

第 18 章 人工神经网络的金融应用

Min Qi

1. 引言

数据驱动建模方法，例如人工神经网络（ANN）方法，在金融中的应用正日益普遍。总的说来，ANN 是一种非线性非参数模型。ANN 使得人们可以充分利用数据，在没有任何限制参数的建模假定下，由数据确定模型的结构和参数。它们在金融领域之所以成为一种吸引人的方法，是因为存在丰富的高质量金融数据，而可检验的金融模型却相对缺乏。随着计算机处理速度的加快和计算成本呈指数递减，这种具有高强度计算的方法变得富有吸引力了。

本文首先对 ANN 作简要介绍，第 3 节阐述它和一些传统统计方法之间的关系。第 4 节提出一些有用的 ANN 建模方法。第 5 节回顾在金融应用中几个主要领域的实证研究。第 6 节给出结论。

2. 人工神经网络

1940 年以来经过积极论和怀疑论之间连续的三个循环之后，对神经网络的研究在过去十年中有突飞猛进的发展。这很大程度上是由于认识到 ANN 有强大的模型识别能力，这使得它在许多应用领域优于其他现有的建模方法。ANN 已经吸引了众多应用领域研究人员的注意，包括信号处理、医疗成像、经济和金融建模（仅仅列举了少数）。同时，认知科学、神经系统科学、心理学、生物学、计算机科学、数学、物理学和统计学研究人员为 ANN 的结构和方法论的发展做出了贡献。许多网络，例如多层前馈网络(multilayer feedforward networks)、重复和统计网络、联合记忆网络和自行组织网络等，已广泛发展，用途各异。现在各种各样有监督或者无监督的学习规则可用于从数据中培育、形成网络。在这些网络中，多层前馈后传网络(multilayer feedforward backpropagation networks)在金融中的应用最为普遍，也是本文研究的中心。对神经网络理论更为广泛的介绍，可参阅 Hecht-Nielsen (1990)、Hertz, Grogh 和 Palmer (1991)、Wasserman (1993)、Bose 和 Liang (1996)。White, Gallant, Hornik, Stinchcombe 和 Wooldridge (1992) 为那些精通神经网络或数理统计的研究人员提供了一本对 ANN 的近似和学习能力进行数学分析的论文集。Gately (1996) 为初学者提供了神经网络应用的一种特别的非技术、一步一步的方法。

2.1. ANN 的结构

受大脑和神经系统研究的启发，神经网络模拟了一个具有许多相对简单的个体单元高度联系、平行计算的结构。个体单元编成三层：输入层、中间层和输出层。前馈网络单方向将输入映射成输出，即从输入层到中间层再到输出层。在中间层和输出层的每一个单元都有一个传递函数(transfer function)，用于传递它接收到的信号。输入层的单元没有传递函数，它们用于向网络分发输入信号。每一个连接(connection)都有一个数值权重，用于修正通过它的信号。

考虑一个具有单一输出单元、 k 个中间层单元和 n 个输入层单元的三层前馈网络（见图 1）。输入层可以用向量 $X = (x_1, x_2, \dots, x_n)'$ 表示，中间层可以用向量 $M = (m_1, m_2, \dots, m_k)'$ 表示，而 y 是输出单元。任一中间层单元接收所有输入单元的权重以及偏倚项（用 x_0 表示， x_0 总是等于 1），并且产生一个输出信号

$$m_j = F\left(\sum \beta_{ij} x_i\right) = F(X' \beta_j), \quad j = 1, 2, \dots, k, \quad i = 0, 1, 2, \dots, n, \quad (2.1)$$

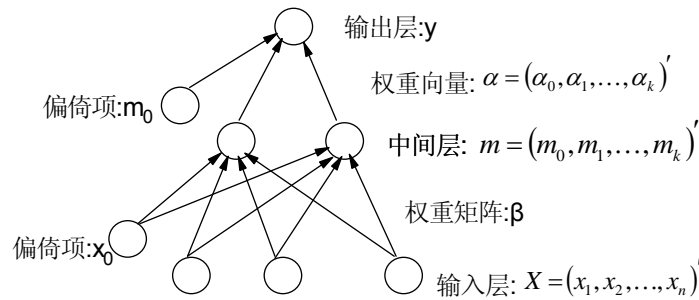


图 1. 三层前馈神经网络

其中 F 是传递函数， x_i 是第 i 个输入信号，而 β_{ij} 是从第 i 个输入单元到第 j 个中间层单元连接的权重。同样，输出单元接收中间层单元输出信号的加权和，并产生信号

$$y = G\left(\sum \alpha_j m_j\right), \quad j = 0, 1, 2, \dots, k, \quad (2.2)$$

其中 G 是传递函数， α_j 是从第 j 个中间层单元到输出单元连接的权重，而 $j=0$ 表示总是等于 1 的偏倚项 m_0 。将 (2.1) 代入 (2.2)，得

$$y = G\left(\alpha_0 + \sum_{j=1}^k \alpha_j F\left(\sum \beta_{ij} x_i\right)\right) = f(X, \theta), \quad (2.3)$$

其中 X 是输入向量， $\theta = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_k, \beta_{01}, \beta_{02}, \dots, \beta_{0k}, \beta_{11}, \beta_{12}, \dots, \beta_{1k}, \dots, \beta_{n1}, \beta_{n2}, \dots, \beta_{nk})'$ 是网络权重向量。 F 和 G 可以取多种函数形式，例如阈值函数（threshold function），函数值为二进制的（ ± 1 ）或（0/1）；S 形（或 logistic）函数，函数值在 0 和 1 之间， $F(a) = G(a) = 1/(1 + \exp(-a))$ 或 $F(a) = a$ （恒等）和 $G(a) = 1/(1 + \exp(-a))$ 。

(2.3) 可以解释为代表所描述的三层前馈神经网络的非线性函数。第 3 节中将阐述，这一表示法包括了许多常见的统计模型，如回归（线性和非线性）、分类（logit, probit）、潜变量模型（MIMIC）、主成分分析以及时间序列分析（ARMA, GARCH）。

用 (2.3) 表示的 ANN 基本结构可以由许多不同方法归纳得到。例如，Poli 和 Jones (1994) 引入一个具有观测噪声和单元间随机连接的多层前馈 ANN。基于噪声的分布假定和连接的随机性，这样的 ANN 可以用卡尔曼过虑过程进行估计，而且已经证实对逻辑图（Logistic Map）生成的混沌时间序列来说，它比牛顿算法具有更高的预测准确度。

2.2. ANN 的学习规则

前一节描述的 ANN 最广泛的估计方法（或者称为学习规则）是误差后向传递(error back propagation) (Rumelhart, Hinton 和 Williams, 1986a, b)，这被认为是二十世纪八十年代中期引发多层神经网络的兴趣再度空前高涨的一个主要原因。Kuan 和 White (1994) 很好地讨论了各种估计方法。后传法是一种递归梯度下降法，它通过向下移动误差曲线的梯度，使得系统的误差平方和达到最小。更明确地说，网络权重向量 θ 的选择将使得损失函数达到最小。

$$\min_{\theta} L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (2.4)$$

其中 N 是样本容量， y_i 是要求的（或目标、实际的）输出值，而 \hat{y}_i 是计算的输出值，

$$\hat{y}_i = f(X_i, \theta) = G \left(\alpha_0 + \sum_{j=1}^k \alpha_j F \left(\sum \beta_{ij} x_{ij} \right) \right), \quad (2.5)$$

梯度下降算法的迭代步长取 θ 到 $\theta + \Delta\theta$ ，而

$$\Delta\theta = -\eta \nabla f(X_i, \theta) (y_i - f(X_i, \theta)) \quad (2.6)$$

其中 $\eta > 0$ 是步幅 (Step Size)，或学习率； $\nabla f(X_i, \theta)$ 是 $f(X_i, \theta)$ 关于 θ (列向量) 的梯度，而链式法则 (chain rule) 用于计算 $\nabla f(X_i, \theta)$ 。

误差曲面是多维的，可能包含多个局部最小值。因此，培育一个网络经常要求以不同的初始权重进行实验，调整学习率，或者增加动力项，以避免陷于局部最优或慢收敛。对于大多数旨在将 ANN 和一些可供选择的模型进行比较的研究，只要 ANN 的表现显著优于对方，就不需要寻求全局最小值。对于那些设法寻求全局最小值的研究，格点搜索是常用方法（例如，参阅 Gorr, Nagin 和 Szczypula, 1994）。其他方法也已提出，如 Bladi 和 Hornik (1989) 发现对应于数据协方差矩阵的第一主成分向量在表面上的投影，误差曲面具有唯一最小值，White, Gallant, Hornik, Stinchcombe 和 Wooldridge (1992) 进一步讨论了全局最优化问题。当达到预定的最大迭代次数或误差目标值，迭代就停止。

2.3. 普遍近似性

ANN 的一个主要的优点是提供输入和输出之间灵活映射的能力。基于 Kolmogorov (1957)，Sprecher (1965)，Lorentz (1976) 和 Hecht-Nielsen (1987, 1990) 的一系列研究，任何连续函数都可以用线性求和以及适当选择的单一非线性函数进行计算。因此，将简单单元安排到一个多层结构中就产生了输入和输出之间的映射关系，它和任何潜在函数关系一致，而不管其“真实”的函数形式。由于输入和输出向量之间具有一般映射关系，故不需要象一般统计和计量建模中那样有不合理的先验约束。

然而，要在输入和输出之间求得一个完美的一般映射关系，就需要正确的传递函数。中间层的 S 形传递函数已经由 Cybenko (1989)，Funahashi (1989)，Hecht-Nielsen (1989)，Hornik 等 (1989) 研究，表明它是合适的。Stinchcombe 和 White (1989) 证明了某些非 S 形函数同样可以使用。这样，ANN 可以被看成是“普遍的近似者”，也就是，给出足够多的中间层单元和适当的调整权重，灵活函数形式可以无限近似任意函数。

3. ANN 和传统统计模型的关系

大部分神经网络的发展主要是由非统计学家取得的。因此，只有很少的统计概念和统计方法应用于这一发展。不过，一些常见的统计模型可以用一般的 ANN 结构来表示，而且许多概念和框架也可以用神经网络的符号来表达（Cheng 和 Titterington, 1994）。另一方面，ANN 可以看成是一类特殊的非线性参数模型，而“学习”对应于模型参数的统计估计。其结果是，非线性模型估计和推断的现代理论可以用于神经网络学习（White, 1989a; Kuan 和 White, 1994）。本节将简要概括 ANN 和一些传统统计方法之间的关系。

3.1. 线性回归

多元线性回归模型可以表示成一个简单的二层前馈网络，它具有线性传递函数 $F(a) = a$ ，Widrow 和 Hoff (1960) 提出的 ADALINE 网络（见图 2），

$$y = \sum_{i=0}^n \beta_i x_i = X'\beta, \quad (3.1)$$

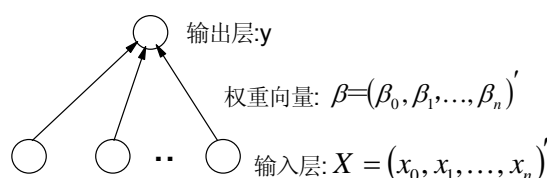


图 2. ADALINE 网络

其中 y 是输出值， $X = (x_0, x_1, \dots, x_n)'$ 是输入向量，而 $\beta = (\beta_0, \beta_1, \dots, \beta_n)'$ 是权重向量。

虽然这样一个网络被证实各种实用领域中非常有用，但是它不能概括或不能很好地表示尚未提出的模式。而且它在计算上比线性回归更难处理。然而，它不需要象线性回归那样对真实数据的形成过程作出同方差性和正交性假定，因此，比经典的线性回归更为稳健。

Widrow 和 Hoff (1960) 的多元适应性线性网络，MADALINE，可以用于表示似无关回归 (Seemingly unrelated regression) 的标准系统 (图 3)：

$$\begin{aligned} y_1 &= \sum_{i=0}^n \beta_{1i} x_i = X'\beta_1, \\ y_2 &= \sum_{i=0}^n \beta_{2i} x_i = X'\beta_2, \\ &\vdots \\ y_k &= \sum_{i=0}^n \beta_{ki} x_i = X'\beta_k, \end{aligned} \quad (3.2)$$

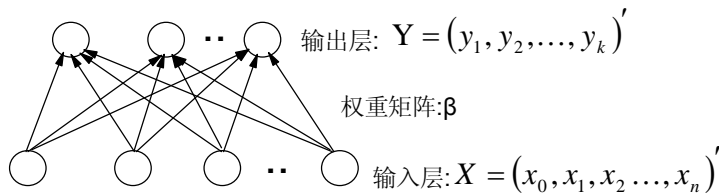


图 3. MADALINE 网络

如果输出的滞后值作为 ADALINE 网络的输入，就得到线性 $AR(d)$ 时间序列方程

$$y_t = \sum_{i=1}^d \beta_i y_{t-i} \tag{3.3}$$

3.2. Logit 和 Probit 模型

在一个具有线性判别传递函数的二层 ADALINE 网络中，直到达到某一临界值水平时单元才有活性，即

$$y = F\left(\sum_{i=0}^n \beta_i x_i\right) \tag{3.4}$$

其中传递函数 $F(a) = 1$ 当 $a > 0$ ， $F(a) = 0$ 当 $a \leq 0$ 。因此输出单元是一个阈单元 (threshold unit)。

具有阈输出单元的网络适用于分类和模式辨别 (Pattern recognition) 问题。由于传递函数 F 可以是任何连续非减函数，所以 F 可以表示成累积分布函数 (cdf)。

当 F 是 logistic 累积分布函数， $F\left(\sum_{i=0}^n \beta_i x_i\right)$ 是常见的二元 logit 模型的条件期望。当 F 是正态累积分布函数， $F\left(\sum_{i=0}^n \beta_i x_i\right)$ 是 probit 模型生成的二元随机变量的条件期望。有关 logit 和 probit 模型更详细的介绍，参阅 Maddala (1983)。

因此，二层神经网络可以表示常见的 logit 和 probit 回归模型，它们涉及到的二元分类或决策，在金融应用中非常流行。但是，由于二层神经网络的局限性，大多数 ANN 的分类应用都使用了一个或者多个中间层。Tam 和 Kiang (1992) 的研究表明，二层 ANN 和线性判别分析的性能非常类似，但是引入了隐蔽层就可以大大提高预测准确度。Ripley (1994) 讨论了更多有关 ANN 和分类的相关方法的研究工作。

3.3. 主成分分析

主成分分析 (PCA) 是进行数据分析的一种常见统计方法，通常用于降低数据矩阵的维数。其目的是在数据空间里寻找一组 m 维正交向量，以尽可能解释数据的变化。特点是 m 小于原始数据的维数，这样，PCA 在保留数据大部分原有信息的基础上，降低了数据矩阵的维数，使得降维后的数据易于处理。更详细的讨论，参阅 Rao (1964)，他研究了在什么情况下，主成分既可以降低数据的维度，又不会损失太多所要寻求的数据信息。

特别地，第一主成分从方差达到最大的方向上取得。第二主成分限制在与第一主成分垂直的子空间中从方差达到最大的方向上取得。第三主成分是与前两个主成分垂直的子空间中从方差达到最大的方向上取得，等等。一般地，第 k 个主成分是沿着完全 (full) 协方

差矩阵第 k 个最大特征值对应的特征向量的方向取得。

几个 ANN 可以实现 PCA (Herz, Grogh 和 Palmer, 1991)。首先考虑一个二层线性前馈网络 (见图 3)，

$$y_j = \sum_{i=1}^n \beta_{ij} x_i = X' \beta_j, \quad (3.5)$$

其中输入向量 $X = (x_1, x_2, \dots, x_n)'$ 是 n 维向量, β_j 是第 j 个输出单元的权重向量。在下面任一学习规则下,

$$\Delta \beta_{ij} = \eta y_j \left(x_i - \sum_{k=1}^j y_k \beta_{ki} \right), \quad (\text{Sanger, 1989}), \quad \text{或} \quad (3.6)$$

$$\Delta \beta_{ij} = \eta y_j \left(x_i - \sum_{k=1}^n y_k \beta_{ki} \right), \quad (\text{Oja, 1989}), \quad (3.7)$$

当达到均衡时, 要求平均权重变化等于 0。这可以表示成

$$\text{mean}(\Delta \beta_j) = C \beta_j - (\beta_j' C \beta_j) \beta_j = 0, \quad (3.8)$$

其中 C 是相关矩阵。因此, 一个均衡的权重向量必须满足

$$C \beta_j = \lambda_j \beta_j, \quad (3.9)$$

于是

$$\lambda_j = \beta_j' C \beta_j = \beta_j' \lambda_j \beta_j = \lambda_j \beta_j' \beta_j \quad (3.10)$$

(3.9) 清楚地表明均衡的 β_j 必须是相关矩阵 C 的特征向量, 而 (3.10) 证明了

$|\beta_j| = 1$ 。同样能够证明 λ_j 是第 j 个最大特征值。

PCA 也可以由一个具有 n 个输入单元、 n 个输出单元和 $m < n$ 个中间层单元的三层线性 ANN, 用自监督后向传递法 (a self-supervised back propagation method) 来实现 (Sanger, 1989)。这一方法的思路是使得目标输出等于输入。当训练集中的输出任意接近于输入时, m 个中间层单元以最先的 m 维主成分的子空间结束。Karhunen 和 Joutsensalo (1995) 推导和讨论了包含非线性的神经 PCA 型学习算法的各种通则。

3.4. 多指标多因素的潜变量模型 (MIMIC 模型)

包含潜变量的因果模型已经广泛应用于社会科学的各个领域, 例如心理学、经济学、教育学等。它们在金融中也有潜在的用途。潜变量是假设的和不可直接观测的变量, 但是和可观测变量存在隐含关系。可观测变量可以是潜变量的果 (“指标”) 或潜变量的因, 或者二者皆是。具有多指标和多因素潜变量的因果模型有时称为 MIMIC 模型。这样一个 MIMIC 模型可以容易地用三层前馈线性 ANN 来表示 (见图 4):

$$M = X' \beta, \quad (4.1)$$

$$Y = M'\alpha, \quad (4.2)$$

其中 β 是输入层和中间层之间连接的权重矩阵， α 是中间层和输出层之间连接的权重矩阵。

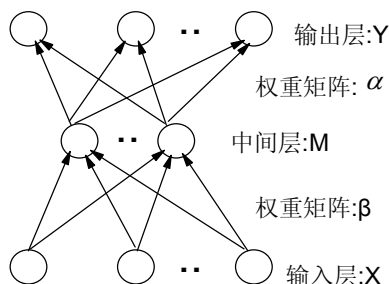


图4. 三层前馈线性ANN

在 (4.1) 中，ANN 的中间层单元， $M = (m_1, m_2, \dots, m_k)'$ （类似于 MIMIC 模型中的潜变量），由 ANN 的输入向量 $X = (x_1, x_2, \dots, x_n)'$ （相当于一组可观测的外生性原因）线性确定。在 (4.2) 中，ANN 的中间层单元线性确定 ANN 的输出单元 $Y = (y_1, y_2, \dots, y_m)'$ ，它相当于一组可观测的内生指标。

在对添加 (4.1) 和 (4.2) 的扰动项作出某些假定，以及对简化型进行某些约束之后，MIMIC 模型可以用最大似然法或某些有限信息方法加以估计（Jöreskog 和 Goldberger, 1975）。虽然培育一个 ANN 的 MIMIC 模型不需要附加约束条件，但是这样一个多层线性网络和二层网络一样具有相同的局限性。只有当输入模式线性独立的时候，它才能运行（Hertz, Grogh 和 Palmer, 1991）。可以表示非线性 MIMIC 模型的多层非线性网络将更引人注目。

4. ANN 的应用和解释

众所周知，神经网络存在着几个方面的局限性可能会限制它的使用。第一，没有正式的理论得以确定最优网络结构，而且输入层和中间层单元的合适个数必须通过实验确定。第二，没有最优算法可以确保得到全局最小值，误差曲面具有多个最小值。第三，ANN 的统计性质通常难以获得，因而不能进行统计推断。第四，培育出的 ANN 模型难以解释。

Cheng 和 Titterington (1994) 概括指出，这些局限性要求在三个主要领域作进一步的研究：（1）现实认知过程的数学建模；（2）网络和神经计算的理论研究；（3）发展进行实际预测和模式辨别的有用工具。虽然前两个领域理所当然地非常重要，但它们不是本文的中心。在这一节，我们将简述一些旨在解决前述局限性的有用的方法和步骤。

4.1. 模型的选择

尽管 ANN 可以是通用的近似形式，但最优网络结构却不是自动确定的。应用中的失败有时出于次优的 ANN 结构。在任何金融应用中，开发一个最优网络，需要（1）识别相关的输入和输出；（2）选择适当的网络结构，包括隐蔽层和隐蔽层单元的必要个数；（3）使用适当的模型评价准则。我们现在逐个阐明这些要点。

4.1.1. ANN 的输入和输出

网络输入和输出变量的选择及数据的质量是 ANN 应用成功的关键。选择在很大程度上取决于 ANN 预期要完成的任务类型，而且或多或少受到建模者关于模型和研究范围辨别力的影响。在实践中，通常用模型的自变量作为网络的输入，而用因变量作为网络的输出。

例如，White (1988) 在一个旨在从经济时间序列中寻求非线性规律的创新性研究中，用 IBM 股票滞后一天的收益率， $r_{t-1}, r_{t-2}, \dots, r_{t-p}$ ，作为网络的输入，而用日期 t 的当日收益率 r_t ，作为网络的输出。这样一个 ANN 的拟合优度，将会就 IBM 股票的日收益率提供支持或者反对有效市场假设以及存在非线性规律的证据。但是，正如作者指出的，为了将研究范围扩展到寻求反对有效市场假设的证据，网络需要精心设计允许额外输入变量，例如交易量、其他股票的价格和交易量、先行指标、宏观经济数据等等。

在 Grudnitski 和 Osburn (1993) 进行的另一个研究中，基于一般经济条件和交易者对期货市场的预期与期货的价格运动有关的认识，他们在 ANN 中使用 24 个输入单元和一个输出单元。输入单元表示六个月输入变量（即，价格变化，价格波动，货币增长率，大投机者、大套期保值者和小交易者 3 种委托的百分比），一次提出四个月的数据。输出是预测月份月中心价格均值的变化。

有时，如果有超乎要求的更多的自变量包括在网络的输入中，则要用到降维方法。我们可以从因变量对一大组自变量的回归中选出较小的一组统计上显著的变量，也可以使用主成分分析和逐步回归方法。例如，Salchenberger, Cinar 和 Lash (1992) 对 29 个金融比率进行了逐步回归，结果筛选出五个变量。然后将这五个金融变量作为神经网络的输入，用于预测储蓄机构倒闭的概率。

为了将输入和输出中量纲的影响减至最小，同时又提高学习算法的有效性，通常对数据集进行标准化处理，使之处于特定的范围，这一范围取决于传递函数。例如，如果一个 ANN 的输出单元具有 S 形或 logistic 传递函数，输出就需要按比例进行处理，使之落在 [0,1] 范围内。否则，落在这一范围之外的目标输出将产生很大的后传误差，而且网络将不能学习包含于特殊培育模式中的输入—输出关系。通常，标准化后的变量具有零均值，单位标准差为 1。

如同任何计量和统计建模一样，数据的质量以及数据集代表总体的适当程度是非常重要的。培育和检验一个 ANN，具有足够的数据也是很重要的。

4.1.2. ANN 的结构

在确定网络的输入和输出层之后，ANN 的结构还是不确定，除非确定了隐蔽层和隐蔽层单元的必要个数。考虑连续值单元的分层网络，隐蔽单元有 logistic 传递函数，输出单元有线性传递函数。大体上，这样一个网络隐含着函数 $y = f(X)$ ，从输入变量

$X = (x_1, x_2, \dots, x_n)'$ ，到输出值 y 。

由于二层 ANN 能力的局限性 (Hertz, Grogh 和 Palmer, 1991)，经常使用有至少一个中间层的网络。Cybenko (1988) 证明了至多有两个隐蔽层的 ANN，若每层有足够单

元，就可以以任意准的准确度近似一组特殊的函数。同样，也已证明只有一个隐蔽层就足以近似任何连续函数（Cybenko, 1989; Hornik, Stinchcombe 和 White, 1989）。许多实证研究，如 Collins, Ghosh 和 Scofield (1988), Dutta 和 Shekhar (1988), Salchenberger, Cinar 和 Lash (1992)（仅仅列举了少数）也证实了这一点。然而，这些结果的正确性，取决于隐蔽单元的适当个数。

隐蔽单元的个数 k 的选择，是一种折衷的表现。如果 k 太小，ANN 不能按要求的准确度近似 $y = f(X)$ 。但是，如果 k 太大，ANN 会过度拟合，而不能推广（或预测）样本外的情况。一个有用的方法是相互确认法（Cross-validation），这种方法选择的中间层单元个数将使得模型在样本外的性能达到最优（White, 1990）。另一个有关的模型选择准则，随机复杂性预测（PSC）（(4.10) 定义）也可以使用（Kuan 和 Liu, 1995）。

Refenes (1995b) 评论了最优网络设计的其他常用方法。这些方法可以分为三种类别。第一种是分析法，即用代数或统计分析来确定先验的隐蔽单元个数。已经提出了几个经验法则，例如，连接的个数不能少于 $0.1T$ ，隐蔽单元的个数是 $(T-1)$ 或 $\log_2 T$ ，这里 T 是样本容量。这种方法的主要问题是进行静态分析，因而只能提供隐蔽单元个数的一个粗略估计。但是，它们还是比网络设计上现行的实验方法好。

第二种类型是结构方法，如级联相关（cascade correlation）（Fahlman 和 Lebiere, 1990），砖瓦结构（tiling）算法（Mezard 和 Nadal, 1989），神经决策树（Gallant, 1986），突跳（upstart）算法（Frean, 1989）和 CLS 过程（Refenes 和 Vithlani, 1991）。这些方法根据需要一个接一个地构筑隐蔽层单元。虽然这些方法保证了网络的收敛，但不能保证普遍性和稳定性。

最后一种类型是网络修剪（network pruning）法，以相反方向运作，修剪网络并去除掉“多余的”或灵敏度最低的连接。这些方法包括网络修剪（Sietsma 和 Dow, 1991）和人工选择（Hergert, Finnoff 和 Zimmermann, 1992）。但是，并不总是可能达到。

4.1.3. ANN 的评价准则

比较可供选择模型的性能并选择一个最佳模型，通常需要确定一个准则。令 $(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N)$ 表示预测值， (y_1, y_2, \dots, y_N) 表示实际值，其中 N 是样本容量。下面列出一些常用的准则。

(1) 均方误差（MSE）和均方根误差（RMSE）：

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (4.1)$$

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (4.2)$$

(2) 平均绝对误差（MAE）和平均绝对百分误差（MAPE）：

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (4.3)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (4.4)$$

(对于有实际值 y_i 等于 0 的样本, MAPE 不存在)。

(3) 可决系数 (R^2)

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}, \quad (4.5)$$

其中 $\bar{y} = \frac{1}{N} \sum y_i$ 。

(4) 皮尔逊 (Pearson) 相关系数 (ρ) : ρ 衡量预测值和实际值之间的线性相关程度,

$$\rho = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum (y_i - \bar{y})^2} \sqrt{\sum (\hat{y}_i - \bar{\hat{y}})^2}}, \quad (4.6)$$

(5) Theil 不等式系数 (U) : U 给出相对于随机游走预测的预测性能,

$$U = \frac{\text{RMSE}}{\sqrt{\frac{1}{N-1} \sum (y_i - y_{i-1})^2}} \quad (4.7)$$

(6) Akaike 信息准则 (AIC) : AIC 调整 MSE 以说明模型的复杂性,

$$\text{AIC} = \text{MSE} \left(\frac{N+k}{N-k} \right), \quad (4.8)$$

其中 k 是模型中自由参数的个数, 或 ANN 中自由权重的个数。

(7) Schwarz 信息准则 (SIC), 或贝叶斯信息准则 (BIC) : SIC 或 BIC 是调整 MSE 以说明模型复杂性的另一种方法,

$$\text{SIC} = \text{BIC} = \ln(\text{MSE}) + \frac{\ln(N)}{N} k \quad (4.9)$$

(8) 预测随机复杂性 (PSC) :

$$\text{PSC} = \frac{1}{N-k} \sum_{i=k+1}^N (y_i - \hat{y}_{ii})^2, \quad (4.10)$$

其中 \hat{y}_{ii} 是基于前 $i-1$ 个观测值参数的预测值。

(9) 方向准确度 (DA) 和混淆比率(confusion rate,CR) :

$$\text{DA} = \frac{1}{N} \sum a_i, \quad (4.11)$$

其中 $a_i = \begin{cases} 1 & \text{当}(y_{i+1} - y_i)(\hat{y}_{i+1} - y_i) > 0, \\ 0 & \text{否则.} \end{cases}$

$$\text{CR} = 1 - \text{DA} \quad (4.12)$$

有时, 需要对可选择模型性能差异的显著性进行检验。T 检验或 Diebold-Mariano 检验 (Diebold 和 Mariano, 1995) 经常用于检验如下零假设, 即两个可选择模型的误差平方没有差异。实际方向和预测方向之间独立性的假设可以用 HM 进行检验 (Henriksson 和

Merton, 1981; Pesaran 和 Timmerman, 1994)。

值得注意的是,任何正确设计和很好培育的 ANN 在样本内的表现,以上述方式评价,往往比相对应的传统统计模型要好得多。了解 ANN 的普遍近似性质之后,这不足为奇。为了避免伪拟合或者过度拟合,用延期(hold-out)样本检验培育的 ANN 是非常重要的,即,使用在培育 ANN 中还未使用的数据来评价培育的 ANN。选择的 ANN 模型有用与否主要取决于其样本外的表现。

Swanson 和 White (1995a, b) 指出,和各种基于样本外预测的模型选择准则,如预测均方误差、预测方向准确度、或基于预测的交易系统获利能力相比较,样本内的 schwarz 信息准则(SIC)并不是样本外表现的可靠导向。假使以样本外表现作为模型选择的准则,用严格未使用的数据集来检验模型也是非常重要的,即,数据在培育和证实中未使用过。否则,样本外预测准确度可能会存在向上偏倚。

4.2. ANN 中的统计推断

很少有关于 ANN 应用的实证研究报告了置信区间或进行假设检验,这是因为 ANN 通常不具备古典的统计性质。然而,如果我们把(2.4)看作非线性最小二乘回归,那么 θ 的估计量就具有非线性最小二乘估计量的统计性质。这样,就可以进行统计推断了。详细请参阅 White (1989a, b), Kuan 和 White (1994)。

Lebaron 和 Weigend (1994) 提出了一种自助(bootstrap)方法,用于确定神经网络预测值的质量和可靠性。虽然这种方法的计算量非常大,但它确实提供了更稳健的预测,以及预测结果的概率分布。在他们对纽约证券交易所日总交易量的多元时间序列预测中,自助方法的结果显示,由培育、相互确认和检验样本之间的不同分离引起的模型性能变差,要显著大于由不同的网络结构和初始权重引起的方差。

4.3. 模型含义

人工神经网络通常被看作“暗箱”,原因是由于函数形式复杂估计模型难以解释。但是,权重、输入和输出之间的关系是清晰定义的,这就允许我们观察这个“暗箱”,并从中发现 ANN 模型的经济含义。沿用 2.1 节图 1 所示三层 ANN 的符号,已经提出几种实用方法用于解释每一个输入变量相对于输出的意义。

(1) 伪权重(pseudo weights)

在一个用 ANN 给看涨期权定价的应用中有五个输入变量, Qi 和 Maddala (1995a) 用输入权重的加权平均数或称为伪权重,来近似一个输入变量对输出的边际贡献。第 i 个输入变量的伪权重定义为

$$PW_i = \sum_{j=1}^k \alpha_j \beta_{ij} = \alpha' \beta_i \quad (4.13)$$

在论文中,他们报告说 PW 的经济含义和看涨期权的性质是一致的。

(2) 输入权重和

由 Sen, Oliver 和 Sen (1995) 提出, Refenes, Zapranis 和 Francis (1995) 采纳了的这一方法用每一个输入变量的输入权重绝对值之和,估计输入变量对结果的影响程度。第 i 个输入变量的输入权重之和(SW)计算如下:

$$SW_i = \sum_{j=1}^k |\beta_{ij}| \quad (4.14)$$

Sen, Oliver 和 Sen (1995) 发现, 在预测公司合并的 logit 分析中显著的所有变量都包括在具有最大输入权重和的五个变量中。

请注意 PW 和 SW 之间的差别。SW 由于取绝对值, 损失了有关输入变量对输出的负效应信息。如果权重都为正, 则 PW 和 SW 对不同输入变量的等级排序应当相同。更为重要的是, Qi (1996) 指出, 当存在实质性的非线性, PW 和 SW 都不再适用, 而模型解释的一个有用工具是灵敏度分析。

(3) 灵敏度分析

灵敏度分析说明的是网络输出对输入变量变化的敏感性。进行灵敏度分析, 首先确定每一个输入变量的最小值、最大值和均值 (或中位数)。每次只改变一个输入变量的值, 而其他输入变量的值保持在它们固定的均值 (或中位数) 上。被动的每一个预测指标, 在它整个取值范围内, 其值按一定数量的相等间隔分布。然后用神经网络模型计算出输出值。神经网络的结果值相对输入变量值绘制的图, 显示当某一特定输入变量变化而其他输入变量固定不变时, 网络输出是如何变化的。这种灵敏度分析已经被 Sen, Oliver 和 Sen (1995) 及 Refenes, Zapranis 和 Francis (1995) 用来增进对其模型的认识。

(4) 敏感性指数 (sensitivity index)

Sen, Oliver 和 Sen (1995) 用敏感性指数考察一个输入变量对输出影响的相对强度。第 i 个输入变量的指数, 是将该输入变量的取值范围分成一定数量 (M) 的相等间隔, 将输出在每一间隔上的变化加以平均:

$$SI_i = \frac{1}{M} \sum_{j=1}^M (\hat{y}_{j+1} - \hat{y}_j) \quad (4.15)$$

敏感性指数提供了衡量输入变量在预测输出上“显著性”的一种方法。Sen 等 (1995) 的结论部分与 logistic 回归一致。

5. 金融应用

ANN 已成功地应用于几个金融领域, 如期权定价、破产预测、汇率预测和股票市场预测。在这一节, 我们将回顾每一领域中某些设计完善、分析仔细的优良实证研究。

5.1. 期权定价

关于神经网络和期权定价已经发表的研究仅仅寥寥数个。期权市场的许多成功和发展可以上溯至创新性的 Black-Scholes 模型和它的扩展。虽然这些参数期权定价公式在它们可以利用的时候受到人们的喜爱, 但是当参数方法失效的时候, 可选择的非参数神经网络就会显示其作用。使用 ANN 已经取得很大的成功。

用神经网络进行期权定价的第一个著名的研究是由 Hutchinson, Lo 和 Poggio (1994) 完成的。首先, 神经网络能从模拟日期权价格的二年培育集 (a two-year training set) 中发现 Black-Scholes 公式的事实, 揭示了神经网络定价公式的潜在价值。期权价格是基于所有 Black-Scholes 假定进行模拟的, 如具有固定均值和波动率的几何布朗运动、固定利率等。结果, 网络公式在样本外期权定价和 delta 套期保值的应用上被证实是成功的。而后, 这个网络被用来对 1987 年—1991 年的 S&P500 期货期权进行定价和 delta 套期保值。结果显示神经网络优于 Black-Scholes 公式。

但是, Hutchinson 等 (1994) 假定存在固定的无风险利率且标的资产具有固定波动率。他们进一步假定标的资产的收益独立于股票价格水平, 因此期权定价公式是资产价格

S 和执行价格 X 的一阶齐次式。这样，他们的网络只有两个输入，S/X 和 T（距到期日的时间），一个输出 C/X，看涨期权价格和执行价格的比值。怀疑这样一个网络是否能捕捉所有期权价格的变动是合情合理的。

另一个用 ANN 进行期权定价的研究是由 Qi 和 Maddala（1995a）完成的。和 Hutchinson, Lo 和 Poggio（1994）不同的是，Qi 和 Maddala 用在期权定价中被认为是重要的变量作为网络的输入，用期权价格作为网络的输出。输入变量包括，标的资产价格（S）、执行价格（X）、无风险利率（r）、距到期日的时间（T）和市场利率（V）。这样一个网络，在 S&P500 指数看涨期权的样本期内和样本期外的表现都优于 Black-Scholes 公式，而且也优于 Hutchinson, Lo 和 Poggio（1994）的报告结果。此外，通过分析网络的权重，Qi 和 Maddala 发现神经网络模型的经济含义和期权价格的性质是一致的，而且发现公开利率是确定期权价格的重要变量。

用 ANN 进行期权定价是发展中的领域。Qi（1995）运用 ANN 研究了看跌——看涨期权的平价关系，发现先前基于传统看跌——看涨期权平价关系的市场无效率的证据可能是夸大的。其他期权数据集和输入变量值得利用 ANN 探索。可以在 Bailey 等（1988）的研究中找到有更多的关于用 ANN 进行期权定价的证据。

5.2. 破产预测

与期权定价相反，人们在破产预测方面已经做了大量的研究工作，使用的标准工具是判别分析（DA）和 logit 模型。在模式匹配的条件下，ANN 的分类和预测能力可以改进相应的传统统计方法。

Tam 和 Kiang（1992）在用 19 个金融比率对 1985 年到 1987 年德克萨斯州银行倒闭进行预测时，将神经网络方法和线性判别分析、logit 模型及其他方法进行了比较。使用 jackknife 方法得到错判率的无偏估计。对原始的后传算法进行修正，使之包括银行倒闭的先验概率和错判成本。修正后的算法允许决策者在 I 型错误（将一个倒闭的银行误判入非倒闭组）和 II 型错误（将非倒闭的银行误判入倒闭组）之间权衡。实证研究结果显示神经网络比其他可选择方法具有更高的预测准确度，而且有一个隐藏层的神经网络优于二层网络。Tam 和 Kiang 还指出在适应性、稳健性和处理多峰分布能力等方面，ANN 为分类技术提供了一个可选择的比较方法。

Salchenberger, Cinar 和 Lash（1992）提出一个用来预测储蓄贷款协会（S&Ls）失败概率的神经网络，用的是能预示机构财务状况恶化的金融变量。与 Tam 和 Kiang（1992）用所有 19 个金融比率作为网络输入不同，Salchenberger 等通过逐步回归将数据维度从 29 降至 5，然后用这五个变量作为网络的输入。尽管如此，结果却是相似的，ANN 的性能和最佳 logit 模型一样好，甚至更好。此外，有些时候，当降低截点(cutoff point)（提高预测失败的概率），I 型错误的减少将伴随着 II 型错误的增加，而 logit 模型比神经网络模型增加得更多。

Altman, Marco 和 Varetto（1994）报告了线性判别分析和 ANN 之间提高准确度及其他有利特性。该研究对 1982 年—1992 年超过 1000 家的健全、脆弱、不健全的意大利工业公司是否处于公司困境进行了判断，建议采用组合方法以精练预测值。

用 ANN 进行破产预测的其他研究还有 Tam 和 Kiang（1990），Odom 和 Sharda（1990），Raghupathi, Schkade 和 Raju（1991），Coats 和 Fant（1992），Huang（1993）和 Podding（1995）。破产预测仅仅是分类问题的一种。其他分类问题还有公司合并预测（Sen, Oliver 和 Sen, 1995），市场反应模型（Dasgupta, Dispensa 和 Ghose, 1994），债券评级（Dutta 和 Shekhar, 1988；Surkan 和 Singleton, 1990；Utans 和 Moody, 1991；Moody 和 Utans, 1995），以及抵押保险（Collins, Ghosh 和 Scofield,

1988)。

5.3. 汇率预测

汇率的不可预测性广为人知。多数不可预测的结论是由线性时间序列方法得出的，因而汇率的线性不可预测性可能应归于线性模型的局限性。自从 1980 年以来，已经发现了非线性的证据。作为一种函数形式灵活的非线性模型，ANN 可以提高预测的准确度。

Kuan 和 Liu (1995) 研究了神经网络在五种货币对美元汇率的样本外预测能力，包括英镑、加拿大元、德国马克、日元和瑞士法郎，数据资料是从 1980 年 3 月 1 日到 1985 年 1 月 28 日纽约外汇市场每日的开盘报价，由 1245 个观测值组成。选择合适网络用的两步骤程序。首先，基于预测随机复杂性 (PSC) 准则 (定义见 4.1.3 节) 选择网络。然后，用递归牛顿算法和非线性最小二乘法估计选择的网络。对日元和英镑而言，ANN 被发现具有显著的市场同步能力，和/或在不同检验期内显著低于随机游走模型的样本外 MSE；然而，对加拿大元和德国马克而言，所选的网络仅有二流的表现。和 Diebold 和 Nason (1990) 的结论相反，结果显示可以利用汇率的非线性来改善点预测和符号预测。同时这一结果也和 Tsibouris (1993) 不同，后者的研究发现 ANN 对汇率变化方向的预测很有用处，但却不便于进行数量大小的预测。

ANN 在汇率上的其他应用还有 Abu-Mostafa (1995)，他报告说具有简单对称点的 ANN 在四个主要外汇市场上的表现在统计上有显著的改善。HsuHsu 和 Tenorio (1995)，用 ANN 来选择预测指标，结果显示预测方向的准确度优于未加工的全体指标。然而，这些研究都没有将模型的性能和基准模型进行比较。虽然他们提供了有用的方法论，但不能作为 ANN 有利于预测汇率的证据。

5.4. 股票市场预测

传统模型，如市场模型、CAPM 和 APT，在增强对股票价格行为的理解上十分有帮助。然而，根据它们在预测股票收益上有限的成功，它们的实际用途往往是有限的。由于 ANN 具有归纳、适应和稳健的性质，大量的研究致力于将 ANN 用于预测股票收益。但迄今为止取得的成功还相当有限。

White (1988) 使用历史数据研究 IBM 股票日收益率的预测能力。虽然发现样本内的拟合 ($R^2=0.175$) 令人惊讶，这和有效市场假设相矛盾，但是样本外实际收益率和预测收益率的相关系数为 -0.0699 (样本内的相关系数是 0.0751)。这一结果并不能为 ANN 的预测能力提供证据，而且目前 ANN 也并不是一个“赚钱的机器”。不过，它有能力捕捉某些股票收益的动态行为。

可是，预测能力问题仍未解决，因为 White 的研究中用的是简单的网络。某些精密的 ANN 结构和学习方法可以改善模型的性能。

Chuah (1993) 用 ANN 预测 1963 年 1 月到 1988 年 12 月纽约证券交易所 (NYSE) 的股票指数收益率，并将网络的预测能力和获利能力与使用相同数据的基准线性模型进行比较。预测能力检验显示网络的预测误差和基准线性模型没有显著的差异，而且网络不具有市场同步能力。获利能力检验考察五年预测期内交易模拟可以产生的收益，并与作为基准的买入持有策略相比较。非线性网络产生 116% 的收益率，而买入持有策略产生 94% 的收益率，线性网络仅产生 38% 的收益率。Qi 和 Maddala (1995b) 对 1959 年 1 月到 1995 年 6 月的 S&P500 指数收益率进行研究也得出了类似的结论。

Refenes, Zapranis 和 Francis (1995) 研究表明在股票收益的多因素动态模型，即 APT 的动态形式中，神经网络是线性回归的上好替代。这一领域的其他研究还有 Kamijo 和

Tanigaw (1990), Schoneburg (1990), Refenes, Zapranis 和 Francis (1994), Haefke 和 Helmenstein (1994, 1995)。在 Trippi 和 Turban (1993) 以及 Refenes (1995a) 的论文中可以找到更多的参考资料。

6. 结论

在本文中，我们简要介绍了 ANN，指出它和一些常见统计模型之间的关系，回顾了某些 ANN 的实际建模方法。我们也评论了金融应用中几个主要领域的实证研究，包括期权定价、汇率预测、破产预测和股票市场预测。

虽然 ANN 在期权定价和分类问题上已经取得了很大的成功，但 ANN 在汇率预测和股票市场预测方面的成果就没有那么引人注目了。当要求拓宽应用范围并发展更好的 ANN 建模方法的时候，也有必要进一步分析改进不力的原因。Ramsey (1995) 指出，开放、非孤立的系统并不总是能被预测的，经济系统封闭和孤立的程度为预测能力提供了实际的限制。不是由于缺乏最优网络结构或学习方法，不可预测的实证证据可能是因为汇率和股票收益是由开放、非孤立经济系统确定的结果。

致谢

我对 G.S.Maddala 有益的讨论和提供本文评述的一些论文表示感谢。我也对 Stephen R.Cosslett, Hongyi Li 和 Yong Yin 帮助我收集论文表示感谢。

参考文献

- Abu-Mostafa, Y. S. (1995). Financial market applications of learning from hints. In: A.-P. refenes, eds., *Neural Networks in the Capital Markets*. John Wiley & Sons, Chichester, 221-232.
- Altman, E., G. Marco, and F. Varetto(1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *J. Banking Finance* 18, 505-529.
- Bailey, D. B., D. M. Thompson and J. L. Feinstein (1988). Option trading using neural networks. In: J. Heralut and N. Giamisas,ed., *Proc. Internat, Workshop on Neural Networks and Their Applications*, Neuro-Nimes, 395-402.
- Baldi, P. and K. Hornik (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks* 2, 53-58.
- Bose, N. K. and P. Liang (1996). *Neural Network Fundamentals with Graphs, Algorithms, and Applications*. McGraw-Hill, New York.
- Cheng, B. and D. Titterington(1994). Neural Networks: A review from a statistical perspective. *Statist. Sci.* 9, 2-54.
- Chuah, K. L. (1993). A nonlinear approach to return predictability in the securities markets using feedforwardneural network. Dissertation, Washington State University.
- Coats, P. and L. Fant (1992). A neural network approach to forecasting financial distress. *J. Business Forecasting* 10, 9--12.
- Collins, E., S. Ghosh and C.Scofield (1988). An application of a multiple neural-network system to

- emulation of mortgage underwriting judgments. Proc. IEEE Internat. Conf. Neural Networks 2, 459-466.
- Cybenko, G. (1988). Continuous valued neural networks with two hidden layers are sufficient. Technical Report, Department of Computer Science, Tufts University, Medford, MA.
- Cybenko, G. (1989). Approximation by superposition of a sigmoid function. Math. of Control Signals and systems 2, 303-314.
- Dasgupta, C. G., G. S. Dispensa and S. Ghose (1994). Comparing the predictive performance of a neural network model with some traditional market response models. Internat. J. Forecasting 10, 235-244.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy, J. Business Econom. Statist. 13, 253-263.
- Diebold, F. X. and J. A. Nason (1990). Nonparametric exchange rate prediction? J. Internat. Econom. 28, 315-332.
- Dutta, S. and S. Shekhar (1988). Bond Rating: A non-conservative application of neural networks. Proc. IEEE Internat. Conf. Neural Networks 2, 443-450.
- Fahlman, S. E. and C. Lebiere (1990). The cascade-correlation learning algorithm. In: D. S. Touretzky, eds. Advances in Neural Information Processing Systems 2. Morgan Kaufmann, San Mateo CA, 525-532.
- Frean M. R. A. (1989). The upstart algorithm: A method for constructing and training feed-forward neural networks. Neural Computation 2, 198-209.
- Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks. Neural Networks 2, 183-192.
- Gallant, S. I. (1986). Three constructive algorithms for neural learning. Proc. 8th Annual Conf. of Cognitive Science Soc.
- Gately, E. (1996). Neural Networks for Financial Forecasting. John Wiley & Sons, New York.
- Gorr, W. L., D. Nagin and J. Szczypula (1994). Comparative study of artificial neural network and statistical models for predicting student grade point average. Internat. J. Forecasting 10, 1-34.
- Grudnitski, G. and L. Osburn (1993). Forecasting S&P and gold futures prices: An application of neural networks. J. Futures Markets 13, 631-643
- Haefke, C. and C. Helmenstein (1994). Stock price forecasting of Austrian initial public offerings using artificial neural networks. Proc. Neural networks Capital Markets.
- Haefke, C. and C. Helmenstein (1995). Predicting stock market averages to enhance profitable trading strategies. Proc. Neural Networks Capital Markets.
- Hecht-Nielsen, R. (1987). Kolmogorov's mapping neural network existence theorem. Proc. IEEE 1st Internat. Conf. Neural Network~ 3, 11-14.
- Hecht-Nielsen, R. (1989). Theory of the back-propagation neural network. Proc. Internat. Joint Conf. Neural Networks, Washington D. C.. IEEE Press, New York, 1, 593-606.
- Hecht-Nielsen, R. (1990). Neurocomputing. Addison-Wesley, MA.
- Henriksson, R. O. and R. C. Merton (1981). On Market timing and investment performance II, Statistical procedures for evaluating forecasting skills. J. Business 54, 513-533.
- Hergert, F., W. Finnoff and H. G. Zimmermann (1992). A comparison of weight elimination methods for reducing complexity in neural networks. Internat. Joint Conf. on Neural Networks, Baltimore, III, 980-987.
- Hertz, J., A. Grogh and R. Palmer (1991). Introduction to the Theory of Neural Computation. Addison-

- Wesley, Redwood City.
- Hornik, K., M. Stinchcombe and H. White (1989). Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359-366.
- Hsu, W., L. S. Hsu and M. F. Tenorio (1995). A neural network procedure for selecting predictive indicators in currency trading. In: A.-P. Refenes, eds., *Neural Networks in the Capital Markets*. John Wiley & Sons, Chichester, 245-257.
- Huang, C. S. (1993) *Neural networks in financial distress prediction: An application to the life insurance industry*. Dissertation, University of Mississippi.
- Hutchinson, J., A. Lo and T. Poggio (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *J. Finance* 99, 851-889.
- Jöreskog, K. G. and A. S. Goldberger (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *J. Amer. Statist. Assoc.* 70, 631-639.
- Kamijo, K.-I. and T. Tanigawa (1990). Stock price recognition - A recurrent neural network approach. *Proc. Internat. Joint Conf. Neural Networks*, San Diego, CA.
- Karhunen, J. and J. Joutsensalo (1995). Generalizations of principal component analysis, optimization problems and neural networks. *Neural Networks* 8, 549-562.
- Kolmogorov, A. N. (1957). On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Dokl. Akad. Nauk USSR* 114, 953-956.
- Kuan, C. and T. Liu (1995). Forecasting exchange rates using feedforward and recurrent neural networks. *J. Appl. Econometrics* 10, 347-364.
- Kuan, C. and H. White (1994). Artificial neural networks: An econometric perspective. *Econometric Rev.* 13, 1-91.
- Lebaron, B. and A. S. Weigend (1994). Evaluating neural network predictors by bootstrapping. University of Wisconsin - Madison, SSRI, Working Paper #9447.
- Lorentz, G. G. (1976). The 13th Problem of Hilbert. *Proc. Symposia Pure Math.*, American Mathematical Society 28.
- Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.
- Mezard, M. and J. Nadal (1989). Learning in feedforward layered network: The tiling algorithm. *J. Physics A* 22, 2191-2203.
- Moody, J. and J. Utans (1995). Architecture selection strategies for neural networks: Application to corporate bond rating prediction. In: A.-P. Refenes, eds., *Neural Networks in the Capital Markets*. John Wiley & Sons, Chichester, 277-300.
- Odom, M. and R. Sharda (1990). A neural network for bankruptcy prediction. *Proc. Internat. Joint Conf. Neural Networks*, San Diego, CA, 2, 163--168.
- Oja, E. (1989). Neural networks, principal components, and subspace, *Internat. J. Neural Systems* 1, 61-68
- Pesaran, M. H. and A. G. Timmerman (1994). A generalization of the non-parametric Henriksson-Merton test of market timing. *Econom. Lett.* 44, 1-7.
- Poddig, T. (1995). Bankruptcy prediction: A comparison with discriminant analysis. In: A.-P. Refenes, eds., *Neural Networks in the Capital Markets*. John Wiley & Sons, Chichester, 311-323.
- Poli I. and R. D. Jones (1994). A neural net model for prediction. *J. Amer. Statist. Assoc.* 89, 117-121.
- Qi, M. (1995). A reexamination of put-call parity on index options: An artificial neural network

- approach. Paper presented at the 3rd ICSA Statistical Conference, Beijing.
- Qi, M. (1996). Applications of generalized nonlinear nonparametric econometric methods (ANNs). Dissertation, The Ohio State University.
- Qi, M. and G. S. Maddala (1995a). Option pricing using ANN: The case of S&P 500 index call options. *Neural Networks in Financial Engineering; Proc. 3rd Internat. Conf. on Neural Networks in the Capital Markets*, London, 78-91.
- Qi, M. and G. S. Maddala (1995b). Economic factors and the stock market: A new perspective. Working Paper, Department of Economics, The Ohio State University.
- Raghupathi, W., L. L. Schkade and B. S. Raju (1991). A neural network approach to bankruptcy prediction. *Proc. IEEE 24th Annual Hawaii Conf. Systems Sciences*.
- Ramsey, J. B. (1995). If nonlinear models cannot forecast, what use are they? Manuscript, New York University.
- Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyaseries A* 26, 329-358.
- Refenes, A.-P. (1995a). eds., *Neural Networks in the Capital Markets*. John Wiley & Sons, Chichester.
- Refenes, A.-P. (1995b). Methods for optimal network design. In: A.-P. Refenes, eds., *Neural Networks in the Capital Markets*. John Wiley & Sons, Chichester, 33-54.
- Refenes, A.-P. and S. Vithlani (1991). Constructive learning by specialization. *Proc. Internat. Conf. Artificial Neural Networks*, Helsinki, Finland.
- Refenes, A.-P., A. D. Zapranis and G. Francis (1994). Stock performance modeling using neural networks: A comparative study with regression models. *Neural Networks* 7, 375-388.
- Refenes, A.-P., A. D. Zapranis and G. Francis (1995). Modeling stock returns in the framework of APT: A comparative study with regression models. In: A.-P. Refenes, eds., *Neural Networks in the Capital Markets*. John Wiley & Sons, Chichester, 101-125.
- Refenes, A.-P., A. D. Zapranis and G. Francis (1994). Stock performance modeling using neural networks: A comparative study with regression models. *Neural Networks* 7, 375-388.
- Ripley, B. (1993). Statistical aspects of neural networks. In: O. E. Barndorff-Nielsen, J. Jensen and W. Kendall, eds. *Networks and Chaos - Statistical and Probabilistic Aspects*. Chapman and Hall, London.
- Ripley, B. (1994). Neural Networks and related methods for classification. *J. Roy. Statist. Soc. SerB* 56, 409-456.
- Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986a). Learning internal representation by error propagation. In: D. E. Rumelhart and J. C. McClelland, ed., *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* 1. MIT Press, Cambridge, 318-362.
- Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986b). Learning internal representation by back-propagating errors. *Nature* 323, 533-536.
- Salchenberger, L., E. Cinar and N. Lash (1992). Neural networks: A new tool for predicting bank failures. *Decision Sciences* 23, 899-916.
- Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks* 2, 459-473
- Schoneburg, E. (1990). Stock price prediction using neural networks: A project report. *Neuro-computing* 2, 17.
- Sen, T. K., R. Oliver and N. Sen (1995). Predicting corporate mergers. In: A.-P. Refenes, eds., *Neural*

- Networks in the Capital Markets. John Wiley & Sons Chichester, 325-340.
- Sietsma, J. and R. F. J. Dow (1991) Creating artificial neural networks that generalize. *Neural Networks* 4, 67-79.
- Singleton, J. and A. Surkan (1991). Modeling the judgment of bond rating agencies: Artificial intelligence applied to finance. *J. Midwest Finance Assoc.* 20, 72-80.
- Sprecher, D. A. (1965). On the structure of continuous functions of several variables. *Trans. Amer. Math. Soc.* 115, 340-355.
- Stinchcombe, M. and H. White (1989). Universal approximation using feedforward networks with non-sigmoid hidden layer activation function. *Proc. Internat. Joint Conf. Neural Networks, San Diego.* IEEE Press, New York, 1,612-617.
- Surkan, A. J. and J. C. Singleton (1990). Neural networks for bond rating improved by multiple hidden layers. *Proc. IEEE Internat. Conf. Neural Networks, San Diego, CA,* 2, 163-168.
- Swanson, N. R. and H. White (1995a). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *J. Business Econom. Statist.* 13, 265-275.
- Swanson, N. R. and H. White (1995b). A model-selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. Working Paper, Department of Economics, Penn State University.
- Tam, K. Y. and Y. M. Kiang (1990). Predicting bank failures: A neural network approach. *Appl Artificial Intelligence* 4, 265-282.
- Tam, K. Y. and Y. M. Kiang (1992). Managerial application of neural networks: The case of bank failure predictions. *Mgmt. Sci.* 38, 926-947.
- Trippi, R. and E. Turban (1993). eds. *Neural Networks in Finance and Investing.* Probus Publishing Company.
- Tsibouris, G. C. (1993). Essays on nonlinear models of foreign exchange. Dissertation, University of Wisconsin-Madison.
- Ujtans, J. and J. Moody (1991). Selecting neural network architectures via the prediction risk: Application to corporate bond rating prediction. *Proc. 1st Internat. Conf. Artificial Intelligence Applications on Wall Street,* IEEE Computer Society Press, Los Alamitos, CA.
- Wasserman, P. (1993). *Advanced Methods in Neural Computing.* Van Nostrand Reinhold, New York.
- White, H. (1988). Economic prediction using neural networks: The case of IBM daily stock returns. *Proc. IEEE Internat. Conf. Neural Networks.*
- White, H. (1989a). Learning in artificial neural networks: A statistical perspective. *Neural Computation* 1, 425-464.
- White, H. (1989b). Some asymptotic results for learning in single hidden-layer feedforward network models. *J. Amer. Statist. Assoc.* 84, 1003-1013.
- White, H. (1990). Connectionist nonparametric regression: Multilayer Feedforward networks can learn arbitrary mappings. *Neural Networks* 3, 535-549.
- White, H., A. R. Gallant, K. Hornik, M. Stinchcombe and J. Wooldridge (1992). eds., *Artificial Neural Networks.' Approximation and Learning Theory.* Blackwell Publishers, Cambridge.
- Widrow, B. and M. E. Hoff (1960). Adaptive switching circuits. *Institute Radio Engineers WESCON Convention Record* 4, 96-104.

