$ 。我们将这些创新组成一个向量  $h_{t+1} = [h_{1,t+1} \dots h_{N,t+1}]$  并定义  $s_{ii,t} = \text{Var}_t(h_{i,t+1})$ ， $s_{ij,t} = \text{Cov}_t(h_{i,t+1}, h_{j,t+1})$ 。则有  $\sum_t = [s_{ij,t}]$  是所有收益的条件协方差矩阵。非常方便地将  $\sum_t$  的非冗余的元素----在主对角线上或下面----构成一向量。执行这个组合的运算就是著名的 vech 运算： $\text{vech}(\sum_t)$  是一个有  $N(N+1)/2$  个元素的向量。

### 多变量 GARCH 模型

我们在单变量文章中处理问题的思想很自然地转向多变量中。单变量  $GARCH(1,1)$  模型 (12.2.6) 的最简单的推广是将  $vech(\sum_t)$  与  $vech(h_t h_t')$  和  $vech(\sum_{t-1})$  联系：

$$vech(\sum_t) = w + \Psi vech(\sum_{t-1}) + vech(h_t h_t') \quad (12.2.22)$$

这里  $w$  是一个有  $N(N+1)/2$  个元素的向量， $\Psi$  和  $\Gamma$  是  $N(N+1)/2 \times N(N+1)/2$  矩阵；所以，这个矩阵的所有参数数目是  $N^2(N+1)^2/2 + N(N+1)/2$ ，是  $N$  的 4 次幂。很明显，这个模型很快就变得难处理了。许多有关多变量  $GARCH$  模型的文章就寻找对一些对 (12.2.22) 的似是而非的限制来减少参数的数量。文章的另外一个重要的目标是找到一些限制来保证协方差矩阵  $\sum_t$  一定为正。这些限制在单变量方程中是相当直接的----比如，单变量  $GARCH(1,1)$  模型中的所有参数必须是正数----但在多变量模型中就没有这么明显了。

Kroner and Ng(1993)提供了一种很好的领先的多变量  $GARCH$  模型的调查方法。第一份详细说明，Bollerslev, Engle, and Wooldridge(1988)的  $VECH$  模型(在  $vech$  运算之后命名)，将协方差矩阵记作一系列单变量  $GARCH$  模型。 $\sum_t$  的每个元素遵循单变量  $GARCH$  模型，由叉积矩阵  $h_t h_t'$  的相应元素所决定。 $\sum_t$  的第  $(i, j)$  个元素由下式给出：

$$s_{ij,t} = w_{ij} + b_{ij} s_{ij,t-1} + a_{ij} h_{it} h_{jt} \quad (12.2.23)$$

这个模型由 (12.2.22) 通过将矩阵  $\Psi$  和  $\Gamma$  成对角线矩阵而得。如果参数矩阵  $[w_{ij}]$ ， $[b_{ij}]$  和  $[a_{ij}]$  都是正定型，则隐含的条件协方差矩阵就通常也为正定型。这个模型对  $\sum_t$  的每个元素都有三个参数，因此，共有  $3N(N+1)/2$  个参数。

第二份详细说明，Engle and Kroner(1995)中的  $BEKK$  模型(在 Bollerslev, Engle, Kraft, 和 Kroner 的一篇早期的工作底稿中命名)通过二次形式(二次齐式)的而不是  $\sum_t$  的单个元素来保证是正定型。模型为：

$$\sum_t = C' C + B' \sum_{t-1} B + A' h_t h_t' A \quad (12.2.24)$$

其中， $C$  是具有  $N(N+1)/2$  个参数的低三角形矩阵， $B$  和  $A$  为每个具有  $N^2$  个参数的平方矩阵，总参数数量为  $(5N^2 + N)/2$ 。

$BEKK$  模型的一个特殊情况是 Engle, Ng, and Rothschild(1990)中的单要素  $GARCH(1,1)$  模型。在这个模型中，我们定义  $N$ -向量  $w$  和标量  $\gamma$ ，有：

$$\sum_t = C' C + H' [b w' \sum_{t-1} w + a (w' h_t)^2] \quad (12.2.25)$$

$C$  如同在前一方程式中所规定。我们可以给这个模型一个标准化的限制；令  $i' w = 1$ ， $i$  是其中的一个向量。向量  $w$  则可被认为是有价证券的加权向量。我们定义  $h_{pt} = w' h_t$ ， $s_{pp,t} = w' \sum_t w$ 。则模型可重新记为：

$$s_{ij,t} = w_{ij} + I_i I_j s_{pp,t}$$

$$s_{pp,t} = w_{pp} + bs_{pp,t-1} + ah_{p,t}^2 \quad (12.2.26)$$

任意两个资产收益的协方差只随有价证券收益的方差运动，有价证券的收益遵循单变量的  $GARCH(1,1)$  模型。单变量  $GARCH(1,1)$  模型是 BEKK 模型的一个特例，矩阵 A 和 B 有秩 1： $A = \sqrt{a}wI'$ ， $B = \sqrt{b}wI'$ 。它有  $(N^2 + 5N + 2)/2$  个自由参数。

这个模型能够被直接延伸到多个要素或高次  $GARCH$  模型。

最后，Bollerslev(1990)提出了一个常数相关模型，其中每个资产收益方差遵循单变量  $GARCH(1, 1)$  模型，任何两个资产的协方差由一常数---相关的系数乘以收益的条件标准偏差：

$$\begin{aligned} s_{ii,t} &= w_{ii} + b_{ii}s_{ii,t-1} + a_{ii}h_{ii}^2 \\ s_{ij,t} &= r_{ij}\sqrt{s_{ii}s_{jj,t}} \end{aligned} \quad (12.2.27)$$

这个模型有  $N(N+5)/2$  个参数。它提供了正定协方差矩阵只要系数  $r_{ij}$  由明确的相关矩阵构成并且参数  $w_{ii}$ ， $a_{ii}$ ， $b_{ii}$  都为正。

为理解这些模型之间的差异，考虑经过相反的大扰动撞击两资产后，资产收益间的条件协方差的变化是有益的。在具有正系数  $a_{ij}$  的 VECH 模型中，负的叉积  $h_{it}h_{jt}$  减少了条件协方差。换句话说，在常数相关模型中，叉积  $h_{it}h_{jt}$  是无关的；任何提高两正相关资产的方差都将提高他们的协方差。在要素模型中， $s_{ij,t}$  只随  $s_{pp,t}$  变化，所以，负的叉积的效果取决于有价证券 p 的权重。

就象在单变量模型中，在多变量  $GARCH$  模型中，收益波动率可能也是持久的。多变量模型在概率上允许一些资产波动率存在共同成分；比如，在一系列波动率中可能存在一个持久的成分，所以一个波动率中的所有变化与另一个相关是短时间的。Bollerslev and Engle (1993)研究了 this 想法，类似于在线性单位根求解过程中的协整概念。

### 多变量随机波动模型

由(12.2.20)限定的单变量随机波动模型也很容易扩展到多变量情况。我们有：

$$h_t = e_t e^{a_t/2} \quad a_t = fa_{t-1} + x_t \quad (12.2.28)$$

$h_t$ ， $e_t$ ， $a_t$  和  $x_t$  是  $(N \times 1)$  向量和  $f$  是  $(N \times N)$  矩阵。这个模型中，矩阵  $f$  有  $N^2$  个参数，协方差矩阵  $e_t$  有  $N(N+1)/2$  个参数，方差矩阵  $h_t$  有  $N(N+1)/2$  个参数，所以总共有  $N(2N+1)$  个参数。没有必要限制  $a_t$  一定为正，很直接地估计的协方差参数的平方根形式来保证隐含的协方差矩阵是正定的。Harvey, Ruiz, and Shephard(1994)提出了这种模型的限制解释，其中  $f$  是对角线的(将参数减少到  $N(N+2)$ )或甚至是单位矩阵(更将参数减少到  $N(N+1)$ )。

即使没有这些特别的限制，理解规定(12.2.28)强加资产收益的常数条件相关也是很重要的。在这方面，它和 Bollerslev ' s 常数相关 GARCH 模型的限制一样，但对任意  $N>3$ ，它就需要更多的参数。

### 条件市场模型

至今，我们所讨论的甚至限制最严的模型都很难应用于大量的横截面数据集，因为它们参数的增长是资产  $N$  的平方。问题是这些模型是以收益的整个条件协方差矩阵作为研究的主要目标。代替的方法是，类似于早期的统计学上的平均值----方差分析的发展，它对条件市场模型发生作用。继续忽略非零平均值收益，得：

$$h_{i,t+1} = b_{it} h_{m,t+1} + x_{i,t+1} \quad (12.2.29)$$

其中， $b_{it} \equiv s_{im,t} s_{mm,t}$  是资产的条件贝塔， $x_{i,t+1}$  是在资产中彼此不相关的特殊扰动。在这个框架下我们就可构造市场收益的条件方差， $s_{mm,t}$ ，成单变量  $GARCH(1,1)$  模型。我们还可构造  $b_{im,t}$  或一样地  $s_{im,t}$  取决于  $s_{mm,t}$ ， $b_{im,t-1}$  和收益  $h_{it}$  和  $h_{mt}$ 。我们可以构造收益的特殊扰动的方差作为另一个单变量  $GARCH(1,1)$  过程。这种模型所隐含的协方差矩阵是正定型的，参数的数目是以速度  $N$  而不是  $N^2$  增长，使得模型适用于大量的资产。Braun, Nelson 和 Sunier(1995)采用了这种逼近，运用  $EGARCH$  函数形式来求解模型的单个成分。

### 12.2.3 一次阶(矩)与二次阶(矩)的联系

我们复习了时间系列中一些特别的复杂的随时间变化的二次阶(矩)模型，它们的一次阶(矩)被假定为常数或 0。但在金融学理论中最重要的是它与资产收益的一次阶与二次阶的关系。因此，我们讨论模型，条件方差收益可能和条件方差或协方差一起变化。

#### $GARCH$ 模型

Engle,Lilien,and Robins(1987)提议增加随时间变化的截距到基础的单变量模型(12.2.2)。记  $r_{t+1}$  为连续的复合资产收益率，是利率的时间系列(由于我们用均值为零的创新)，我们有：

$$r_{t+1} = m_t + s_t e_{t+1}, \quad m_t = g_0 + g_1 s_t^2 \quad (12.2.30)$$

和以前一样， $e_{t+1}$  是 IID 随机变量， $s_t^2$  遵循任何  $GARCH$  过程。 $GARCH$ -in-mean 或  $GARCH$ -M 模型使得收益的条件均值与条件方差是线性的。它可用极大似然估计，尽管我们不知道模型是否满足极大似然估计的渐进正态的匀称条件。

$GARCH$ -M 模型也能被详细化，条件平均值在条件标准偏差而不是条件方差。它被 Bollerslev,Engle 和 Wooldridge(1988)和其他人员推广到多变量的

情况，但参数数目随着资产数目的增加而飞快增长，因此，它比较适合少量的资产。

### 工具变数逼近

作为 GARCH-M 模型的代替法，campbell(1987)和 Harvey(1989,1991)提出了联系一次阶和二次阶通过 GMM 来估计参数。这些作者以一个市场收益模型开始，这个模型中，期望的市场收益的方差是线性的，有条件对一些向量  $H_t$ ， $H_t$  包含 L 指令或预测变量：

$$E[r_{m,t+1} | H_t] = g_0 + g_1 \text{Var}[r_{m,t+1} | H_t] \quad (12.2.31)$$

Campbell 和 Harvey 假设条件期望收益在指令和定义误差是线性的。

$$u_{m,t+1} \equiv r_{m,t+1} - H_t b_m$$

$$e_{m,t+1} \equiv r_{m,t+1} - g_0 - g_1 (r_{m,t+1} - H_t b_m)^2 \quad (12.2.32)$$

$b_m$  表示市场收益的回归系数的向量 on the instruments。误差  $u_{m,t+1}$  是市场收益与指令的线性联合之间的差额，而误差  $e_{m,t+1}$  则是市场收益与线性函数  $u_{m,t+1}^2$  之间的差额。模型(12.2.31)隐含着误差  $u_{m,t+1}$  和  $e_{m,t+1}$  都与指令  $H_t$  互不相关。有 L 个指令，则有 2L 个不相关变量来估计 L+2 个参数 ( $g_0, g_1$  和  $b_m$  中的 L 个参数)。这样 GMM 提供了参数估计和模型限制的矫正检验。

这种方法能轻松地推广到包括其他资产的情况，资产的期望收益应该满足以下条件：

$$E[r_{i,t+1} | H_t] = g_0 + g_1 \text{Cov}[r_{i,t+1}, r_{m,t+1} | H_t] \quad (12.2.33)$$

如果有 N 个资产，我们定义向量  $r_{t+1} \equiv [r_{1,t+1}, \dots, r_{N,t+1}]'$ 。则  $r_{t+1}$  的条件期望值为

$$E[r_{t+1} | H_t] = H_t B, \quad \text{其中, } B \text{ 为有 } NL \text{ 个参数的矩阵。我们定义误差:}$$

$$u_{t+1} \equiv r_{t+1} - H_t B$$

$$e_{t+1} \equiv r_{t+1} - g_0 - g_1 (r_{t+1} - H_t B)(r_{m,t+1} - H_t b_m) \quad (12.2.34)$$

我们得到 2NL 个额外的正交性条件来求解 NL+1 个额外参数。式(12.2.33)中共有 2(N+1)L 个正交性条件，N(L+1)+L+2 个参数。这样只要两个或更多指令的出现这个模型就确定了。

Harvey(1989)更深地将这个模型推广到允许随时间变化的风险价格。他将 (12.2.33) 替换为：

$$E[r_{i,t+1} | H_t] = g_0 + g_{1t} \text{Cov}[r_{i,t+1}, r_{m,t+1} | H_t] \quad (12.2.35)$$

$g_{1t}$  随时间而变化，但对所有资产都一样。Since (12.2.35) holds for market portfolio itself,

$$g_{1t} = \frac{E[r_{m,t+1} | H_t] - g_0}{\text{Var}[r_{m,t+1} | H_t]} \quad (12.2.36)$$

Harvey 用这个来估计模型。他将(12.2.36)代入(12.2.35)，两边乘以  $Var[r_{m,t+1} | H_t]$ ，又根据  $E[r_{m,t+1} | H_t] = H_t b_m$ ， $E[r_{t+1} | H_t] = H_t B$  构造一个新的误差向量：

$$v_{t+1} = (r_{m,t+1} - H_t B_m)^2 (H_t B - g_0 \mathbf{i}) - (r_{t+1} - H_t B)(r_{m,t+1} - H_t b_m)(H_t b_m - g_0) \quad (12.2.37)$$

Harvey 将(12.2.37)中的  $v_{t+1}$  替代(12.2.34)中的  $e_{t+1}$ ，并将(12.2.32)中的误差  $e_{m,t+1}$  去掉。这就给出了一个少 L 个正交条件和 1 个参数(由于模型中剔除了  $g_1$ ) 的系统。Overidentifying restrictions 的数目减少了 L-1。Harvey(1989)发现当以美国股票指数作为市场有价证券时，风险的价格会变化。然而，他也拒绝模型的过拟合限制。Harvey(1991)使用了世界股票指数作为市场有价证券也得到同样结果。

### 条件 CAPM 和无条件 CAPM

等式(12.2.35)和(12.2.36)可以写成

$$E[r_{i,t+1} | H_t] = g_0 + b_{it} I_t \quad (12.2.38)$$

其中， $b_{it} \equiv Cov[r_{i,t+1}, r_{m,t+1} | H_t] / Var[r_{m,t+1} | H_t]$ ，资产 i 的条件贝塔， $I_t \equiv E[r_{m,t+1} | H_t] - g_0$ ，市场上超过无风险收益的超额收益的期望值。

Jagannathan and Wang(1996)强调，CAPM 的条件 version 不需要第 5 章所讨论的无条件 CAPM。如果我们取(12.2.38)的无条件期望，得：

$$E[r_{i,t+1}] = g_0 + (E[b_{it}]) (E[I_t]) + Cov(b_{it}, I_t) \quad (12.2.39)$$

其中， $E[I_t]$  是市场上超额收益的无条件期望值。 $E[b_{it}]$  是条件贝塔的无条件期望值，尽管差别很小，但不需要和无条件贝塔一样。最重要的是，条件贝塔和式(12.2.39)中的期望超额市场收益  $I_t$  之间的协方差。资产的市场风险越高其贝塔就越高，就比无条件 CAPM 有更高的无条件平均收益。Jagannathan 和 Wang(1996)提出当股票市场的期望超额收益比较高时，小股票的贝塔值会上升，这种作用导致了小股票的高平均收益。他们尽管没有直接构造小股票贝塔的时间变化模型，但提出了一些间接的证据。

### 波动率创新和收益创新

实验研究没有确切证据证明股票收益的高波动率时期就是股票收益的高期望时期。一些文章提出了一些较弱的证据表达了这种关系(见 Bollerslev, Engle, and, Wooldridge[1988], French, Schwert, and Stambaugh[1987], and Harvey[1989])，但其他文章将短期名义利率作为一个工具，发现收益的平均值与波动率之间是负相关的(见 Campbell[1987] and Glosten, Jagannathan, and Runkle[1993])。

如同 French, Schwert, 和 Stambaugh(1987)所强调的那样, 有更有力的证据表明正的波动率创新与负的收益创新是相关的。我们已经讨论不对称 GARCH 模型怎么适合这种关系。在更高层次来说, 它能用两种方法中的一个来表示。一种可能是负的扰动影响波动率。Black(1976)提出的杠杆作用假设, 认为当杠杆公司的总价值下降时, 资产净值会变得更小。由于资产净值承担了公司的所有风险, 资产净值的波动率百分比应该提高。甚至一个公司在财务上没有运用债务杠杆, 但如果公司对工人或供应厂商有固定的承诺, 这也可能发生。尽管有一些真实的事例, 但使用实际杠杆估计, 也很难计算收益的波动率相关性的等级(见 Christie[1982]和 Schwert[1989])。

另外的一种解释是因果关系起了另外作用: 正的波动率扰动降低收益。Campbell 和 Hentschel(1992)称之为波动率反应假设。如果当波动率提高时股票的期望收益增加, 并且期望的股息不变, 则当波动率提高时, 股票价格应该下降。Campbell 和 Hentschel 通过对收益(7.2.26)应用对数线性近似建立公式模型:

$$r_{t+1} = E_t[r_{t+1}] + h_{d,t+1} - h_{r,t+1} \quad (12.2.40)$$

其中,  $h_{d,t+1} \equiv E_{t+1}[\sum_{j=0}^{\infty} r^j \Delta d_{t+1+j}] - E_t[\sum_{j=0}^{\infty} r^j \Delta d_{t+1+j}]$  表示式(7.2.25)中未来股息期望值的变化。 $h_{r,t+1} \equiv E_{t+1}[\sum_{j=1}^{\infty} r^j r_{t+1+j}] - E_t[\sum_{j=1}^{\infty} r^j r_{t+1+j}]$  表示未来收益的期望值的变化。

Campbell 和 Hentschel 构造股息消息变量  $h_{d,t+1}$  是均值为 0 的 GARCH(1,1) 过程:  $h_{d,t+1} \sim (0, s_t^2)$ , 其中,  $s_t^2 = w + b s_{t-1}^2 + a h_{d,t}^2$ 。他们将股息消息的方差与期望收益成线性关系:  $E_t[r_{t+1}] = g_0 + g_1 s_t^2$ 。这些假设隐含着修订后的所有未来收益的期望值是现在波动率扰动 ( $h_{d,t+1}^2 - s_t^2$ ) 的倍数:

$$\begin{aligned} h_{r,t+1} &= E_{t+1}[\sum_{j=1}^{\infty} r^j r_{t+1+j}] - E_t[\sum_{j=1}^{\infty} r^j r_{t+1+j}] \\ &= q (h_{d,t+1}^2 - s_t^2) \end{aligned} \quad (12.2.42)$$

其中,  $q = g_1 r a / (1 - r(a + b))$ 。当  $g_1$  很大时,  $a$  很大时(扰动严重影响未来波动率),  $a + b$  很大时(波动率扰动对期望收益有持久的影响), 系数  $q$  也很大(期望的收益紧跟波动率变化)。代入(12.2.40), 得收益的隐含方程:

$$r_{t+1} = g_0 + g_1 a_t^2 + h_{d,t+1} - q(h_{d,t+1}^2 - s_t^2) \quad (12.2.42)$$

这并不是 GARCH 过程, 而是 GARCH 过程的二次式。它意味着收益是负扰动的, 因为  $h_{d,t+1}$  的大的负的实现会经过二次式会被放大, 而大的正的实现会被抑制。直观的是两边任何大的扰动都会提高期望的未来波动率和要求的收益, 减少现在股票的收益。逆转地, “没有消息是好消息”; 如果  $h_{d,t+1} = 0$ , 则降低了期望的未来的波动率和提高现在的股票收益。Campbell and Hentschel 在估计

模型(12.2.42)时找到了比单单估计一个标准 GARCH-M 模型更有力的证据风险  $g_1$  的正价格。这个结果表示波动率反馈效应和杠杆作用导致了股票市场波动率的不对称行为。

### 12.3 非参数估计

在一些金融文章中，我们可能发现两变量  $Y$  和  $X$  之间的函数关系，而没有结构模型来限制关系的参数。在这些情况下，我们可以用非参数的估算技术来描述没有依靠任何非线性关系的说明的非线性的广泛的范围。和第 12.1 和 12.2 章中介绍的相当高的结构或参数逼近来估计非线性法形成对比的是，非参数的估计法对非线性的性质不需太多的假设。然而，这并不是没有代价的----非参数化估计是强数据要求，而且对太小的取样效果不大。此外，非参数估计特别易于过拟合，这是一个在统计学上不容易克服的问题(见下面第 12.5 章)。

也许最常用的非参数估计是光滑估计，通过复杂的方法平均化数据来减少测量误差。核回归，正交系列扩张，映射追击，最近邻域估计，平均导数估计，样条函数，和人工神经网络都是光滑估计的例子。为明白平均的意义，我们假设现在希望估计两个变量  $X_t$  和  $Y_t$  之间的关系，两变量满足：

$$Y_t = m(X_t) + e_t \quad t=1, \dots, T \quad (12.3.1)$$

其中  $m(\cdot)$  是任意固定的但未知的非线性函数， $\{e_t\}$  是均值为 0 的 IID 过程。

假设  $X_{t_0} = x_0$  时计算  $m(\cdot)$  在特定日期  $t_0$  的值，如果对这个测量值  $X_{t_0}$ ，我们可以得到变量的重复独立测量值  $Y_{t_0}$ ，令  $Y_{t_0}^1 = y_1, \dots, Y_{t_0}^n = y_n$ 。则  $m(\cdot)$  在点  $x_0$  的估计值为：

$$\hat{m}(x_0) = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n [m(x_0) + e_t^i] \quad (12.3.2)$$

$$= m(x_0) + \frac{1}{n} \sum_{i=1}^n e_t^i \quad (12.3.3)$$

根据极大值法则，当  $n$  足够大时，(12.3.3)中第二项可忽略。

当然，如果  $\{Y_t\}$  是一个时间系列，对给定  $X_t$ ，我们没有足够重复测量值。然而，如果我们假设函数  $m(\cdot)$  充分光滑，时间系列测量值  $X_t$  邻近  $x_0$ ，则其对应的值  $Y_{t_0}$  也接近  $m(x_0)$ 。换句话说，如果  $m(\cdot)$  足够光滑，则存在一足够小的  $x_0$  的邻域，使得  $m(x_0)$  恒定且可用  $Y_t$  的平均值来估计，与接近  $x_0$  的  $X_t$  系列相一致。

$X_t$  的值越接近  $x_0$ ，则对应的  $Y_t$  的平均值越接近  $m(x_0)$ 。This argues for a weighted average of the  $Y_t$  's ,where the weights decline as the  $X_t$  's get farther away from  $x_0$ 。这种估计  $m(x)$  的加权平均过程就是光滑的本质。更正式地，对任意  $x$ ， $m(x)$  的光滑估计可表示为：

$$\hat{m}(x) \equiv \frac{1}{T} \sum_{i=1}^T w_{i,T}(x) Y_i \quad (12.3.4)$$

其中，当系列  $X_i$  接近  $x$  时，加权系数  $\{w_{i,T}(x)\}$  很大，当  $X_i$  远离  $x$  时，很小。

在实施这个方法前，我们首先要定义什么是“近”和“远”。如果我们选择  $x$  的太大的邻域来计算平均值，则得到的加权平均值会太光滑，得不到  $m(\cdot)$  的真实的非线性。如果我们选择太小的邻域，加权平均值会显得太易变，反映了和  $m(\cdot)$  中的变化一样的噪音。因此，加权系数  $\{w_{i,T}(x)\}$  应认真选择，以平衡上述两种考虑。我们在第 12.3.1 到 12.3.3 和 12.5 中讨论它及相关主题。

### 12.3.1 核回归

一种估计  $m(\cdot)$  的重要的光滑技术是核回归。在核回归模型中，加权函数  $w_{i,T}(x)$  由概率密度函数  $K(x)$  构造而成， $K(x)$  也称为核：

$$K(x) \geq 0 \quad \int K(u) du = 1 \quad (12.3.5)$$

尽管  $K(x)$  是概率密度函数，但它在接下来的分析中并不用概率法则 --- 它仅仅是计算加权平均值的一种方便的方法，并不意味着  $X$  遵循  $K(x)$  (参数假设)。

通过设定变量  $h > 0$  来重新调节核，如果我们定义如下，则改变  $h$ ，我们可以改变它的 spread：

$$K_h(u) \equiv \frac{1}{h} K(u/h) \quad \int K_h(u) du = 1 \quad (12.3.6)$$

我们定义加权函数如下：

$$w_{i,T}(x) \equiv K_h(x - X_i) / g_h(x) \quad (12.3.7)$$

$$g_h(x) \equiv \frac{1}{T} \sum_{i=1}^T K_h(x - X_i) \quad (12.3.8)$$

如果  $h$  非常小，则平均值会在  $X_i$  's 的相当小的邻域中，如果  $h$  非常大，则平均值会在  $X_i$  's 的相当大的邻域中。因此，调整光滑参数  $h$ ，也称为带宽，就可以控制平均数的程度。将(12.3.8)代入(12.3.4)得到 Nadaraya-Watson 核心估计  $\hat{m}_h(x)$ ：

$$\hat{m}_h(x) = \frac{1}{T} \sum_{i=1}^T w_{i,T}(x) Y_i = \frac{\sum_{i=1}^T K_h(x - X_i) Y_i}{\sum_{i=1}^T K_h(x - X_i)} \quad (12.3.9)$$

Figure 12.4 Simulation of  $Y_i = \sin(X_i) + 0.5e_i$

在确定规律性条件下，核  $K$  的样条和加权系数的数量和行为随着例子的范围增大而增大，也许可以发现， $\hat{m}(x)$  经几种方式渐进地收敛于  $m(\cdot)$  (详见

H Härdle [1990])。大部分核都有这种收敛的性质，但在本章的剩余部分和我们的实验例子中，我们都用最普通的核，高斯核(the Gaussian kernel)：

$$K_h(x) = \frac{1}{h\sqrt{2p}} e^{-\frac{x^2}{2h^2}} \quad (12.3.10)$$

### 核回归的说明

为说明核回归在描述非线性关系方面的能力，我们将光滑技术应用到由 Monte Carlo 模拟构成的人工数据集。用  $\{X_i\}$  表示 500 个观测值的系列，取值在 0 与  $2p$  之间，间隔均匀。 $\{Y_i\}$  通过以下非线性关系与  $\{X_i\}$  相关：

$$Y_i = \sin(X_i) + 0.5e_i \quad (12.3.11)$$

其中， $\{e_i\}$  是 IID 伪随机标准正态的变量系列。使用模型数据  $\{X_i, Y_i\}$  (见图 12.4)，根据核回归，我们可以估计条件期望  $E[Y_i | X_i] = \sin(X_i)$ 。为做这些，我们将 Nadaraya-Watson 估计(12.3.9)with Gaussian 核心应用这些数据，并使宽带参数  $h$  在  $0.1\hat{s}_x$  和  $0.5\hat{s}_x$  之间变化，其中， $\hat{s}_x$  是  $\{X_i\}$  的标准偏差。通过  $h$  在标准偏差的单位区间变动，我们通过  $X_i$  本身的标准偏差来将说明变量  $X_i$  标准化，就如(12.3.10)所建议。

对  $h$  的每个值，我们取核估计作为  $X_i$  的函数，这些点如图 12.5a 到 12.5c 所示。观测带宽为  $0.1\hat{s}_x$ ，发现核估计波浪起伏----带宽太小，无法提供足够的局部均值回收于  $\sin(X_i)$ 。核估计获得了数据的循环的性质，它也发现了由噪音引起的随机方差，随着带宽的增加，它可能消除，从而扩大了局部均值的范围。图 12.5b 显示了带宽为  $0.3\hat{s}_x$  的核估计，显得更光滑和更紧密真实的条件期望。

随着带宽的增加，局部均值超过了连续的范围，核估计的可变性( $x$  的函数)减少了。图 12.5c 显示了带宽为  $0.5\hat{s}_x$  的核估计，由于一些 sine 函数的真实的差异随着噪音一起消除了，显得太光滑了。在极限情况下，核估计接近  $\{Y_i\}$  的样本均值，并且作为  $X_i$  的函数， $Y_i$  的可变性都消失了。

### 12.3.2 最理想的带宽选择

很显然，从第 12.3.1 章的例子中，我们可以发现在核回归的应用中，选择合适的带宽是很重要的。选择最佳的带宽有很多种方法，最普通的是互相证实理论(theory of cross-validation)，主要是因为它的强度和渐进理想(详见 Härdle [1990,Chapter5])。在这种方法中，带宽的选择为了缩小核估计的加权平均平方差。特别地，取  $T$  个观测值的样本  $\{X_i, Y_i\}_{i=1}^{t=T}$ ，令：

$$\hat{m}_{h,j}(X_j) = \frac{1}{T} \sum_{t \neq j} w_{t,T}(X_j) Y_t \quad (12.3.12)$$

这是基于扣除第  $j$  个观察值的数据集的核估计，估计第  $j$  个观察值  $X_j$  的值。

于是互相证实函数  $CV(h)$  定义如下：

$$CV(h) = \frac{1}{T} \sum_{t=1}^T [Y_t - \hat{m}_{h,t}(X_t)]^2 d(X_t) \quad (12.3.13)$$

其中  $d(X_t)$  是非负加权函数用来减少边界效应 (Härdle [1990, p.162] 中有更详尽的阐述)。函数  $CV(h)$  称为互相证实函数，因为通过  $T$  个二次抽样  $\{X_t, Y_t\}_{t \neq j}$  证实了核估计适合  $\{Y_t\}$ ，每次抽样省略一个观测值。最佳带宽就是使得函数最小化的那个。

### 12.3.3 平均导数估计

在许多金融应用中，我们希望  $Y_t$  与几个变量  $X_{1t}, \dots, X_{kt}$  非参数相关。例如，我们希望用几个因素的线性函数构造出股票和债券的预期收益：市场收益，利差，股息收益等等 (见 Lo and MacKinlay [1996])。毫无疑问，这种想法比第 12.3.1 章中例子中的单变量更具有雄心。为明白为什么，我们假设有五个独立变量的例子，不失一般性，可令这五个变量取值单位区间  $[0, 1]$ 。即使我们将每个变量的集合分成十个相等的区间，这也会产生总数为  $10^5 = 100,000$  个邻域，每个带宽为 0.10；这样，我们就至少需要 100,000 个观测值来保证每个邻域的每个数据点的平均值。这个由维数引起的祸根只能通过对非线性加限制才能解决。

例如，假设  $X_{it}$  的线性组合与  $Y_t$  非参数相关。当提供足够的结构允许估计使用合理的取样范围时，有利于捕获重要的非线性性质。特别地，考虑以下多变量的非线性模型：

$$Y_t = m(X_t') + e_t, \quad E(e_t | X_t) = 0 \quad (12.3.14)$$

其中  $X_t = [X_{1t}, \dots, X_{kt}]'$  是一个  $(k \times 1)$  向量， $m(\cdot)$  是一任意固定的非线性函数。函数  $m(\cdot)$  可通过以下两步估计：(1) 通过平均导数估计  $\hat{b}$  来估计  $b$ ；(2) 通过  $Y_t$  对  $X_t' \hat{b}$  的核回归来估计  $m(\cdot)$ 。

Stoker (1986) 发现，如果下列两个条件之一成立的话，(12.3.14) 中系数  $b$  可通过一般的最小二乘法转化为比例因子来估计：(1)  $X_t$ 's 是多变量正态向量；或，更一般，(2) 对  $i=1, \dots, k$ ， $E[X_{it} | X_t' b]$  是线性的。如果上述两个条件有一个成立，Stoker (1986) 提出了独特的估计法，平均导数估计法 (Average derivative estimators)，可以一贯地估计  $b$  (见 Stoker [1992])。

平均导数估计法是基于以下事实， $m(\cdot)$  关于  $X_i$  的导数的期望值与  $\mathbf{b}$  是均衡的：

$$E\left[\frac{\partial m}{\partial X_i}\right] = E[m'(X_i, \mathbf{b})] \mathbf{b} \propto \mathbf{b} \quad (12.3.15)$$

因此，平均导数估计法等价于对比例因子  $\mathbf{b}$  的估计，由于它可能包容在  $m(\cdot)$  中并可用核回归估计，这个比例因子和我们的目标不相关，。

有几个平均导数估计变量：直接的，间接的，和斜率估计。Stoker(1991, Theorem 1) 表明它们都是渐进等价的；然而，Stoker(1992, Chapter 3) 赞成间接斜率估计(ISE)有两个原因。首先，如果  $Y_i$  与  $X_i$  确实是线性相关，间接斜率估计还是无偏见的，然而，其他方法却不一定。第二，由于 ISE 的比例形式，间接斜率估计法对其非参数成分估计要求更简单(见下)。

直观地，间接斜率估计  $\hat{\mathbf{b}}_{ISE}$  开拓了这个事实，未知的参数向量  $\mathbf{b}$  与因变量  $Y$  和独立变量  $X_i$  的边际密度的对数的负导数之间的协方差是均衡的，用  $l(\cdot)$  表示。因此，通过计算  $Cov(Y, l(\cdot))$ ，我们得到比例因子  $\mathbf{b}$  的前后一致的估计。通过计算  $Y$  与样本对  $l(\cdot)$  的副本之间的协方差可以计算协方差。

更普遍地， $\hat{\mathbf{b}}_{ISE}$  也可记为  $Y_i$  对  $X_i$  经工具矩阵  $H$  的回归工具变量(IV)估计(见附录 Section A.1)：

$$\hat{\mathbf{b}}_{ISE} = (H'X)^{-1}H'Y \quad (12.3.16)$$

其中  $Y \equiv [Y_1 \dots Y_T]'$ ，

$$H \equiv \begin{bmatrix} 1 & I_b(X_1)\hat{l}(X_1)' \\ \vdots & \vdots \\ 1 & I_b(X_i)\hat{l}(X_i)' \\ \vdots & \vdots \\ 1 & I_b(X_T)\hat{l}(X_T)' \end{bmatrix}, \quad X \equiv \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_T \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix} \quad (12.3.17)$$

$\hat{l}(\cdot)$  是  $X_i$  的边缘密度的对数的负导数， $I_b(X)$  是用比固定常数  $b$  更小的估计边缘密度修正样本的一部分的指标函数：

$$I_b(x) \equiv 1[\hat{f}(x) > b] \quad (12.3.18)$$

在大部分的实验应用中，常数  $b$  限定在样本的 1%到 5%之间也是修正过了的。

为得到  $\hat{l}(\cdot)$ ，如果  $f(x)$  表示  $X_i$  的边际密度，则  $f(x)$  的高斯核估计表示如下：

$$\hat{f}(x) = \frac{1}{T} \frac{1}{h^k} \sum_{i=1}^T K\left(\frac{X - X_i}{h}\right) \quad (12.3.19)$$

$$\text{其中，} K\left(\frac{X - X_i}{h}\right) \equiv \prod_{i=1}^k K_i\left(\frac{x_i - X_{it}}{h}\right) \quad (12.3.20)$$

$$= (2p)^{-k/2} \exp\left[-\frac{1}{2h^2}(X - X_t)'(X - X_t)\right] \quad (12.3.21)$$

因此，得

$$\hat{f}(x) = \frac{1}{T} \frac{1}{h^{k+1}} \sum_{t=1}^T K\left(\frac{X - X_t}{h}\right) \quad (12.3.22)$$

$$= -\frac{1}{T} \frac{1}{h^{k+1}} \sum_{t=1}^T K\left(\frac{X - X_t}{h}\right)(X - X_t/h) \quad (12.3.23)$$

于是，我们定义  $\hat{l}(X)$  为：

$$\hat{l}(X) \equiv -\frac{\hat{f}'(X)}{\hat{f}(X)} \quad (12.3.24)$$

尽管  $f(x)$  的多变量性质，但在核估计 (12.3.19) 中依然只存在一个简单的带宽来调整。象在单变量的情况一样，带宽控制着局部均值的程度，但现在超过多维的邻域。作为一个实践问题，局部平均值过程的数字性质通过在计算  $\hat{f}(\cdot)$  前，对所有的  $X_{it}$ s 进行标准偏差来标准化来提高。然后，乘以每个  $b_i$ ，对应的  $X_{it}$  的标准偏差来解除标准化。

#### 12.3.4 应用：估计状态价格密度

在不确定性下的投资活动中最重要的先进理论之一是 Arrow(1964) 和 Debreu(1959) 的时间状态优先选择模型，在他们的文章中，他们介绍了简单的证券，每只股票在固定的自然状态支付 \$1，其他时间没有。就是现在所说的 Arrow-Debreu 证券，他们是基本模块，我们从他们那衍生出许多现在的不确定性环境下的经济平衡理解。

在实践中，由于真实的 Arrow-Debreu 证券并不在任何一个交易所交易。Arrow-Debreu 证券价格无法观测。然而，运用非参数技术----尤其是多变量核回归--- Ait-Sahalia 和 Lo(1996) 发展了对这些证券的估计方法，即连续状态情况下的状态价格密度 (SPD)。SPD 包括了涉及经济活动中的风险资产的定价 (和套利交易) 的信息财富。原则上，它可以用在其它资产的定价，甚至那些现在还没有交易的资产。(例如见 Ait-Sahalia 和 Lo[1995])。

更重要的是，SPD 包括了许多有关偏好和资产定价动力学的成分。例如，如果在资产定价的数据生成程序中加入参数限制，SPD 估计可以用在资产定价的平衡模型中的典型的代理人的仁义偏好的推理上。(见 Bick[1990] and He and Leland[1993])。同样地，如果加上专门的偏好，SPD 估计可以用来推断资产定价的数据生成程序(例如，见 Kerman and Kani[1994]，Dupire[1994]，Jackwerth and Rubinstein[1985]，Longstaff [1992,1994]，Rady[1994]，Rubinstein[1985]，and Shimko [1991,1993])。

事实上，Rubinstein[1985]已经发现，以下任意两个条件隐含着三个条件：(1)典型的代理人偏好；(2)资产定价动力学；和(3)SPD。

### 状态价格密度的定义

为正式地定义 SPD，考虑标准的动力交换经济(见第八章)，某一证券在时间  $t$  的均衡价格与时间  $T$  的清算盈利  $Y(C_T)$  --- 总消费  $C_T$  的函数，的关系是：

$$P_t = E[Y(C_T)M_{t,T}] \quad , \quad M_{t,T} \equiv \frac{d^{T-t}U'(C_T)}{U'(C_t)} \quad (12.3.25)$$

其中， $M_{t,T}$  表示时间  $t$  到  $T$  之间的边际替代率， $d$  表示时间偏好率。这个均衡资产定价关系使得证券现在的价格与未来盈利的期望值的贴现值相等，以随机贴现因子贴现。

Lucas(1978)发现，(12.3.25)并没有隐含要求  $\{P_t\}$  满足鞅过程。支持了 Leroy's(1973)的在合理决定资产价格中，鞅过程既非必要也不是充分条件的观点。然而，假设未来消费的条件分布密度函数表示为  $f_t(\cdot)$ ，(12.3.25)的条件期望也能表示如下：

$$E_t[Y(C_T)M_{t,T}] = \int Y(C_T) \frac{d^{T-t}U'(C_T)}{U'(C_t)} f_t(C_T) dC_T \quad (12.3.26)$$

$$= e^{-r_{t,T}(T-t)} \int Y(C_T) f_t^*(C_T) dC_T \quad (12.3.27)$$

$$= e^{-r_{t,T}(T-t)E_t^*[Y(C_T)]} \quad (12.3.28)$$

$$\text{其中，} f_t^*(C_T) \equiv \frac{M_{t,T} f_t(C_T)}{\int M_{t,T} f(C_T) dC_T} \quad , \quad (12.3.29)$$

$r_{t,T}$  表示在  $T$  时刻提供一单位消费的资产，在时间  $t$  到  $T$  之间以连续的复利计算的净收益率；即，它是无风险资产收益。

欧拉等式(Euler equation)表明资产现在的价格可以表示为贴现的期望盈利，贴现率为无风险利率(详见第八章)。然而，期望值取决于 SPD  $f^*$ ，加权替换的边际率的概率密度函数，而不是未来消费的起始概率密度函数  $f$ 。在连续时间假定下， $f^*$  也称为风险中性价格密度(Cox and Ross[1976])或等价鞅测度(Harrison and Kreps [1979])。

一旦  $f_t^*$  确定，它可以用简单的  $T$  时刻的清算盈利，任意的消费  $C_T$  的函数，来表示任何资产在日期  $t$  的定价。SPD 也提供了在第八章中所讨论的偏好基础的等价模型与第九章讨论的套利基础导数定价模型之间的联系。事实上，所有的金融证券---不管是否是派生的---其价格隐含着 Arrow-Debreu 证券价格，并且这些价格可以用来测量其他所有不管怎么复杂的证券的价格。

### SPD 的导数定价

在一般规则性条件下，我们可以将  $f^*$  展开为  $t$  和  $T$  的函数，因此，简单 SPD  $f^*(C_T; t, T)$  可以用来对在将来  $T \geq t$  的时间有清算盈利  $Y(C_T)$  的资产在时期  $t$  的定价(见脚注 16)：

$$P_t = e^{-r_t(T-t)} \int Y(C_T) f^*(C_T; t, T) dC_T \quad (12.3.30)$$

and we shall adopt this convention for notational simplicity。例如， $T$  日期总消费  $C_T$ ，执行价格为  $B$  的欧式看涨期权的盈利函数为： $Y(C_T) = \max[C_T - X, 0]$ ，则  $t$  时刻的价格  $G_t$  为：

$$G_t = e^{-r_t(T-t)} \int \max(C_T - X, 0) f^*(C_T; t, T) dC_T \quad (12.3.31)$$

甚至最复杂的路径独立衍生证券根据(12.3.30)也可以定价和套期保值。例如，有一具有非常非线性盈利函数的证券：

$$Y(C) = \frac{a-b}{1+\exp[-b(C-a)]} + b \quad a>0, \quad b<0 \quad (12.3.32)$$

$$a \equiv c + \frac{1}{b} \log(-a/b) \quad (12.3.33)$$

盈利函数的图形是光滑的，看涨差价期权，以较低执行价购买一份看涨期权，以较高执行价卖出一份看涨期权。(见图 12.5) 和 Cox and rubinstein [1985, Chapter]。

从价格导数中求得 SPD。

Ross(1976)，Banzand Miller(1978)和 Breeden and Litzenberger(1978)发现，期权价格与 SPD 之间有比(12.3.30)提出的更紧密的关系。特别地，他们发现赎回价格函数  $G_t$  对执行价格  $X$  的二阶导数应该等于 SPD：

$$\frac{\partial^2 G_t}{\partial X^2} = e^{-r_t(T-t)} f^* \quad (12.3.34)$$

因此，在期权定价公式里的是 SPD  $f^*$ 。

用(12.3.34)估计 SPD，我们假设一看涨期权定价公式。尽管存在许多参数定价公式(见 Hull [1993, Chapter 17])，Ait-sahalia and Lo(1996)构造了一个非参数定价公式，更少的限制---首先光滑和弱依赖---对潜在的资产定价的数据生成过程。然而，就如 Black 和 Scholes(1973)和 Merton(1973)里的参数公式提供了极大的方便，如果参数假设(即几何布朗运动)满足的话。非参数理论坚强地反对这些假设。由于有一些实验证据对这些假设存在疑问，至少对股票指数，非参数方法可能有一些重要的优势。

存在可观测的看涨期权价格  $\{G_t, X_t, t_i\}$  (其中  $t_i \equiv T - t_i$ )，标的资产价格  $\{P_t\}$ ，无风险利率  $\{r_t\}$ ，我们可以构造光滑的非参数看涨期权价格函数：

$$\hat{G}(P, X, t, r_t) = \hat{E}[G | P, X, t, r_t] \quad (12.3.35)$$

应用多变量核  $K$ ，形成  $d=4$  的单变量核的乘积：

$$K_h(P, X, \mathbf{t}, r_t) \equiv k_{h_p}(P)k_{h_x}(X)k_{h_t}(\mathbf{t})k_{h_r}(r_t) \quad (12.3.36)$$

因此，得：

$$\hat{G}(P, X, \mathbf{t}, r_t) = \frac{\sum_{i=1}^n k_{h_p}(P - P_i)k_{h_x}(X - X_i)k_{h_t}(\mathbf{t} - \mathbf{t}_i)k_{h_r}(r - r_i)C_i}{\sum_{i=1}^n k_{h_p}(P - P_i)k_{h_x}(X - X_i)k_{h_t}(\mathbf{t} - \mathbf{t}_i)k_{h_r}(r - r_i)} \quad (12.3.37)$$

期权的德尔塔  $\Delta$  和 SPD 估计由  $\hat{P}$  的微分得到：

$$\hat{\Delta}(P, X, \mathbf{t}, r_t) = \frac{\partial \hat{G}(P, X, \mathbf{t}, r_t)}{\partial P} \quad (12.3.38)$$

$$\hat{f}^*(P_T | P, \mathbf{t}, r_t) = e^{rt} \left[ \frac{\partial^2 \hat{G}(P, X, \mathbf{t}, r_t)}{\partial X^2} \right]_{X=P_T} \quad (12.3.39)$$

和对实物看涨期权价格函数的光滑假设一样，在对数据生成过程的标准常规假设下，*Ait-sahalia* 和 Lo(1996)发现，期权价格的估计，期权的德尔塔  $\Delta$  和 SPD 是相容的和渐进正态。他们提供了清楚的渐进的方差表达式。

应用 SPD，任何具有特征  $\mathbf{z}$  和  $T = t + t$  时刻的盈利函数  $Y(\mathbf{z}, P_T)$  的衍生证券在任何时刻  $t$  都可通过以下函数定价：

$$G(P, \mathbf{z}, \mathbf{t}, r_t) = e^{-rt} \int_0^\infty Y(\mathbf{z}, P_T) \hat{f}^*(P_T | P, \mathbf{t}, r_t) dP_T \quad (12.3.40)$$

如果，盈利函数  $Y(\cdot)$  是  $P_T$  的两次微分，则有

$$G(P, \mathbf{z}, \mathbf{t}, r_t) = e^{-rt} \int_0^\infty Y(\mathbf{z}, P_T) \hat{f}^*(P_T | P, \mathbf{t}, r_t) dP_T \quad (12.3.41)$$

$$= \int_0^\infty Y(\mathbf{z}, P_T) \frac{\partial^2 \hat{G}}{\partial P_T^2} dP_T \quad (12.3.42)$$

$$= \int_0^\infty \frac{\partial^2 Y(\mathbf{z}, P_T)}{\partial P_T^2} \hat{G} dP_T \quad (12.3.43)$$

对  $\hat{G}$  的积分而不是对其二次求导，提高了估计的收敛速度----  $\hat{G}$  以  $n^{1/2}h^{4/2}$  的收敛而它的对光滑函数  $P_T$  的积分收敛速度为  $n^{1/2}h^{3/2}$ ，而， $\hat{G}$  的二次导数的收敛速度为  $n^{1/2}h^{8/2}$ ，对光滑函数  $P_T$  的积分收敛速度为  $n^{1/2}h^{6/2}$ 。在对盈利函数的二次导数的积分中，得到收敛速度为  $h^{3/2}$  的因子----如果存在的话----against  $G$  instead of integrating the payoff function itself against the second derivative of  $G$ 。

*Ait-sahalia* and Lo(1996)应用估计法来对 S&P 500 看涨和看跌期权来定价和德尔塔套利交易，通过在 January 4, 1993 到 December 31, 1993 之间芝加哥期权交易所的日常数据，总共 14431 个观察值。SPD 的估计展示了负的挠和过度峭度，历史股票收益的一般特征(见第 1 章)。同样，不象许多参数期权定价模型那样，SPD 生成的期权定价公式能够描述持久的波动率的“笑容”和其他市场价格的实验特征。

## 12.4 人工神经网络

最近，在工程学和工商界颇受注意的代替非参数回归的是人工神经网络。人工神经网络也可以看作是一种非参数技术，因此，这些模型在第 12.3 节中十分自然。然而，由于最初他们是从生物现象中得到----特别是从神经细胞的生理学中得来----他们成为一个独立的，独特的和萌芽的文章的一部分(见 Hertz, Krogh, and Palmer[1991], Hutchinson, Lo, and Poggio[1994], Poggio and Girosi[1990], and White[1992] for overviews of this literature)。

为强调人工神经网络的一般非参数起源，我们在本节中列举了三种网络，统称为学习网络(见 Barron and Barron [1988])。在第 12.4.1 中，我们介绍了多层感知器，可能是最近的文章中最流行的人工神经网络类型---这就是“神经网络”的意思。在第 12.4.2 和 12.4.3 中，我们列举了其他有网络解释的两种技术：射线基础函数和映射追逐回归。

### 12.4.1 多层感知器

可能人工神经网络的最简单的例子是 McCulloch and Pitts(1943)中的二元模型(binary threshold model)，其中，输出变量  $Y$  只取值 0 或 1，与输入变量  $X_j$ ， $j=1, \dots, J$  的集合  $J$  是非线性关系：

$$Y = g\left(\sum_{j=1}^J \mathbf{b}_j X_j - \mathbf{m}\right) \quad (12.4.1)$$

$$g(u) = \begin{cases} 1 & u \geq 0 \\ 0 & u < 0 \end{cases} \quad (12.4.2)$$

Figure 12.7 Binary Threshold Model

根据(12.4.1)，每一输入变量  $X_j$  通过系数  $\mathbf{b}_j$  加权，称为连接强度，然后将所有输入变量加权相加。如果加权总和超过了阈  $\mathbf{m}$ ，则人工神经元通过激活函数  $g(\cdot)$  接通或激活；否则它保持休眠状态。这种简单的网络经常如图 12.7 所示，输入层与输出层相接。二元模型的推广形成了现在大部分的人工神经网络模型的基础。特别地，为得到连续的输出值，Heaviside 激活函数(12.4.2)被逻辑函数所代替(见图 12.8)：

$$g(u) = \frac{1}{1 + e^{-u}} \quad (12.4.3)$$

Figure 12.8 Comparison of Heaviside and Logistic Activation Functions

Figure 12.9 Multilayer Perceptron with a Single Hidden Layer

同样，由于  $m$  经常可用来构造一个非零激活级模型，不失一般性，定义起始值  $X_1 = 1$ ，在输入连接强度的相反数  $-b_1$  为激活级时，我们令  $m$  为 0。

但是，二元模型的最重要的拓展可能是在输入层和输出层之间引入隐藏层。特别地，令

$$Y = h\left(\sum_{k=1}^K a_k g(\mathbf{b}_k' X)\right) \quad (12.4.4)$$

$$\mathbf{b}_k = [b_{k1} b_{k2} \cdots b_{kj}]', \quad X = [X_1 X_2 \cdots X_j]'$$

其中， $h(\cdot)$  另一个任意的非线性激活函数。在这种情况下，输入层与多个隐藏单位相连，在每个隐藏单位里他们是加权的(不同的)，通过同一激活函数  $g(\cdot)$ 。每个隐藏单位的输出值则又加权-----这次是通过  $a_k$  ---求和并通过第二个激活函数  $h(\cdot)$  转换。这么一种网络构造是多层感知器 (MLP) 的例子---只有单一(隐藏)层----可能是人工神经网络在最近的应用中，最常用的类型。与图 12.7 对比。这种多层感知器有更复杂的网络拓扑(见图 12.9)。这很显然可推广到更多的隐藏层，即多层感知器。

对给定的输入和输出系列  $\{X_t, Y_t\}$ ，MLP 近似值总计来估计 MLP 网络的参数---向量  $\mathbf{b}_k$  和标量  $a_k$ ， $k = 1, \dots, K$  ----典型地最小化输出值和网络的平方偏差的总和，即  $\sum_t [Y_t - \sum_k a_k g(\mathbf{b}_k' X)]^2$ 。在文章的术语中，参数估计的过程称为训练网络。这比它看起来更少自负---参数估计的早期理论是后向传播的，这确实是一种模仿学习行为(虽然是一个非常简单的模型)。然而，White(1992)引用了许多的后向传播的实验缺点(数值不稳定性、偶尔不收敛，等等)，因此估计(12.4.4)的参数的首选模型是非线性最小平方法。

甚至，单一的隐藏层 MLP(12.4.4) 具有普通近似性质 (universal approximation property)：它可以将任意非线性函数近似为有合适数目隐藏单位的任意精确度(见 White [1992])。然而，一些非参数估计技术也具有单变量近似性质，包括第 12.3 节中的非参数回归估计和第 12.4.2 节和 12.4.3 节中的技术。当然，这也没有告诉我们实践中这些技术的性能，对给定的数据集，可能存在一种技术在精确度或其他方面比另一种更具优势。

MLP 的最主要的优势在于通过相当简单的函数的组合将复杂的非线性关系表达出来。这种设定导致了它自己自然地并行处理，尽管现在没有金融应用开

发 MLP 的这种性质，但这可能很快就会改变，就象并行处理软件和硬件变得更广泛应用一样。

为介绍 MLP 模型，我们将它应用于(12.3.11)所产生的人工数据集。对 MLP 所定义的一个隐藏层和 5 个隐藏单元，通过  $q(\cdot)$  作用密度函数，我们得到下面模型：

$$\begin{aligned} \hat{Y}_t = & 5.282 - 14.576g(-1.472 + 1.869X_t) - 5.411g(-2.628 + 0.641X_t) \\ & - 3.071g(13.288 - 2.347X_t) + 6.320g(-2.009 + 4.009X_t) \\ & + 7.892g(-3.316 + 2.484X_t) \end{aligned} \quad (12.4.5)$$

其中， $g(u) = 1/(1 + e^{-u})$ 。将这个模型如图 12.10 所示并与第 12.3.1 节中描述的核估计相比较，尽管(12.4.5)看起来一点都不象正弦函数，然而，MLP 执行起来相当理性并容易估计。

Figure 12.10 MLP(1,5) Model of  $Y_t = \sin(X_t) + 0.5e_t$

#### 12.4.2 射线基础函数

射线基础函数(RBFs)首先是用来解决插值问题 --- 确定一条准确的曲线通过一个点集(见 Powell[1987]for a review)。最近，许多研究人员将 RBFs 延伸到更一般的近似中(见 Broomhead and Lowe[1988], Moody and Darken [1989], and Poggio and dirosi[1990])。特别地，Poggio and Girosi(1990)介绍了 RBFs 怎样起源于古典的常规问题，其中，一些未知函数  $Y = m(X)$  由一些数据集  $(X_p, Y_p)$  来近似和一些光滑约束。在多重回归类推中， $d$  维向量  $X_t$  是解释性(独立)变量， $Y_t$  是因变量， $m(\cdot)$  可能是非线性函数，是  $Y_t$  对  $X_t$  的条件期望值。则有：

$$Y_t = m(X_t) + e_t \quad E[e_t | X_t] = 0 \quad (12.4.6)$$

规划(或非参数估计)问题则可以看作将下面目标函数最小化：

$$V(m) \equiv \sum_{t=1}^T (\| \hat{Y}_t - m(X_t) \|^2 + I \| Dm(X_t) \|^2) \quad (12.4.7)$$

其中， $\| \cdot \|$  是某个向量模， $D$  是微分算子。(12.4.7)求和式子中的首项是  $m(X_t)$  和  $Y_t$  之间的差额，第二项是补偿函数，是  $m(\cdot)$  光滑性上的减函数。 $I$  控制光滑度与拟合度之间的权衡。

在其最一般形式，在确定条件下(例如，见 Poggio and Girosi[1990])，(12.4.7)的最小化的解决方法由下面表达式给出：

$$\hat{m}(X_t) = \sum_{k=1}^K b_k \hat{m}_k (\| X_t - U_k \|) + D(X_t) \quad (12.4.8)$$

其中,  $\{U_k\}$  是  $d$  维向量中心(类似于样条函数的节),  $\{b_k\}$  是标量系数,  $\{\hat{m}_k\}$  是标量函数,  $D(\cdot)$  是一多项式,  $K$  比样本中的观察值数目  $T$  少得多。这些近似值被 Poggio and Girosi(1990)称为 hyperbasis functions, 与样条紧密相关, 与核估计和其他非参数估计一样光滑。

我们现在的目的是, 通过一个  $(d \times d)$  加权矩阵  $W$ , 取向量模为加权欧几里得(Euclidean)形式, 多项式也为线性或常数项, 得  $\hat{m}(\cdot)$  的设定:

$$\hat{m}(X_t) = a_0 + a_1 + \sum_{k=1}^K b_k \hat{m}_k(X_t - U_k)' W' W(X_t - U_k) \quad (12.4.9)$$

其中,  $a_0$  和  $a_1$  是多项式  $D(\cdot)$  的系数。Miccheli(1986)发现一大类基础函数  $\hat{m}(\cdot)$  是合适的, 但最常用的是高斯  $e^{-x/s^2}$  和多元二次式  $\sqrt{x+s^2}$ 。

这种网络可以推广到任何真实价值的输出, 但在应用中, 我们对要求的输出范围有一些先验的知识, 应用非线性转换函数到产出上反映这种知识在计算上更有效率。在我们导出定价模型的应用中就是这样, 其中, 一些 RBF 网络会被产出 S 形所扩张。将范围  $(-\infty, \infty)$  转化为  $(0,1)$ 。特别地, 扩张的网络成为  $g(\hat{m}(x))$  形式, 其中,  $g(u) = 1/(1+e^{-u})$ 。和 MLP 一样, 对给定的输入与产出的集合  $(X_p, Y_t)$  的 RBF 近似, 包括估计 RBF 网络的参数---- $d(d+1)/2$  个独特的矩阵  $W'W$  输入, 中心集  $\{U_k\}$  的  $dk$  个元素,  $d+k+1$  个系数  $a_0$ 、 $a_1$  和  $\{b_k\}$  ----典型地通过非线性最小二乘法, 即, 将  $\sum_t [Y_t - m(X_t)]^2$  最小数值化。

### 12.4.3 映射追逐回归

映射追逐是在统计领域中通过观察低维映射来分析高维数据集的理论。Friedman and Stuetzle (1981)发展非线性回归问题形式称为映射追逐回归 (PPR)。与 MLP 类似, PPR 模型是数据映射的集合, 即估计系数与数据的内积, 但与 MLP 不同的是他们也从数据中估计非线性联合函数。根据 12.4.1 的符号, 单变量产出的 PPR 公式可表示如下:

$$m(X_t) = a_0 + \sum_{k=1}^K a_k m_k(b_k' X_t) \quad (12.4.10)$$

其中, 函数  $m_k(\cdot)$  从数据中估计得来(典型的光滑),  $\{a_k\}$  和  $\{b_k\}$  是系数,  $K$  是映射的数目,  $a_0$  一般取值为产出  $m(X_t)$  的均值。PPR, RBF 和 MLP 网络的相似之处在(12.4.10)中很明白地体现出来。

### 12.4.4 学习网络的局限

尽管学习网络在近似非线性函数方面有许多优点, 它们也有一些不可忽视的局限。特别地, 目前在给定的应用中没有普遍接受的程序来构造网络体系,

即隐藏层的数目，隐藏单位的数目，激活函数的设定等等。尽管一些主要的规则在偶然的实验观察中出现，但他们最多还仅是启发式地。

在训练网络的过程中，困难也增加了。典型地，通过求平方差的和的最小值得到网络参数，但由于这些设定固有的非线性属性，这个目标函数不会是球形凸起，可能会有许多局部最小。

最后，传统的统计推理技术，例如显著性检验，由于层的嵌套，经常不能应用于网络模型。例如，如果(12.4.4)中， $a_k$ 有一个为0，则隐藏层间的连接强度 $b_k$ 是不明的。因此，甚至简单的显著性检验和可靠区间也要求复杂的持续假设(maintained hypothesis)的组合来正确解释。

#### 12.4.5 应用：学习 BS 公式(Black-Scholes formula)

给定神经网络的指数和挠性来近似复杂的非线性关系，这是一个在衍生证券方面很自然的应用，衍生证券的定价公式是高度非线性甚至他们在密闭的区间才有效。特别地，Hutchinson,Lo,and Poggio(1994)提出以下挑战：如果期权价格确实由 Black-Scholes 公式决定，那么神经网络能否“掌握” Black-Scholes 公式？用更标准的统计术语来说：通过学习网络，Black-Scholes 公式能否进行非参数化估计，在实践中保证足够的精确度。

Hutchinson,Lo,and Poggio(1994)面对这个挑战，执行 Monte Carlo 模拟实验，对人工生成的 Black-Scholes 期权价格所附有的各种神经网络与 Black-Scholes 公式进行比较，进行分析和样本外保值实验，看看他们有多接近。即使只有六月的日常数据集，学习网络定价公式可以相当精确地近似 Black-Scholes 公式。

特别地，他们开始估计一个两年期的日常股票价格的样本，根据芝加哥期权交易所的规则，根据 Black-Scholes 公式所确定的价格，创造一个每天期权的横截面。由于网络是对这个样本培养的，他们将这个两年期股票和(多重)期权价格作为培养路径。给定一个模拟日常股票价格的培养路径 $\{P(t)\}$ ，根据芝加哥期权交易所的规则，他们构造相应的期权价格路径来介绍股票期权。

图 12.11 所示的典型的训练路径。由于期权产生于一个特定样本路径是(随机的)股票价格路径的函数，这个数据矩阵的大小(期权的数目和数据点的总数目)随着样本路径而变化。对它们的训练集，每个样本路径的期权数目从 71 到 91 变化，平均值为 81。数据点的总数目变化范围为 5227 到 6847，平均值为 6001。

从神经网络得到的非线性模型中获得的期权价格和德尔塔的估计，很难将它与真实的 Black-Scholes 值区分开来。一个 RBF 网络的估计与误差的例子如图 12.12 所示。这个特别的 RBF 网络的估计等式是：

$$\begin{aligned} \widehat{G/X} = & -0.06 \sqrt{\left[ \begin{array}{c} P/X - 1.35 \\ t - 0.45 \end{array} \right] \begin{pmatrix} 59.79 & -0.03 \\ -0.03 & 10.24 \end{pmatrix} \begin{pmatrix} P/X - 1.35 \\ t - 0.45 \end{pmatrix} + 2.55} \\ & -0.03 \sqrt{\left[ \begin{array}{c} P/X - 1.18 \\ t - 0.24 \end{array} \right] \begin{pmatrix} 59.79 & -0.03 \\ -0.03 & 10.24 \end{pmatrix} \begin{pmatrix} P/X - 1.18 \\ t - 0.24 \end{pmatrix} + 1.97} \\ & +0.03 \sqrt{\left[ \begin{array}{c} P/X - 0.98 \\ t + 0.20 \end{array} \right] \begin{pmatrix} 59.79 & -0.03 \\ -0.03 & 10.24 \end{pmatrix} \begin{pmatrix} P/X - 0.98 \\ t + 0.20 \end{pmatrix} + 0.00} \\ & +0.10 \sqrt{\left[ \begin{array}{c} P/X - 1.05 \\ t + 0.10 \end{array} \right] \begin{pmatrix} 59.79 & -0.03 \\ -0.03 & 10.24 \end{pmatrix} \begin{pmatrix} P/X - 1.05 \\ t + 0.10 \end{pmatrix} + 1.62} \\ & +0.14P/X - 0.24t - 0.01 \end{aligned} \quad (12.4.11)$$

其中， $t \equiv T - t$ 。从(12.4.11)观察得到，RBF 模型的中心不受输入范围的限制，事实上，在本例中，也不在第三和第四个中心。这些网络中的最大的误差可能出现在平价期权到期执行时的结点上，同样沿着样本点的边界。

尽管学习网络的价格的精确度毫无疑问大家很感兴趣，但这不足够保证非参数近似的实验的恰当。特别地，用期权保值的能力同样重要，由于以套利为基础的定价公式的存在断定依靠在动态的套利策略中复制期权的能力(见第九章中的讨论)。这附加的约束对程序化技术提供了额外的动力，特别地，在 Hutchinson, Lo, and Poggio(1994)中的 RBF 网络。

特别地，德尔塔套利策略要求对标的的定价公式的导数有精确的近似，这种要求直接导致了调整技术如 RBF 网络对光滑的要求。Hutchinson, Lo, and Poggio(1994)发现 RBF 和 MLP 网络都对模拟 Black-Scholes 数据提供了极好的套利能力，和在 S&P500 期权期货上一样的实验应用，在一些情况下，胜过 Black-Scholes 公式(公式是在德尔塔套利执行是连续的假设下求导的，尽管这些模拟假设日常德尔塔套利)。

当参数求导定价公式有效时，他们是首选的，Hutchinson, Lo, and Poggio(1994)的结果表明当参数理论失效时，非参数学习网络替代可以有效代替。虽然他们的发现是有前途的，但我们无法得出结论，即这种近似在一般情况能否应用----他们的模拟仅仅注意了 Black-Scholes 模型，他们的实验应用也只包括了一些简单的工具和时期，从 1987 到 1991 之间的 S&P 期权期货。

然而，这种一般的近似对未来的研究引向许多有前途的方向。在这方面的最重要的进步是附加输入值的设定，参数模型不容易捕获的输入值即市场收益，一般市场波动率和其他经济条件测度。

学习网络的说明中对合适统计推论的需求激发了其他的研究方向。首先，我们要求一些理论来相配这种网络结构----非线性的单位，基础函数的类型等等----对手上的特殊的(明确的)数据集，最佳的，首选的，自动的形式。

第二，样本大小和近似误差之间的关系应该研究，解析地或通过附加 Monte Carlo 模拟实验。也许一些数据依赖的度量可以构造出来，则可以实时地提供近似误差的估计，在同样的情况下，典型的统计估计的标准差也可获得。

最后，对更好的执行测度的需求是清楚的。然而，典型的 goodness-of-fit 测量如  $R^2$  确实提供了一些对模型选择的指导，他们仅仅是不完全的执行测量。而且，自由度的概念对非线性模型不再是明确的，这对所有统计学上的拟合测度都有暗示。

## 12.5 过拟合和数据窥察

尽管本章中所讨论的每个非线性理论都有自己的成本和好处，过拟合和数据窥察问题对他们的影响都一样。当一个模型拟合得“太好”就产生过拟合，也就是说，模型描述随机噪音和真实的非线性一模一样。启发式地，产生过拟合的根源在于太少的“自由度”或相对数据点的数目有太多的参数，典型的征兆是极好的样本内拟合(in-sample fit)极差的样本外表现(out-sample performance)。数据窥察是一个会导向极好但虚假的样本外表现的问题。当我们忽略了一个事实，那就是许多设定搜索被引向得到拟合数据的模型的最后的设定，数据窥察偏差会上升。甚至，存在模型事实上是错误的，但通过搜索足够长的各种数据集和/或参数值，我们可能发现一些组合会拟合数据。然而，这种拟合是虚假的和仅仅是我们广泛的查找过程的一个征兆。

不幸的是，由于这个提高他们的过程和产生真正的实验发现的过程是一样的，对这两个问题没有任何简单的矫正。产生这两个问题的根源在于没有能力执行控制实验，从而，我们对数据的理解严重地依靠实验推理。随着统计推理的形式，存在误差的边缘，并且这个边缘对我们处理数据和修正模型中的小的变化都很敏感。

而且，有许多方法来减轻过拟合和数据窥察的影响。例如，系统设定研究可能明确地计算出来，就象在 Lo and MacKinlay(1990b)。在这样的例子中，用不正确的统计分布来推理可以预防找到不存在的显著的结果。细致的样本外

表现评价可能发现过拟合问题，并且如果相当少的样本外检验被执行，或如果他们被引向不同(弱相关)的数据，这样就会减少数据窥察的影响。

但减小过拟合和窥察影响的最有效的方法可能是通过先验理论考虑对设定查找增加一些限制。这些考虑可能是具有很好铰接的经济行为数学模型，或由心理现象引起的行为模型，或以判断，直觉和过去经验为基础的主要的启发标准。然而，所有的这些根源也在一定程度上受到过拟合和数据窥察的影响----推理所不可避免的问题----他们较少受影响和提供更少数据依赖的模型生效方式。

所有的这些都表明，对模型的先验框架或设定在面临数据前是需要的。随同我们设法捕获的现象的种类，和查找中应用的相关变量，提出这样的设定，则突然产生虚假的成功模型的机会就减少了。

## 12.6 结论

很明显，在金融应用中，非线性起了很显著的作用，归功于大量数据集的计算能力和可行性的提高。不象前面章节所说的材料，本章所讨论的一些想法尚未成型，更多的试探性的。不用多久，许多应用的技术会被改进，一些会被放弃。然而，提高研究方向的能力是很重要的，open questions to be addressed. 特别是在研究的时期。

尽管我们讨论非线性模型的灵活性，但他们也确实存在一些局限性。与线性模型相比，他们更难精确估计，对非正常值更敏感，数字上更不稳定，更易发生过拟合和数据窥察。与一般的想法相反，非线性模型要求更多的经济结构和先验的考虑，而不是更少。他们的解释经常要求更多的努力和注意。然而，非线性是经济生活普遍存在的一个事实，在许多金融应用中，非线性的根源和性质可以很容易地鉴别或，至少，以某种特征化的形式表现出来。在这种情况下，本章介绍的技术是金融计量学家的有利武器。

## 第 12 章的问题

12.1 在数字计算机上执行的大部分伪随机的数据(数字)产生器是乘法线性同余发生器(multiplicative linear congruential generators(MLCG))，其中， $X_n = (aX_{n-1} + c) \text{ 模 } m$ ， $a$  是某一“选择好的”乘数， $c$  是任意的常数， $m$  等于或略小于在计算机语言中表示的最大的整数。(例如，令

$a=1664525, c=0, m=2^{32}$ )。对比 MLCC 数目，考虑下面两个非线性递归式：帐篷图(见第 12.1.1)和逻辑图，分别为：

$$X_n = \begin{cases} 2X_{n-1} & X_{n-1} < \frac{1}{2} \\ 2(1-X_{n-1}) & X_{n-1} \geq \frac{1}{2} \end{cases} \quad X_0 \in (0,1) \quad (12.6.1)$$

$$X_n = 4X_{n-1}(1-X_{n-1}), \quad X_0 \in (0,1) \quad (12.6.2)$$

这些递归式是混沌系统的例子，展示对初始条件和显著的复杂动力学行为的极端敏感度。

12.1.1 什么是伪随机数据生成器的好的性质，如何比较实际应用中的生成器的特性(不是理论上)？

12.1.2 对帐篷和逻辑图，执行各种 Monte Carlo 模型与 MLCC 比较来决定哪一个是更好的伪随机数据生成器。哪一个更好，为什么？在决定使用哪个标准前，你要考虑使用伪随机数据生成器的应用种类。提示：用 1.99999999 代替(12.6.1)中的 2，而用 3.99999999 代替(12.6.2)中的 4----额外(特别)的信用：解释原因。

12.2 对 S&P500 指数从 1926.1 到 1994.12 间，每个月的收益，估计一多层感知器模型，使用 5 个时滞收益作为输入量和一个有 10 个单位的隐藏层。Calculate the in-sample root-mean-squared-error(RMSE)of the one-step-ahead forecast of your model and compare it to the corresponding out-of-sample results for the test period 1986:1to 1994:12. 你能解释执行中的区别吗？

12.3 用核回归来估计从 1965 年 1 月到 1994 年 12 月间的 IBM 与 S&P500 的每月收益之间的关系。怎样从核估计的结果中计算出常规的贝塔来？至少需要构造两个测度通过普通最小二乘法来捕获核估计的递增值。