# 半-非参数模型的大样本筛分（Sieve）估计方法*

陈晓红
耶鲁大学经济系
纽黑文，康涅狄格州，美国
电子邮箱: xiaohong.chen@yale.edu

初稿：2002 年 11 月　终稿：2006 年 6 月

**摘要**

在当代经济学研究中，研究者经常会发现参数模型的限制过于严格，一旦模型设定不准确就容易出现严重的估计偏误。相比之下，半参数模型更加灵活和稳健；但它也有潜在的问题，例如可能会引入非紧凑的无限维参数空间，或者优化问题不再适定。面对这些问题，筛分（Sieve）方法通过在一系列近似参数空间（即 sieves）上最大化经验目标函数提供了一种简便的解决方案。这种方法的核心在于：筛近似空间相比较原参数空间更简单易于分析，同时筛近似空间在原参数空间上是稠密的，这就使得上述优化问题从不适定变成适定的。通过选择不同的筛空间和经验标准，筛分方法在估计具有（或不具有）内生性以及潜在异质性的复杂半非参数模型时非常灵活，具有广泛的适用范围。筛分方法可以很简便地整合来自经济学理论的先验信息和条件约束，例如单调性、凸性、可加性、可乘性、排除性和非负性等。此外，筛分方法可以同时估计半非参数模型中的参数部分和非参数部分，并且在多数情况下这两部分的筛估计量都可以达到最优收敛速度。

本文描述了如何使用筛分方法估计半非参数计量模型。我们介绍了关于筛分方法估计量的大样本特征的一般结果，包括筛分方法极值估计量的一致性，筛分方法 M-估计量的收敛速度，回归函数的级数估计的逐点正态性，无限维未知参数（函数）的平滑泛函的筛分估计量的 $\sqrt{n}$ - 渐近正态特征和有效性等。本文提供了具体实例来阐述这些结果。

关键词: 筛极值估计法，级数，筛最小距离，半参数两步估计法，半非参数模型下的内生性

JEL: C13, C14, C20.

*作者感谢 C. Ai, J. Heckman, B. Honore, J. Huang, G. Imbens, R. Matzkin, W. Newey, J. Powell 以及 H. White 提出的宝贵建议，感谢 J. Huang 分享他在凹拓展线形模型上的研究，同时感谢两位匿名审稿人提出的修改意见。作者感谢 K. Hyndman, A. Ingster, M. Kredler, D. Pouzo 和 R. Sela 的校对工作，M. Garibotti, D. Pouzo 和 V. Tsyrennikov 的数据模拟工作，以及其他于 2002 年秋季、2003 年秋季、2005 年春季和秋季在纽约大学参加过“计量经济学前沿讲座”并阅读过本文早期版本的博士研究生。本文获得了来自美国国家科学基金会和纽约大学 C.V. Starr Center 的资助。文责自负。

# 目录

# 1 前言

在理论和应用计量经济学中，半参和非参建模技术越来越受到研究者的关注。[1] 究其原因，部分来自于经济学理论很少给出变量之间的函数关系，也不会设定残差项的分布密度。半非参数模型越来越受欢迎的另一个原因是随着时代的进步收集和分析大型经济数据集的计算成本在下降。Barnett et al. (1991) 中的全部章节和 Engle and McFadden (1994) 编写的计量经济学手册（Handbook of Econometrics）第四卷中的部分章节已经总结回顾了截至上世纪九十年代中期计量经济学在半参和非参模型上的进展。[2]更近一点的研究包括 Horowitz（1998），他使用核方法 (Kernel) 分别估计了四类具有代表性的的半参数计量经济学模型。此外，Pagan and Ullah (1999), Haerdle et al. (2004) 和 Li and Racine (2006) 系统总结了当前影响较大的应用核方法、局部线性回归和级数估计方法来实现半参和非参计量模型的估计和检验的相关理论和实证研究。本文将回顾近年来关于使用筛分方法（Sieve）估计半非参数模型的大样本理论的一些研究进展 (Grenander, 1981)。

半非参数模型涉及无限维参数空间中的未知参数（函数）。因此，使用有限样本估计这类模型在计算上往往会非常困难。此外，即使可以解决在无限维参数空间上最大化经验目标函数的问题，所得到的估计量也很可能会具有例如不一致和/或非常慢的收敛速度的劣大样本性质。这是因为在无限维非紧凑空间上的优化问题可能不适定。为了解决这个问题，筛分方法在一系列显著简化的有限维参数空间（称之为 “筛”，sieves）上求解目标函数最大化问题。为了确保得到一致的统计量，我们要求随着样本量的增长这些筛也愈加复杂，这样一来当样本量足够大时筛在原始参数空间上是稠密的。[3]

非参数或半参数模型中的无限维未知参数通常可以被视为具有特定性质（如二阶导数有界、单调、凹函数等）的某函数空间的成员。因此，我们可以借助许多在数学和计算机科学中已经开发出来的具有确定性的近似方法来挑选合适的、简便易算的、更接近未知目标函数的筛（sieves）。例如，我们可以使用基于幂级数、傅立叶级数、样条函数或其他基函数的线性生成空间来构造筛或近似空间；参考 Judd (1998，第 6、12 章) 借助筛分方法得到经济和金融问题的数值解。由于我们可以用有限维参数来刻画这些已知近似空间，在使用筛分法估计非参或半参模型时本质上是将无限维未知参数估计问题转化为有限维参数估计问题。然而，为了获得估计量的理想理论性质，用来定义近似空间的未知参数个数的增长速度不能过快，需要随样本量上升而缓慢增加。正是这个特征使得筛分法超越了采用固定有限维参数空间的经典参数方法，具有更好的灵活性和稳健性。

筛分方法的一大优势在于非常易于实施。当未知函数非线性地进入目标函数（或矩条件），或满足一些已知的限制，例如单调性、凹性、可加性、可乘性或排除性时，或者当研究者已经知道误差分布满足某些尾部特征（如长尾）时，筛分方法应用起来就更加简便。通过选择不同的目标函数和筛，筛分方法提供了一种灵活可计算的适用于具有（或不具有）特定约束、内生性和潜在异质性的复杂半非参数模型的估计方法。此

[1]在本文中，如果一个经济学模型中的所有未知参数的取值范围（即参数空间）是有限维的，则称该经济学模型为 “参数模型”；如果其未知参数的参数空间是无限维的，则称之为 “非参数模型”；如果我们感兴趣的参数的取值范围是有限维的，但多余参数的参数空间是无限维的，则称之为 “半参数模型”；如果我们感兴趣的参数中有一部分是有限维的，而另外一部分是无限维的，则称该模型为 “半非参数模型”。

[2]参阅 Newey and McFadden (1994), Andrews (1994a), Powell (1994), Hardle and Linton (1994), Matzkin (1994), Manski (1994) 及其他相关文章。

[3]这些术语在之后的两个章节中会更加清楚。

外，它可以同时估计半非参数模型中的参数部分和非参数部分，并且估计量通常可以达到两部分的最优收敛速度。我们将在后续章节中举例说明这一点。

虽然筛分方法易于实现，筛分估计量也通常具有理想的大样本特性，但其理论性质不能通过直接应用现有的经典参数模型理论来证明。任何关于筛分法的大样本理论都不仅应该考虑由于我们用更简单的筛分空间替代原始参数空间所导致的近似误差，还要注意控制随着样本量变大而逐渐增长的筛分参数空间的复杂程度。因此，筛分方法的大样本性质通常难以证明，这可能部分地解释了为什么当前使用这种技术的计量经济学应用比使用核方法的更少。然而，我们应该注意到，筛分估计方法同许多计量经济学的标准估计方法（如基于级数的方法）自洽，后者往往是筛分法的某种特殊形式。因此，在以往文献中其实已经出现过一些关于筛分方法大样本性质的结果，却并没有提到“筛”这个词。

在本文中，我们将介绍关于筛分法的大样本估计理论的一般结果，并举例说明如何应用这些结果。受篇幅所限，我们只选择一部分相对容易理解又不失一般性的半非参计量应用来介绍和总结筛分方法的相关理论。对于文中没有详细介绍的理论结果，文末给出了参考文献。

本文的其余部分安排如下。在第 2 节中，我们首先介绍半非参数计量经济学模型的几个例子。然后定义筛分极值估计及几种特殊情况，包括筛分 M 估计、筛分极大似然估计（MLE）、筛分广义最小二乘法（GLS）、筛分最小距离（MD）等。本文使用范例来具体说明各种目标函数。此外，我们介绍受到普遍关注的“级数”估计量，它是当目标函数为凹形且筛空间为有限维线性时所获得的筛分 M 估计量。[4] 此后，我们将讨论计量经济学中用到的经典函数空间和筛分空间，并通过一个蒙特卡洛研究来展示具体如何实现筛极值估计。[5]第 3 节主要介绍无限维未知参数的筛分估计的大样本特性。我们首先为一般的筛分极值估计法给出全新的当原始参数空间不紧凑、问题本身可能不适定时的一致性定理。该定理在两个层面上意味着筛分 M 估计量和筛分 MD 估计量均满足一致性特征。然后，我们给出关于筛分 M 估计量的收敛速度的结果，并通过实例说明如何应用这一结果。我们还回顾了级数估计量的收敛速度和逐点渐近正态性结果。在第 4 节中，我们介绍了关于未知无限维参数的平滑泛函的筛估计量的 $\sqrt{n}$ - 渐近正态性的一般结果，其中 $n$ 表示样本大小。我们首先讨论流行的半参数模型两步估计法，其中第一步可以通过任何非参方法（如核、局部线性回归和筛分方法）估计未知函数（大多是无限维），第二步通过广义矩方法（GMM）来估计有限维未知参数。值得注意的是，上述第二步中 GMM 估计量的 $\sqrt{n}-$ 渐近正态性定理同现有半参估计理论有些许不同。接下来，我们回顾了未知函数的平滑泛函的筛 M 估计的渐近正态性以及筛 MLE 估计的半参有效性。最后，我们介绍了近期关于半非参数条件矩模型中有限维参数部分的筛分 MD 估计的理论；在这类模型中未知函数可以取决于内生变量。第 5 节介绍了有关使用筛分方法做统计推断的其他专题，由于篇幅所限这些专题未能在本文中进行详细讨论。

在本文中我们一直假设存在一个完备的概率空间，数据 $\{Z_t = (Y_t', X_t')' : t \geq 1\}$ 是严格平稳遍历的，[6] 同时所有的概率计算都是在真实概率测度 $P_o$ 下完成的。对于随机变量 $V_n$ 以及正数 $b_n$, $n \geq 1$, 我们定义 $V_n = O_P(b_n)$ 为 $\lim_{c\to\infty} \limsup_n P(|V_n| \geq cb_n) = 0$, 定义 $V_n = o_P(b_n)$ 为对所有 $c > 0$, $\lim_n P(|V_n| \geq cb_n) =$

[4]这个级数估计量的定义与目前的计量经济学文献略有不同。

[5]关于如何构造非半参统计量的更多细节请参考 Ichimura and Todd (2006)。

[6]在本章中，符号 $'$ 表示向量转置。参见 Hansen（1982），White（1984）或 Wooldridge（1994）对于严格平稳遍历过程的定义。我们做出这个假设来简化文章中的表达。参见 White 和 Wooldridge（1991）关于一般从属异构过程的筛分极值估计法。

0 。表达式 $\text{plim}_{n\to\infty}V_n = 0$ 同时意味着 $V_n = o_P(1)$ (也就是说, $V_n$ 依概率收敛到 0 )。类似的, $V_n = o_{a.s.}(1)$ 意味着 $V_n$ 几乎处处收敛到 0。对于两列正数 $b_{1n}$ 和 $b_{2n}$, 表达式 $b_{1n} \asymp b_{2n}$ 意味着 $b_{1n}/b_{2n}$ 上、下均有界, 且这些界独立于 $n$。

# 2 筛分估计方法：例子，定义和筛

正如引言中所提到的，筛分方法有两个关键组成部分：目标函数和筛参数空间（即一系列近似空间）。目标函数和筛参数空间都可以很灵活。特别是，几乎所有在 Newey 和 McFadden（1994）中提到的经典目标函数，只要满足可识别 (Identification) 的要求就可以被应用在筛分估计方法中。因此，筛分方法区别于传统方法主要在于筛参数空间的选择上，我们将在本节对此进行讨论。

## 2.1 半非参数计量经济学模型的实证范例

在计量经济学中，我们难以列出所有现有和潜在的半非参数模型及其在实证问题中的应用。在本小节中，我们主要介绍三个实证方面的例子；其他例子可以在 Manski (1994)，Powell (1994)，Matzkin (1994)，Horowitz (1998)，Pagan and Ullah (1999)，Blundell and Powell (2003) 等文章以及相关专题下的其他研究中找到。

**例 2.1. （具有不可观察异质性的单期久期模型）** 经济学中关于求职性失业（Flinn and Heckman，1982），工作岗位流转（Jovanovic，1979），劳动力供给（Heckman and Willis，1977）及其他相关研究的经典单期模型通常意味着（给定个体异质性）久期的结构化条件概率分布服从某种具体的函数形式。具体而言，用 $G(\tau|u,x)$ 表示给定一维不可观测的异质性 $U = u$ 和多维可观测的异质性 $X = x$ 下久期 $T$ 的结构化分布函数。给定 $X = x$，可观测到的持续时间的分布如下：

$$F(\tau|x) = \int G(\tau|u,x)dh(u),$$

这里我们将不可观测的一维异质性 $U$ 作为一个具有分布函数 $h(\cdot)$ 的随机因子建模。我们通过一系列相互独立且同分布的观测 $\{T_i, X_i\}_{i=1}^n$ 来确定 $F(\tau|x)$。理论模型通常会给出 $G$ 的取决于有限维未知 $\beta$ 的参数函数形式。用 $g(\cdot|\beta,u,x)$ 表示 $G(\cdot|\beta,u,x)$ 的概率密度函数。传统的参数最大似然估计方法假设不可观测的异质性服从某基于未知有限维参数 $\gamma$ 的概率分布 $h_\gamma$。在该假设下，这一方法通过 $\arg\max_{\beta,\gamma}\frac{1}{n}\sum_{i=1}^n \log\{\int g(T_i|\beta,u,X_i)dh_\gamma(u)\}$ 来估计上述未知参数 $\beta,\gamma$。

Heckman and Singer(1984) 指出，理论和实例研究都表明在这类模型中如果错误地设定了无法直接观测的异质性 $u$ 的分布，那么结构化参数 $\beta$ 的极大似然估计将会是不一致的。为解决这一问题，他们提出以下半非参数单期持续时间模型

$$F(\tau|\beta,h,x) = \int G(\tau|\beta,u,x)dh(u), \tag{2.1}$$

这里并不假设不可观测的异质性 $U$ 服从某种特定的概率分布。Heckman and Singer (1984) 给出了 $(\beta',h)$ 的可识别性结果，并提出使用筛分极大似然估计法来同时估计 $(\beta',h)$。他们的研究还表明这一估计是一致

的。

Heckman-Singer 模型是一大类半非参数模型的典型范例，这类模型以半非参形式设定可被观测到的经济变量的条件分布。研究者可以通过假设误差项与自变量之间互相独立来得到上述半非参条件分布；这类假设经常出现在包括离散选择模型、转换模型、样本选择模型、混合模型、数据随机删失、非线性测量误差等模型中。更一般地，我们可以考虑基于分位数独立性、对称性约束或其他对分布的定性限制的半非参数模型。关于这方面的研究可以参考 Horowitz (1998)，Manski (1994)，Powell (1994) 和 Bickel et al. (1993) 等文。

**例 2.2.（恩格尔曲线的形状不变系统）** Blundell et al. (2003) 已经证明，满足 Slutsky 对称性条件并允许在给定年份让人口因素影响预算支出份额的恩格尔曲线系统必须采取以下形式：

$$Y_{1\ell i} = h_{1\ell}(Y_{2i} - h_0(X_{1i})) + h_{2\ell}(X_{1i}) + \varepsilon_{\ell i}, \quad \ell = 1, ..., N,$$

上述公式中 $Y_{1\ell i}$ 是第 $i$ 个家户在 $\ell$ 商品上的预算份额, $Y_{2i}$ 是第 $i$ 个家户的总非耐用消费品支出的自然对数，$X_{1i}$ 是一个由家户 $i$ 的可能影响其非耐用品支出的人口因素所组成的向量。这里需要注意 $h_0(X_{1i})$ 并不随商品种类变化而改变，在消费者需求文献中这被统称为等价量表。通过引用大量实证证据和许多现有研究，Blundell et al. (2003) 发现通行的 $h_{1\ell}(\cdot)$ 满足线性或二次型的假设是不充分的，并且现有消费者需求理论对 $h_{1\ell}(\cdot)$ 的形式仅提出了以下纯非参数模型设定：

$$E[Y_{1\ell i} - \{h_{1\ell}(Y_{2i} - h_0(X_{1i})) + h_{2\ell}(X_{1i})\}|X_{1i}, Y_{2i}] = E[\varepsilon_{\ell i}|X_{1i}, Y_{2i}] = 0, \tag{2.2}$$

这里 $h_{1\ell}$, $h_{2\ell}$ 以及 $h_0$ 全都是未知函数。为了识别所有满足 (2.2) 的未知函数 $\theta = (h_0, h_{11}, ..., h_{1N}, h_{21}, ..., h_{2N})'$，我们可以只假设在 $h_{1\ell}, \ell = 1, ..., N$ 中，至少有一个函数是非线性的，同时假设在 $X_1$ 的支集中存在某 $x_1^*$ 满足 $h_{2\ell}(x_1^*) = 0$, $\ell = 1, ..., N$。

尽管理论给出了识别结果，但是当 $X_{1i}$ 包含太多的家户人口变量时（例如，$\dim(X_{1i}) \geq 3$），对未知函数 $h_0, h_{21}, ..., h_{2N}$ 的非参估计会遇到“维度诅咒”的问题，导致估计结果不准确。因此，实证研究人员必须给模型添加更多的结构。使用英国家庭支出调查（FES）数据，Blundell et al. (1998) 发现以下半非参数设定比较合理：

$$E[Y_{1\ell i} - \{h_{1\ell}(Y_{2i} - g(X_{1i}'\beta_1)) + X_{1i}'\beta_{2\ell}\}|X_{1i}, Y_{2i}] = 0, \tag{2.3}$$

这里 $h_{1\ell}$, $\ell = 1, ..., N$ 仍然是未知函数，但是现在 $h_0(X_{1i}) = g(X_{1i}'\beta_1)$，$h_{2\ell}(X_{1i}) = X_{1i}'\beta_{2\ell}$ 两函数都只取决于有限维未知参数 $\beta_1$ 和 $\beta_{2\ell}$，而不取决于任何其他未知变量。这时我们关注的参数为 $\theta = (\beta_1', \beta_{21}', ..., \beta_{2N}', h_{11}, ..., h_{1N})'$。Blundell et al (1998) 使用核方法估计了上述半非参模型；此外，Blundell et al (2001) 还使用筛分方法再次估计了这一模型。

上面两种模型设定 (2.2) 和 (2.3) 均假设非耐用品总支出 $Y_{2i}$ 是外生的。然而，实证研究显示这一假设并不成立。考虑到非耐用品总支出的内生性，Blundell et al. (2001) 考虑了以下半非参数工具变量（IV）回归：

$$E[Y_{1\ell i} - \{h_{1\ell}(Y_{2i} - g(X_{1i}'\beta_1)) + X_{1i}'\beta_{2\ell}\}|X_{1i}, X_{2i}] = 0, \tag{2.4}$$

这里我们关心的参数仍然是 $\theta = (\beta_1', \beta_{21}', ..., \beta_{2N}', h_{11}, ..., h_{1N})'$，同时 $X_{2i}$ 是第 $i$ 个家户户主的总收入，它被用作非耐用品总支出 $Y_{2i}$ 的工具变量。他们通过筛分方法重新估计了上述支出模型，并以实证结果证明了使用半非参模型处理总支出内生性问题的重要性。

**例 2.3 （基于消费理论的资产定价模型）**基于消费理论的标准资产定价模型假定在零时刻，消费者最大化其总效用函数的预期现值 $E_0\{\sum_{t=0}^{\infty}\delta^t u(C_t)\}$，这里 $\delta$ 是时间折现因子，$u(C_t)$ 是 $t$ 期的效用。基于消费理论的资产定价模型来源于代表性消费者最优消费选择问题的一阶条件。这些一阶条件为消费和资产回报的跨期边际替代比率的联合分布赋予了特定约束。这些条件表明，对任何在 $t+1$ 期具有总回报率 $R_{\ell,t+1}$ 的可交易资产 $\ell$ 而言，下面的欧拉（Euler）方程成立：

$$E\left(M_{t+1}R_{\ell,t+1} \mid \mathbf{w}_t\right) = 1, \qquad \ell = 1, ..., N, \tag{2.5}$$

方程中 $M_{t+1}$ 表示消费的跨期边际替代率，$E(\cdot|\mathbf{w}_t)$ 表示给定 $t$ 期信息集（由 $\mathbf{w}_t$ 生成的 $\sigma-$ 域）的条件期望。更一般地，满足方程（2.5）的任何非负随机变量 $M_{t+1}$ 统称为随机折现因子（SDF）；参考 Hansen and Richard (1987) 和 Cochrane (2001)。

Hansen and Singleton（1982）假定 $t$ 期的效用服从指数形式 $u(C_t) = [(C_t)^{1-\gamma} - 1]/[1-\gamma]$，其中 $\gamma$ 是每个时期效用函数的曲率参数，这意味着随机折现因子满足 $M_{t+1} = \delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}$；欧拉方程为:

$$E\left(\delta_o\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_o} R_{\ell,t+1} - 1 \mid \mathbf{w}_t\right) = 0, \quad \ell = 1, ..., N, \tag{2.6}$$

可以使用 Hansen (1982) 提出的广义矩估计方法 (GMM) 估计未知参数 $\delta_o, \gamma_o$。然而，实证上已经拒绝了这种经典的基于指数效用函数的资产定价模型 (2.6)。

许多后续论文试图通过引入耐用品、习惯形成或不可分割的偏好设定来放宽模型 (2.6) 对指数效用函数的依赖，以更好地适应真实数据。第一类这方面的尝试提出比 $M_{t+1} = \delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}$ 更灵活的参数型随机折现因子。参见 Eichenbaum and Hansen (1990), Constantinides (1990), Campbell and Cochrane (1999)。第二类文章则通过数个状态变量的非参函数来刻画随机折现因子 $M_{t+1}$。这方面研究可以参见 Gallant and Tauchen (1989), Newey and Powell (1989), 以及 Bansal and Viswanathan (1993)。最近，Chen and Ludvigson (2003) 通过半参方式设定 $M_{t+1}$ 以方便加入一些偏好参数。具体而言，他们的研究将指数效用函数设定与非参数内部习惯形成相结合：$E_0\{\sum_{t=0}^{\infty}\delta^t[(C_t - H_t)^{1-\gamma} - 1]/[1-\gamma]\}$。这里 $H_t = H(C_t, C_{t-1}, ..., C_{t-L})$ 是 $t$ 期的习惯水平。$H(\cdot)$ 是一个当期和过去消费水平的一阶齐次未知函数，并可以表达为 $H(C_t, C_{t-1}, ..., C_{t-L}) = C_t h_o\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)$，注意这里 $h_o(\cdot)$ 是未知的。很明显需要加入下列条件：$0 \le h_o(\cdot) < 1$ 以满足 $0 \le H_t < C_t$。以下外部习惯设定是 Chen and Ludvigson（2003）模型的一种特殊情况：

$$E\left(\delta_o\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_o} \frac{\left(1 - h_o\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right)\right)^{-\gamma_o}}{\left(1 - h_o\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)\right)^{-\gamma_o}} R_{\ell,t+1} - 1 \mid \mathbf{w}_t\right) = 0, \tag{2.7}$$

对 $\ell = 1, ..., N$，这里 $\gamma_o > 0$, $\delta_o > 0$ 是未知的一维偏好参数，$h_o(\cdot) \in [0, 1)$ 是未知函数，$H_{t+1} = C_{t+1} h_o\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right)$ 是 $t+1$ 期的习惯水平。Chen 和 Ludvigson（2003）应用筛分方法估计该模型

及其扩展应用，在这些扩展中研究允许未知形式的内部习惯形成。他们使用季度数据做了实证研究，其结果支持了“灵活的非线性内部习惯形成”假说。

**半非参条件矩模型** 我们注意到，例 2.2 和 2.3 以及许多其他经济模型都包含以下形式的半非参数条件矩约束

$$E[\rho(Z_t;\theta_o)|X_t]=0,\quad \theta_o\equiv(\beta_o',h_o')', \tag{2.8}$$

其中，$\rho(\cdot;\cdot)$ 是一个残差项函数的列向量，其函数形式取决于未知参数 $\theta\equiv(\beta',h')'$；$\{Z_t'=(Y_t',X_t')\}_{t=1}^n$ 是观测数据，$Y_t$ 是内生变量向量，$X_t$ 是条件变量向量。这里 $E[\rho(Z_t,\theta)|X_t]$ 表示给定 $X_t$ 时 $\rho(Z_t,\theta)$ 的条件期望，同时模型假定给定 $X_t$ 时 $Y_t$ 的真实条件分布未知（作为多余函数处理）。我们关注的参数 $\theta_o\equiv(\beta_o',h_o')'$ 包括有限维未知参数向量 $\beta_o$ 和无限维未知函数 $h_o(\cdot)=(h_{o1}(\cdot),...,h_{oq}(\cdot))'$，其中 $h_{oj}(\cdot)$ 的自变量可以取决于 $Y$，$X$，已知的指标函数 $\delta_j(Z,\beta_o)$（除了 $\beta_o$ 未知），其他未知函数 $h_{ok}(\cdot)$（$k\neq j$），或者未被观测的随机变量。受资产定价和理性预期模型的启发，Hansen（1982，1985）为平稳遍历时间序列数据（通常 $Z_t'=(Y_t',X_t')$，并且 $X_t$ 包括 $Y_t$ 的滞后项及其他在 $t$ 期已知的变量）研究了条件矩限制 $E[\rho(Z_t;\beta_o)|X_t]=0$（不包含未知的 $h_o$）Chamberlain (1992), Newey and Powell (2003), Ai and Chen (2003) 以及 Chen and Pouzo (2006) 为独立同分布数据研究了一般性的条件矩限制 $E[\rho(Z_t;\beta_o,h_o)|X_t]=0$。

由（2.8）给出的半非参数条件矩模型可以分为两个广泛的子类。第一个子类由“无内生性模型”组成，意思是 $\rho(Z_t,\theta)-\rho(Z_t,\theta_o)$ 不依赖于任何内生变量（$Y_t$）；因此我们可以通过最大化总体目标函数 $Q(\theta)=-E[\rho(Z_t,\theta)'\{\Sigma(X_t)\}^{-1}\rho(Z_t,\theta)]$（$\Sigma(X_t)$ 是一个正定加权矩阵）来识别真实参数 $\theta_o$。第二个子类由“含内生性模型”组成，意思是 $\rho(Z_t,\theta)-\rho(Z_t,\theta_o)$ 取决于内生变量（$Y_t$）。这时可以通过最大化下列目标函数来识别真实参数 $\theta_o$：

$$Q(\theta)=-E[m(X_t,\theta)'\{\Sigma(X_t)\}^{-1}m(X_t,\theta)]\quad \text{其中}\quad m(X_t,\theta)\equiv E[\rho(Z_t,\theta)|X_t].$$

尽管第一个子类可以被视为第二个子类的一种特殊情况，但当 $\theta$ 包含未知函数时，为第一子类条件矩模型中识别的 $\theta$ 的非参数估计量推导各类大样本渐近性质要相对容易得多。第一个子类包括许多在计量经济学中得到广泛关注的半非参回归模型，比如例 2.2 中的模型设定 (2.2) 和 (2.3) 就属于这一子类。此外，Engle et al. (1986) 及 Robinson (1988) 中的半线性回归模型 $E[Y_i-X_{1i}'\beta_o-h_o(X_{2i})|X_{1i},X_{2i}]=0$，Powell et al. (1989), Ichimura (1993) 和 Klein and Spady (1993) 的指数回归模型 $E[Y_i-h_o(X_i'\beta_o)|X_i]=0$，Chen and Tsay (1993), Cai et al. (2000) 和 Chen and Conley (2001) 中的可变系数模型 $E[Y_i-\sum_{j=1}^q h_{oj}(D_{ji})X_{ji}\mid(D_{ki},X_{ki}),k=1,...,q]=0$，以及 Horowitz and Mammen (2004) 的带有已知关联函数 ($F$) 的可加模型 $E[Y_i-F(\sum_{j=1}^q h_{oj}(X_{ji}))|X_{1i},...,X_{qi}]=0$ 也都属于第一个子类。

第二个子类则包括例 2.2 中的模型设定（2.4），例 2.3，半非参数资产定价和理性预期模型，以及具有灵活的参数设定的联立方程模型等。这个子类的一个具有代表性同时具有一定难度的例子是纯非参数工具变量 (IV) 回归 $E[Y_{1i}-h_o(Y_{2i})|X_i]=0$，这类模型在 Newey and Powell (2003), Darolles et al. (2002), Blundell et al. (2001), Hall and Horowitz (2005) 以及 Carrasco et al. (2006) 中都有研究过。一个更困难的例子是非参数工具变量分位数回归问题 $E[1\{Y_{1i}\leq h_o(Y_{2i})\}-\gamma|X_i]=0$，其中已知 $\gamma\in(0,1)$，Chernozhukov et al. (2006), Horowitz and Lee (2006) 以及 Chen and Pouzo (2006) 研究了这类分位数回归模型。更多的例子请

参考 Blundell and Powell (2003), Florens (2003), Newey and Powell (1989), Carrasco et al. (2006) 和 Chen and Pouzo (2006) 等文章。

## 2.2 筛分极值估计的定义

### 2.2.1 不适定的问题与适定的问题，筛分极值估计

用 $\Theta$ 表示一个具有（伪）度量 $d$ 的无限维参数空间。典型的半非参数计量经济学模型设定有一个总体目标函数 $Q:\Theta\to\mathcal{R}$，并设定仅在（伪）真实参数 $\theta_o\in\Theta$ 下该目标函数取得最大值。[7]我们通过识别计量模型来选择 $Q(\cdot)$ 同时得到 $\theta_o$ 的存在性。（伪）真实参数 $\theta_o\in\Theta$ 是未知的，但同联合概率分布 $P_o(z_1,\cdots,z_n)$ 有关。基于这个分布可以得到一个样本量为 $n$ 的观测数据 $\{Z_t\}_{t=1}^n$，$Z_t\in\mathcal{R}^{d_z}$，$1\le d_z<\infty$。我们用 $\widehat{Q}_n:\Theta\to\mathcal{R}$ 表示一个经验标准，这个标准对所有 $\theta\in\Theta$ 是数据 $\{Z_t\}_{t=1}^n$ 的可测函数，并随着样本量 $n\to\infty$ 按照某种定义（该定义在 3.1 节会更加清楚）收敛到 $Q$。一种常见的估计 $\theta_o$ 的方式是在 $\Theta$ 上最大化 $\widehat{Q}_n$；最大化者 $\arg\sup_{\theta\in\Theta}\widehat{Q}_n(\theta)$（假定存在）则一般被称为极值估计。参考 Amemiya (1985, 第 4 章), Gallant and White (1988b), Newey and McFadden (1994) 和 White (1994)。

当 $\Theta$ 是无限维，并且对于（伪）度量 $d$ 可能不紧凑时，[8]在 $\Theta$ 上最大化 $\widehat{Q}_n$ 可能是不适定的；或者尽管最大化者 $\arg\sup_{\theta\in\Theta}\widehat{Q}_n(\theta)$ 存在，它也很难计算，并且可能具有不好的大样本特性，例如不一致性和/或非常慢的收敛速度等。出现这些困难是因为在无限维非紧凑空间上的优化问题可能不再是适定的。在本文中，如果对所有 $\Theta$ 中的序列 $\{\theta_k\}$，只要 $Q(\theta_o)-Q(\theta_k)\to 0$，就有 $d(\theta_o,\theta_k)\to 0$，我们就称该优化问题是“适定的”；反之，如果 $Q(\theta_o)-Q(\theta_k)\to 0$，但是 $d(\theta_o,\theta_k)\not\to 0$，则称之为“不适定的”。[9]

对于一个给定的半非参模型，假设在参数空间 $\Theta$ 中 $\theta_o$ 唯一地最大化 $Q(\theta)$。那么这个问题是否适定取决于伪度量 $d$ 的选择。这是因为无限维空间 $\Theta$ 上的不同度量可能彼此不等价。[10]特别的，有可能某种定义在 $\Theta$ 上的标准范数（例如 $||\theta_o-\theta||_s$）在 $Q(\theta_o)-Q(\theta)$ 中并不连续，导致在 $||\cdot||_s$ 下问题不适定；但可能在 $\Theta$ 上存在另一种伪度量（例如 $||\theta_o-\theta||_w$）在 $Q(\theta_o)-Q(\theta)$ 中连续，这时在 $||\cdot||_w$ 下问题就是适定的。通常情况下这个度量是要比 $||\cdot||_s$ 更弱（也就是说，由 $||\theta_o-\theta||_s\to 0$ 可以推导出 $||\theta_o-\theta||_w\to 0$）。更多这方面讨论请参考 Ai and Chen (2003, 2004a)。[11]

无论半非参数问题适定与否，筛分方法都提供了一种广泛适用的方法来应对在无限维空间 $\Theta$ 上最大化 $\widehat{Q}_n$ 所面临的问题：在被 Grenander (1981) 称之为“筛”的一系列相比原空间 $\Theta$ 更简单但在其中稠密的近似空间上求解上述最大化问题。比较常见的筛具有下列性质：紧凑的，非递减的（$\Theta_n\subseteq\Theta_{n+1}\subseteq\cdots\subseteq\Theta$），以及对任意的 $\theta\in\Theta$ 都存在 $\pi_n\theta\in\Theta$ 满足随着 $n\to\infty$，有 $d(\theta,\pi_n\theta)\to 0$（这里 $\pi_n$ 可以看作从 $\Theta$ 到 $\Theta_n$ 的一个投影映射）。

---

[7]尽管在本文中我们称 $\theta_o$ 为“真实”参数，事实上它可能是伪真实参数，这取决于计量模型的设定和 $Q$ 的选择。参考 Ai and Chen (2004a) 中关于估计设定错误的半非参模型的讨论）。

[8]在无限维度量空间 $(\mathcal{H},d)$ 中，紧集是一个 $d$ 闭合和完全有界的集合。（集合“完全有界”的定义是：对任何 $\varepsilon>0$，可以用有限多个半径为 $\varepsilon$ 的开球覆盖整个集合）。只有在有限维欧几里德空间中，$d$ 闭合和有界集才是紧凑的。注意这里“完全有界”和“有界”的区别。

[9]参考 Carrasco et al. (2006) 和 Vapnik (1998) 对线性非参模型中不适定求逆问题的讨论。

[10]这与所有范数在有限维欧几里得空间上均等价不同。

[11]在 Ai and Chen (2003) 中，他们使用较弱的伪度量来推导由模型 $E[\rho(Z_t;\beta_o,h_o)|X_t]=0$ 识别的 $\beta_o$ 的估计量 $\widehat{\beta}$ 的 $\sqrt{n}$-正态性。他们的讨论包括了 $h_o(\cdot)$ 取决于内生变量 $Y$，以及在标准均方误差度量 $\sqrt{E[h(Y)-h_o(Y)]^2}$ 下估计问题可能不适定的几种情况。

近似筛分极值估计量 $\hat{\theta}_n$ 是在筛空间 $\Theta_n$ 上最大化 $\widehat{Q}_n(\theta)$ 的统计量，也即：

$$\widehat{Q}_n(\hat{\theta}_n) \geq \sup_{\theta\in\Theta_n} \widehat{Q}_n(\theta) - O_P(\eta_n), \quad \text{同时}\eta_n \to 0 \text{ 随着}n \to \infty. \tag{2.9}$$

当 $\eta_n = 0$ 时，我们称在 (2.9) 中的 $\hat{\theta}_n$ 为 "确" 筛分极值估计。[12]我们可以清楚看到，筛分极值估计方法包含了标准的极值估计方法（只需对所有 $n$ 将 $\Theta_n = \Theta$）。

**备注 2.1.** White and Wooldridge (1991, 定理 2.2) 证明过在下列充分条件满足的前提下，在 (2.9) 中的 $\hat{\theta}_n$ 是存在的并且是可测的：(i) 对所有 $\theta \in \Theta_n$，$\widehat{Q}_n(\theta)$ 是数据 $\{Z_t\}_{t=1}^n$ 的可测函数；(ii) 对任何数据 $\{Z_t\}_{t=1}^n$，$\widehat{Q}_n(\theta)$ 在测度 $d(\cdot,\cdot)$ 下在 $\Theta_n$ 上是上连续的；(iii) 筛空间 $\Theta_n$ 在测度 $d(\cdot,\cdot)$ 下是紧凑的。因此，在余下章节中我们假设 (2.9) 中的 $\hat{\theta}_n$ 存在并可测。

在半非参计量模型中，$\theta_o \in \Theta$ 可以被分解为两部分 $\theta_o = (\beta_o', h_o')' \in B \times \mathcal{H}$，其中 $B$ 表示一个有限维的紧凑参数空间，$\mathcal{H}$ 是一个无限维的参数空间。这时，筛空间可以是 $\Theta_n = B \times \mathcal{H}_n$，这里 $\mathcal{H}_n$ 是 $\mathcal{H}$ 的一个筛。在上述筛空间下得到的 (2.9) 中的估计量 $\hat{\theta}_n = (\widehat{\beta}_n, \widehat{h}_n)$ 有时被称为联立（或联合）筛分极值估计量。对于半非参计量模型，我们也可以使用 "近似筛分极值估计" 法估计 $(\beta_o, h_o)$，这种方法分为以下两步：

**第一步：** 对任意固定的 $\beta \in B$，计算 $\widehat{Q}_n(\beta, \widetilde{h}(\beta)) \geq \sup_{h\in\mathcal{H}_n} \widehat{Q}_n(\beta, h) - O_P(\eta_n)$，这里 $\eta_n = o(1)$；

**第二步：** 使用求解 $\widehat{Q}_n(\widehat{\beta}, \widetilde{h}(\widehat{\beta})) \geq \max_{\beta\in B} \widehat{Q}_n(\beta, \widetilde{h}(\beta)) - O_P(\eta_n)$ 得到的 $\widehat{\beta}_n$ 估计 $\beta_o$，然后用 $\widehat{h}_n = \widetilde{h}(\widehat{\beta}_n)$ 估计 $h_o$。

根据半非参数模型的具体结构，有时候上述筛极值估计方法可能更易于计算。

### 2.2.2 筛分 M-估计法

当 $\widehat{Q}_n(\theta)$ 可以用以下样本平均值来表示时：

$$\sup_{\theta\in\Theta_n} \widehat{Q}_n(\theta) = \sup_{\theta\in\Theta_n} \frac{1}{n}\sum_{t=1}^{n} l(\theta, Z_t),$$

其中 $l : \Theta \times \mathcal{R}^{d_z} \to \mathcal{R}$ 是基于单一观测的目标函数；我们也称 (2.9) 的解 $\hat{\theta}_n$ 为近似 "筛分极大似然 (M-) 估计"。[13] 这包括筛极大似然估计（MLE），筛最小二乘法（LS），筛广义最小二乘法（GLS）和筛分分位数回归作为特例。

**例 2.1 续：** Heckman and Singer (1984) 使用筛分极大似然估计法在例 2.1 中的半参数设定（2.1）下估计了未知真实参数 $\theta_o = (\beta_o', h_o)' \in \Theta$：

$$\sup_{\theta\in\Theta_n} \widehat{Q}_n(\theta) = \sup_{\beta\in B, h\in\mathcal{H}_n} \frac{1}{n}\sum_{i=1}^{n} \log\left(\int g(T_i|\beta, u, X_i) dh(u)\right),$$

这里随着 $n \to \infty$，筛空间 $\mathcal{H}_n$ 在 $\mathcal{R}$ 上定义的概率分布函数空间中变得稠密。

[12]由于筛空间 $\Theta_n$ 的复杂度随样本量增加而上升，很明显在 $\Theta_n$ 上最大化 $\widehat{Q}_n(\theta)$ 的问题不需要精确解；一个在 (2.9) 中的近似最大化者 $\hat{\theta}_n$ 就足以保证一致性。参见 3.1 中的一致性定理。

[13]我们的定义沿用了 Newey 和 McFadden（1994）。一些统计学家如 Birgé and Massart（1998）称之为筛最小对比估计。

**例 2.2 续：** 可以用筛分非线性最小二乘法估计例 2.2 中的非参外生支出模型 (2.2)：

$$\sup_{\theta\in\Theta_n}\widehat{Q}_n(\theta)=\sup_{h\in\mathcal{H}_n}\frac{-1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{N}\left[Y_{1\ell i}-\{h_{1\ell}(Y_{2i}-h_0(X_{1i}))+h_{2\ell}(X_{1i})\}\right]^2,$$

这里 $\theta=h=(h_0,h_{11},...,h_{1N},h_{21},...,h_{2N})'$ 是未知参数，$\Theta_n=\mathcal{H}_n=\mathcal{H}_{0,n}\times\prod_{\ell=1}^{N}\mathcal{H}_{1\ell,n}\times\prod_{\ell=1}^{N}\mathcal{H}_{2\ell,n}$ 是筛空间，[14]同时我们在筛空间 $\mathcal{H}_{2\ell,n}$（$\ell=1,...,N$）上施加识别条件 $h_{2\ell}(x_1^*)=0$。例 2.2 的半非参数外生支出模型（2.3）也可以通过非线性最小二乘法估计：

$$\sup_{\theta\in\Theta_n}\widehat{Q}_n(\theta)=\sup_{\beta\in B,h\in\mathcal{H}_n}\frac{-1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{N}[Y_{1\ell i}-\{h_{1\ell}(Y_{2i}-g(X_{1i}'\beta_1))+X_{1i}'\beta_{2\ell}\}]^2,$$

这里 $\theta=(\beta',h')'=(\beta_1',\beta_{21}',...,\beta_{2N}',h_{11},...,h_{1N})'$ 代表未知参数，$\Theta_n=B\times\mathcal{H}_n=B_1\times\prod_{\ell=1}^{N}B_{2\ell}\times\prod_{\ell=1}^{N}\mathcal{H}_{1\ell,n}$ 是筛空间。

更一般地，我们可以应用筛分广义最小二乘法

$$\sup_{\theta\in\Theta_n}\widehat{Q}_n(\theta)=\sup_{\theta\in\Theta_n}\frac{-1}{n}\sum_{i=1}^{n}\rho(Z_i,\theta)'\{\Sigma(X_i)\}^{-1}\rho(Z_i,\theta)$$

来估计条件矩限制 (2.8) 下的第一个子类的所有模型，其中 $\rho(Z_i,\theta)-\rho(Z_i,\theta_o)$ 并不依赖于内生变量 $Y_i$，这里 $\Sigma(X_i)$ 是正定加权矩阵函数（例如单位矩阵）。有关在这类方法中如何选择最优加权矩阵，请参考第 4.3 节中的备注 4.3。

### 2.2.3 级数估计，凹型广义线性模型

在本文中，我们称筛 M 估计法的一种特殊情况为 “级数” 估计，这种筛 M 估计法的目标函数 $\widehat{Q}_n(\theta)=\frac{1}{n}\sum_{t=1}^{n}l(\theta,Z_t)$ 是凹函数，其筛空间 $\Theta_n$ 是 “有限维线性” 的空间。如果对任意 $\theta_1,\theta_2\in\Theta$ 和纯量 $\tau\in(0,1)$ 都有 $\widehat{Q}_n(\tau\theta_1+(1-\tau)\theta_2)\geq\tau\widehat{Q}_n(\theta_1)+(1-\tau)\widehat{Q}_n(\theta_2)$，我们就称该目标函数 $\widehat{Q}_n$ 是凹函数。当然，这个定义只有当参数空间 $\Theta$ 是凸（对任意 $\theta_1,\theta_2\in\Theta$ 和任意纯量 $\tau\in(0,1)$，有 $\tau\theta_1+(1-\tau)\theta_2\in\Theta$）的时才有意义。如果一个筛 $\Theta_n$ 是由有限多个已知基函数线性张成的空间，我们称 $\Theta_n$ 是有限维线性的。具体例子可以参考 2.3.1 节。

虽然我们对级数估计的定义可能与当前计量经济学文献中的定义有所出入，但它与统计学文献中的 “凹型广义线性模型” 的筛 M 估计密切相关。参考 Hansen (1994), Stone et al. (1997), 以及 Huang (2001)。考虑一个在 $\mathcal{Z}$ 中取值的随机变量 $Z$，其中 $\mathcal{Z}$ 是一个任意的集合。$Z$ 的概率密度 $p_o(z)$ 取决于一个真实但未知的参数 $\theta_o$。所有的凹扩展线性模型均有以下三个特征：(1) 一个 (可能是无限维) 的线性参数空间 $\Theta$；(2) 在单个观测 ($z$) 上的目标函数是 $\theta$ 的凹函数，也即对任意给定的 $\theta_1,\theta_2\in\Theta$，任意纯量 $\tau\in(0,1)$ 和任意 $z\in\mathcal{Z}$，有 $l(\tau\theta_1+(1-\tau)\theta_2,z)\geq\tau l(\theta_1,z)+(1-\tau)l(\theta_2,z)$。(3) 总体目标函数 $Q(\theta)=E[l(\theta,Z)]$ 是严格凹的。也即对任意纯量 $\tau\in(0,1)$ 和任意给定的不同函数 $\theta_1,\theta_2\in\Theta$，有 $E[l(\tau\theta_1+(1-\tau)\theta_2,Z)]>\tau E[l(\theta_1,Z)]+(1-\tau)E[l(\theta_2,Z)]$。

[14]在本文中 $\prod_{\ell=1}^{N}\mathcal{H}_{\ell,n}$ 代表笛卡尔积 $\mathcal{H}_{1,n}\times\cdots\times\mathcal{H}_{N,n}$。

凹型广义线性模型的筛 M 估计可以通过在一个有限维线性筛空间 $\Theta_n$ 上无约束地最大化 $\widehat{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l(\theta, Z_t)$ 来实现。这样所得到的估计量在本文中被称为级数估计量。因此，对于相同的凹目标函数，如果筛空间 $\Theta_n$ 是有限维线性的（例如在 2.3.1 和 2.3.2 小节中出现的几种情况），则筛 M 估计量是级数估计量；反之，如果筛空间 $\Theta_n$ 不是有限维线性的（例如在 2.3.3 和 2.3.4 小节中出现的几种情况），则筛 M 估计量就不是级数估计量。虽然这个级数估计量的定义看起来限制较多，但是在第 3 节中，它将使得描述统计量的大样本性质更加容易。

对于级数估计，目标函数的凹性起着核心作用。特别地，在估计中使用的筛空间不需要紧凑，并且可以是任何不受限制的有限维线性空间。使用这种筛不仅有利于简化估计量的计算过程，而且可以便利地在非参数多元回归框架中讨论正交投影和方差分解（例如可加性）的泛函分析；参考 Stone (1985, 1986), Andrews and Whang (1990), Huang (1998a)。

为了将级数估计应用于半非参数模型，我们需要首先找到一个识别未知参数的凹目标函数。下面我们举几个这方面的例子。

**例 2.4 多元线性回归** 我们考虑一个未知的多元条件矩函数 $\theta_o(\cdot) = h_o(\cdot) = E(Y|X = \cdot)$。这里这里 $Z = (Y, X)$，$Y$ 是一维标量，$X$ 有支集 $\mathcal{X}$（后者是 $\mathcal{R}^d$, $d \geq 1$ 的有界子集）。假设 $h_o \in \Theta$, 这里 $\Theta$ 是函数 $h$ 空间的一个线性子空间，$h$ 满足 $E[h(X)^2] < \infty$。定义 $l(h, Z) = -[Y - h(X)]^2$ 和 $Q(\theta) = -E\{[Y - h(X)]^2\}$; 这两个函数都是 $h$ 的凹函数，并且 $Q$ 是 $h \in \Theta$ 的严格凹函数。

我们用 $\{p_j(X), j = 1, 2, ...\}$ 表示可以近似任何实值且平方可积的函数 $X$ 的一系列已知基函数；关于这类基函数的具体例子可以参考章节 2.3.1 或 Newey (1997)。那么

$$\Theta_n = \mathcal{H}_n = \{h : \mathcal{X} \to \mathcal{R}, h(x) = \Sigma_{j=1}^{k_n} a_j p_j(x) : a_1, ..., a_{k_n} \in \mathcal{R}\}, \tag{2.10}$$

上式中随着 $n \to \infty$，缓慢地 $\dim(\Theta_n) = k_n \to \infty$，这里 $\Theta_n$ 是 $\Theta$ 的有限维线性筛。$\widehat{h} = \arg\max_{h \in \mathcal{H}_n} \frac{-1}{n}\sum_{t=1}^{n}[Y_t - h(X_t)]^2$ 是条件矩 $h_o(\cdot) = E(Y|X = \cdot)$ 的级数估计量。此外，这个级数估计量 $\widehat{h}$ 有简单的解析式：

$$\widehat{h}(x) = p^{k_n}(x)'(P'P)^{-} \sum_{i=1}^{n} p^{k_n}(X_i)Y_i, \quad x \in \mathcal{X}, \tag{2.11}$$

这里 $p^{k_n}(X) = (p_1(X), ..., p_{k_n}(X))'$，$P = (p^{k_n}(X_1), ..., p^{k_n}(X_n))'$，$(P'P)^{-}$ 代表 Moore-Penrose 广义逆矩阵。在（2.11）中的估计量 $\widehat{h}$ 被称为级数最小二乘估计量或线性筛最小二乘估计量。

**例 2.5 多元分位回归** 选取 $\alpha \in (0, 1)$。我们考虑估计满足 $E[1\{Y \leq h_o(X)\}|X] = \alpha$ 的未知多元 $\alpha$ 分位函数 $\theta_o(\cdot) = h_o(\cdot)$。这里 $Z = (Y, X)$，$X$ 的支集 $\mathcal{X}$ 是 $\mathcal{R}^d$，$d \geq 1$ 的有界子集。假设 $h_o \in \Theta$，这里 $\Theta$ 是满足 $E[h(X)^2] < \infty$ 的函数 $h$ 空间的一个线性子空间。选择 $l(h, Z) = [1\{Y \leq h(X)\} - \alpha][Y - h(X)]$，[15] 同时 $Q(\theta) = E\{[1\{Y \leq h(X)\} - \alpha][Y - h(X)]\}$，那么两个函数都是 $h$ 的凹函数，并且 $Q$ 是 $h \in \Theta$ 的严格凹函数。

选取 $\Theta_n = \mathcal{H}_n$ 作为一个有限维线性筛，例如 (2.10) 中给出的例子。那么 $\widehat{h} = \arg\max_{h \in \mathcal{H}_n} \frac{1}{n}\sum_{t=1}^{n}[1\{Y_t \leq h(X_t)\} - \alpha][Y_t - h(X_t)]$ 是条件分位函数 $h_o$ 的级数估计量。

[15] 这是 Koenker and Bassett (1978) 文中的"check" 函数。

**例 2.6 对数密度估计** 我们用 $f_o$ 表示 $Z$ 在 $\mathcal{Z}$ 上的真实未知概率密度函数。假设我们希望估计对数密度 $\log f_o$。由于 $\log f_o$ 满足非线性约束 $\int_{\mathcal{Z}} \exp\{\log f_o(z)\}dz = 1$，我们可以将其改写为 $\log f_o = h_o - \log \int_{\mathcal{Z}} \exp h_o(z)dz$，并将 $h_o$ 作为某线性空间的未知函数来处理。由于对任何常数 $c$ 都有 $\log f_o = [h_o + c] - \log \int_{\mathcal{Z}} \exp[h_o(z) + c]dz$，我们需要作位置标准化处理来确保可以识别 $h_o$，例如 $\int_{\mathcal{Z}} h(z)dz = 0$ 或者在某个特定的 $z^* \in \mathcal{Z}$ 上有 $h(z^*) = 0$。这样我们可以唯一地确定 $h$ 并使映射 $h \mapsto \log f$ 成为双射。因此，我们假设 $h_o \in \Theta$，其中 $\Theta$ 是一个满足 $E[h(Z)^2] < \infty$ 和 $\int_{\mathcal{Z}} h(z)dz = 0$ 条件的实值函数 $h$ 的空间的线性子空间。在某观测 $Z$ 上的对数似然函数是 $l(h, Z) = h(Z) - \log \int_{\mathcal{Z}} \exp h(z)dz$。Stone (1990) 已经证明 $l(h, Z)$ 是凹的，并且 $Q(\theta) = E\{h(Z) - \log \int_{\mathcal{Z}} \exp h(z)dz\}$ 是 $h \in \Theta$ 的严格凹函数。

我们用 $\{p_j(Z), j = 1, 2, ...\}$ 表示可以近似任意实值平方可积函数 $Z$ 的一系列已知基函数。那么

$$\Theta_n = \mathcal{H}_n = \{h : \mathcal{Z} \to \mathcal{R}, h(z) = \Sigma_{j=1}^{k_n} a_j p_j(z) : \int_{\mathcal{Z}} h(z)dz = 0, a_1, ..., a_{k_n} \in \mathcal{R}\},$$

这里随着 $n \to \infty, \dim(\Theta_n) = k_n \to \infty$ 但相对缓慢；$\Theta_n$ 是 $\Theta$ 的有限维线性筛，同时 $\widehat{h} = \arg\max_{h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^n \left[h(Z_i) - \log \int_{\mathcal{Z}} \exp h(z)dz\right]$ 是对数密度函数 $h_o$ 的级数估计量。

很容易看出，可以以相同的方式估计对数条件密度和对数谱密度；参考 Stone (1994) 和 Kooperberg et al. (1995b)。

**例 2.7 条件风险函数估计** 考虑一个正的存活时间 $T$，一个正的终止时间 $C$，观测到的时间 $Y = \min(T, C)$，以及在 $\mathcal{X}$ 上取值的协变量向量 $X$。我们用 $Z = (X', Y, 1(T \le C))'$ 表示一个观测。假设给定 $X$ 下 $T$ 和 $C$ 条件独立，同时对已知的给定正常数 $\tau_0$ 有 $\Pr(C \le \tau_0) = 1$。让 $f_o(\tau|x)$ 和 $F_o(\tau|x)$, $\tau > 0$ 分别表示给定 $X = x$ 下 $T$ 的未知的真实条件密度函数和条件分布函数。那么比率 $f_o(\tau|x)/[1 - F_o(\tau|x)]$, $\tau > 0$，被称作给定 $X = x$ 下的 $T$ 的条件风险函数。我们要估计对数条件风险函数 $h_o(\tau, x) = \log\{f_o(\tau|x)/[1 - F_o(\tau|x)]\}$。由于在 $Z$ 观测下的似然函数等于

$$[f(Y|X)]^{1(T \le C)}[1 - F(Y|X)]^{1(T > C)} = [\exp\{h(Y, X)\}]^{1(T \le C)} \exp\left(-\int_0^Y \exp\{h(\tau, X)\}d\tau\right),$$

在单个观测下的对数似然函数由下式给出：

$$l(h, Z) = 1(T \le C)h(Y, X) - \int_0^Y \exp\{h(\tau, X)\}d\tau.$$

Kooperberg et al. (1995a) 证明了 $l(h, Z)$ 是 $h$ 的凹函数和 $Q(\theta) = E\{l(h, Z)\}$ 是 $h$ 的严格凹函数。

假设 $h_o \in \Theta$，其中 $\Theta$ 是一个满足条件 $E[h(Y, X)^2] < \infty$ 的实值函数 $h$ 的空间的线性子空间。用 $\{p_j(Y, X), j = 1, 2, ...\}$ 表示可以近似任意实值平方可积函数 $f(Y, X)$ 的一系列已知基函数。那么

$$\Theta_n = \mathcal{H}_n = \{h : (0, \tau_0] \times \mathcal{X} \to \mathcal{R}, h(\tau, x) = \Sigma_{j=1}^{k_n} a_j p_j(\tau, x) : a_1, ..., a_{k_n} \in \mathcal{R}\},$$

这里随着 $n \to \infty$，缓慢地 $\dim(\Theta_n) = k_n \to \infty$。$\Theta_n$ 是 $\Theta$ 的有限维线性筛。$\widehat{h} = \arg\max_{h \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^n \left[1(T_i \le C_i)h(Y_i, X_i) - \int_0^{Y_i} \exp\{h(\tau, X_i)\}d\tau\right]$ 是对数条件风险函数 $h_o$ 的级数估计量。

最后，我们应当指出并不是所有半非参数 M 估计问题都可以被重新参数化为级数估计问题。例如，例 2.2 中的非参外生支出模型 (2.2) 并不属于凹拓展线性模型，这是由于在这类模型下未知函数 $h_0(X_1)$ 作为

自变量以非线性的形式进入另一个未知函数 $h_{1\ell}(Y_2 - h_0(X_1)), \ell = 1, ..., L$。然而，如上一节所述，该模型仍然可以通过一般筛 M 估计方法估计。

### 2.2.4 筛最小距离（MD）估计

当 $-\widehat{Q}_n(\theta)$ 可以表示为与 0 的距离的平方时，我们称 (2.9) 的解 $\hat{\theta}_n$ 为近似筛最小距离（MD）估计。一个具有代表性的二次形式为

$$\sup_{\theta\in\Theta_n} \widehat{Q}_n(\theta) = \sup_{\theta\in\Theta_n} -\frac{1}{n}\sum_{t=1}^{n} \widehat{m}(X_t,\theta)'\{\widehat{\Sigma}(X_t)\}^{-1}\widehat{m}(X_t,\theta) \tag{2.12}$$

这里依概率 $\widehat{m}(X_t, \theta_o) \to 0$。上式中 $\widehat{m}(X_t, \theta)$ 是某固定有限维的矩约束函数的非参估计。$\widehat{\Sigma}(X_t)$ 是对跟 $\widehat{m}(X_t, \theta)$ 具有同样维度的加权矩阵的非参估计。引入加权矩阵 $\widehat{\Sigma}$ 主要出于效率的考虑，[16] 并且依概率 $\widehat{\Sigma}(X_t) \to \Sigma(X_t)$，这里 $\Sigma(X_t)$ 是跟 $\widehat{\Sigma}(X_t)$ 维度相同的正定矩阵。无论 $\rho(Z_t, \theta) - \rho(Z_t, \theta_o)$ 是否取决于内生变量 $Y_t$，我们都可以应用筛 MD 标准 (2.12) 来估计所有满足条件矩约束 $E[\rho(Z, \theta_o)|X] = 0$ 的模型。特别的，$\widehat{m}(X_t, \theta)$ 可以是任何条件均值函数 $m(X_t, \theta) = E[\rho(Z, \theta)|X = X_t]$ 的非参估计。参考 Newey and Powell (1989, 2003) 和 Ai and Chen (1999, 2003)。

另一类典型的二次型是筛广义矩条件（GMM）标准：

$$\sup_{\theta\in\Theta_n} \widehat{Q}_n(\theta) = \sup_{\theta\in\Theta_n} -\widehat{g}_n(\theta)'\widehat{W}\widehat{g}_n(\theta) \tag{2.13}$$

这里依概率 $\widehat{g}_n(\theta_o) \to 0$。$\widehat{g}_n(\theta)$ 是维度随样本量增长的整体矩条件的样本平均，$\widehat{W}$ 是同 $\widehat{g}_n(\theta)$ 一样具有增长的维度的随机加权矩阵。跟上面的情况一致，这里引入 $\widehat{W}$ 是出于效率的考虑，并且依概率 $\widehat{W} - W_n \to 0$，其中 $W_n$ 是同 $\widehat{W}$ 一样维度随样本量增长的正定矩阵。

注意当且仅当下列数目增加的非条件矩约束成立时，才有 $E[\rho(Z, \theta_o)|X] = 0$：

$$E[\rho(Z_t, \theta_o)p_{0j}(X_t)] = 0,\ j = 1, 2, ..., k_{m,n}, \tag{2.14}$$

这里 $\{p_{0j}(X), j = 1, 2, ..., k_{m,n}\}$ 是一系列随着 $k_{m,n} \to \infty$ 可以近似任何实值平方可积函数的已知基函数。让 $p^{k_{m,n}}(X) = (p_{01}(X), ..., p_{0k_{m,n}}(X))'$。很显然我们可以通过筛 GMM 标准 (2.13) 使用 $\widehat{g}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \rho(Z_t, \theta) \otimes p^{k_{m,n}}(X_t)$ 来估计条件矩约束 (2.8) $E[\rho(Z, \theta_o)|X] = 0$。

我们不仅可以使用筛 MD(2.12) 和筛 GMM(2.13) 估计所有条件矩约束 (2.8) 下的模型，这两种方法本身也是紧密相关的。例如，当应用筛 MD(2.12) 方法时，我们可以使用级数最小二乘估计量 (2.15) 来估计条件矩函数 $m(X, \theta) = E[\rho(Z, \theta)|X]$:

$$\widehat{m}(X, \theta) = \sum_{j=1}^{n} \rho(Z_j, \theta)p^{k_{m,n}}(X_j)'(P'P)^{-}p^{k_{m,n}}(X), \tag{2.15}$$

上式中 $P = (p^{k_{m,n}}(X_1), ..., p^{k_{m,n}}(X_n))'$，同时随着 $n \to \infty$，缓慢地有 $k_{m,n} \to \infty$。这里 $(P'P)^{-}$ 是 Moore-Penrose 广义逆矩阵。如果使用单位矩阵作为加权矩阵 $\widehat{\Sigma}(X_t) = I$，上述筛 MD 估计 (2.12) 则成为下列筛

[16] 参考 Ai and Chen (2003) 或本文 4.3 节关于半参估计效率的讨论。

GMM 估计 (2.13):

$$\min_{\theta\in\Theta_n}\left(\sum_{i=1}^n \rho(Z_i,\theta)\otimes p^{k_{m,n}}(X_i)\right)' \left(I\otimes(P'P)^-\right)\left(\sum_{i=1}^n \rho(Z_i,\theta)\otimes p^{k_{m,n}}(X_i)\right), \tag{2.16}$$

这里 $\otimes$ 表示 Kronecker 乘积，具体内容可以参考 Ai and Chen (2003)。

**例 2.2 续** 例 2.2 中的半非参内生支出模型 (2.4) 可以通过筛 MD(2.12) 方法估计，其中 $\widehat{m}(X_i,\theta) = (\widehat{m}_1(X_i,\theta),...,\widehat{m}_N(X_i,\theta))'$，以及

$$\widehat{m}_\ell(X_i,\theta) = \sum_{j=1}^n [Y_{1\ell j} - \{h_{1\ell}(Y_{2j} - g(X_{1j}'\beta_1)) + X_{1j}'\beta_{2\ell}\}]p^{k_{m,n}}(X_j)'(P'P)^- p^{k_{m,n}}(X_i),$$

式中 $\theta = (\beta',h')' = (\beta_1',\beta_{21}',...,\beta_{2N}',h_{11},...,h_{1N})'$ 是未知参数向量，$\Theta_n = B\times\mathcal{H}_n = B_1\times\prod_{\ell=1}^N B_{2\ell}\times\prod_{\ell=1}^N \mathcal{H}_{1\ell,n}$ 是筛空间；具体细节可以参考 Blundell et al. (2001)。

**例 2.3 续** 例 2.3 中的半非参外部习惯模型 (2.7) 可以通过筛 GMM(2.16) 方法估计，其中 $\rho(Z_t,\theta) = (\rho_1(Z_t,\theta),...,\rho_N(Z_t,\theta))'$，以及

$$\rho_\ell(Z_t,\theta) = \delta\left(\frac{C_t}{C_{t+1}}\right)^\gamma \frac{\left(1-h\left(\frac{C_t}{C_{t+1}},...,\frac{C_{t+1-L}}{C_{t+1}}\right)\right)^{-\gamma}}{\left(1-h\left(\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right)\right)^{-\gamma}} R_{\ell,t+1} - 1, \quad \ell = 1,...,N,$$

$$Z_t = \left(\frac{C_t}{C_{t+1}},...,\frac{C_{t+1-L}}{C_{t+1}},\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t},R_{1,t+1},...,R_{N,t+1},X_t\right), \quad X_t = \mathbf{w}_t,$$

式中 $\theta = (\beta',h)' = (\delta,\gamma,h)'$ 是未知参数向量，$\Theta_n = B\times\mathcal{H}_n = B_\delta\times B_\gamma\times\mathcal{H}_n$ 是筛空间，这里要求筛空间 $\mathcal{H}_n$ 满足 $0\le h<1$。很显然，这个模型 (2.7) 也可以用筛 MD 方法估计 (2.12)，其中 $\widehat{m}(X_t,\theta) = \widehat{m}(\mathbf{w}_t,\theta)$ 是 $E[\rho(Z_t,\theta)|X_t=\mathbf{w}_t]$ 的 (例如级数 LS 估计量 (2.15)) 非参估计量；参考 Chen and Ludvigson (2003)。[17]

## 2.3 典型函数空间和筛空间

这一章我们将介绍一些常用的筛，其近似性质在近似理论的相关数学文献中已经广泛讨论过。

### 2.3.1 典型的平滑函数和（有限维）线性筛

我们首先回顾非参数估计文献中使用最广泛的平滑函数；参考 Stone (1982, 1994), Robinson (1988), Newey (1997) 和 Horowitz (1998)。现在我们假设 $\mathcal{X} = \mathcal{X}_1\times...\times\mathcal{X}_d$ 是紧区间 $\mathcal{X}_1,\ldots,\mathcal{X}_d$ 的笛卡尔积。让 $0<\gamma\le 1$。对于一个定义在 $\mathcal{X}$ 上的实值函数 $h$，如果存在正数 $c$ 满足对任何 $x,y\in\mathcal{X}$ 都有 $|h(x)-h(y)|\le c|x-y|_e^\gamma$，则称 $h$ 满足指数为 $\gamma$ 的 Hölder 条件。这里 $|x|_e = \left(\sum_{l=1}^d x_l^2\right)^{1/2}$ 是 $x=(x_1,\ldots,x_d)\in\mathcal{X}$ 的欧几里得范数。给定一个 $d$-元组非负整数 $\alpha=(\alpha_1,\ldots,\alpha_d)$，称 $[\alpha] = \alpha_1+\cdots+\alpha_d$，同时让 $D^\alpha$ 表示差分算子：

$$D^\alpha = \frac{\partial^{[\alpha]}}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}}.$$

[17]还有矩程序的半非参数递归方法，这类方法使我们能够估计具有潜变量的非线性时间序列模型。参考 Chen and White (1998, 2002), Pastorello et al. (2003) 和 Linton and Mammen (2005)。

用 $m$ 表示非负整数，让 $p=m+\gamma$。如果一个定义在 $\mathcal{X}$ 上的实值函数 $h$ 是 $m$ 次连续可导，并且 $D^\alpha h$ 对所有 $[\alpha]=m$ 的 $\alpha$ 都满足指数为 $\gamma$ 的 Hölder 条件，则称 $h$ 为 $p$-平滑。

我们称定义在 $\mathcal{X}$ 上的所有 $p$-平滑实值函数类为 $\Lambda^p(\mathcal{X})$（这被称为 Hölder 类），称定义在 $\mathcal{X}$ 上的所有 $m-$ 次连续可导实值函数为 $C^m(\mathcal{X})$。定义一个平滑度为 $p=m+\gamma$ 的 Hölder 球为

$$\Lambda_c^p(\mathcal{X})=\left\{h\in C^m(\mathcal{X}):\ \sup_{[\alpha]\le m}\sup_{x\in\mathcal{X}}|D^\alpha h(x)|\le c,\ \sup_{[\alpha]=m}\sup_{x,y\in\mathcal{X},x\neq y}\frac{|D^\alpha h(x)-D^\alpha h(y)|}{|x-y|_e^\gamma}\le c\right\}.$$

由于可以用各类线性筛很好地近似拟合 $p$-平滑函数，这类 Hölder（或 $p$-平滑）函数在计量经济学中非常常见。

如果一个筛是有限多个已知基函数的线性生成空间，则我们称其为“(有限维) 线性筛”。线性筛包括幂级数、傅里叶级数、样条和小波等，共同形成了一大类可用于筛极值估计的筛。下面我们给出定义在 $\mathcal{X}=[0,1]$ 上的单变量函数的线性筛的具体例子。

**多项式** 用 $\mathrm{Pol}(J_n)$ 表示定义在 $[0,1]$ 上的次数小于或等于 $J_n$ 的多项式空间

$$\mathrm{Pol}(J_n)=\left\{\sum_{k=0}^{J_n}a_kx^k,\ x\in[0,1]:\ a_k\in\mathcal{R}\right\}.$$

**三角多项式** 用 $\mathrm{TriPol}(J_n)$ 表示定义在 $[0,1]$ 上的级数小于或等于 $J_n$ 的三角多项式空间

$$\mathrm{TriPol}(J_n)=\left\{a_0+\sum_{k=1}^{J_n}[a_k\cos(2k\pi x)+b_k\sin(2k\pi x)],\ x\in[0,1]:\ a_k,b_k\in\mathcal{R}\right\}.$$

用 $\mathrm{CosPol}(J_n)$ 表示定义在 $[0,1]$ 上的级数小于或等于 $J_n$ 的余弦多项式空间

$$\mathrm{CosPol}(J_n)=\left\{a_0+\sum_{k=1}^{J_n}a_k\cos(k\pi x),\ x\in[0,1]:\ a_k\in\mathcal{R}\right\}.$$

用 $\mathrm{SinPol}(J_n)$ 表示定义在 $[0,1]$ 上的级数小于或等于 $J_n$ 的正弦多项式空间

$$\mathrm{SinPol}(J_n)=\left\{\sum_{k=1}^{J_n}a_k\sin(k\pi x),\ x\in[0,1]:\ a_k\in\mathcal{R}\right\}.$$

我们注意到上述三类三角多项式各自适合近似某类函数：经典的三角多项式筛 $\mathrm{TriPol}(J_n)$ 适用于近似定义在 $[0,1]$ 上的周期函数；余弦筛 $\mathrm{CosPol}(J_n)$ 适用于近似定义在 $[0,1]$ 上的非周期函数；正弦筛 $(J_n)$ 适用于近似在边界点处趋向于 0 的筛（即 $h(0)=h(1)=0$）。

**单变量样条** 用 $J_n$ 表示正整数，同时用 $t_0,t_1,\ldots,t_{J_n},t_{J_n+1}$ 表示满足 $0=t_0<t_1<\cdots<t_{J_n}<t_{J_n+1}=1$ 的实数。我们分割 $[0,1]$ 为 $J_n+1$ 个子区间 $I_j=[t_j,t_{j+1}),\ j=0,\ldots,J_n-1$，并且有 $I_{J_n}=[t_{J_n},t_{J_n+1}]$。假设节点 $t_1,\ldots,t_{J_n}$ 具有有界的网格比：

$$\frac{\max_{0\le j\le J_n}(t_{j+1}-t_j)}{\min_{0\le j\le J_n}(t_{j+1}-t_j)}\le c\ \text{对某些常数}c>0. \tag{2.17}$$

让 $r\ge1$ 代表一个整数。一个在 $[0,1]$ 上的函数是 $r$ 阶样条，或等价的，是具有节点 $t_1,\ldots,t_{J_n}$ 的 $m=r-1$ 度样条，如果以下条件成立：(i) 在每个区间 $I_j,\ j=0,\ldots,J_n$ 上是 $m$ 度或更低度数的多项式。(ii)（对于

$m \geq 1$）在 $[0,1]$ 上 $(m-1)$-次连续可导。这类样条函数构成了 $J_n + r$ 维度的线性空间。具体这方面的讨论，可以参考 de Boor (1978) 和 Schumaker (1981)。对于确定的整数 $r \geq 1$，我们让 $\mathrm{Spl}(r, J_n)$ 表示 $r$ 阶（或 $m = r-1$ 度）样条空间，其节点 $J_n$ 满足 (2.17)。由于

$$\mathrm{Spl}(r, J_n) = \left\{ \sum_{k=0}^{r-1} a_k x^k + \sum_{j=1}^{J_n} b_j [\max\{x - t_j, 0\}]^{r-1},\ x \in [0,1] :\ a_k, b_j \in \mathcal{R} \right\},$$

我们也称 $\mathrm{Spl}(r, J_n)$ 为 $m \equiv r-1$ 度的多项式样条筛。

在本文中，$L_2(\mathcal{X}, leb)$ 表示满足 $\int_{\mathcal{X}} |h(x)|^2 dx < \infty$ 的实值函数 $h$ 的空间。

**小波** 选取正整数 $m \geq 0$。如果一个实值函数 $\psi$ 满足如下三个条件，我们称其为 $m$ 度“母小波”：(i) 对于 $0 \leq k \leq m$，有 $\int_{\mathcal{R}} x^k \psi(x) dx = 0$；(ii) 随着 $|x| \to \infty$，$\psi$ 自身及其所有小于等于 $m$ 阶的导数都迅速减小；(iii) $\{2^{j/2}\psi(2^j x - k) : j, k \in \mathbb{Z}\}$ 形成一个 $L_2(\mathcal{R}, leb)$ 上的 Riesz 基，其具体含义是 $\{2^{j/2}\psi(2^j x - k) : j, k \in \mathbb{Z}\}$ 的线性生成空间是 $L_2(\mathcal{R}, leb)$ 的稠密子集，同时存在正的常数 $c_1 \leq c_2 < \infty$，使得对于所有双无限平方可加序列 $\{a_{jk} : j, k \in \mathbb{Z}\}$ 满足

$$c_1 \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a_{jk}|^2 \leq \left\| \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_{jk} 2^{j/2} \psi(2^j x - k) \right\|^2_{L_2(\mathcal{R}, leb)} \leq c_2 \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a_{jk}|^2$$

如果一个尺度函数 $\phi$ 满足以下三个条件，我们就称其为“父小波”：（i）$\int_{\mathcal{R}} \phi(x) dx = 1$；（ii）随着 $|x| \to \infty$，$\phi$ 自身及其所有小于等于 $m$ 阶的导数都迅速减小；（iii）$\{\phi(x-k) : k \in \mathbb{Z}\}$ 形成 $L_2(\mathcal{R}, leb)$ 的闭子空间的一个 Riesz 基。

**正交小波** 给定正整数 $m \geq 0$，存在支撑在紧集上的 $m$ 度父小波 $\phi$ 和 $m$ 度母小波 $\psi$ 满足对于任意整数 $j_0 \geq 0$、任意 $L_2(\mathcal{R}, leb)$ 中的函数 $g$ 都有如下小波 $m-$ 正则多解析度展开：

$$g(x) = \sum_{k=-\infty}^{\infty} a_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=-\infty}^{\infty} b_{jk} \psi_{jk}(x), \quad x \in \mathcal{R},$$

其中

$$\begin{aligned} a_{jk} &= \int_{\mathcal{R}} g(x) \phi_{jk}(x) dx, \quad \phi_{jk}(x) = 2^{j/2} \phi(2^j x - k), \quad x \in \mathcal{R}, \\ b_{jk} &= \int_{\mathcal{R}} g(x) \psi_{jk}(x) dx, \quad \psi_{jk}(x) = 2^{j/2} \psi(2^j x - k), \quad x \in \mathcal{R}, \end{aligned}$$

以及 $\{\phi_{j_0 k}, k \in \mathbb{Z}; \psi_{jk}, j \geq j_0, k \in \mathbb{Z}\}$ 是 $L_2(\mathcal{R}, leb)$ 的一个标准正交基[18]。参考 Meyer (1992, 定理 3.3)。

对于 $j \geq 0$ 以及 $0 \leq k \leq 2^j - 1$，我们用下式表示定义在 $[0,1]$ 上的周期化小波

$$\phi^*_{jk}(x) = 2^{j/2} \sum_{l \in \mathbb{Z}} \phi(2^j x + 2^j l - k), \quad \psi^*_{jk}(x) = 2^{j/2} \sum_{l \in \mathbb{Z}} \psi(2^j x + 2^j l - k), \quad x \in [0,1].$$

[18] $\int_{\mathcal{R}} \psi_{jk}(x) \psi_{jk}(x) dx = 1$；同时对于 $j \neq j'$ 或 $k \neq k'$，有 $\int_{\mathcal{R}} \psi_{jk}(x) \psi_{j'k'}(x) dx = 0$；此外 $\int_{\mathcal{R}} \phi_{j_0 k}(x) \phi_{j_0 k}(x) dx = 1$，以及对于 $k'$，有 $\int_{\mathcal{R}} \phi_{j_0 k}(x) \phi_{j_0 k'}(x) dx = 0$；此外，对于 $j \geq j_0$，有 $\int_{\mathcal{R}} \phi_{j_0 k}(x) \psi_{jk'}(x) dx = 0$。

对于 $j_0 \geq 0$，集合 $\{\phi^*_{j_0k}, k = 0, \cdots, 2^{j_0} - 1; \psi^*_{jk}, j \geq j_0, k = 0, \ldots, 2^j - 1\}$ 是 $L_2([0,1], leb)$ 的一个标准正交基（参考 Daubechies, 1992）我们考虑由这个小波基张成的有限维线性空间。对一个整数 $J_n > j_0$，定义

$$\text{Wav}(m, 2^{J_n}) = \left\{ \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0k}\phi^*_{j_0k}(x) + \sum_{j=j_0}^{J_n-1} \sum_{k=0}^{2^j-1} \beta_{jk}\psi^*_{jk}(x),\, x \in [0,1] :\, \alpha_{j_0k}, \beta_{jk} \in \mathcal{R} \right\}$$

或者等价地 (参考 Meyer, 1992),

$$\text{Wav}(m, 2^{J_n}) = \left\{ \sum_{k=0}^{2^{J_n}-1} \alpha_k\phi^*_{J_nk}(x),\, x \in [0,1] :\, \alpha_k \in \mathcal{R} \right\}.$$

**张量** (Tensor) **积空间** 我们用 $\mathcal{U}_\ell$, $1 \leq \ell \leq d$ 表示在欧几里得空间上的紧凑集合，用 $\mathcal{U} = \mathcal{U}_1 \times ... \times \mathcal{U}_d$ 表示它们的笛卡儿积。对于 $1 \leq \ell \leq d$，我们用 $\mathbb{G}_\ell$ 表示一个定义在 $\mathcal{U}_\ell$ 上的函数的线性空间，其中任何一个 $\mathbb{G}_\ell$ 都可以是上面提到的筛空间。$\mathbb{G}_1, \ldots, \mathbb{G}_d$ 的张量乘积 $\mathbb{G}$ 定义为由函数 $\prod_{\ell=1}^d g_\ell(x_\ell)$ 张成的在 $\mathcal{U}$ 上的函数空间，其中对 $1 \leq \ell \leq d$ 有 $g_\ell \in \mathbb{G}_\ell$。值得注意的是 $\dim(\mathbb{G}) = \prod_{\ell=1}^d \dim(\mathbb{G}_\ell)$。张量乘积是一种由单变量函数的线性筛生成多元函数的线性筛的标准方法。因为其简单又易于实施，线性筛被广泛的应用在具体问题中。此外，线性筛也可以用来近似 Hölder 空间 $\Lambda^p(\mathcal{X})$ 中的函数。下文中我们用 $\theta$ 表示一个实值函数，其定义域 $\mathcal{X} \subset \mathcal{R}^d$ 有界。用 $\|\theta\|_\infty \equiv \sup_{x\in\mathcal{X}} |\theta(x)|$ 表示 $\theta$ 的 $L_\infty$ 范数，用 $\|\theta\|_{2,leb} \equiv \{\int_{\mathcal{X}} [\theta(x)]^2 dx/\text{vol}(\mathcal{X})\}^{1/2}$ 表示相对 $\mathcal{X}$ 的勒贝格测度的 $L_2$ 比例范数。定义在 $L_\infty(\mathcal{X}, leb)$-范数和 $L_2(\mathcal{X}, leb)$-范数下的 $\theta_o \in \Lambda^p(\mathcal{X})$ 的筛近似误差为:

$$\rho_{\infty n} \equiv \inf_{g\in\Theta_n} ||g - \theta_o||_\infty \quad \text{and} \quad \rho_{2n} \equiv \inf_{g\in\Theta_n} ||g - \theta_o||_{2,leb}.$$

显然有 $\rho_{2n} \leq \rho_{\infty n}$。对于多元函数 $\theta_o \in \Theta = \Lambda^p([0,1]^d)$，我们考虑其张量积线性筛空间 $\Theta_n$，这里我们对常用的一元线性近似空间 $\Theta_{n1}, \ldots, \Theta_{nd}$ 作张量积构造 $\Theta_n$。让 $\dim(\Theta_n) = k_n$，同时 $[p]$ 为满足 $[p] < p$ 的最大的整数。那么我们得到下列 $\theta_o \in \Lambda^p([0,1]^d)$ 的张量积筛近似误差收敛速度:

**多项式** 如果每个 $\Theta_{n\ell} = \text{Pol}(J_n)$，那么 $\rho_{\infty n} = O(J_n^{-p}) = O(k_n^{-p/d})$（参考 Timan, 1963 的 5.3.2 节）。

**三角多项式** 如果 $\theta_o$ 可以被扩展为周期性函数, 同时如果每个 $\Theta_{n\ell} = \text{TriPol}(J_n)$, 那么 $\rho_{\infty n} = O(J_n^{-p}) = O(k_n^{-p/d})$（参考 Timan, 1963 的 5.3.1 节）

**样条函数** 如果每个 $\Theta_{n\ell} = \text{Spl}(r, J_n)$，其中 $r \geq [p] + 1$，那么 $\rho_{\infty n} = O(J_n^{-p}) = O(k_n^{-p/d})$（参考 Schumaker, 1981 的 (13.69) 和定理 12.8）。

**正交小波** 如果每个 $\Theta_{n\ell} = \text{Wav}(m, 2^{J_n})$，其中 $m > p$，那么 $\rho_{\infty n} = O(2^{-pJ_n}) = O(k_n^{-p/d})$（参考 Meyer, 1992 的命题 2.5）

### 2.3.2 加权平滑函数类和（有限维）线性筛

在半非参数计量经济学应用中，有时我们所关注的未知参数是具有无界支集的函数。在这里我们提出两个有限维线性筛，它们可以很好地近似具有无界支集的函数。下文中我们用 $L_p(\mathcal{X}, \omega)$, $1 \leq p < \infty$ 表示满足对给定的平滑加权函数 $\omega : \mathcal{X} \mapsto (0, \infty)$ 有 $\int_{\mathcal{X}} |h(x)|^p \omega(x) dx < \infty$ 的实值函数 $h$ 的空间。

**Hermite 多项式** Hermite 多项式级数 $\{H_k : k = 1, 2, .....\}$ 是有 $\omega(x) = \exp\{-x^2\}$ 的 $L_2(\mathcal{R}, \omega)$ 空间的一个标准正交基，可以通过对多项式级数 $\{x^{k-1} : k = 1, 2, .....\}$ 使用 Gram-Schmidt 方法展开得到，其中

内积的定义为 $\langle f,g\rangle_\omega = \int_{\mathcal{R}} f(x)g(x)\exp\{-x^2\}dx$。也就是说，$H_1(x) = 1/\sqrt{\int_{\mathcal{R}}\exp\{-x^2\}dx} = \pi^{-1/4}$，以及对所有 $k \geq 2$，

$$H_k(x) = \frac{x^{k-1} - \sum_{j=1}^{k-1}\langle x^{k-1}, H_j\rangle_\omega H_j(x)}{\sqrt{\int_{\mathcal{R}}[x^{k-1} - \sum_{j=1}^{k-1}\langle x^{k-1}, H_j\rangle_\omega H_j(x)]^2\exp\{-x^2\}dx}}.$$

用 HPol($J_n$) 表示定义在 $\mathcal{R}$ 上的 $J_n$ 阶或更低的 Hermite 多项式空间：

$$\text{HPol}(J_n) = \left\{\sum_{k=1}^{J_n+1} a_k H_k(x)\exp\{-\frac{x^2}{2}\},\ x\in\mathcal{R}:\ a_k\in\mathcal{R}\right\}.$$

那么随着 $J_n \to \infty$，在 $L_2(\mathcal{R}, leb)$ 中的任何函数都可以用 HPol($J_n$) 来近似。

当 HPol($J_n$) 筛被用作近似一个未知的 $\sqrt{\theta_o}$，其中 $\theta_o$ 是定义在 $\mathcal{R}$ 上的概率密度函数，这时相应的筛最大似然估计量在计量经济学中也被称作 SNP。参考 Gallant and Nychka (1987), Gallant and Tauchen (1989) 以及 Coppejans and Gallant (2002)。

**Laguerre 多项式** Laguerre 多项式序列 $\{L_k : k = 1, 2, .....\}$ 是 $L_2([0,\infty),\omega)$（其中 $\omega(x) = \exp\{-x\}$）的一个标准正交基。我们可以通过对多项式系列 $\{x^{k-1} : k = 1, 2, .....\}$ 应用 Gram-Schmidt 方法来得到这一基函数系列，其中内积的定义为 $\langle f,g\rangle_\omega = \int_0^\infty f(x)g(x)\exp\{-x\}dx$。用 LPol($J_n$) 表示定义在 $[0,\infty)$ 上的阶数小于或等于 $J_n$ 的 Laguerre 多项式空间：

$$\text{LPol}(J_n) = \left\{\sum_{k=1}^{J_n+1} a_k L_k(x)\exp\{-\frac{x}{2}\},\ x\in[0,\infty):\ a_k\in\mathcal{R}\right\}.$$

那么随着 $J_n \to \infty$，任意在 $L_2([0,\infty), leb)$ 中的函数都可以用 LPol($J_n$) 筛来近似。

### 2.3.3 其他平滑函数类和（有限维）非线性筛

非线性筛也可以用作筛极值估计。计量经济学中常见的一类非线性筛是单隐层前馈人工神经网络(ANN)。这里我们介绍三种典型的 ANN 类别。其他类别可以参考 Hornik et al. (1994)。

**Sigmoid ANN** 定义：

$$\text{sANN}(k_n) = \left\{\sum_{j=1}^{k_n}\alpha_j S(\gamma_j^{'}x + \gamma_{0,j}) : \gamma_j\in\mathcal{R}^d, \alpha_j, \gamma_{0,j}\in\mathcal{R}\right\},$$

其中 $S : \mathcal{R}\to\mathcal{R}$ 是一个 sigmoid 激活函数，也就是满足 $\lim_{u\to-\infty} S(u) = 0$ 和 $\lim_{u\to\infty} S(u) = 1$ 的有界非减函数。一些常见的 sigmoid 激活函数包括

- 海维赛德 (heaviside) $S(u) = 1\{u \geq 0\}$;
- 逻辑（logistic）$S(u) = 1/(1+\exp\{-u\})$;
- 双曲正切（hyperbolic tangent）$S(u) = (\exp\{u\} - \exp\{-u\})/(\exp\{u\} + \exp\{-u\})$;
- 高斯 S 形（Gaussian sigmoid）$S(u) = (2\pi)^{-1/2}\int_{-\infty}^{u}\exp(-y^2/2)dy$;

- 余弦平压（cosine squasher）$S(u) = \frac{1+\cos(u+3\pi/2)}{2} 1\{|u| \leq \pi/2\} + 1\{u > \pi/2\}$.

用 $\mathcal{X}$ 表示在 $\mathcal{R}^d$ 中的一个紧集，用 $C(\mathcal{X})$ 表示从 $\mathcal{X}$ 到 $\mathcal{R}$ 的连续函数映射空间。Gallant and White (1988a) 首先证明了在上确界范数下，使用 cosine squasher 激活函数的 sANN 筛在 $C(\mathcal{X})$ 中稠密。Cybenko (1989) 和 Hornik et al. (1989) 证明了在上确界范数下，使用任意 S 形激活函数的 sANN$(k_n)$ 在 $C(\mathcal{X})$ 中稠密。

让 $\mathcal{H} = \{h \in L_2(\mathcal{X}, leb) : \int_{\mathcal{R}^d} |w||\widetilde{h}(w)|dw < \infty\}$。这意味着当且仅当其平方可积且傅里叶变换 $\widetilde{h}$ 具有有限的一阶矩时，$h \in \mathcal{H}$。这里 $\widetilde{h}(w) \equiv \int \exp(-iwx)h(x)dx$ 是 $h$ 的傅里叶变换。Barron (1993) 证明了对任意 $h_o \in \mathcal{H}$，在 $L_2(\mathcal{X}, leb)$-范数下的 sANN$(k_n)$ 筛近似误差速率 $\rho_{2n}$ 不慢于 $O([k_n]^{-1/2})$，之后 Makovoz (1996) 将使用 heaviside sigmoid 函数的 sANN$(k_n)$ 的速率改进为 $O([k_n]^{-1/2-1/(2d)})$，Chen and White (1999) 则将使用一般 sigmoid 函数的 sANN$(k_n)$ 的速率改进为 $O([k_n]^{-1/2-1/(d+1)})$。

**一般 ANN** 定义:

$$\text{gANN}(k_n) = \left\{ \sum_{j=1}^{2^r k_n} \alpha_j [\max\{|\gamma_j|_e, 1\}]^{-m} \psi(\gamma_j' x + \gamma_{0,j}) : \gamma_j \in \mathcal{R}^d, \alpha_j, \gamma_{0,j} \in \mathcal{R} \right\},$$

这里 $\psi : \mathcal{R} \to \mathcal{R}$ 是任意非固定阶多项式的激活函数。特别的，我们常常用 $\psi$ 表示在 Hölder 空间 $\Lambda^m(\mathcal{R})$ 中，并且对某个 $r \geq m$ 满足 $0 < \int_{\mathcal{R}} |D^r \psi(x)|dx < \infty$ 的平滑函数。这一定义包括了上述所有 sigmoid 激活函数（$m = 0$ 和 $r = 1$），更多例子可以参考 Hornik et al. (1994)。

让 $\mathcal{H} = \{h \in L_2(\mathcal{X}, \mu) : h(x) = \int \exp(ia'x)d\sigma_h(a), \int_{\mathcal{R}^d} [\max\{|a|_e, 1\}]^{m+1} d|\sigma_h|_{tv}(a) < \infty\}$，其中 $\sigma_h$ 是一个复数值（相对实数值）测度，并且 $|\sigma_h|_{tv}$ 表示 $\sigma_h$ 的总变差。用 $W_2^m(\mathcal{X}, \mu)$ 代表加权 Sobolev 函数空间，其中的函数及其所有偏导数（最高到 $m$ 次）都在某有限测度 $\mu$ 下 $L_2(\mathcal{X}, \mu)$ 可积。众所周知在 $\mathcal{H}$ 中的函数同时也属于 $W_2^m(\mathcal{X}, \mu)$。用 $||h||_{m,\mu} = \{\int h(x)^2 d\mu(x) + \int |D^m h(x)|_e^2 d\mu(x)\}^{1/2}$ 表示加权 Sobolev 范数。Hornik et al. (1994) 证明了对任意 $h_o \in \mathcal{H}$，在加权 Sobolev 范数 $(||\cdot||_{m,\mu})$ 下 gANN$(k_n)$ 筛近似误差速率不慢于 $O([k_n]^{-1/2})$，这一结果之后由 Chen and White (1999) 改进为 $O([k_n]^{-1/2-1/(d+1)})$。

**高斯径向基 ANN** 让 $\mathcal{X} = \mathcal{R}^d$。定义:

$$\text{rbANN}(k_n) = \left\{ \alpha_0 + \sum_{j=1}^{k_n} \alpha_j G\left( \frac{\{(x-\gamma_j)'(x-\gamma_j)\}^{1/2}}{\sigma_j} \right) : \gamma_j \in \mathcal{R}^d, \alpha_j, \sigma_j \in \mathcal{R}, \sigma_j > 0 \right\},$$

其中 $G$ 是标准高斯密度函数。用 $W_1^m(\mathcal{X})$ 表示 Sobolev 函数空间，其中的函数以及它们的所有偏导数（最高到 $m$ 次）都是 $L_1(\mathcal{X}, leb)$ 可积。Meyer (1992) 证明了 rbANN$(k_n)$ 在平滑函数类 $W_1^m(\mathcal{X})$ 中稠密。Girosi (1994) 证明了对任意 $h_o \in \mathcal{H}$，在 $L_2(\mathcal{X}, leb)$ 范数下的 rbANN$(k_n)$ 筛近似误差速率 $\rho_{2n}$ 不慢于 $O([k_n]^{-1/2})$，这一结果后来由 Chen et al. (2001) 改进为 $O([k_n]^{-1/2-1/(d+1)})$。

其他关于非线性筛的例子包括使用数据导向节点选择（或自由节点）的样条筛，以及具有阈值的小波筛。非线性筛更灵活，可以获得比线性筛更好的近似性质；具体对比可以参考 Chen and Shen (1998)。

### 2.3.4 无限维（非线性）筛和惩罚方法

上文中列出的有限维截断级数是最常用的筛分空间。然而，筛极值估计的一般理论也可以允许无限维的筛分空间。例如，考虑平滑函数类 $\Theta = \Lambda^p(\mathcal{X})$，其中 $\mathcal{X} = [0,1]$, $p > 1/2$。很清楚可以看出任意函数 $\theta \in \Theta$ 都可以表示为一个无限傅里叶级数 $\theta(x) = \sum_{k=1}^{\infty}[a_k \cos(kx) + b_k \sin(kx)]$，并且其具有 $\gamma \in (0, p]$ 分数阶幂的导数也可以用傅里叶级数来定义：

$$\theta^{(\gamma)}(x) = \sum_{k=1}^{\infty} k^{\gamma} \left[ (a_k \cos\frac{\pi\gamma}{2} + b_k \sin\frac{\pi\gamma}{2}) \cos(kx) + (b_k \cos\frac{\pi\gamma}{2} - a_k \sin\frac{\pi\gamma}{2}) \sin(kx) \right].$$

类似的，任意函数 $\theta \in \Theta = \Lambda^p(\mathcal{X})$ 及其分数阶导数都可以表示为无限样条和小波级数，具体可以参考 Meyer (1992)。对某 $p > 1/2$ 及整数 $q \geq 1$，定义 $\text{pen}(\theta) = (\int_{\mathcal{X}} |\theta^{(p)}(x)|^q dx)^{1/q}$。那么我们可以选择筛 $\Theta_n = \{\theta \in \Theta : \text{pen}(\theta) \leq b_n\}$，其中随着 $n \to \infty$，缓慢地有 $b_n \to \infty$，具体可以参考 Shen (1997)。这里 $q$ 的选择一般情况下跟目标函数 $\widehat{Q}_n(\theta)$ 相关，例如对条件矩回归 (Wahba, 1990) 会选择 $q = 2$，(Koenker et al., 1994) 中选择 $q = 1$，对分位回归 (Koenker and Mizera, 2003) 则选择总变差范数。

一般地，如果参数空间 $\Theta$ 是例如 Hölder, Sobolev 或 Besov 空间的典型函数空间，那么任意 $\theta \in \Theta$ 函数都可以表示为某已知 Riesz 基 $\{B_k(\cdot)\}_{k=1}^{\infty}$ 的无限级数。一个无限维筛空间可能采取下列形式：

$$\Theta_n = \{\theta \in \Theta : \theta(\cdot) = \sum_{k=1}^{\infty} a_k B_k(\cdot),\ \text{pen}(\theta) \leq b_n\} \quad \text{其中} b_n \to \infty \text{ 缓慢地}, \tag{2.18}$$

$\text{pen}(\theta)$ 是平滑（或粗糙）惩罚项。

**备注 2.2.** 当 $\widehat{Q}_n(\theta)$ 是凹的且 $\text{pen}(\theta)$ 是凸的时，筛极值估计量 $\sup_{\theta \in \Theta_n} \widehat{Q}_n(\theta)$（$\Theta_n$ 由 (2.18) 给出）同下列“惩罚极值估计量”等价

$$\max_{\theta \in \Theta} \{\widehat{Q}_n(\theta) - \lambda_n \text{pen}(\theta)\} \tag{2.19}$$

其中我们选择拉格朗日乘子 $\lambda_n$ 以满足 $\text{pen}(\widehat{\theta}) = b_n$。参考 Eggermont and LaRiccia (2001, 章节 1.6)。

### 2.3.5 保形筛

有很多筛是可以保持未知目标函数的（已知）形状的，比如非负性、单调性和凸性。有关保形样条和多项式筛可以参考 DeVore (1977a, 1977b)；保形小波筛可以参考 Anastassiou and Yu (1992a, 1992b) 以及 Dechevsky and Penev (1997)。这里我们介绍一种保形筛。

**基数 B-样条小波** 定义阶数 $r \geq 1$ 的基数 B-样条小波为

$$B_r(x) = \frac{1}{(r-1)!} \sum_{j=0}^{r} (-1)^j \binom{r}{j} [\max(0, x-j)]^{r-1}, \tag{2.20}$$

该函数支集为 $[0, r]$，以 $r/2$ 为中心对称，并且是最高次为 $r-1$ 的分段多项式。它满足对所有 $x \in \mathcal{R},$，有 $B_r(x) \geq 0$, $\sum_{k=-\infty}^{+\infty} B_r(x-k) = 1$，该性质对同未知目标函数的形状保持一致非常关键。它的导数满足 $\frac{\partial}{\partial x} B_r(x) = B_{r-1}(x) - B_{r-1}(x-1)$。关于如何用递归方法构造基数 B-样条小波以及其具体性质，可以参考 Chui (1992, 第 4 章)。

我们可以通过如下方法为 $L_2(\mathcal{R}, leb)$ 构造基数 B-样条小波基。让 $\phi_r(x) = B_r(x)$ 表示父小波（或尺度函数）。那么存在“唯一”母小波函数 $\psi_r$，其最小支集为 $[0, 2r-1]$，具体表达式为

$$\psi_r(x) = \sum_{\ell=0}^{3r-2} q_\ell B_r(2x-\ell), \quad q_\ell = (-1)^\ell 2^{1-r} \sum_{j=0}^{r} \left( \begin{array}{c} r \\ j \end{array} \right) B_{2r}(\ell+1-j).$$

让

$$\phi_{r,jk}(x) = 2^{j/2} B_r(2^j x - k), \quad \psi_{r,jk}(x) = 2^{j/2} \psi_r(2^j x - k), \quad x \in \mathcal{R},$$

那么对整数 $j_0 \geq 0$, $\{\phi_{r,j_0 k}, k \in \mathbb{Z}; \psi_{r,jk}, j \geq j_0, k \in \mathbb{Z}\}$ 是 $L_2(\mathcal{R}, leb)$ 的一个 Riesz 基。此外，任意 $L_2(\mathcal{R}, leb)$ 中的函数 $g$ 都有如下样条-小波 $m = r-1$ 正则多尺度展开：

$$g(x) = \sum_{k=-\infty}^{\infty} a_{j_0 k} 2^{j_0/2} B_r(2^{j_0} x - k) + \sum_{j=j_0}^{\infty} \sum_{k=-\infty}^{\infty} b_{jk} \psi_{r,jk}(x), \quad x \in \mathcal{R},$$

参考 Chui (1992, 章节 6)。对于一个正整数 $J_n > j_0 = 0$，设定

$$\text{SplWav}(r-1, 2^{J_n}) = \left\{ \sum_{k=-\infty}^{\infty} a_{0k} B_r(x-k) + \sum_{j=0}^{J_n-1} \sum_{k=-\infty}^{\infty} \beta_{jk} \psi_{r,jk}(x),\, x \in \mathcal{R} :\ a_{0k}, \beta_{jk} \in \mathcal{R} \right\}$$

或等价地，[19]

$$\text{SplWav}(r-1, 2^{J_n}) = \left\{ \sum_{k=-\infty}^{\infty} \alpha_k 2^{J_n/2} B_r(2^{J_n} x - k),\, x \in \mathcal{R} :\ \alpha_k \in \mathcal{R} \right\}.$$

任意在 $\mathcal{R}$ 上的非递减连续函数可以通过 $\text{SplWav}(r-1, 2^{J_n})$ 筛来近似，其中 $\{\alpha_k\}$ 是一列非递减的数组。特别的，让

$$\text{MSplWav}(r-1, 2^{J_n}) = \left\{ g(x) = \sum_{k=-\infty}^{\infty} \alpha_k 2^{J_n/2} B_r(2^{J_n} x - k + \frac{r}{2}) : \alpha_k \leq \alpha_{k+1} \right\}$$

表示单调样条小波筛。那么对定义在 $\mathcal{R}$ 上的任意有界非减连续函数 $\theta_o$，使用上确界范数下的 $\text{MSplWav}(r-1, 2^{J_n})$ $(r \geq 1)$ 筛近似误差速度是 $O(2^{-J_n})$；对定义在 $\mathcal{R}$ 上的任意有界非减连续函数 $\theta_o$，使用上确界范数下的 $\text{MSplWav}(r-1, 2^{J_n})$ $(r \geq 2)$ 筛近似误差速度是 $O(2^{-2J_n})$；参考 Anastassiou and Yu (1992a)。

### 2.3.6 筛空间的选择

筛空间 $\Theta_n = B \times \mathcal{H}_n$ 的选择取决于它对 $\Theta = B \times \mathcal{H}$ 的近似程度以及是否令 $\max_{\theta \in \Theta_n} \widehat{Q}_n(\theta)$ 易于计算。

一般的，当筛空间 $\Theta_n = B \times \mathcal{H}_n$ 是无约束的有限维线性空间时，计算 $\max_{\theta \in \Theta_n} \widehat{Q}_n(\theta)$ 会较为简便。此外，如果目标函数 $\widehat{Q}_n(\theta)$ 是凹函数，我们可以直接应用本文 2.2.2 节中提到的凹拓展线性模型的筛级数估计量。

然而，在决定是否选择某个筛空间时我们不能只关心计算是否简便。这是由于筛估计量的大样本性质也取决于所选择的筛的近似特征。不幸的是，相比非线性筛空间，有限维线性筛并不总是具有更好的近似特征。例如，让我们来考虑估计一个多元条件矩函数 $h_o(\cdot) = E[Y_t | X_t = \cdot] \in \Theta$。用 $\Theta_n$ 表示筛空间。那么

[19]有关近似 $\mathcal{R}$ 上的二次可导函数的这类筛的一般性质可以参考 Chen et al. (1998)。

$\widehat{\theta} = \widehat{h} = \arg\max_{h\in\Theta_n} \frac{-1}{n}\sum_{t=1}^{n}[Y_t - h(X_t)]^2$ 是 $h_o$ 的筛 M-估计量。如果 $\Theta = \Lambda^p([0,1]^d)$ $(p > d/2)$ 是一个 $p-$ 平滑函数空间，那么我们可以选择章节 2.3.1 中提到的任何有限维线性筛空间作为 $\Theta_n$。这时得到的估计量 $\widehat{h}$ 是一个级数估计量。然而，如果如 2.3.3 中定义的 $\Theta = W_1^1([0,1]^d)$，这时选择非线性高斯径向基 ANN 作为筛空间 $\Theta_n$ 会更好，所得到的估计量仍然是筛 M-估计量，但不再是级数估计量。更多例子请参考本文第三章。

筛 $\Theta_n$ 对 $\Theta$ 的近似效果往往取决于 $\Theta$ 空间内函数的支集，平滑度，形状约束，以及由计量模型决定的函数结构，比如可加性，非负性，排除性约束等。例如，Hermite 多项式筛可以近似具有无界支集和相对薄的尾部的多元未知平滑密度函数，但幂级数筛和傅立叶级数筛却很难做到这一点。这就是为什么 Gallant and Nychka (1987) 选取 Hermite 多项式筛 MLE 估计量，因为他们希望近似具有无界支集的多元未知平滑密度函数，并且能够包括多元正态分布（薄尾）在其中。另一个例子是一阶单调样条筛可以很好地近似任何有界单调但不可导函数，而三阶基数 B-样条小波筛可以很好地近似任何有界单调可导函数。在例 2.1 中，Heckman and Singer (1984, 300-301) 不想对潜在随机因子的分布函数 $h(\cdot)$ 添加任何假设，于是他们用一阶单调样条筛来近似这个分布函数。为了估计在完全非参标量扩散模型下的条件期望算子的第一特征函数，Chen et al. (1998) 使用了保形三阶基数 B-样条小波筛来近似未知的第一特征函数，这是由于已经知道该函数是单调且二次连续可导。最后一个例子是 Chen and Ludvigson (2003)。例 2.3 介绍了用筛 MD 方法估计半非参外部习惯模型 (2.7)。文章使用了具有逻辑激活函数的 sANN 筛来近似未知习惯函数 $H(C_t, C_{t-1}, ..., C_{t-L}) = C_t h\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)$。这部分是由于当 $L \geq 3$ 时，未知平滑函数 $h : \mathcal{R}^L \to [0,1)$ 可以用 sANN 筛很好地近似；部分是由于当使用具有逻辑激活函数的 sANN 筛来近似 $h\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)$ 时，添加习惯约束 $0 \leq H(C_t, C_{t-1}, ..., C_{t-L}) < C_t$ 相对更简单。

为了使筛分估计具有足够快的收敛速度，选择具有良好近似误差率和受控制的复杂度的筛非常重要。[20]虽然如此，在一些计量经济学应用中，关于未知函数我们仅仅知道其平滑度和支集，这种情况下具体筛空间的选择不再重要，更关键的是选择筛空间以达到理想的近似误差收敛速度。

## 2.4 蒙特卡洛（Monte Carlo）研究

为了说明如何实现筛极值估计方法，我们在本节介绍一个用 Matlab 和 Fortran 实现的 Monte Carlo 模拟研究。真实模型是：$Y_1 = X_1\beta_o + h_{o1}(Y_2) + h_{o2}(X_2) + U$ 其中 $\beta_o = 1$, $h_{o1}(Y_2) = 1/[1 + \exp\{-Y_2\}]$, $h_{o2}(X_2) = \log(1 + X_2)$。我们假设 $Y_2$ 是内生的，同时 $Y_2 = X_1 + X_2 + X_3 + R \times U + e$，这里 $R = 0.9$ (强相关) 或 0.1 (弱相关)。假设自变量 $X_1$, $X_2$, $X_3$ 相互独立并且在 $[0,1]$ 上均匀分布，$e$ 跟 $(X, U)$ 独立且满足正态分布，其均值为 0，方差为 0.1。(我们也尝试过 $E[e^2] = 0.05, 0.25$，模拟结果同 $E[e^2] = 0.1$ 下得到的结论非常类似，所以我们不再赘述)

给定 $X = (X_1, X_2, X_3)'$，$U$ 的条件分布遵从正态分布，其均值为 0，方差为 $(X_1^2 + X_2^2 + X_3^2)/3$。让 $Z = (Y_1, Y_2, X')'$。我们生成一个样本量为 1000 的随机样本数据 $\{Z_i\}_{i=1}^n$。假设一个计量经济学家观察到这

[20]这一点从稍后第 3 节讨论的大样本理论可以清楚地看出。

个模拟的数据集 $\{Z_i\}_{i=1}^n$，希望通过这个数据来估计满足下列条件矩约束的真实参数 $\theta_o = (\beta_o, h_{o1}, h_{o2})'$：

$$E[Y_{1i} - \{X_{1i}\beta_o + h_{o1}(Y_{2i}) + h_{o2}(X_{2i})\}|X_i] = 0. \tag{2.21}$$

上述模型概括了 Ai and Chen (2003) 中提到的部分线性工具变量回归模型 $E[Y_1 - \{X_1\beta_o + h_{o1}(Y_2)\}|X] = 0$，并将其推广为部分可加工具变量回归模型。由于 $h_{o1}(Y_2)$ 是内生变量 $Y_2$ 的未知函数，这两种模型都属于不适定求逆问题。

让 $\rho(Z,\theta) = Y_1 - \{X_1\beta + h_1(Y_2) + h_2(X_2)\}$ 其中 $\theta = (\beta, h_1, h_2)'$。我们称参数 $\theta_o = (\beta_o, h_{o1}, h_{o2})'$ 被识别，如果满足仅当 $\theta = \theta_o$ 时，才有 $E[\rho(Z,\theta)|X] = 0$。识别 $\theta_o$ 的一个充分条件是 $Var(X_1) > 0$，$h_1(y_2)$ 是满足 $\sup_{y_2}|h_1(y_2)| \leq 1$ 的一个有界函数，以及 $h_2(0.5) = \log(3/2)$。特别地，我们假设 $\theta_o = (\beta_o, h_{o1}, h_{o2})' \in \Theta = B \times \mathcal{H}_1 \times \mathcal{H}_2$，其中 $B$ 是 $\mathcal{R}$ 上的一个紧区间，$\mathcal{H}_1 = \{h_1 \in C^2(\mathcal{R}) : \sup_{y_2}|h_1(y_2)| \leq 1, \int [D^2 h_1(y_2)]^2 dy_2 < \infty\}$ 同时 $\mathcal{H}_2 = \{h_2 \in C^2([0,1]) : h_2(0.5) = \log(3/2), \int [D^2 h_2(x_2)]^2 dx_2 < \infty\}$。

由于这类模型 (2.21) 属于满足 $E[\rho(Z,\theta_o)|X] = 0$ 的第二子类条件矩约束模型 (2.8)，我们可以应用筛 MD 标准 (2.12) 来估计 $\theta_o = (\beta_o, h_{o1}, h_{o2})$。我们选择 $\Theta_n = B \times \mathcal{H}_{1n} \times \mathcal{H}_{2n}$ 作为筛空间，其中

$$\mathcal{H}_{1n} = \left\{ h_1(y_2) = \Pi_1' B^{k_{1,n}}(y_2) : \int [D^2 h_1(y_2)]^2 dy_2 \leq c_1 \log n \right\},$$

$B^{k_{1,n}}(y_2)$ 可以是具有等间距节点（根据 $Y_2$ 数据的分位数）的多项式样条基，或者三阶基数 B-样条基，或者 Hermite 多项式基，[21] $\dim(\Pi_1) = k_{1,n}$ 是 $h_1$ 的未知筛系数的个数。类似的，

$$\mathcal{H}_{2n} = \left\{ h_2(x_2) = \Pi_2' B^{k_{2,n}}(x_2) : \int [D^2 h_2(x_2)]^2 dx_2 \leq c_2 \log n,\ h_2(0.5) = \log(3/2) \right\},$$

$B^{k_{2,n}}(x_2)$ 可以是具有等间距节点（根据 $X_2$ 数据的分位数）的多项式样条基，或者三阶基数 B-样条基，$\dim(\Pi_2) = k_{2,n}$ 是 $h_2$ 的未知筛系数的个数。在这个研究中，我们尝试 $k_{1,n} = 4, 5, 6, 8$ 和 $k_{2,n} = 4, 5, 6$。

这里我们只考虑使用单位矩阵作为加权 $\widehat{\Sigma}(X) = I$ 的筛 MD(2.12) 估计，[22] 同时使用级数最小二乘估计量 $\widehat{m}(X,\theta)$ 来估计条件均值函数 $E[\rho(Z,\theta)|X],$，这是目标函数为

$$\min_{\beta\in B, h_1\in\mathcal{H}_{1n}, h_2\in\mathcal{H}_{2n}} \frac{1}{n}\sum_{i=1}^n \{\widehat{m}(X_i,\theta)\}^2, \quad \text{with}$$

$$\widehat{m}(X,\theta) \quad = \quad \sum_{j=1}^n [Y_{1j} - \{X_{1j}\beta + h_1(Y_{2j}) + h_2(X_{2j})\}] p^{k_{m,n}}(X_j)'(P'P)^- p^{k_{m,n}}(X),$$

在具体模拟中 $p^{k_{m,n}}(X)$ 为四阶多项式样条筛，基函数包括 $\{1, X_1, X_1^2, X_1^3, X_1^4, [\max(X_1 - 0.5, 0)]^4, X_2, X_2^2, X_2^3, X_2^4, [\max(X_2 - 0.5, 0)]^4, X_3, X_3^2, X_3^3, X_3^4, [\max(X_3 - 0.1, 0)]^4, [\max(X_3 - 0.25, 0)]^4, [\max(X_3 - 0.5, 0)]^4, [\max(X_3 - 0.75, 0)]^4, [\max(X_3 - 0.90, 0)]^4, X_1X_3, X_2X_3, X_1[\max(X_3 - 0.25, 0)]^4, X_2[\max(X_3 - 0.25, 0)]^4, X_1[\max(X_3 - 0.75, 0)]^4, X_2[\max(X_3 - 0.75, 0)]^4\}$。我们注意到上述方法等价于有约束条件的两阶段最小二乘法 (2SLS)：共有 $k_{m,n} = 26$ 个工具变量和 $\dim(\Theta_n) = 1 + k_{1,n} + k_{2,n}$ $(< k_{m,n})$ 未知参数：

$$\min_{\beta\in B, h_1\in\mathcal{H}_{1n}, h_2\in\mathcal{H}_{2n}} [\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\Pi]' P(P'P)^- P'[\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\Pi],$$

[21] 更多关于 $\mathcal{H}_{1n}$ 选择的讨论可以参考 Blundell et al. (2001)。

[22] 使用最优加权矩阵的筛 MD 估计具体细节见本文 4.3 节或 Ai and Chen (2003)。

其中 $\mathbf{Y}_1 = (Y_{11}, ..., Y_{1n})'$, $\mathbf{X}_1 = (X_{11}, ..., X_{1n})'$, $\Pi = (\Pi_1', \Pi_2')'$, $\mathbf{B}_1 = (B^{k_{1,n}}(Y_{21}), ..., B^{k_{1,n}}(Y_{2n}))'$, $\mathbf{B}_2 = (B^{k_{2,n}}(X_{21}), ..., B^{k_{2,n}}(X_{2n}))'$ 以及 $\mathbf{B} = (\mathbf{B}_1', \mathbf{B}_2')'$。

由于 $\rho(Z, \theta)$ 是 $\theta = (\beta, h_1, h_2)'$ 的线性函数，在这个模型中联立筛 MD 估计法等价于代换（profile）筛 MD 估计法。我们可以先计算一个 $h_1(y_2) + h_2(x_2)$ 的代换筛估计量。也就是说，对任意 $\beta$，我们通过解决下面的带函数平滑度约束条件的优化问题来计算筛系数 $\Pi$:

$$\min_{\Pi: \int [D^2 h_\ell(y)]^2 dy \le c_\ell \log n, \ell = 1,2} [\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\Pi]' P(P'P)^- P' [\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\Pi] \tag{2.22}$$

对某上界 $c_\ell > 0, \ell = 1, 2$。用 $\widetilde{\Pi}(\beta)$ 表示 (2.22) 的解，$\widetilde{h}_1(y_2; \beta) + \widetilde{h}_2(x_2; \beta) = (B^{k_{1,n}}(y_2)', B^{k_{2,n}}(x_2)')\widetilde{\Pi}(\beta)$ 表示 $h_1(y_2) + h_2(x_2)$ 的代换筛估计量。接下来，我们通过解决下列 2SLS 问题得到的 $\widehat{\beta}_{iv}$ 来估计 $\beta$:

$$\min_\beta [\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\widetilde{\Pi}(\beta)]' P(P'P)^- P' [\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\widetilde{\Pi}(\beta)]. \tag{2.23}$$

最后，我们通过下式来估计 $h_{o1}(y_2) + h_{o2}(x_2)$

$$\widehat{h}_1(y_2) + \widehat{h}_2(x_2) = (B^{k_{1,n}}(y_2)', B^{k_{2,n}}(x_2)')\widetilde{\Pi}(\hat{\beta}_{iv}),$$

并且通过位置约束 $h_2(0.5) = \log(3/2)$ 去估计 $h_{o1}$ 和 $h_{o2}$:

$$\widehat{h}_{2,iv}(x_2) = B^{k_{2,n}}(x_2)'\widetilde{\Pi}_2(\hat{\beta}_{iv}) - B^{k_{2,n}}(0.5)'\widetilde{\Pi}_2(\hat{\beta}_{iv}) + \log(3/2),$$

$$\widehat{h}_{1,iv}(y_2) = B^{k_{1,n}}(y_2)'\widetilde{\Pi}_1(\hat{\beta}_{iv}) + B^{k_{2,n}}(0.5)'\widetilde{\Pi}_2(\hat{\beta}_{iv}) - \log(3/2).$$

我们注意到尽管该模型 (2.21) 属于文献中难以处理的不适定求逆问题，然而上述代换筛 MD 方法非常容易计算，事实上，$\hat{\beta}_{iv}$ and $\widetilde{\Pi}(\hat{\beta}_{iv})$ 都有解析解。具体而言，我们注意到 (2.22) 等价于

$$\min_{\Pi, \lambda_\ell} (\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\Pi)' P(P'P)^- P' (\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\Pi) + \sum_{\ell=1}^{2} \lambda_\ell \{\Pi_\ell' C_\ell \Pi_\ell - c_\ell \log n\},$$

其中对 $\ell = 1, 2$，有 $C_\ell = \int [D^2 B^{k_{\ell,n}}(y)][D^2 B^{k_{\ell,n}}(y)]' dy$, $\Pi_\ell' C_\ell \Pi_\ell = \int [D^2 h_\ell(y)]^2 dy$，以及 $\lambda_\ell \ge 0$ 是拉格朗日乘子。然而，我们并不希望指定上界 $c_\ell > 0, \ell = 1, 2$；相反的我们选择较小的惩罚权重 $\lambda_1, \lambda_2$ 并解决下列优化问题:

$$\min_{\Pi} (\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\Pi)' P(P'P)^- P' (\mathbf{Y}_1 - \mathbf{X}_1\beta - \mathbf{B}\Pi) + \sum_{\ell=1}^{2} \lambda_\ell \Pi_\ell' C_\ell \Pi_\ell \tag{2.24}$$

让 $C(\lambda_1, \lambda_2) = \begin{bmatrix} \lambda_1 C_1 & 0 \\ 0 & \lambda_2 C_2 \end{bmatrix}$ 作为平滑度惩罚矩阵。最小化问题 (2.24) 有很简单的解析解:

$$\widetilde{\Pi}(\beta) = \left(\mathbf{B}' P(P'P)^- P' \mathbf{B} + C(\lambda_1, \lambda_2)\right)^- \mathbf{B}' P(P'P)^- P' [\mathbf{Y}_1 - \mathbf{X}_1\beta] = W[\mathbf{Y}_1 - \mathbf{X}_1\beta],$$

这里 $W = (\mathbf{B}' P(P'P)^- P' \mathbf{B} + C(\lambda_1, \lambda_2))^- \mathbf{B}' P(P'P)^- P'$。将 $\widetilde{\Pi}(\beta)$ 代入 2SLS 问题 (2.23)，我们得到

$$\hat{\beta}_{iv} = [\mathbf{X}_1'(I - \mathbf{B}W)' P(P'P)^- P' (I - \mathbf{B}W)\mathbf{X}_1]^{-1} \mathbf{X}_1'(I - \mathbf{B}W)' P(P'P)^- P' (I - \mathbf{B}W)\mathbf{Y}_1,$$

以及 $\widetilde{\Pi}(\hat{\beta}_{iv}) = W[\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}_{iv}]$。

在不同的相关系数 $R = 0.9, 0.1$ 和 $0.0$ 下，根据上述设计共生成了 1000 个数据点。我们应用了筛 MD 方法，为了简便起见在具体实施中设定加权矩阵 $\widehat{\Sigma}(X) = I$ 以及惩罚权重为 $\lambda_1 = 0.005$ (或 0.001) and $\lambda_2 = 0.0001$ (或 0)。我们记录了估计系数。接下来，我们生成 1000 个新数据并重新计算估计参数。重复上述过程 400 次，表格 1 和 2 统计了所得到的 $\beta_o$ 估计量的均值和标准误差。为了更好的衡量筛 MD 统计量在拟合非参数部分 $h_{o1}(Y_2)$ 和 $h_{o2}(X_2)$ 的表现，我们在表格中包括了 400 次模拟的积分平方偏差 (IBias$^2$) 和积分均值平方偏差 (IMSE)。[23]表格 1 比较了不同程度内生性和选择不同 $h_1(Y_2)$ 的筛时估计量的表现情况。表格 2 总结了当 $R = 0.9$ 时，$h_1(Y_2)$ 和 $h_2(X_2)$ 的估计量对筛项数和惩罚参数的敏感程度。我们还在图 1 中分析了强相关性 ($R = 0.9$) 下 $h_{o1}(Y_2)$ 和 $h_{o2}(X_2)$ 估计函数情况，其中实线代表真实函数，虚线代表筛 MD（或筛 IV）估计。

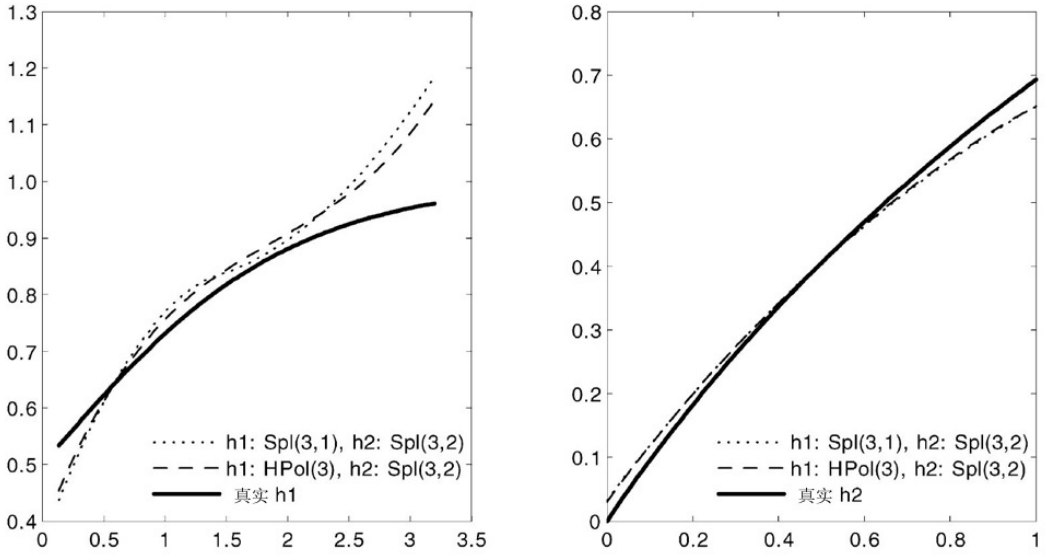


图 1: 真实函数及估计，$R = 0.9$，$\lambda_1 = 0.001$，$\lambda_2 = 0.0001$

通过表格 1 和 2 以及图 1 不难看出，即便有强相关性，$\beta_o$ 和 $h_{o2}(X_2)$ 的筛 MD 估计都表现的很好。我们发现 $\beta_o$ 和 $h_{o2}(X_2)$ 的筛 IV 估计对于惩罚系数 $\lambda_1$, $\lambda_2$ 和 $h_{o1}(Y_2)$ 的筛基的选择都不敏感，结果在不同参数取值下相对稳健。$h_{o1}(Y_2)$ 的筛 IV 估计同样对不同的筛基选择并不敏感，尽管在强相关性下它对惩罚系数 $\lambda_1$ 稍微更敏感。由于在强相关性条件下，$h_{o1}(Y_2)$ 的估计问题是一个难解的不适定求逆问题，随着 $\lambda_1$ 变小，$h_{o1}()$ 的积分平方偏差变化不大，然而 $h_{o1}()$ 的积分方差增长显著。我们使用其他筛基（如三阶基数 B-样条）、不同筛项个数以及不同惩罚程度所得到的蒙特卡洛结果跟上文中得到的差异不显著。这些结果也与 Blundell et al. (2001) 中得到的结果一致。

[23]表 1 中的 IBias$^2(h_1)$ 和 IMSE($h_1$。) 计算方法如下：用 $\hat{h}_i$ 表示从 $i-$th 观测得到的 $h_{o1}$ 的估计量，那么 $\overline{h}(y) = \sum_{i=1}^{400} \hat{h}_i(y)/400$ 是通过 400 个数据集逐点平均得到的。我们计算逐点平方误差 $[\overline{h}(y) - h_{o1}(y)]^2$，以及逐点方差 $400^{-1}\sum_{i=1}^{400}[\hat{h}_i(y) - \overline{h}(y)]^2$。则积分平方误差可以通过在 $\underline{y}$ ($Y_2$ 的 2.5% 分位数) 到 $\overline{y}$($Y_2$ 的 97.5% 分位数) 上计算数值积分逐点平方误差得到。IMSE 可以通过类似的计算方法得到。

表 1: 不同内生性，$h_2$ 用 Spl(3,2) , $k_{2n}$=5, $\lambda_2$=0.0001.

| $R$ | $\beta$ | SE($\beta$) | IBias$^2$($h_1$) | IMSE($h_1$) | IBias$^2$($h_2$) | IMSE($h_2$) |
|---|---|---|---|---|---|---|
| | | Spl(3,2) | $k_{1n}$=5 | $\lambda_1$=0.005 | | |
| 0.0 | 1.0081 | 0.0909 | 0.0003 | 0.0427 | 0.0000 | 0.0026 |
| 0.1 | 1.0021 | 0.0907 | 0.0003 | 0.0446 | 0.0000 | 0.0026 |
| 0.9 | 0.9404 | 0.0947 | 0.0148 | 0.0926 | 0.0003 | 0.0030 |
| | | Spl(3,1) | $k_{1n}$=4 | $\lambda_1$=0.001 | | |
| 0.0 | 1.0076 | 0.0891 | 0.0002 | 0.0225 | 0.0000 | 0.0025 |
| 0.1 | 1.0010 | 0.0886 | 0.0002 | 0.0229 | 0.0000 | 0.0025 |
| 0.9 | 0.9398 | 0.0941 | 0.0160 | 0.0623 | 0.0003 | 0.0029 |
| | | HPol(4) | $k_{1n}$=5 | $\lambda_1$=0.005 | | |
| 0.0 | 1.0089 | 0.0906 | 0.0003 | 0.0395 | 0.0000 | 0.0026 |
| 0.1 | 1.0029 | 0.0901 | 0.0003 | 0.0397 | 0.0000 | 0.0026 |
| 0.9 | 0.9418 | 0.0948 | 0.0121 | 0.0830 | 0.0003 | 0.0030 |
| | | HPol(3) | $k_{1n}$=4 | $\lambda_1$=0.001 | | |
| 0.0 | 1.0078 | 0.0890 | 0.0002 | 0.0202 | 0.0000 | 0.0025 |
| 0.1 | 1.0012 | 0.0885 | 0.0002 | 0.0205 | 0.0000 | 0.0025 |
| 0.9 | 0.9401 | 0.0941 | 0.0112 | 0.0546 | 0.0003 | 0.0029 |

## 2.5 筛分方法在计量经济学中的应用列表（初步）

在本节最后，我们举出一些筛极值估计法在计量经济学中的应用。[24] 绝大部分现有应用都属于微观计量经济学的范畴。Elbadawi et al. (1983) 研究了需求弹性的傅里叶级数最小二乘估计。Cosslett (1983) 提出二元选择模型的非参最大似然估计。Heckman and Singer (1984) 考虑了久期模型的筛最大似然估计，其中未知残差项的分布可以用一阶样条来近似。该估计方法也应用在 Cameron and Heckman (1998) 的生命周期教育问题中。Duncan (1986) 使用样条筛最大似然方法估计了一个截尾回归模型。Hausman and Newey (1995) 考虑了消费者剩余的幂级数和样条级数最小二乘估计方法。Hahn (1998) 以及 Imbens et al. (2005) 在平均处理效应模型的两步法有效估计中使用了幂级数和样条函数。Newey et al. (1999) 和 Pinkse (2000) 考虑了联立方程的三角系统的级数估计。为了估计将 Heckman's (1979) 样本选择模型进行半参数泛化后得到的模型，Gallant and Nychka (1987) 提出 Hermite 多项式筛最大似然估计；此外 Newey (1988) 和 Das et al. (2003) 应用级数最小二乘估计方法。近来，Newey (2001) 使用筛 MD 方法来估计一个非线性测量误差模型。Blundell et al. (2001) 使用代换筛 MD 方法估计具有非参内生支出的保形恩格尔曲线。Coppejans (2001) 使用筛极大似然估计了二元选择模型。Khan (2005) 使用筛最小二乘法来估计具有未知异方差性的 probit 二元选择模型。Hirano et al. (2003) 提出使用筛 logistic 回归法估计处理效应模型的倾向指数。Mahajan (2004) 使用筛 MLE 方法估计了一个具有自变量错误分类问题的二元半参单指标模型。Chen et al. (2004a)

[24] 虽然我们的关注点仅限于经济学科，但是仍然无法覆盖筛分方法在计量经济学中的所有现有应用。任何遗漏仅因为所知有限完全出于无意。

表 2: 不同惩罚程度和筛项，R=0.9

| $(\lambda_1, \lambda_2)$ | $\beta$ | SE($\beta$) | IBias$^2$($h_1$) | IMSE($h_1$) | IBias$^2$($h_2$) | IMSE($h_2$) |
|---|---|---|---|---|---|---|
| $h_1$ 和 $h_2$ 用 Spl(3,1), $k_{1n}$=$k_{2n}$=4 | | | | | | |
| (0.001,0.0) | 0.9366 | 0.0941 | 0.0176 | 0.0612 | 0.0003 | 0.0018 |
| (0.05,0.001) | 0.9324 | 0.0867 | 0.0185 | 0.0568 | 0.0003 | 0.0016 |
| $h_1$ 和 $h_2$ 用 Spl(3,3), $k_{1n}$=$k_{2n}$=6 | | | | | | |
| (0.001,0.0) | 0.9451 | 0.0984 | 0.0124 | 0.1594 | 0.0003 | 0.0032 |
| (0.05,0.001) | 0.9441 | 0.0954 | 0.0125 | 0.0720 | 0.0003 | 0.0028 |

研究了半非参多元连接函数（copula）模型的筛 MLE 估计法。Chen et al. (2005) 则借助样条筛估计了具有辅助样本的非线性非传统测量误差模型。Chen et al. (2004b) 证明了他们的估计方法对非经典测量误差，数据缺失和处理效应的一般非线性广义矩方法模型是半参有效率的。Hu and Schennach (2006) 应用筛 MLE 方法使用工具变量估计了非线性非经典测量误差模型。Brendstrup and Paarsch (2004) 则使用 Hermite 和 Laguerre 多项式筛 MLE 方法来估计序贯不对称英式拍卖。Bierens (2006) 和 Bierens and Carvalho (2006) 应用 Legendre 多项式筛 MLE 分别估计了一个区间截尾混合比例风险模型和一个竞争的累犯风险模型。

在时间序列计量经济学中，筛分方法同样得到广泛应用。Engle et al. (1986) 使用部分线性样条回归预测了电力需求。Engle and Gonzalez-Rivera (1991) 使用筛 MLE 来估计 ARCH 模型，其中用一阶样条筛来近似标准化时间序列残差的未知密度。Gallant and Tauchen (1989) 和 Gallant et al. (1991) 使用 Hermite 多项式筛 MLE 来研究资产定价和外汇汇率。Gallant and Tauchen (1996, 2004) 提议结合 Hermite 多项式筛和模拟的矩估计方法来有效解决许多复杂的具有潜变量的资产定价模型；他们的方法已经广泛应用在实证金融中。Bansal and Viswanathan (1993), Bansal et al. (1993) 和 Chapman (1997) 考虑对整个随机折现因子（或定价核）使用一些宏观因子的函数做筛近似分析。White (1990) 和 Granger and Terasvirta (1993) 建议使用 sigmoid ANN 筛进行非参最小二乘预测。Hutchinson et al. (1994) 使用径向基 ANN 对期权进行估价。Chen et al. (2001) 使用部分线性 ANN 和 ridgelet 筛方法预测美国通货膨胀。McCaffrey et al. (1992) 通过 ANN 筛估计了混沌系统的 Lyapunov 指数。[25] Chen and Ludvigson (2003) 使用 sigmoid ANN 筛来估计消费资产定价模型中的未知习惯函数。Polk et al. (2003) 应用 sigmoid ANN 来计算检验股票回报可预测性的条件分位数。Chen et al. (1998) 使用保形样条-小波筛从离散时间低频观测中估计完全非参数标量扩散模型的特征函数。Chen and Conley (2001) 使用同样的筛估计了一个具有灵活条件期望和条件协方差的空间时间模型。Phillips (1998) 应用标准正交基来分析伪回归问题。Engle and Rangel (2004) 提出一个新的样条 GARCH 模型来测度非条件波动率，并将这一模型应用在 50 年 50 个国家股票交易市场日交易数据中。更多金融时间序列模型可以参考 Fan and Yao (2003)。

[25] 他们的工作同 Gallant and White (1992) 使用 ANN 筛对多元未知回归方程的导数估计非常接近。Shintani and Linton (2003) 使用 ANN 筛提出一个针对混沌的非参检验。

# 3 未知函数筛估计的大样本性质

我们已经知道筛分方法具有一般性且易于实现。在本节中，我们将首先在较弱的正则性条件下证明筛极值方法可以一致地估计有限维和无限维未知参数。然而，对于计量经济学和统计推断，研究者还想了解在有限数据下一致筛估计量的精确程度及其极限分布。不幸的是，对于未知函数的筛极值估计量尚不存在逐点极限分布的一般理论。文献中关于密度和最小二乘（LS）回归函数的序列估计量的逐点极限分布已经有一些结果，我们将在本节结尾对这些结果进行回顾。但是我们还是有一些现成的结果可以借鉴。我们已经对未知函数的平滑泛函的筛估计量的 $\sqrt{n}-$ 渐近正态性建立了成熟的理论[26] 。

正如我们将在第 4 节中看到的，为了在半非参数模型中得出参数分量的筛估计量的 $\sqrt{n}-$ 渐近正态性和半参数效率，非参数分量的筛估计量收敛到真实未知函数的速度应当比（在某种度量下）$n^{-1/4}$ 快。这就要求即使未知函数是多余参数（即不是研究者感兴趣的参数），也需要得到无限维未知函数的筛分估计量的收敛速度。此外，当未知函数也是非参数或半非参数模型中研究者感兴趣的参数时，收敛速率将为判断有限样本下筛估计量的准确程度提供有用信息。不幸的是，到目前为止，还没有关于未知函数的一般筛极值估计量的收敛速率的统一理论。[27] 尽管如此，筛 M 估计量的收敛速度理论目前已经非常成熟。

我们首先在 3.1 节中提供一个关于一般筛极值估计的新的一致性定理。然后，我们回顾现有的关于未知函数筛 M 估计量的收敛速度和逐点渐近分布的结果。我们在 3.2 节中总结了现有的关于未知函数一般筛 M 估计量的收敛速度结果，并用两个例子说明如何验证为得到该一般结果所需要的技术条件。虽然序列级数估计是筛 M 估计的一种特殊情况，但由于其特殊的性质（即凹目标函数和有限维线性筛空间），我们可以在另一套充分条件下推导出序列级数估计量的收敛速度；我们将在 3.3 节中详细讨论这个问题。3.4 节介绍了在最小二乘回归函数的一种特殊情况下级数估计量的逐点正态性的一些现有结果。

## 3.1 筛极值估计量的一致性

对于无限维且可能不紧凑的参数空间 $\Theta$，Geman and Hwang (1982) 得到了独立同分布数据下筛 MLE 估计的一致性；White and Wooldridge (1991) 证明了在非独立和异质数据下筛极值估计的一致性。对于无限且紧凑的参数空间 $\Theta$，Gallant (1987) 和 Gallant and Nychka (1987) 证明了筛 M-估计的一致性；Newey and Powell (2003) 和 Chernozhukov et al. (2006) 证明了筛分 MD 估计的一致性。接下来，我们为允许有非紧凑的无限维参数 $\Theta$ 并适用于不适定的半非参数问题的近似筛极值估计量提出新的一致性定理。[28]

用 $d(\cdot,\cdot)$ 表示 $\Theta$ 上的（伪）度量。特别地，当 $\Theta = B \times \mathcal{H}$，其中 $B$ 是某欧几里得空间的子集，$\mathcal{H}$ 是某具范函数空间的子集时，我们可以用 $d(\theta,\widetilde{\theta}) = |\beta - \widetilde{\beta}|_e + \|h - \widetilde{h}\|_{\mathcal{H}}$，这里 $|\cdot|_e$ 表示欧几里得范数，$\|\cdot\|_{\mathcal{H}}$ 是

[26] 关于“平滑泛函”的定义可以参考本文第四章。这里只要知道常规有限维参数和未知函数的平均导数都是平滑函数的例子。

[27] 据我们所了解到的，目前有一篇未发表的论文 (Chen and Pouzo, 2006) 得到了满足半非参条件矩模型 $E[\rho(Z,\beta_o,h_o(\cdot))|X] = 0$ 的参数 $\theta_o = (\beta_o, h_o)$ 的筛 MD 估计 $\widehat{\theta}_n$ 的收敛速度，模型中未知的 $h_o(\cdot)$ 可以取决于内生变量 $Y$ 或潜变量。更早一些时候，Ai and Chen (2003) 在弱度量下得到了快于 $n^{-1/4}$ 的收敛速度。此外，还有一些论文在特定模型下得到了 $h_o$ 的筛 MD 估计的收敛速度；模型 $E[Y_1 - h_o(Y_2)|X] = 0$ 可以参考 Blundell et al. (2001) 和 Hall and Horowitz (2005)。Van der Vaart and Wellner (1996, 定理 3.4.1) 为筛极值估计量给出了一个抽象的收敛速率。然而，他们的条件排除了不适定的半非参问题，同时需要过程 $\sqrt{n}(\widehat{Q}_n - Q)$ 的收敛速率满足一个极大不等式，目前并不存在理论证明一般的目标函数 $\widehat{Q}_n$ 是否满足上述要求。因此，公平地说目前缺乏关于筛极值估计量的收敛速度的一般理论。

[28] 根据 Stinchcombe (2002) 的一个最新定理，筛极值估计的一致性是一般属性。

函数空间 $\mathcal{H}$ 上定义的范数。例如，如果 $\mathcal{H}=C^m(\mathcal{X})$ 具有有界 $\mathcal{X}$，我们可以让 $\|h\|_{\mathcal{H}}$ 为 $\|h\|_\infty$ 或 $\|h\|_{2,leb}$。

**条件 3.1.** *(识别) (i)* $Q(\theta_o)>-\infty$, 如果 $Q(\theta_o)=+\infty$ 那么对所有 $\theta\in\Theta_k\backslash\{\theta_o\}$ 和所有 $k\geq 1$, 有 $Q(\theta)<+\infty$ *; (ii)* 存在不增加的正函数 $\delta()$ 和一个正函数 $g()$ 使得对所有 $\varepsilon>0$ 和所有 $k\geq 1$,

$$Q(\theta_o)-\sup_{\{\theta\in\Theta_k:d(\theta,\theta_o)\geq\varepsilon\}}Q(\theta)\geq\delta(k)g(\varepsilon)>0$$

**条件 3.2.** *(筛空间)* 对所有 $k\geq 1$ 有 $\Theta_k\subseteq\Theta_{k+1}\subseteq\Theta$ *;* 同时存在一系列 $\pi_k\theta_o\in\Theta_k$ *such* 满足随着 $k\to\infty$, 有 $d(\theta_o,\pi_k\theta_o)\to 0$。

**条件 3.3.** *(连续性) (i)* 对每个 $k\geq 1$, $Q(\theta)$ 在度量 $d(\cdot,\cdot)$ 下在 $\Theta_k$ 中上连续; *(ii)* $|Q(\theta_o)-Q(\pi_{k(n)}\theta_o)|=o(\delta(k(n)))$。

**条件 3.4.** *(紧筛空间)* 筛空间 $\Theta_k$ 在度量 $d(\cdot,\cdot)$ 下紧凑。

**条件 3.5.** *(筛一致收敛) (i)* 对所有 $k\geq 1$, $plim_{n\to\infty}\sup_{\theta\in\Theta_k}|\widehat{Q}_n(\theta)-Q(\theta)|=0$*; (ii)* $\widehat{c}(k(n))=o_P(\delta(k(n)))$ 其中 $\widehat{c}(k(n))\equiv\sup_{\theta\in\Theta_{k(n)}}|\widehat{Q}_n(\theta)-Q(\theta)|$*; (iii)* $\eta_{k(n)}=o(\delta(k(n)))$。

**定理 3.1.** 让 $\hat{\theta}_n$ 表示在 *(2.9)* 中定义的近似筛极值估计量。如果条件 *3.1 - 3.5* 成立, 则有 $d(\hat{\theta}_n,\theta_o)=o_P(1)$。

证明. 根据备注 2.1, $\widehat{\theta}_n$ 是适定和可测的。对所有 $\varepsilon>0$, 在条件 3.3(i) 和 3.4 下 $\sup_{\{\theta\in\Theta_{k(n)}:d(\theta,\theta_o)\geq\varepsilon\}}Q(\theta)$ 存在。根据定义，我们有对所有 $\varepsilon>0$,

$$\Pr\left(d(\widehat{\theta}_n,\theta_o)>\varepsilon\right)\leq\Pr\left(\sup_{\{\theta\in\Theta_{k(n)}:d(\theta,\theta_o)\geq\varepsilon\}}\widehat{Q}_n(\theta)\geq\widehat{Q}_n\left(\pi_{k(n)}\theta_o\right)-O(\eta_{k(n)})\right)\leq P_1+P_2,$$

其中

$$\begin{aligned}P_1&\equiv\Pr\left(\sup_{\{\theta\in\Theta_{k(n)}:d(\theta,\theta_o)\geq\varepsilon\}}\left|\widehat{Q}_n(\theta)-Q(\theta)\right|>\widehat{\nu}(k(n))\right)\\&\leq\Pr\left(\sup_{\theta\in\Theta_{k(n)}}\left|\widehat{Q}_n(\theta)-Q(\theta)\right|>\widehat{\nu}(k(n))\right),\end{aligned}$$

以及

$$\begin{aligned}P_2&\equiv\Pr\left(\sup_{\{\theta\in\Theta_{k(n)}:d(\theta,\theta_o)\geq\varepsilon\}}Q(\theta)\geq Q\left(\pi_{k(n)}\theta_o\right)-2\widehat{\nu}(k(n))-O(\eta_{k(n)})\right)\\&=\Pr\left(2\widehat{\nu}(k(n))+\{Q(\theta_o)-Q\left(\pi_{k(n)}\theta_o\right)\}+O(\eta_{k(n)})\geq Q(\theta_o)-\sup_{\{\theta\in\Theta_{k(n)}:d(\theta,\theta_o)\geq\varepsilon\}}Q(\theta)\right)\end{aligned}$$

令 $\widehat{\nu}(k(n))=\widehat{c}(k(n))$ 那么根据 $\widehat{c}(k(n))$ 的定义和条件 3.5(i) 可以得到 $P_1=0$; 通过条件 3.1 和 3.5(ii) 推出 $P_2\leq\Pr[2\widehat{c}(k(n))+\{Q(\theta_o)-Q(\pi_{k(n)}\theta_o)\}+O(\eta_{k(n)})\geq\delta(k(n))g(\varepsilon)]\to 0$。证毕。 □

**备注 3.1 (1).** 定理 3.1 适用于适定和不适定的半非参数模型。当问题（例如非参 IV 回归 $E[Y_1 - h_o(Y_2)|X] = 0$）不适定时，研究者可以用 $\liminf_k \delta(k) = 0$，这时条件 3.1(ii), 3.3(ii) 和 3.5(ii)(iii) 仍然满足。关于其他适用于非适定问题的筛极值估计的一致性理论请参考 Chen and Pouzo (2006)。

**(2)** 如果 $\liminf_k \delta(k) > 0$，那么在 $\eta_{k(n)} = o(1)$ 下条件 3.5(iii) 自动满足，条件 3.5(ii) 可以由 3.5(i) 推出，同时条件 3.3(ii) 可以从条件 3.2 和 **3.3(ii)'** 得到：$Q(\theta)$ 在 $\theta_o \in \Theta$ 处连续。

**(3)** 定理 3.1 是 White and Wooldridge (1991) 文中引理 2.6 的一个拓展。在条件 3.4, 3.5(i) 和条件 3.1', 3.2' 以及 3.3'（定义如下）下，由该引理可以推出 $d(\widehat{\theta}_n, \theta_o) = o_P(1)$：

**条件 3.1'.** (i) $Q(\theta)$ 在 $\theta_o \in \Theta$ 处连续，$Q(\theta_o) > -\infty$; (ii) 对所有 $\varepsilon > 0$，$Q(\theta_o) > \sup_{\{\theta \in \Theta : d(\theta, \theta_o) \geq \varepsilon\}} Q(\theta)$。

**条件 3.2'.** 对所有正整数 $k \geq 1$，有 $\Theta_k \subseteq \Theta_{k+1} \subseteq \Theta$；对所有 $\theta \in \Theta$，存在 $\pi_k \theta \in \Theta_k$ 使得随着 $k \to \infty$，$d(\theta, \pi_k \theta) \to 0$。

**条件 3.3'.** 对每一个 $k \geq 1$，(i) $\widehat{Q}_n(\theta)$ 对所有 $\theta \in \Theta_k$ 是数据 $\{Z_t\}_{t=1}^n$ 的可测函数；(ii) 对任意数据 $\{Z_t\}_{t=1}^n$，$\widehat{Q}_n(\theta)$ 在度量 $d(\cdot, \cdot)$ 下在 $\Theta_k$ 中上连续。

注意在条件 3.2 成立时，从条件 3.1'(ii) 和 $\delta(k) = const. > 0$ 可以推出 3.1(ii) 成立，所以备注 3.1(2) 成立，有 $d(\widehat{\theta}_n, \theta_o) = o_P(1)$。然而，当 $\Theta$ 是一个非紧凑无限维参数空间时，3.1'(ii) 可能在某些不适定的半非参模型中不成立。

**(4)** 条件 3.1' 可以由 **3.1"** 推出：(i) 在 $d(\cdot, \cdot)$ 下 $\Theta$ 紧凑，并且 $Q(\theta)$ 在 $\Theta$ 上上连续；(ii) $Q(\theta)$ 在 $\Theta$ 上的 $\theta_o$ 处取得唯一最大值，$Q(\theta_o) > -\infty$。根据定理 3.1，我们得到：在条件 3.1", 3.2, 3.4 和 3.5(i) 下，$d(\widehat{\theta}_n, \theta_o) = o_P(1)$。这一结论同 Newey and Powell (2003) 中的引理 A.1 以及 Chernozhukov et al. (2006) 非常相似。

**备注 3.2.** 如果 $\hat{\theta}_n$ 满足 $\widehat{Q}_n(\hat{\theta}_n) \geq \sup_{\theta \in \Theta_n} \widehat{Q}_n(\theta) - O_{a.s.}(\eta_n)$，那么 $d(\widehat{\theta}_n, \theta_o) = o_{a.s.}(1)$ 在条件 3.1 - 3.4 和 **3.5"** 下：(i) 对所有 $k \geq 1$，$\sup_{\theta \in \Theta_k} |\widehat{Q}_n(\theta) - Q(\theta)| = o_{a.s.}(1)$；(ii) $\widehat{c}(k(n)) = o_{a.s.}(\delta(k(n)))$；(iii) $\eta_{k(n)} = o(\delta(k(n)))$。这将 Gallant's (1987) 关于近似筛极值估计的定理拓展到几乎处处收敛，同时也适用于非紧凑无限维 $\Theta$ 和不适定的半非参模型。

注意当 $\Theta_k = \Theta$ 是紧凑的，定理 3.1 的条件成为 Newey and McFadden (1994) 和 White (1994) 文中参数极值估计一致性的标准假设。对于半非参数模型，整个参数空间 $\Theta$ 包含无限维的未知函数，通常也是非紧凑的。虽然如此，研究者可以很简单的构造紧凑的近似参数空间（筛）$\Theta_k$。此外，我们可以相对容易地验证紧凑筛空间的一致收敛性,[29] 虽然当空间 $\Theta$ 太 "大" 或 "复杂" 时，"$\text{plim}_{n \to \infty} \sup_{\theta \in \Theta} |\widehat{Q}_n(\theta) - Q(\theta)| = 0$" 可能并不成立。

我们现在回顾一些函数类别复杂度的概念。让 $L_r(P_o)$, $r \in [1, \infty)$ 表示具有有限 $r$ 阶矩的实值随机变量空间，用 $\|\cdot\|_r$ 表示 $L_r(P_o)$ 范数。用 $\mathcal{F}_n = \{g(\theta, \cdot) : \theta \in \Theta_n\}$ 表示实值、$L_r(P_o)$ 可测、由 $\theta \in \Theta_n$ 索引的函数类。一种 $\mathcal{F}_n$ 类的复杂度概念是 "$L_r(P_o)$-无括覆盖数目"，其定义为可以覆盖 $\mathcal{F}_n$ 的 $w$-球 $\{\{f : ||f - g_j||_r \leq w\}, ||g_j||_r < \infty, j = 1, \ldots, N\}$ 的最小数目，一般写作 $N(w, \mathcal{F}_n, ||\cdot||_r)$。类似的，我们可以定义 $N(w, \mathcal{F}_n, ||\cdot||_{n,r})$ 为 $L_r(P_n)$-（随机）无括覆盖数目，其中 $\|\cdot\|_{n,r}$ 表示 $L_r(P_n)$-范数，$P_n$ 是随机样本 $\{Z_i\}_{i=1}^n$ 的实证测度。有时候 $\mathcal{F}_n$ 的覆盖数目可以随着 $n$ 增长增加到无穷大；基于此，我们考虑另一种被称

[29]我们可以修改 Newey（1991）中的推论 2.2 的证明或 Andrews（1992）中的引理 1 的证明，为（就 3.3（i）和 3.4 而言）条件 3.5(i) 和 $\Theta_k$ 上的逐点收敛提供充分条件。

为 “$L_r(P_o)$-无括度量熵” 的复杂度概念：$H(w,\mathcal{F}_n,||\cdot||_r)\equiv\log(N(w,\mathcal{F}_n,||\cdot||_r))$，以及 “$L_r(P_n)$-（随机）无括度量熵”，$H(w,\mathcal{F}_n,||\cdot||_{n,r})\equiv\log(N(w,\mathcal{F}_n,||\cdot||_{n,r}))$。关于度量熵的具体讨论可以参考 Pollard (1984), Andrews (1994a), van der Vaart and Wellner (1996) 和 van de Geer (2000)。

当函数类 $\Theta$ 从度量熵的角度来看太过复杂时，其在整个参数空间 $\Theta$ 上的一致收敛可能不成立，但在筛空间 $\Theta_n$（即条件 3.5(i)）仍可以成立。例如，当 $\widehat{Q}_n(\theta)=n^{-1}\sum_{t=1}^n l(\theta,Z_t)$ 和 $\{Z_t\}_{t=1}^n$ 是独立同分布，$E\{\sup_{\theta\in\Theta_n}|l(\theta,Z_t)|\}<\infty$，那么条件 3.5(i) 被满足当且仅当对所有 $w>0$，$H(w,\{l(\theta,\cdot):\theta\in\Theta_n\},||\cdot||_{n,1})=o_P(n)$；参考 Pollard (1984)。当空间 $\Theta$ 是无穷维且非有界时，$H(w,\{l(\theta,\cdot):\theta\in\Theta\},||\cdot||_{n,1})=O_P(n)$ 可能发生，因此 $\sup_{\theta\in\Theta}|\widehat{Q}_n(\theta)-Q(\theta)|\neq o_P(1)$。这种情况下，在整个参数空间 $\Theta$ 上求取最大值 $\arg\sup_{\theta\in\Theta}\widehat{Q}_n(\theta)$ 得到的极值估计量可能不存在或不一致。

定理 3.1 的条件 3.1 - 3.4 是基本的正则性条件；在具体的应用中，我们可以为条件 3.5 提供更一般的充分条件。在下面的备注中我们讨论筛 M 和筛 MD 估计量的一致性结果。用 $N(w,\Theta_n,d)$ 表示在度量 $d$ 下可以覆盖筛空间 $\Theta_n$ 的 $w-$ 半径球的最低数量。

**备注 3.3.** （筛 M-估计量 $\hat{\theta}_n=\arg\sup_{\theta\in\Theta_n}n^{-1}\sum_{t=1}^n l(\theta,Z_t)-o_P(1)$ 的一致性）：假设条件 3.2 和 3.4 成立，条件 3.1 中的 $Q(\theta)=E\{l(\theta,Z_t)\}$，$\liminf_{k(n)}\delta(k(n))>0$，同时 $E\{l(\theta,Z_t)\}$ 在 $\theta=\theta_o\in\Theta$ 处连续，那么在下列条件 3.5M 下有 $d(\widehat{\theta}_n,\theta_o)=o_P(1)$：

**条件 3.5M.** (i) $\{Z_t\}_{t=1}^n$ 独立同分布，$E\{\sup_{\theta\in\Theta_n}|l(\theta,Z_t)|\}$ 有界；(ii) 存在一个有限的 $s>0$ 和满足 $E\{U(Z_t)\}<\infty$ 的随机变量 $U(Z_t)$ 使得 $\sup_{\theta,\theta'\in\Theta_n:d(\theta,\theta')\leq\delta}|l(\theta,Z_t)-l(\theta',Z_t)|\leq\delta^s U(Z_t)$；(iii) 对所有 $\delta>0$，有 $\log N(\delta^{1/s},\Theta_n,d)=o(n)$。

备注 3.3 可以从定理 3.1 和 Pollard（1984）中的定理 II.24 直接得出。这是因为从条件 3.5M (i) 和 (ii) 可以推出 $H(w,\{l(\theta,\cdot):\theta\in\Theta_n\},||\cdot||_{n,1})\leq\log N(\delta^{1/s},\Theta_n,d)$，因此从条件 3.5M 是 3.5(i) 的充分条件。关于更多条件 3.5 的更一般的充分假设可以参考 White and Wooldridge (1991, 定理 2.5) 和 Ai and Chen (2004a, 引理 A.1)。

**备注 3.4.** （筛 MD 估计量 $\hat{\theta}_n=\arg\inf_{\theta\in\Theta_n}\frac{1}{n}\sum_{t=1}^n\widehat{m}(X_t,\theta)'\{\widehat{\Sigma}(X_t)\}^{-1}\widehat{m}(X_t,\theta)+o_P(1)$ 的一致性）：假设条件 3.2 和 3.4 成立；仅当 $\theta=\theta_o\in\Theta$ 时 $m(X_t,\theta)\equiv E\{\rho(Z_t,\theta)|X_t\}=0$；对所有 $X_t$，$m(X_t,\theta)$ 在度量 $d(\cdot,\cdot)$ 下在 $\theta_o$ 上连续；以及 $\liminf_{k(n)}\delta(k(n))>0$。那么在下列条件 3.5MD 下有 $d(\widehat{\theta}_n,\theta_o)=o_P(1)$：

**条件 3.5MD.** (i) $\{Z_t\}_{t=1}^n$ 独立同分布，$E\{\sup_{\theta\in\Theta_n}|m(X_t,\theta)'m(X_t,\theta)|\}$ 有界；(ii) 存在有限的 $s>0$ 和满足 $E\{[U(X_t)]^2\}<\infty$ 的 $U(X_t)$ 使得 $\sup_{\theta,\theta'\in\Theta_n:d(\theta,\theta')\leq\delta}|m(X_t,\theta)-m(X_t,\theta')|\leq\delta^s U(X_t)$; (iii) 对所有 $\delta>0$ 满足 $\log N(\delta^{1/s},\Theta_n,d)=o(n)$；(iv) 对一个正定有限矩阵 $\Sigma(X_t)$，一致地在 $X_t$ 上有 $\widehat{\Sigma}(X_t)=\Sigma(X_t)+o_P(1)$；(v) 一致地在 $\theta\in\Theta_n$ 上 $\frac{1}{n}\sum_{i=1}^n|\widehat{m}(X_i,\theta)-m(X_i,\theta)|^2=o_P(1)$ 。

备注 3.4 的证明可以参考 Chen and Pouzo (2006)；他们还为筛 MD 估计量 $\hat{\theta}_n$ 的一致性提供了在不要求 $\liminf_{k(n)}\delta(k(n))>0$ 时的充分条件。此外，当 $\widehat{\Sigma}(X_t)$ 和 $\widehat{m}(X_t,\theta)$ 分别是 $\Sigma(X_t)$ 和 $m(X_t,\theta)$ 的核和级数估计时，Newey and Powell (2003) 和 Ai and Chen (1999, 2003, 2004a) 提供了条件 3.5MD(iv) 和 (v) 的更一般的充分条件。

最后，定理 3.1 也适用于在误判的半非参数模型中得到某些伪真值的筛极值估计的收敛速度；具体应用

可以参考 Ai and Chen (2004a) 中的引理 3.1。

### 3.2 筛 M 估计量的收敛速度

现在已经有许多关于未知函数的筛 M 估计量的收敛速度的结论。对独立同分布数据，Van de Geer (1995) 得到了筛 LS 回归的收敛速度。Shen and Wong (1994), 和 Birgé and Massart (1998) 得到了一般筛 M 回归的收敛速度。Van de Geer (1993) 和 Wong and Shen (1995) 得到了筛 MLE 的收敛速度。关于时间序列数据，Chen and Shen (1998) 得到了稳健 $\beta$-混合模型的筛 M 估计量的收敛速度。[30] 有关收敛速度的一般理论在技术上相对复杂，且依赖于经验过程理论。在本节中，我们提供了一个简单版本的筛 M 估计的收敛速率结果，其条件很容易验证。如果读者对筛 M 估计的收敛速度的一般理论感兴趣，可以参考 Shen and Wong (1994), Wong and Shen (1995) 以及 Birgé and Massart (1998)。

回到 $\theta_o \in \Theta$ 以及近似筛 M 估计 $\widehat{\theta}_n$ 解决：

$$n^{-1}\sum_{t=1}^{n} l(\hat{\theta}_n, Z_t) \geq \sup_{\theta\in\Theta_n} n^{-1}\sum_{t=1}^{n} l(\theta, Z_t) - O_P(\varepsilon_n^2) \quad \text{with} \quad \varepsilon_n \to 0。 \tag{3.1}$$

用 $d(\theta_o, \theta)$ 表示一个 $\Theta$ 上的（伪）度量，其满足 $d(\theta_o, \hat{\theta}_n) = o_P(1)$。定义 $K(\theta_o, \theta) \equiv E(l(\theta_o, Z_t) - l(\theta, Z_t))$。[31] 让 $||\theta_o - \theta||$ 表示在 $\Theta$ 上的度量，其使得对所有 $\theta \in \Theta$ 有 $||\theta_o - \theta|| \leq const.d(\theta_o, \theta)$，以及对满足 $d(\theta_o, \theta) = o(1)$ 的 $\theta \in \Theta$ 有 $||\theta_o - \theta|| \asymp K^{1/2}(\theta_o, \theta)$。我们之后会给出在 $||\theta_o - \theta||$ 下的筛估计 $\hat{\theta}_n$ 的收敛速度，由此可以自动得出 $\overline{d}(\theta_o, \hat{\theta}_n)$ 的上界，其中 $\overline{d}$ 是定义在 $\Theta$ 上的满足 $\overline{d}(\theta_o, \theta) \leq const.K^{1/2}(\theta_o, \theta)$ 的度量。

为了使得 $\widehat{\theta}_n$ 在度量 $||\theta_o - \hat{\theta}_n||$ 下以较快的速度收敛到 $\theta_o$，不仅需要筛估计误差速度 $||\theta_o - \pi_n\theta_o||$ 以足够快的速度收敛到 0，还需要筛空间 $\Theta_n$ 不能够太复杂。我们已经介绍了 $L_r(P_o)-$ 无括覆盖数目（度量熵）作为一种衡量 $\mathcal{F}_n = \{g(\theta, \cdot) : \theta \in \Theta_n\}$ 函数类复杂度的方式；现在我们考虑另一种复杂度的测度。用 $\mathcal{L}_r$ 表示在 $\|\cdot\|_r$ 范数下的 $\mathcal{F}_n$ 的完备。对任意给定的 $w > 0$，如果存在满足 $\max_{1\leq j\leq N} \|g_j^u - g_j^l\|_r \leq w$ 的一系列函数对 $\{g_1^l, g_1^u, \cdots, g_N^l, g_N^u\} \subset \mathcal{L}_r$，并且对任意 $g \in \mathcal{F}_n$，存在一个满足 $g_j^l \leq g \leq g_j^u$ $a.e.-P_o$ 的 $j \in \{1, \cdots, N\}$，那么这类函数对的最小数目 $N_{[]}(w, \mathcal{F}_n, ||\cdot||_r) \equiv \min(N : \{g_1^l, g_1^u, \cdots, g_N^l, g_N^u\})$ 就被称为 $L_r(P_o)-$ 带括覆盖数目。类似地，$H_{[]}(w, \mathcal{F}_n, ||\cdot||_r) \equiv \log(N_{[]}(w, \mathcal{F}_n, ||\cdot||_r))$ 被称为 $L_r(P_o)-$ 函数类 $\mathcal{F}_n$ 的带括度量熵。更多细节可以参考 Pollard (1984), Andrews (1994a), van der Vaart and Wellner (1996) 以及 van de Geer (2000)。

我们现在讨论 Chen and Shen (1996) 文中得到的在独立同分布假设下的相关结果；对于平稳 $\beta$-混合过程下的结论，参考 Chen and Shen (1998)；对于一致混合过程，参考 Chen and White (1999)。[32]

**条件 3.6.** $\{Z_t\}_{t=1}^n$ 是独立同分布 *(i.i.d.)* 或 $m$-相关序列。

**条件 3.7.** *存在 $C_1 > 0$ 满足对所有 $\varepsilon > 0$,*

$$\sup_{\{\theta\in\Theta_n : ||\theta_o - \theta|| \leq \varepsilon\}} \mathrm{Var}\left(l(\theta, Z_t) - l(\theta_o, Z_t)\right) \leq C_1\varepsilon^2$$

[30] 必须承认这里不可能提及筛 M 估计收敛速度的所有现有结果。关于特殊筛的收敛速度的论文已经有很多，例如 Stone 和他的合作者在多项式样条回归和密度估计方面的工作，细节可以参考本文章节 3.3；以及 Donoho, Johnstone 和其他研究者在小波方面的研究 (参考 Donoho et al. 1995); Barron (1993), White (1990) 及其他研究者在神经网络方面的研究。

[31] 如果目标函数是对数似然函数，那么 $K(\theta_o, \theta)$ 就是 Kullback-Leibler 信息。

[32] 对于不同的应用于非线性时间序列模型的非参方法的描述可以参考 Fan and Yao (2003)。

**条件 3.8.** 对任意 $\delta > 0$，存在常数 $s \in (0,2)$ 满足

$$\sup_{\{\theta \in \Theta_n : ||\theta_o - \theta|| \le \delta\}} |l(\theta, Z_t) - l(\theta_o, Z_t)| \le \delta^s U(Z_t),$$

其中对某 $\gamma \ge 2$ 有 $E([U(Z_t)]^\gamma) \le C_2$。

条件 3.6 和 3.7 意味着在 $\theta_o$ 的一个邻域中，$\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^n (l(\theta, Z_t) - l(\theta_o, Z_t))\right)$ 的表现与 $||\theta_o - \theta||^2$ 类似。条件 3.8意味着，在 $\theta_o$ 的某个（较小）局部邻域中 $l(\theta, Z_t)$ 在度量 $||\theta_o - \theta||$ 下在 $\theta_o$ 处“连续”；这局部等价于 $K^{1/2}$。此外，可以通过分析目标函数的具体形式很容易地验证条件 3.7 和 3.8。

定义 $\mathcal{F}_n = \{l(\theta, Z_t) - l(\theta_o, Z_t) : ||\theta_o - \theta|| \le \delta, \theta \in \Theta_n\}$，对某常数 $b > 0$，让[33]

$$\delta_n = \inf\{\delta \in (0,1) : \frac{1}{\sqrt{n}\delta^2} \int_{b\delta^2}^{\delta} \sqrt{H_{[]}(w, \mathcal{F}_n, ||\cdot||_2)}dw \le const\}.$$

为了计算 $\delta_n$，只需要知道 $H_{[]}(w, \mathcal{F}_n, ||\cdot||_2)$ 的上界；这方面已经有一些现成的结论。例如，根据 Ossiander (1987) 的引理 2.1 我们知道 $H_{[]}(w, \mathcal{F}_n, ||\cdot||_2) \le H(w, \mathcal{F}_n, ||\cdot||_\infty)$。此外，条件 3.8 意味着

$$H_{[]}(w, \mathcal{F}_n, ||\cdot||_2) \le \log N(w^{1/s}, \Theta_n, ||\cdot||)$$

对于如 2.3.1 中列出的那些有限维线性筛，我们有 $\log N(\epsilon, \Theta_n, ||\cdot||) \le const.\dim(\Theta_n)\log(\frac{1}{\epsilon})$ (参考 Chen and Shen, 1998)；对于神经网络和脊非线性筛我们有 $\log N(\epsilon, \Theta_n, ||\cdot||) \le const.\dim(\Theta_n)\log(\frac{\dim(\Theta_n)}{\epsilon})$ (参考 Chen and White, 1999).

**定理 3.2.** 用 $\hat{\theta}_n$ 表示 *(3.1)* 中定义的近似筛 $M$ 估计量。如果条件 3.6–3.8 满足，那么

$$||\theta_o - \hat{\theta}_n|| = O_P(\varepsilon_n), \quad \text{其中} \quad \varepsilon_n = \max\{\delta_n, ||\theta_o - \pi_n\theta_o||\}$$

我们注意到随着筛 $\Theta_n$ 的复杂程度增加，$\delta_n$ 会逐渐变大，它可以被理解为标准差；确定的近似误差 $||\theta_o - \pi_n\theta_o||$ 则与之相反，其随着 $\Theta_n$ 的复杂度上升而变小，它可以被理解成误差。我们可以通过选择 $\Theta_n$ 的复杂程度使得 $\delta_n \asymp ||\theta_o - \pi_n\theta_o||$，从而达到最优收敛速度。

Chen and Shen (1998) 已经通过三个例子展示了如何应用这个定理的时间序列版本：他们首先考虑应用神经网络筛，小波筛或样条筛的多元非参回归；其次是使用样条或傅里叶级数筛的部分可加时间序列模型；最后通过单调样条筛估计一个具有未知链接的转化模型。Chen and White（1999）通过神经网络筛考察了一个时间序列非参数条件分位数回归，并使用同样的筛估计了一个多变量条件密度。Chen and Conley (2001) 将上述定理应用在一个具有灵活空间条件协方差的可变系数 VAR 模型中。接下来我们用两个例子介绍如何验证定理 3.2 的条件。

[33]在 Chen and Shen (1998, p.297) 中有一个笔误，在 $\delta_n$ 的定义里的“sup” 应该替换成 “inf”。虽然如此，在 Chen and Shen (1998) 文中 $\delta_n$ 的所有其他计算仍然正确。

### 3.2.1 例：具有单调约束的可加均值回归

假设 i.i.d. 数据 $\{Y_t, X_t' = (X_{1t}, ..., X_{qt})\}_{t=1}^n$ 是由下列模型生成的：

$$Y_t = h_{o1}(X_{1t}) + ... + h_{oq}(X_{qt}) + e_t, \quad E[e_t|X_t] = 0$$

这里 $\theta_o = (h_{o1}, ..., h_{oq})' \in \Theta = \mathcal{H}$ 为我们关注的参数；$\mathcal{H} = \mathcal{H}^1 \times \cdots \times \mathcal{H}^q$ 将在假设 3.1 中具体介绍。简单起见，我们假设对 $j = 1, ..., q$ 有 $\dim(X_j) = 1$，$\dim(X) = q$ 以及 $\dim(Y) = 1$。我们通过在筛 $\Theta_n = \mathcal{H}_n$ 上最大化目标函数 $\widehat{Q}_n(\theta) = n^{-1}\sum_{t=1}^n l(\theta, Z_t)$（其中 $l(\theta, Z_t) = -(1/2)[Y_t - \sum_{j=1}^q h_j(X_{jt})]^2$，$Z_t = (Y_t, X_t')'$）来估计回归方程 $\theta_o(X) = \sum_{j=1}^q h_{oj}(X_{jt})$。让 $\|\theta - \theta_o\|^2 = \mathrm{E}(\theta(X_t) - \theta_o(X_t))^2 = E\{\sum_{j=1}^q [h_j(X_{jt}) - h_{oj}(X_{jt})]\}^2$。

**假设 3.1.** *(i)* $h_{o1} \in \mathcal{H}^1 = C([b_{11}, b_{21}]) \cap \{h : 非递减\}$; *(ii)* 对于 $j = 2, ..., q$ 有 $h_{oj} \in \mathcal{H}^j = \Lambda_{c_j}^{p_j}([b_{1j}, b_{2j}])$ 以及 $p_j > 1/2$；同时对于某已知的 $x_j^* \in (b_{1j}, b_{2j})$ 有 $h_{oj}(x_j^*) = 0$。

**假设 3.2.** $\sigma^2(X) \equiv E[e^2|X]$ 有界。

假设 3.1 (ii) 是模型识别的充分条件，假设 3.2 则是许多论文中都会要求的简单正则性条件；参考 Newey (1997)。

我们选择的筛具有以下形式 $\mathcal{H}_n = \mathcal{H}_n^1 \times \cdots \times \mathcal{H}_n^q$。首先让 $\mathcal{H}_n^1$ 为一个保形筛，例如 2.3.5 节中的单调样条小波筛 $\mathrm{MSplWav}(r_1 - 1, 2^{J_{1n}})$，其中 $r_1 \geq 1$，$k_{1n} = 2^{J_{1n}}$。对于 $j = 2, ..., q$，我们让 $\mathcal{H}_n^j = \{h_j \in \Theta_{jn} : h_j(x_j^*) = 0, ||h_j||_\infty \leq c_j\}$，这里 $\Theta_{jn}$ 可以是 2.3.1 节中提到的任何有限维线性筛，比如 $\Theta_{jn} = \mathrm{Pol}(k_{jn})$ 或 $\mathrm{TriPol}(k_{jn})$ 或 $\mathrm{Spl}(r_j, k_{jn})$ $(r_j \geq [p_j] + 1)$，或 $\mathrm{Wav}(m_j, 2^{J_{jn}})$ $(m_j > p_j$ 以及 $k_{jn} = 2^{J_{jn}})$。

在下面的结论中我们令 $p_1 = 1$ 和 $p = \min\{p_1, p_2, \cdots, p_q\}$。

**命题 3.1.** $\widehat{\theta}_n$ 为一个筛 $M$ 估计。假设假设 *3.1* 和 *3.2* 成立。定义 $k_{jn} = O(n^{1/(2p_j+1)})$，$j = 1, ..., q$。那么 $\|\widehat{\theta}_n - \theta_o\| = O_P(n^{-p/(2p+1)})$，其中 $p = \min\{p_1, \cdots, p_q\}$。

证明. 定理 3.2 可以直接用来证明这个命题。很容易看到 $K(\theta_o, \theta) \asymp \|\theta - \theta_o\|^2$. 假设条件 3.6 成立。现在我们逐条验证条件 3.7 和 3.8成立。由于 $l(\theta, Z_t) - l(\theta_o, Z_t) = (\theta - \theta_o)[e_t + (\theta_o - \theta)/2]$，我们有

$$\begin{aligned} E[l(\theta, Z_t) - l(\theta_o, Z_t)]^2 &\leq 2E(\sigma^2(X_t)[\theta_o(X_t) - \theta(X_t)]^2) + (1/2)E([\theta_o(X_t) - \theta(X_t)]^4) \\ &\leq const.\|\theta - \theta_o\|^2 + (1/2)E([\theta_o(X_t) - \theta(X_t)]^4). \end{aligned}$$

根据 Gabushin (1967) 的定理 1（当 $p$ 是整数）和 Chen and Shen (1998) 的引理 2（对任意 $p > 0$），我们有 $\|\theta - \theta_o\|_\infty \leq c\|\theta - \theta_o\|^{2p/(2p+1)}$。因此

$$\begin{aligned} E([\theta_o(X_t) - \theta(X_t)]^4) &\leq \sup_x [\theta(x) - \theta_o(x)]^2 E([\theta_o(X_t) - \theta(X_t)]^2) \\ &\leq C\|\theta - \theta_o\|^{2(1+[2p/(2p+1)])}. \end{aligned}$$

因此对所有 $\varepsilon \leq 1$ 条件 3.7满足。另一方面，

$$|l(\theta, Z_t) - l(\theta_o, Z_t)| \leq \|\theta - \theta_o\|_\infty [|e_t| + (\|\theta_o\|_\infty + \|\theta\|_\infty)/2] \quad a.s..$$

使用 Chen and Shen (1998) 的引理 2 我们可以看到条件 3.8 在 $s = 2p/(2p+1)$, $U(Z_t) = |e_t| + const.$ 和 $\gamma = 2$ 时满足。

为了应用定理 3.2, 我们仍然需要计算确定性近似误差率 $||\theta_o - \pi_n\theta_o||$ 和函数类 $\mathcal{F}_n = \{l(\theta, Z_t) - l(\theta_o, Z_t) : \|\theta - \theta_o\| \leq \delta, \theta \in \Theta_n\}$ 的带括度量熵 $H_{[]}(w, \mathcal{F}_n, ||\cdot||_2)$。根据定义, $||\theta_o - \pi_n\theta_o|| \leq const. \max\{\|h_{oj} - \pi_n h_{oj}\|_\infty : j = 1, ..., q\}$。定义 $C = \sqrt{E\{U(Z_t)^2\}}$, 那么对于所有 $0 < \frac{w}{C} \leq \delta < 1$, $H_{[]}(w, \mathcal{F}_n, ||\cdot||_2) \leq \sum_{j=1}^q \log N(\frac{w}{C}, \mathcal{H}_n^j, ||\cdot||_\infty)$。

最后一部分计算取决于筛的选择。首先, 根据 Anastassiou and Yu (1992a) 有 $\|h_{o1} - \pi_n h_{o1}\|_\infty = O((k_{1n})^{-1})$; 并且根据 Lorentz (1966), 对 $j = 2, ..., q$ 有 $\mathcal{H}^j = \Lambda_{c_j}^{p_j}$, $\|h_{oj} - \pi_n h_{oj}\|_\infty = O((k_{jn})^{-p_j})$。其次, 根据 van de Geer (2000) 的引理 2.5, 对 $j = 1, 2, ..., q$, 有 $\log N(\frac{w}{C}, \mathcal{H}_n^j, ||\cdot||_\infty) \leq const \times k_{jn} \times \log(1 + \frac{4c_j}{w})$。因此 $\delta_n$ 解决

$$\begin{aligned}\frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{H_{[]}(w, \mathcal{F}_n, ||\cdot||_2)}dw &\leq \frac{1}{\sqrt{n}\delta_n^2}\max_{j=1,...,q}\int_{b\delta_n^2}^{\delta_n}\sqrt{k_{jn} \times \log(1 + \frac{4c_j}{w})}dw \\ &\leq \frac{1}{\sqrt{n}\delta_n^2}\max_{j=1,...,q}\sqrt{k_{jn}} \times \delta_n \leq const\end{aligned}$$

而且上述问题的解 $\delta_n \asymp \max_{j=1,...,q}\sqrt{\frac{k_{jn}}{n}}$。根据定理 3.2, $\|\widehat{\theta}_n - \theta_o\| = O_P\left(\max_{j=1,...,q}\{(k_{jn})^{-p_j}, \delta_n\}\right)$。通过选择 $k_{jn} = O(n^{1/(2p_j+1)})$, $j = 1, ..., q$, 我们得到 $\|\widehat{\theta}_n - \theta_o\| = O_P(n^{-p/(2p+1)})$, 其中 $p = \min\{p_1, \cdots, p_q\} > 0.5$。这意味着 $\|\widehat{h}_j - h_{oj}\|_2 = O_P(n^{-p/(2p+1)})$, $j = 1, ..., q$。 □

**备注 3.5. (1)** 由于假设 3.1 中列出的参数空间 $\mathcal{H} = \mathcal{H}^1 \times \cdots \times \mathcal{H}^q$ 在 $||\cdot||$ 范数下紧凑, 我们可以使用原参数空间 $\mathcal{H}$ 作为筛空间 $\mathcal{H}_n$。再次应用定理 3.2 , 注意到近似误差 $\|\pi_n\theta_o - \theta_o\| = 0$, 我们得到 $\|\widehat{\theta}_n - \theta_o\| = O_P(\delta_n)$, 其中 $\delta_n$ 是以下问题的解:

$$\begin{aligned}&\frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{\sum_{j=1}^q \log N(w, \mathcal{H}^j, ||\cdot||_\infty)}dw \\ \leq\ &\frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{\sum_{j=1}^q \left(\frac{c_j}{w}\right)^{1/p_j}}dw \quad \text{根据 Birman and Solomjak (1967)} \\ \leq\ &\frac{1}{\sqrt{n}\delta_n^2}\max_{j=1,...,q} const. (\delta_n)^{1-\frac{1}{2p_j}} \leq const.\end{aligned}$$

当 $\delta_n = O(n^{-p/(2p+1)})$ 其中 $p = \min\{p_1, \cdots, p_q\} > 0.5$ 时上式成立。然而, 现在还不清楚在给定有限数据集时如何在整个参数空间 $\mathcal{H}$ 上实现上述优化过程。

**(2)** 假设在命题 3.3 中我们用 $h_{o1} \in \Lambda_{c_1}^{p_1}([b_{11}, b_{21}])$ 替换假设 3.1(i), 并令 $\mathcal{H}_n^1 = \text{Pol}(k_{1n})$, 或 $\text{TriPol}(k_{1n})$, 或 $\text{Spl}(r_1, k_{1n})$, 其中 $r_1 \geq [p_1] + 1$, 或 $\text{Wav}(m_1, 2^{J_{1n}})$ 其中 $m_1 > p_1$, $2^{J_{1n}} = k_{1n}$。定义 $p = \min\{p_1, \cdots, p_q\} > 0.5$。那么我们有 $\|\widehat{h}_j - h_{oj}\|_2 = O_P(n^{-p/(2p+1)})$, $j = 1, ..., q$。进一步的, 对某整数 $m \geq 1$ 令 $||D^m\widehat{h}_j - D^m h_{oj}||_2 = \{E[D^m\widehat{h}_j(X_{jt}) - D^m h_{oj}(X_{jt})]^2\}^{1/2}$。如果 $p > m \geq 1$ 那么对 $j = 1, ..., q$ 有 $||D^m\widehat{h}_j - D^m h_{oj}||_2 = O_P(k_{jn}^{-(p-m)}) = O_P(n^{-(p-m)/(2p+1)})$。这一收敛速度达到了 Stone (1982) 中得到的最优收敛速度。

### 3.2.2 例：多元分位数回归

假设 i.i.d. 数据 $\{Y_t, X_t\}_{t=1}^n$ 由下列模型生成：

$$Y_t = \theta_o(X_t) + e_t, \quad P[e_t \le 0|X_t] = \alpha \in (0,1),$$

其中 $X_t \in \mathcal{X} = \mathcal{R}^d$, $d \ge 1$。我们通过在 $\Theta_n$ 上最大化目标函数 $\widehat{Q}_n(\theta) = n^{-1}\sum_{t=1}^n l(\theta, Y_t, X_t)$ 来估计条件分位函数 $\theta_o(\cdot)$；目标函数中 $l(\theta, Y_t, X_t) = \{1(Y_t < \theta(X_t)) - \alpha\}[Y_t - \theta(X_t)]$。选择 $\|\theta - \theta_o\|^2 = \mathrm{E}(\theta(X_t) - \theta_o(X_t))^2$ 以及令 $W_1^1(\mathcal{X})$ 为 2.3.3 节中定义的 Sobolev 空间。

**假设 3.3.** $\theta_o \in \Theta = W_1^1(\mathcal{X})$。

**假设 3.4.** 用 $f_{e|X}$ 表示给定 $X_t$ 的 $e_t$ 的条件密度，该条件密度满足 $0 < \inf_{x\in\mathcal{X}} f_{e|X=x}(0) \le \sup_{x\in\mathcal{X}} f_{e|X=x}(0) < \infty$ 和随着 $|z| \to 0$，$\sup_{x\in\mathcal{X}} |f_{e|X=x}(z) - f_{e|X=x}(0)| \to 0$。

我们已经知道有限维线性筛的张量乘积（如 2.3.1 中的例子）无法很好地近似 $W_1^m(\mathcal{X})$, $m \ge 1$ 中的函数，因此，基于这类线性筛的估计量的收敛速度将比基于非线性筛的要慢；具体例子可以参考如 Chen and Shen (1998, 命题 1 情形 1.3(ii))。对于时间序列回归模型，Chen and White (1999), Chen et al. (2001) 已经证明了用神经网络筛近似 $W_1^m(\mathcal{X})$ 中的函数有更快的收敛速度。因此为了估计未知的 $\theta_o \in W_1^1(\mathcal{X})$，我们考虑如下高斯径向 ANN 筛 $\Theta_n$：

$$\Theta_n = \left\{ \alpha_0 + \sum_{j=1}^{k_n} \alpha_j G\left( \frac{\{(x-\gamma_j)'(x-\gamma_j)\}^{1/2}}{\sigma_j} \right), \sum_{j=0}^{k_n} |\alpha_j| \le c_0,\ |\gamma_j| \le c_1, 0 < \sigma_j \le c_2 \right\},$$

其中 $G$ 是标准高斯密度函数。

**命题 3.2.** 用 $\widehat{\theta}_n$ 表示筛 $M$ 估计。如果假设 3.3 和 3.4 成立。令 $k_n^{2(1+1/(d+1))}\log(k_n) = O(n)$。那么 $\|\widehat{\theta}_n - \theta_o\| = O_P([n/\log n]^{-(1+2/(d+1))/[4(1+1/(d+1))]})$。

证明. 定理 3.2 可以直接用来证明这个结果。条件 3.6 直接假设成立。根据在条件密度 $f_{e|X}$ 上的假设，很容易验证 $K(\theta_o, \theta) \asymp \mathrm{E}(\theta(X_t) - \theta_o(X_t))^2$；细节可以参考 Chen and White (1999, p686-687)。现在我们验证条件 3.7 和 3.8。注意到 $|l(\theta, Y_t, X_t) - l(\theta_o, Y_t, X_t)| \le \max(\alpha, 1-\alpha)|\theta(X_t) - \theta_o(X_t)|$, 我们有

$$\mathrm{Var}\left(l(\theta, Y_t, X_t) - l(\theta_o, Y_t, X_t)\right) \le E[l(\theta, Y_t, X_t) - l(\theta_o, Y_t, X_t)]^2 \le E[\theta(X_t) - \theta_o(X_t)]^2,$$

因此条件 3.7 成立。此外，我们有

$$\sup_{\{\theta\in\Theta_n : \|\theta-\theta_o\|\le\delta\}} |l(\theta, Y_t, X_t) - l(\theta_o, Y_t, X_t)| \le \sup_{\{\theta\in\Theta_n : \|\theta-\theta_o\|\le\delta\}} |\theta(X_t) - \theta_o(X_t)|,$$

以及由 Gabushin (1967) 的定理 1 得到的 $\|\theta - \theta_o\|_\infty \le c\|\theta - \theta_o\|^{2/3}$。因此，条件 3.8 在 $s = 2/3$, $U(X_t) \equiv c$ 下满足。

现在根据 Chen et al. (2001) 中得到的结果，我们有 $\|\theta_o - \pi_n\theta_o\| \le const.(k_n)^{-1/2-1/(d+1)}$ 和 $\log N(w, \Theta_n, ||\cdot||_\infty) \le const.k_n \log(\frac{k_n}{w})$。选择 $k_n^{2(1+1/(d+1))}\log(k_n) = O(n)$，我们应用定理 3.2很容易看出以下结果：$\|\widehat{\theta}_n - \theta_o\| = O_P([n/\log n]^{-(1+2/(d+1))/[4(1+1/(d+1))]})$。 □

## 3.3 级数估计量的收敛速度

在本小节中，我们将给出凹扩展线性模型的级数估计量的收敛速度。之前讨论过，在该框架下参数空间 $\Theta$ 往往是一个线性空间。该空间一般具有下列特征：是平方可积函数空间的子空间；样本目标函数 $\widehat{Q}_n(\theta) = n^{-1}\sum_{i=1}^n l(\theta, Z_i)$ 几乎处处是 $\theta \in \Theta$ 的凹函数；整体目标函数 $Q(\theta) = E[l(\theta, Z_i)]$ 是 $\theta \in \Theta$ 的严格凹函数。这里我们给出的结果主要来自 Huang (1998a, 2001) 和 Newey (1997)。

本小节中，$\{Z_i\}_{i=1}^n$ 是 i.i.d.，$\theta$ 表示具有有界定义域 $\mathcal{X} \subset \mathcal{R}^d$ 的实值函数。我们用 $\|\widehat{\theta} - \theta_o\|$ 来测度 $\widehat{\theta}$ 与 $\theta_o$ 之间的距离。

**条件 3.9.** 对任意 *Lebesgue* 平方可积函数 $\theta$ 有 $\|\theta\| \asymp \|\theta\|_{2,leb}$。

在例 2.4 中的多元 LS 回归里，$\theta_o(X) = E[Y|X]$，很自然地我们可以选 $\|\theta\| = ||\theta||_2 = \{E[\theta(X)^2]\}^{1/2}$ 作为范数。如果 $X$ 的密度是严格有上下界且下界大于 0，那么条件 3.9 得到满足。一般而言，范数 $\|\cdot\|$ 的选择需要取决于实际应用和数据生成过程。

我们在线性筛空间上添加以下条件。

**条件 3.10.** 有限维线性筛空间，$\Theta_n$，理论上可以被识别；也就是说，任何满足 $\|\theta\| = 0$ 的 $\theta \in \Theta_n$ 意味着处处 $\theta(u) = 0$。

在条件 3.9下，使用如 2.3.1 中提到的常见线性近似空间都可以满足条件 3.10。

**条件 3.11.** $\theta_o = \arg\max_\Theta E[l(\theta, Z)]$ 满足 $\|\theta_o\|_\infty \leq K_o < \infty$。

**条件 3.12.** 对任意有界函数 $\theta_1, \theta_2 \in \Theta$, $E[l(\theta_1 + \tau(\theta_2 - \theta_1), Z)]$ 在 $\tau \in [0,1]$ 上两次连续可导。对任意常数 $0 < K < \infty$, 有 $\frac{\partial^2}{\partial \tau^2} E[l(\theta_1 + \tau(\theta_2 - \theta_1), Z)] \asymp -\|\theta_2 - \theta_1\|^2$ 对于满足 $\|\theta_1\|_\infty \leq K$ 和 $\|\theta_2\|_\infty \leq K$ 和 $0 \leq \tau \leq 1$ 的 $\theta_1, \theta_2 \in \Theta$ 成立。

给定上述条件，我们可以定义 $\overline{\theta}_n \equiv \arg\max_{\theta \in \Theta_n} E[l(\theta, Z)]$，很容易看出 $||\overline{\theta}_n - \theta_o|| \asymp \inf_{\theta \in \Theta_n} ||\theta - \theta_o||$。

**条件 3.13.** 对任意函数对 $\theta_1, \theta_2 \in \Theta_n$, $l(\theta_1 + \tau(\theta_2 - \theta_1), Z)$ 在 $\tau$ 上两次连续可导。此外，*(i)*

$$\sup_{g \in \Theta_n} \frac{\left| \frac{\partial}{\partial \tau} l(\overline{\theta}_n + \tau g, Z)\Big|_{\tau=0} \right|}{\|g\|} = O_P\left(\sqrt{\frac{\dim(\Theta_n)}{n}}\right);$$

*(ii)* 对任意常数 $0 < K < \infty$, 存在 $c > 0$ 使得 $\frac{\partial^2}{\partial \tau^2} l(\theta_1 + \tau(\theta_2 - \theta_1), Z) \leq -c\|\theta_2 - \theta_1\|^2$ 对任意满足 $\|\theta_1\|_\infty \leq K$ 和 $\|\theta_2\|_\infty \leq K$ 以及 $0 \leq \tau \leq 1$ 的 $\theta_1, \theta_2 \in \Theta_n$ 都成立；一个例外是在某个随着 $n \to \infty$ 其概率趋向于 *0* 的事件上，这时上述定义的 $c$ 不再存在。

定义 $k_n = \dim(\Theta_n)$, $A_n \equiv \sup_{\theta \in \Theta_n, \|\theta\|_{2,leb} \neq 0} (||\theta||_\infty / ||\theta||_{2,leb})$ 以及 $\rho_{2n} \equiv \inf_{\theta \in \Theta_n} \|\theta - \theta_o\|_{2,leb}$。在条件 3.9 - 3.11 下，我们有 $\rho_{2n} \asymp \inf_{\theta \in \Theta_n} \|\theta - \theta_o\|$。下面的结果是 Huang (2001) 文中关于凹拓展线性模型的筛估计量的一个特例。

**定理 3.3.** 假设条件 3.9–3.13 成立。让 $\lim_{n\to\infty} A_n\rho_{2n} = 0$ 和 $\lim_{n\to\infty} A_n^2 k_n/n = 0$。那么级数估计量 $\widehat{\theta}$ 唯一确定地存在，该事件随着 $n \to \infty$ 依概率趋向于 *1*，并且

$$\|\widehat{\theta} - \theta_o\| = O_P\Big(\sqrt{\frac{k_n}{n}} + \rho_{2n}\Big)$$

定理 3.3 可以看作是定理 3.2在 $\delta_n \asymp \sqrt{\frac{k_n}{n}}$ 和 $||\pi_n\theta_o - \theta_o|| \asymp \rho_{2n}$ 下的一种特殊情况。具体而言，首先注意在条件 3.9 - 3.11 下存在一个基本唯一确定的元素 $\pi_n\theta_o \in \Theta_n$ 满足 $||\pi_n\theta_o - \theta_o|| = \inf_{\theta\in\Theta_n} ||\theta - \theta_o||$ 与 $||\pi_n\theta_o - \theta_o|| \asymp ||\pi_n\theta_o - \theta_o||_{2,leb} \asymp \rho_{2n}$ 是近似误差率。其次，在凹拓展线性模型中对有限维线性筛 $\Theta_n$ 而言我们有 $\log N(w, \Theta_n, ||\cdot||_\infty) \le const.k_n \log(\frac{1}{w})$，因此 $\delta_n \asymp \sqrt{\frac{k_n}{n}}$。

常数 $A_n \ge 1$ 是有限维线性筛空间 $\Theta_n$ 不规则性的一种测度。由于我们要求 $\Theta_n$ 在理论上可以被识别，同时 $\Theta_n$ 中的函数是有界的，所以 $A_n$ 是有限的。事实上，让 $\{\phi_j, j = 1, \ldots, k_n\}$ 表示一组 $\Theta_n$ 的相对理论内积的标准正交基。那么，根据 Cauchy–Schwarz 不等式，$A_n \le \{\sum_{j=1}^{k_n} \|\phi_j\|_\infty^2\}^{1/2} < \infty$。很明显对所有 $\theta \in \Theta_n$ 都有 $\|\theta\|_\infty \le A_n\|\theta\|_{2,leb}$。我们通常在常见的近似空间中选择线性筛空间，例如第 2.3.1 节所述的近似空间；这样通过直接应用近似理论文献中的已有结果可以容易地获得相关常数 $A_n$。以下是几个例子。

**多项式**如果 $\Theta_n = \mathrm{Pol}(J_n)$ 和 $\mathcal{X} = [0, 1]$，那么 $A_n \asymp J_n$ (参考 DeVore and Lorentz, 1993 定理 4.2.6)。

**三角多项式**如果 $\Theta_n = \mathrm{TriPol}(J_n)$ 和 $\mathcal{X} = [0, 1]$，那么 $A_n \asymp J_n^{1/2}$ (参考 DeVore and Lorentz, 1993 定理 4.2.6)。

**单元样条**如果 $\Theta_n = \mathrm{Spl}(r, J_n)$ 和 $\mathcal{X} = [0, 1]$, 那么 $A_n \asymp J_n^{1/2}$ (参考 DeVore and Lorentz, 1993 定理 5.1.2)。

**正交小波**如果 $\Theta_n = \mathrm{Wav}(m, 2^{J_n})$ 和 $\mathcal{X} = [0, 1]$, 那么 $A_n \asymp 2^{J_n/2}$ (参考 Meyer, 1992 引理 2.8 )。

**张量积空间**让 $\Theta_n$ 表示 $\Theta_{n1}, \ldots, \Theta_{nd}$ 的张量积。张量积线性筛空间 $\Theta_n$ 的常数 $A_n$ 可以通过其成分空间的相应常数得到。令 $a_{n\ell} = \sup_{\theta\in\Theta_{n\ell}, \|\theta\|_{2,leb}\neq 0}(||\theta||_\infty/||\theta||_{2,leb})$，$1 \le \ell \le d$。Huang (1998a) 证明了 $A_n \le const. \prod_{\ell=1}^d a_{n\ell}$。

我们将上述结论应用于例 2.4 的多元 LS 回归。

**假设 3.5.** *(i)* $X$ 具有紧凑支集 $\mathcal{X}$ 并有在 $\mathcal{X}$ 上的严格有界正密度函数，这里 $\mathcal{X} \subset \mathcal{R}^d$ 是紧凑区间 $\mathcal{X}_1, ..., \mathcal{X}_d$ 的笛卡尔积；*(ii)* $Var(Y|X = \cdot)$ 在 $\mathcal{X}$ 上有界；*(iii)* $h_o(\cdot) = E[Y|X = \cdot] \in \Lambda^p(\mathcal{X})$ 其中 $p > d/2$.

定理 3.3 可以处理一般有限维线性筛空间 $\Theta_n$。为简单起见，我们在这里只考虑例 2.4 中当筛空间 $\Theta_n$ 是常见一元线性近似空间 $\Theta_{n1}, \ldots, \Theta_{nd}$ 的张量积的情形。那么 $k_n = \dim(\Theta_n) = \prod_{\ell=1}^d \dim(\Theta_{n\ell})$。

**命题 3.3.** 给定假设 3.5 成立，用 $\widehat{h}_n$ 表示例 *2.4* 中 $h_o$ 的级数估计量，其中筛 $\Theta_n$ 定义为一元筛空间 $\Theta_{n1}, \ldots, \Theta_{nd}$ 的张量积。对 $\ell = 1, ..., d$,

- 如果 $\Theta_{n\ell} = \mathrm{Pol}(J_n)$, $p > d$ 和 $J_n^{3d}/n \to 0$，那么 $\|\widehat{h}_n - h_o\| = O_P(\sqrt{J_n^d/n} + J_n^{-p})$;
- 如果 $\Theta_{n\ell} = \mathrm{TriPol}(J_n)$, $p > d/2$ 与 $J_n^{2d}/n \to 0$ 那么 $\|\widehat{h}_n - h_o\| = O_P(\sqrt{J_n^d/n} + J_n^{-p})$;
- 如果 $\Theta_{n\ell} = \mathrm{Spl}(r, J_n)$ 其中 $r \ge [p]+1$, $p > d/2$ 以及 $J_n^{2d}/n \to 0$, 那么 $\|\widehat{h}_n - h_o\| = O_P(\sqrt{J_n^d/n} + J_n^{-p})$.

令 $J_n = O(n^{1/(2p+d)})$，那么 $\|\widehat{h}_n - h_o\| = O_P(n^{-p/(2p+d)})$。

我们注意到，这个命题也可以作为 Newey（1997）中定理 1 的直接结果。[34] 选择合适的 $J_n \asymp n^{1/(2p+d)}$ 使得在方差 $(J_n^d/n)$ 和偏差平方 $(J_n^{-2p})$ 中取得平衡：$J_n^d/n \asymp J_n^{-2p}$。得到的收敛速度 $n^{-2p/(2p+d)}$ 在回归和密度估计中实际上是最优的：根据 Stone ( 1982)，没有其他估计可以一致地在 $p$-平滑函数类上取得更快地收敛速度。收敛速度取决于两个量：目标函数 $\theta_o$ 的平滑度 $p$，以及目标函数的定义域的维度 $d$。这里需要注意维度 $d$ 对收敛速度的影响：给定平滑度 $p$，定义域维度越大，收敛速度越慢; 此外，当维数趋于无穷大时收敛速度趋向于 0。这为“维度的诅咒”现象提供了一个数学上的解释。在未知多元函数上添加可加性会使得估计量具有更快的收敛速度；参考 3.2.1 小节，Stone (1985, 1986), Andrews and Whang (1990), Huang (1998b) 以及 Huang et al (2000)。

## 3.4 级数最小二乘（LS）估计量的逐点渐近正态性

到目前为止，我们已经建立了相对完整的筛 M 估计量收敛速度理论。然而，相对应的渐近分布理论还不完整，需要更多这方面的工作。所有现有已知结论都是针对密度和 LS 回归函数的级数估计量的。级数 LS 估计量的渐近正态性在 Andrews (1991b), Gallant and Souza (1991), Newey (1994b, 1997), Zhou et al. (1998), 和 Huang (2003) 中都已经研究过。Stone (1990) 和 Strawderman and Tsiatis (1996) 已经分别在密度估计和风险估计方面给出了多项式样条估计量的渐近正态性。[35]

在本小节中我们关注例 2.4。也就是说，我们假设数据 $\{Z_i = (Y_i, X_i')'\}_{i=1}^n$ 是 i.i.d.，以及我们关注的参数 $\theta_o(\cdot) = h_o(\cdot) = E[Y|X = \cdot]$，是具有有界定义域 $\mathcal{X} \subset \mathcal{R}^d$ 的实值函数。

### 3.4.1 样条级数 LS 估计量的渐近正态性

这里我们介绍 Huang (2003) 关于样条级数 LS 估计量的逐点渐近正态性的结论。

**假设 3.6.** *(i)* $Var(Y|X = \cdot)$ *在* $\mathcal{X}$ *上严格为正; (ii)*

$$\sup_{x\in\mathcal{X}} E\left[\{Y - h_o(X)\}^2 \times 1(|Y - h_o(X)| > \lambda) \mid X = x\right] \to 0 \quad \text{随着} \lambda \to \infty.$$

下文中 $\Phi(\cdot)$ 表示标准正态分布函数，同时 $\mathrm{SD}(\hat{h}(x)|X_1, \ldots, X_n) = \{\mathrm{Var}(\hat{h}(x)|X_1, \ldots, X_n)\}^{1/2}$。

**定理 3.4.** *[Huang 2003]* 给定假设 3.5 和 3.6 成立。用 $\widehat{h}_n$ 表示例 *2.4* 中的 $h_o$ 的级数估计，其中所使用的筛 $\Theta_n$ 是一元样条筛空间 $\Theta_{n\ell} = \mathrm{Spl}(r, J_n)$, $r \geq [p] + 1$, $1 \leq \ell \leq d$ 的张量积。如果 $\lim_{n\to\infty} J_n^d \log n/n = 0$ 与 $\lim_{n\to\infty} J_n/n^{1/(2p+d)} = \infty$，那么

$$\Pr\left(\hat{h}(x) - h_o(x) \leq t \times \mathrm{SD}(\hat{h}(x)|X_1, \ldots, X_n)\right) \to \Phi(t), \quad t \in \mathcal{R}.$$

[34] 命题 3.6 是在 $||\cdot||_2-$ 范数下有关 LS 回归的收敛速度的。此外也有一些在 $||\cdot||_\infty-$ 范数下的结论；可以参考如 Stone (1982), Newey (1997) 和 de Jong (2002)。

[35] 关于同上述问题紧密相关的平滑样条分位数估计量的渐近正态性可以参考 Portnoy (1997)。

例如定理 3.4之类的渐近分布结果可以用来构造渐近置信区间。用 $\widehat{\mathrm{SD}}(\hat{h}(x)|X_1,\ldots,X_n)$ 表示 $\mathrm{SD}(\hat{h}(x)|X_1,\ldots,X_n)$ 的一致估计；其具体表达式可以参考 Andrews (1991b) 和 Newey (1997)。令 $\hat{h}^l_\alpha(x)=\hat{h}(x)-z_{1-\alpha/2}\widehat{\mathrm{SD}}(\hat{h}(x)|X_1,\ldots,X_n)$ 和 $\hat{h}^u_\alpha(x)=\hat{h}(x)+z_{1-\alpha/2}\widehat{\mathrm{SD}}(\hat{h}(x)|X_1,\ldots,X_n)$, 其中 $z_{1-\alpha/2}$ 是标准正态分布的 $(1-\alpha/2)$ 分位数。如果定理 3.4的条件都满足，那么 $[\hat{h}^l_\alpha(x),\hat{h}^u_\alpha(x)]$ 是 $h_o(x)$ 的渐近 $1-\alpha$ 置信区间；也就是说，$\lim_{n\to\infty}P(\hat{h}^l_\alpha(x)\le h_o(x)\le\hat{h}^u_\alpha(x))=1-\alpha$。

回顾张量积样条筛 $\Theta_n$, $k_n=\dim(\Theta_n)\asymp J_n^d$。如果 $h_o(\cdot)$ 是 $p$-平滑，那么张量积样条筛有偏差阶为 $J_n^{-p}\asymp k_n^{-p/d}$。定理 3.4中的条件 $\lim_{n\to\infty}J_n/n^{1/(2p+d)}=\infty$ 意味着偏差项相对估计的标准差而言可以渐近近似忽略。条件 $\lim_{n\to\infty}k_n/n^{d/(2p+d)}=\infty$ 常常被称作平滑不足（或过度拟合）；也就是说，在平滑不足情况下所使用的筛参数 $(k_n)$ 的个数超过了 Stone (1982) 文中达到最优收敛速度的筛参数的个数。

### 3.4.2 级数 LS 估计量的泛函的渐近正态性

我们现在回顾 Newey (1997) 中关于级数 LS 估计量的泛函的渐近正态性的结果。令 $a:\Theta\to\mathcal{R}$ 表示一个已知泛函，我们希望估计 $a(h_o)$，其中 $h_o(\cdot)=E[Y|X=\cdot]\in\Theta$。之前我们已经得到 $\widehat{h}(\cdot)=p^{k_n}(\cdot)'(P'P)^-\sum_{i=1}^n p^{k_n}(X_i)Y_i$ 是 $h_o(\cdot)$ 的级数 LS 估计量，其中 $p^{k_n}(X)$ 是有限维线性筛 (2.10)（参考例 2.4)。那么自然地可以选择 $a(\widehat{h})$ 作为 $a(h_o)$ 的一个估计量。

选择一个非负整数 $s\ge 0$，并且在 $\Theta$ 上定义一个强范数为 $\|h\|_{s,\infty}=\max_{[\gamma]\le s}\sup_{x\in\mathcal{X}}|D^\gamma h(x)|$。此外，令 $\zeta_0(k_n)\equiv\sup_{x\in\mathcal{X}}|p^{k_n}(x)|_e$, $\zeta_s(k_n)\equiv\max_{[\gamma]\le s}\sup_{x\in\mathcal{X}}|D^\gamma p^{k_n}(x)|_e$, 其中 $|\cdot|_e$ 是欧几里得范数。

**假设 3.7.** *(i)* $Var(Y|X=\cdot)$ *在* $\mathcal{X}$ *上严格为正;* $\sup_{x\in\mathcal{X}}E\left[\{Y-h_o(X)\}^4|X=x\right]<\infty$*; (ii)* $E[p^{k_n}(X)p^{k_n}(X)']$ *的最小特征值在* $k_n$ *上一致地严格为正; (iii) 对于非负整数* $s\ge 0$ *存在* $\alpha>0$, $\beta^*_{k_n}$ *such that* $\inf_{g\in\Theta_n}\|g-h_o\|_{s,\infty}=\|p^{k_n}(\cdot)'\beta^*_{k_n}-h_o(\cdot)\|_{s,\infty}=O(k_n^{-\alpha})$。

**假设 3.8.** *下列两个条件至少一个成立: (i)* $\lim_{n\to\infty}k_n\{\zeta_0(k_n)\}^2/n=0$, *同时* $a(h)$ *是* $h\in\Theta$ *的线性泛函; (ii) 在假设 3.7 中的* $s$ *有* $\lim_{n\to\infty}k_n^2\{\zeta_s(k_n)\}^4/n=0$, *同时存在* $h\in\Theta$ *的线性函数* $D(h;\widetilde{h})$ *满足对于某* $c_1,c_2,\varepsilon>0$ *和对所有满足* $||\widetilde{h}-h_o||_{s,\infty}<\varepsilon$ *和* $||\overline{h}-h_o||_{s,\infty}<\varepsilon$ *的* $\widetilde{h},\overline{h}$, *下列为真*

$$\begin{aligned}|a(h)-a(\widetilde{h})-D(h-\widetilde{h};\widetilde{h})| &\le c_1\{||h-\widetilde{h}||_{s,\infty}\}^2;\ \text{与}\\ |D(h;\widetilde{h})-D(h;\overline{h})| &\le c_2||h||_{s,\infty}||\widetilde{h}-\overline{h}||_{s,\infty}。\end{aligned}$$

**假设 3.9.** *(i) 存在正整数* $c$ *满足* $|D(h;h_o)|\le c||h||_{s,\infty}$（$s$ *同假设 3.7 中的定义)；(ii) 存在* $h_n\in\Theta_n$ *使得* $E[h_n(X)^2]\to 0$, *但* $D(h_n;h_o)$ *不等于 0。*

假设 3.7(iii) 是在强范数 $||h||_{s,\infty}$ 下的在筛近似误差上的条件。假设 3.8 意味着 $a(h)$ 在 $h$ 上 Frechet 可导；定义导数所用的范数为 $||h||_{s,\infty}$。假设 3.9 意味着导数 $D(h;h_o)$ 在范数 $||h||_{s,\infty}$ 下连续，但在均方范数 $||h||_2=\{E[h(X)^2]\}^{1/2}$ 下不连续。这一均方不连续特征说明估计量 $a(\widehat{h})$ 对 $a(h_o)$ 并不是 $\sqrt{n}-$ 一致的；更多细节讨论可以参考 Newey (1997)。接下来我们令 $\Sigma=E[p^{k_n}(X)p^{k_n}(X)'Var(Y|X)]$,

$$A=\frac{\partial a(p^{k_n}(X)'\beta)}{\partial\beta}|_{\beta^*_{k_n}}\quad\text{与}\quad V_{k_n}=A'\{E[p^{k_n}(X)p^{k_n}(X)']\}^{-1}\Sigma\{E[p^{k_n}(X)p^{k_n}(X)']\}^{-1}A。$$

我们让 $\xrightarrow{d}$ 表示依分布收敛，$\mathcal{N}(0,1)$ 表示一个由标准正态分布生成的标量随机变量。

**定理 3.5.** *[Newey 1997]* 给定假设 3.5*(i)(ii), 3.7 - 3.9* 成立。用 $\widehat{h}_n$ 表示例 *2.4* 中的 $h_o$ 的级数估计量，其中所用筛 $\Theta_n$ 为线性筛 *(2.10)*。如果 $\lim_{n\to\infty}\sqrt{n}k_n^{-\alpha}=0$，那么

$$\sqrt{\frac{n}{V_{k_n}}}\Big(a(\widehat{h})-a(h_o)\Big)\xrightarrow{d}\mathcal{N}(0,1)$$

我们注意到对线性泛函 $a(h_o)=h_o(x)$ 而言，该理论意味着任何满足假设 3.5(i)(ii), 3.7, 3.8(i) 和 3.9(ii) 的级数 LS 估计量 $\widehat{h}(x)$ 都有逐点渐近正态性。当我们将范围收窄到 $h_o(x)$ 的张量积样条级数估计量时，那么假设 3.8(i) 要求 $\lim_{n\to\infty}k_n^2/n=0$，这比定理 3.4中要求的 $\lim_{n\to\infty}k_n\log n/n=0$ 更严格。然而，定理 3.4仅适用于样条级数 LS 估计量，与之相比 Newey (1994b, 1997) 的结果更有一般性。

本节中得到的正态分布结果仅在 i.i.d. 数据下成立。对于基于时间序列非独立观测下的级数 LS 估计量的线性泛函的渐近正态性可以参考 Andrews (1991b)。

# 4 半参模型有限维参数部分的筛估计的大样本性质

在第二章中的一般筛极值估计框架下，一个模型通常有一个参数向量 $\theta=(\beta,h)$，其中 $\beta$ 代表有限维未知参数向量，$h$ 则是无限维未知参数。当 $\beta$ 和 $h$ 都是我们关心的对象时，我们称这个模型是 “半-非参数” 模型。当 $h$ 是多余参数时，沿用 Powell (1994) 及其他文献中的命名方式，我们称该模型是 “半参数” 模型。

对于弱相关观测，半参数模型可以分为以下两类：(i) 不能以 $\sqrt{n}-$ 收敛速度估计 $\beta$，也就是说 $\beta$ 具有 0 信息界；参考 van der Vaart (1991); (ii) 可以以 $\sqrt{n}-$ 收敛速度估计 $\beta$。第一类模型其实应该属于非参模型。然而，由于这类模型仍然可以用筛分方法来估计，我们可以应用一般筛估计的收敛速度结果来为这类参数 $\beta$ 得到慢于 $\sqrt{n}-$ 的收敛速度。到目前为止，几乎没有关于 $\beta$ 的筛估计是否可以达到最佳收敛速度及其极限分布的研究。值得一提的是针对 Heckman and Singer's (1984) 模型，Ishwaran (1996a) 证明了 $\beta$ -参数不能以 $\sqrt{n}-$ 速度估计；但是 Ishwaran (1996b) 构造了 $\beta$ 的另一个估计量，该估计量可以达到最优收敛速度，然而它并不是筛 MLE 估计量。在 Ishwaran (1996a, b) 之前，Honore (1990, 1994) 巧妙提出一个 $\beta$ 的估计量并计算了其收敛速度；该估计量也不是筛 MLE。到目前为止 Heckman and Singer (1984) 中使用的 MLE 估计量能否达到 Ishwaran 的最优收敛速度仍然是未解之谜。[36]

现在已经有大量的关于第 (ii) 类模型中 $\beta$ 的半参估计的文献；具体总结可以参考 Bickel et al. (1993), Newey and McFadden (1994), Powell (1994), Horowitz (1998) 以及 Pagan and Ullah (1999)。绝大部分这方面的结果是使用 “两步法” 得到的：第一步用 $\widehat{h}$ 非参地估计 $h$，第二部用 M-估计量、广义矩估计 GMM 或更一般的 MD-估计量来估计 $\beta$，这里我们用第一步得到的 $\widehat{h}$ 替换未知 $h$。此外也有一些同时估计 $\beta$ 和 $h$ 一般方法。例如，筛同步方法通过在筛参数空间 $\Theta_n=B\times\mathcal{H}_n$ 上最大化样本目标函数 $\widehat{Q}_n(\beta,h)$ 来共同估计 $\beta$

[36]计量经济学中在第 (i) 类模型的一些特定形式下还有其他重要的结论。例如，Manski (1985) 为中位数为 0 的二元选择模型提出了最大 score 估计量；Kim and Pollard (1990) 为 Manski 的估计量证明了 $n^{1/3}$ 一致性；Horowitz (1992) 在 Manski 的估计量的基础上提出了平滑最大 score 估计量，并证明了他的估计量以快于 $n^{1/3}$ 的速度收敛到真值而且其渐近分布为正态分布；Andrews and Schafgans (1998) 为 Heckman 的样本选择模型提出了慢于 $\sqrt{n}$ 收敛速度的核估计量；Honore and Kyriazidou (2000) 为离散选择动态面板数据模型提出了慢于 $\sqrt{n}$ 收敛速度的核估计量；更多例子可以参考 Powell (1994), Horowitz (1998), Pagan and Ullah (1999)。

和 $h$。较早的筛 MLE 在计量经济学中的应用使用了这种方法，例如 Duncan (1986) 和 Gallant and Nychka (1987)。

在 4.1 小节中我们回顾 $\beta$ 的两步法估计的 $\sqrt{n}-$ 渐近正态性方面的现有理论。4.2 小节总结在 $\beta$ 的筛联立 M-估计的 $\sqrt{n}-$ 渐近正态性和效率方面研究的最新进展。4.3 小节关注 $\beta$ 的联立筛 MD 估计的 $\sqrt{n}-$ 渐近正态性和估计效率。

## 4.1 半参两步法估计量

计量经济学中已经有一些关于半参数两步法的一般理论的论文。Andrews (1994b) 提出 $\beta$ 的 MINPIN 估计量；这是一种极值估计量，其经验目标函数取决于第一步 $h$ 的非参估计。Andrews (1994b) 还为这个 $\beta$ 的 MINPIN 估计量给出了一系列相对宽泛的充分条件来保证其 $\sqrt{n}-$ 正态性。Ichimura and Lee (2006) 则提出了一系列相对细致的充分条件以保证未知参数 $\beta$ 的半参数两步法 M-估计量的 $\sqrt{n}-$ 正态性。Newey (1994a), Pakes and Olley (1995), 和 Chen et al. (2003) 研究了未知参数 $\beta$ 的半参数两步法 GMM 估计量的一些特性。除了提供了计算两步法第二步未知参数 $\beta$ 的估计量的渐近方差的方法外，Newey (1994a) 证明了第二步中对 $\beta$ 的估计及其渐近方差并不取决于第一步非参估计所选用的具体方法，而只取决于第一步估计的收敛速度。

### 4.1.1 渐近正态性

接下来我们陈述两个主要结果，这两个结果都是从 Chen et al. (2003) 中的结论略作修改得到的。在 Chen et al. (2003) 文章中，经验目标函数可以是 $\beta$ 和 $h$ 上的非平滑函数。用 $M : B \times \mathcal{H} \mapsto \mathcal{R}^{d_m}$ 表示一个非随机，值域为向量空间的可测函数，其中 $B$ 是 $\mathcal{R}^{d_\beta}$ 的紧凑子集，且 $d_m \geq d_\beta$。识别假设是在 $\beta = \beta_o \in B$ 处有 $M(\beta, h_o(\cdot, \beta)) = 0$ 以及对所有 $\beta \neq \beta_o$ 有 $M(\beta, h_o(\cdot, \beta)) \neq 0$。我们用 $\beta_o \in B$ 和 $h_o \in \mathcal{H}$ 表示真实未知有限维和无限维参数，其中函数 $h_o \in \mathcal{H}$ 可以取决于（未知）参数 $\beta$ 和观测到的数据 $Z$。为了表达上的便利，我们通常略去 $h_o$ 的自变量；因此：$(\beta, h) \equiv (\beta, h(\cdot, \beta))$, $(\beta, h_o) \equiv (\beta, h_o(\cdot, \beta))$ 与 $(\beta_o, h_o) \equiv (\beta_o, h_o(\cdot, \beta_o))$。我们假设 $\mathcal{H}$ 是赋有伪-度量 $||\cdot||_{\mathcal{H}}$ 的函数向量空间；该伪-度量是关于 $\beta$ 的上确界范数度量和关于所有其他自变量的某种伪度量。假设存在一个随机的向量值域函数 $M_n : B \times \mathcal{H} \to \mathcal{R}^{d_m}$ 取决于观测到的数据 $\{Z_i : i = 1, \ldots, n\}$, 使得 $M_n(\beta, h_o)' W M_n(\beta, h_o)$ 同 $M(\beta, h_o)' W M(\beta, h_o)$ 很接近，其中 $W$ 是某对称正定矩阵（一般称作加权矩阵）。假设在每个 $\beta$ 下都有一个真实未知参数 $h_o(.)$ 的初始非参数估计量 $\widehat{h}(.)$ 。用 $W_n$ 表示满足 $W_n - W = o_P(1)$ 的随机（或确定的）加权矩阵。那么我们可以用解决如下样本最小距离问题的 $\widehat{\beta}$ 来估计 $\beta_o$：[37]

$$\min_{\beta \in B} M_n(\beta, \widehat{h})' W_n M_n(\beta, \widehat{h}) \tag{4.1}$$

对任意 $\beta \in B$，如果 $\{h + \tau(\overline{h} - h) : \tau \in [0, 1]\} \subset \mathcal{H}$ 和 $\lim_{\tau \to 0}[M(\beta, h + \tau(\overline{h} - h)) - M(\beta, h)]/\tau$ 存在，我们称 $M(\beta, h)$ 是 $h$ 的在 $[\overline{h} - h]$ 方向上的路径可导函数；我们用 $\Gamma_2(\beta, h)[\overline{h} - h]$ 表示这个极限。

**定理 4.1.** *假设 $\beta_o \in int(B)$ 满足 $M(\beta_o, h_o) = 0$, $\widehat{\beta} - \beta_o = o_P(1)$, $W_n - W = o_P(1)$, 以及：*

[37] 关于估计量 $\widehat{\beta} - \beta_o = o_P(1)$ 的一致性特征，可以参考 Chen et al. (2003) 的定理 1。

*(4.1.1)* 对于某正序列 $\delta_n = o(1)$，有 $\|M_n(\widehat{\beta},\widehat{h})\| = \inf_{||\beta-\beta_o||\le\delta_n} \|M_n(\beta,\widehat{h})\| + o_P(1/\sqrt{n})$。

*(4.1.2) (i)* $M(\beta,h_o)$ 在 $\beta$ 上的常偏导数 $\Gamma_1(\beta,h_o)$ 在 $\beta_o$ 的某个领域中存在并在 $\beta=\beta_o$ 处连续；*(ii)* 矩阵 $\Gamma_1 \equiv \Gamma_1(\beta_o,h_o)$ 满足 $\Gamma_1'W\Gamma_1$ 是非奇异矩阵。

*(4.1.3)* $M(\beta,h_o)$ 在所有方向（表示为 $[h-h_o]$ 上的路径导数 $\Gamma_2(\beta,h_o)[h-h_o]$ 存在并对满足 $||\beta-\beta_o|| = o(1)$ 的所有 $\beta$ 以及满足 $||h-h_o||_{\mathcal{H}} = o(1)$ 的所有 $h$ 有：

$$||\Gamma_2(\beta,h_o)[h-h_o] - \Gamma_2(\beta_o,h_o)[h-h_o]|| \le ||\beta-\beta_o|| \times o(1)$$

（注：这里只要 *4.1.4* 或 *4.1.4'* 之中有一个成立即可。）*(4.1.4)* 对所有满足 $||\beta-\beta_o|| = o(1)$ 的 $\beta$ 有 $||M(\beta,\widehat{h}) - M(\beta,h_o) - \Gamma_2(\beta,h_o)[\widehat{h}-h_o]|| = o_P(n^{-1/2})$ 。

*(4.1.4)' (i)* 存在某非负常数 $c \ge 0, \epsilon \in (0,1]$ 使得对所有满足 $||\beta-\beta_o|| = o(1)$ 的 $\beta$ 和所有满足 $||h-h_o||_{\mathcal{H}} = o(1)$ 的 $h$ 有

$$||M(\beta,h) - M(\beta,h_o) - \Gamma_2(\beta,h_o)[h-h_o]|| \le c||h-h_o||_{\mathcal{H}}^{1+\epsilon}$$

以及 *(ii)* $c||\widehat{h}-h_o||_{\mathcal{H}}^{1+\epsilon} = o_P(n^{-1/2})$。

*(4.1.5)* 对所有满足 $\delta_n = o(1)$ 的正数序列 $\{\delta_n\}$ 有

$$\sup_{\|\beta-\beta_o\|<\delta_n, \|h-h_o\|_{\mathcal{H}}<\delta_n} \frac{\|M_n(\beta,h) - M(\beta,h) - M_n(\beta_o,h_o)\|}{n^{-1/2} + ||M_n(\beta,h)|| + ||M(\beta,h)||} = o_P(1)$$

*(4.1.6)* 对某个有限矩阵 $V_1$，有 $\sqrt{n}\{M_n(\beta_o,h_o) + \Gamma_2(\beta_o,h_o)[\widehat{h}-h_o]\} \xrightarrow{d} \mathcal{N}[0,V_1]$。

那么 $\sqrt{n}(\widehat{\beta}-\beta_o) \xrightarrow{d} \mathcal{N}[0,(\Gamma_1'W\Gamma_1)^{-1}\Gamma_1'WV_1W\Gamma_1(\Gamma_1'W\Gamma_1)^{-1}]$。

**备注 4.1.** 可以沿用 Chen et al. (2003) 定理 2 的证明来得到上述定理 4.1。注意条件 (4.1.4) 可以由 (4.1.4)' 推出；同时当 $\epsilon = 1$ 时条件 (4.1.4)' 就等同于 Newey (1994a) 和 Chen et al. (2003) 文章中要求的条件。当 $M(\beta,h)$ 是 $h$ 的高度非线性函数并且/或者当 $h(\cdot,\beta)$ 的自变量"·"具有无界支集时，那么条件 (4.1.4)'(i) 在 $\epsilon = 1$ 下可能不再成立，但条件 (4.1.4)' 在 $0 < \epsilon < 1$ 下通常仍然成立。适用于非经典误差模型和丢失数据问题的两步法 GMM 估计中有关上述讨论的例子可以参考 Chen et al. (2004b)。根据条件 (4.1.4)'(ii) $||\widehat{h}-h_o||_{\mathcal{H}} = o_P(n^{-1/2(1+\epsilon)})$，$\epsilon$ 越小，$\widehat{h}$ 收敛到 $h_o$ 的速度就越需要更快。在极端情况 $||\widehat{h}-h_o||_{\mathcal{H}} = O_P(n^{-1/2})$ 下（这时 $h_o$ 是概率分布函数），条件 4.1.4 可以由以下条件推出：*(4.1.4)" (i)* 对所有满足 $||\beta-\beta_o|| = o(1)$ 的 $\beta$ 和所有满足 $||h-h_o||_{\mathcal{H}} = o(1)$ 的 $h$，有 $||M(\beta,h) - M(\beta,h_o) - \Gamma_2(\beta,h_o)[h-h_o]|| = ||h-h_o||_{\mathcal{H}} \times o(1)$；(ii) $||\widehat{h}-h_o||_{\mathcal{H}} = O_P(n^{-1/2})$。

许多计量经济学模型对应 $M(\beta,h) = E[m(Z_i,\beta,h)]$, $M_n(\beta,h) = n^{-1}\sum_{i=1}^n m(Z_i,\beta,h)$, 其中 $m : \mathcal{R}^{d_z} \times B \times \mathcal{H} \to \mathcal{R}^{d_m}$ 是一个可测的向量值函数并满足 $E[m(Z_i,\beta,h_o(\cdot,\beta))] = 0$ 当且仅当 $\beta = \beta_o \in B$，后者为 $\mathcal{R}^{d_\beta}$ 的子集。这种情况下，Chen et al. (2003) 中的定理 3 为 i.i.d. 数据 $\{Z_i\}$ 下的随机等度连续条件 (4.1.5) 提供了一系列很容易验证的充分条件。下面的引理将他们的结论拓展到严平稳过程上。令 $\mathcal{F} = \{m(z,\beta,h) : \beta \in B, h \in \mathcal{H}\}$ 表示以 $(\beta,h)$ 索引的可测函数类；用 $H_{[]}(w,\mathcal{F},||\cdot||_r)$ 表示 $\mathcal{F}$ 的 $L_r(P_o)$-度量带括熵。

**引理 4.1.** 假设 $\{Z_t : t \geq 1\}$ 是严平稳过程，$M(\beta,h) = E[m(Z_t,\beta,h)]$ 和 $M_n(\beta,h) = n^{-1}\sum_{i=1}^n m(Z_t,\beta,h)$；$m(Z_t,\beta,h(\cdot))$ 中 $h(\cdot)$ 的自变量只取决于 $\beta$ 和有限多个 $Z_t$。假设 $m = (m_1,\ldots,m_{d_m})'$ 中的每个成分 $m_j$ 都满足：

*(4.2.1)* $m_j(\cdot,\beta,h)$ 在 $\beta,h$ 上局部一致 $L_r(P_o)$-连续。具体而言，对所有 $(\beta,h) \in B \times \mathcal{H}$，所有较小的正值 $\delta = o(1)$，以及某些常量 $s_j \in (0,1]$，$K_j > 0$ 和 $r \geq 1$ 下式成立：

$$\left( E\left[ \sup_{(\beta',h'):||\beta'-\beta||<\delta, ||h'-h||_{\mathcal{H}}<\delta} |m_{lcj}(Z,\beta',h') - m_{lcj}(Z,\beta,h)|^r \right] \right)^{1/r} \leq K_j \delta^{s_j}$$

那么：*(i)* 对 $j = 1,...,d_m$ 有 $H_{[]}(w,\mathcal{F}_j,||\cdot||_r) \leq \log N([\frac{\varepsilon}{2K_j}]^{1/s_j}, B, ||\cdot||) + \log N([\frac{\varepsilon}{2K_j}]^{1/s_j}, \mathcal{H}, ||\cdot||_{\mathcal{H}})$。

更进一步地，假设

*(4.2.2)* $B$ 是 $\mathcal{R}^{d_\beta}$ 的一个紧凑子集。对 $j = 1,\ldots,d_m$ 有 $\int_0^\infty \sqrt{\log N(\varepsilon^{1/s_j},\mathcal{H},||\cdot||_{\mathcal{H}})}d\varepsilon < \infty$。

*(4.2.3)* 或者 $\{Z_t\}_{t=1}^n$ 是 *i.i.d.* 数据同时 *(4.2.1)* 在 $r \geq 2$ 下成立，或者 $\{Z_t\}_{t=1}^n$ 是具有混合衰减速率满足对某 $r > 2$ 有 $\sum_{t=1}^\infty t^{2/(r-2)}\beta_t < \infty$ 的 *beta*-混合过程，并且条件 *(4.2.1)* 在 $r > 2$ 下成立。

那么：*(ii)* 对所有满足 $\delta_n = o(1)$ 的正数 $\delta_n$ 有

$$\sup_{\|\beta-\beta_o\|<\delta_n, \|h-h_o\|_{\mathcal{H}}<\delta_n} \|M_n(\beta,h) - M(\beta,h) - \{M_n(\beta_o,h_o) - M(\beta_o,h_o)\}\| = o_P(n^{-1/2}) \tag{4.2}$$

证明. 我们已经由 Chen et al. (2003) 的定理 3 得到结果 (i)。在 i.i.d. 数据下的结果 (ii) 也可以由 Chen et al. (2003) 的定理 3 直接得到。对于平稳 bate-混合过程，条件 (4.2.1) - (4.2.3) 意味着 $\int_0^\infty \sqrt{H_{[]}(w,\mathcal{F},||\cdot||_r)}dw < \infty$，其中 $r > 2$。这和 $\sum_{t=1}^\infty t^{2/(r-2)}\beta_t < \infty$ 一起满足了 Doukhan et al. (1995) 论文中有关平稳 beta-混合过程的 Donsker 定理所要求的全部假设。这也就证明了随机等度连续 (4.2) 中的结果 (ii)。 □

Chen et al. (2003) 中的定理 3 和引理 4.2 都是将 Andrews (1994a) 文中定义的“第二类”和“工具变量类”的相关结论从 $\beta \in B$ 拓展到 $(\beta,h) \in B \times \mathcal{H}$ 上。条件 (4.2.1) 允许 $(\beta,h)$ 上的非连续矩函数，例如 $(\beta,h)$ 的符号函数和指示函数。

给定 Newey (1994a), Chen et al. (2003) 中的结论和定理 4.1，上述两步法里第一步估计 $h$ 时具体选择什么方法应当主要取决于该方法实施的难易程度。最近，Lee (2003) 提出了一个适用于部分线性分位数回归 $Y_t = X_{0t}'\beta_o + h_o(X_{1t}) + e_t$, $P[e_t \leq 0|X_t] = \alpha \in (0,1)$ 的两步 $\sqrt{n}$ 渐近正态的和有效的 $\beta$ 估计量，其中第一步是 $Y_t$ 在 $X = (X_0',X_1')'$ 上的高维度核分位数回归。Chen et al. (2003) 考虑将 Lee 的模型变为具有内生自变量的部分线性分位回归，研究了相关估计量的具体性质。他们使用两步 GMM 方法提出另一个 $\beta$ 的 $\sqrt{n}$-渐近正态估计量，其中第一步非参数估计只考虑 $h(X_{1t})$。我们可以将他们的模型拓展为一个部分可加分位数回归模型：

$$Y_t = X_{0t}'\beta_o + \sum_{j=1}^q h_{oj}(X_{jt}) + e_t, \quad P[e_t \leq 0|X_t] = \alpha \in (0,1)。$$

如果已知 $h_{o1},...,h_{oq}$，那么可以基于矩条件 $E[m(Z_i,\beta,h_o)] = 0$ 当且仅当 $\beta = \beta_o$，$m(Z_i,\beta,h_o) = X_0\{\alpha - 1(Y \leq X_{0t}'\beta + \sum_{j=1}^q h_{oj}(X_{jt})\})$ 来估计 $\beta_o$。很明显，为了应用半参数两步 GMM 方法（即最大化样本矩

$n^{-1}\sum_{i=1}^{n} m(Z_i,\beta,\widehat{h}))$ 来估计 $\beta$，如果 $\widehat{h}=(\widehat{h}_1,...,\widehat{h}_q)$ 是一个筛估计量，那么这一方法将会大大简化。例如 $\widehat{h}$ 可以是在固定 $\beta$ 下求解 $\max_{h\in\mathcal{H}_n}\widehat{Q}_n(\beta,h)=n^{-1}\sum_{t=1}^{n} l(\beta,h,Y_t,X_t)$ 得到的，其中

$$l(\beta,h,Y_t,X_t)=\{1(Y_t<X_{0t}'\beta+\sum_{j=1}^{q}h_j(X_{jt}))-\alpha\}[Y_t-X_{0t}'\beta-\sum_{j=1}^{q}h_j(X_{jt})],$$

上式中 $\mathcal{H}_n=\mathcal{H}_n^1\times\cdots\times\mathcal{H}_n^q$ 同在 3.2.1 节中的定义一致，而并没有采用高维核分位数回归方法。Andrews (1994b), Newey (1994a, b), Newey et al. (1999) 和 Das et al. (2003) 在具有可加 LS 回归的两步法估计中也为第一步估计提出了同样的建议。

现在也已经有大量的关于通过各种两步法来实现 $\beta$ 的有效估计的一般理论文献。例如，$\beta$ 的代换 MLE 估计（可以看作 Andrews (1994b) MINPIN 方法的重要子类）可以达到估计效率的上界；具体可以参考例如 Severini and Wong (1992), Ai (1997) 以及 Murphy and van der Vaart (2000)。其他可以实现同等效率的两步估计方法包括基于有效 score 方程的估计法，参考 Bickel et al. (1993) 和 Newey (1990a)；以及最优加权 GMM 方法，参考 Newey (1990a, b, 1993)。其他例子可以参考 Powell (1994) 以及 Pagan and Ullah (1999)。很显然，我们可以将这些方法同第一步中未知函数 $h$ 的基于筛分方法的非参估计结合起来。

## 4.2 筛联立 M 估计

现在关于 $\beta$ 和 $h$ 的筛联立 M 估计的一般理论文章还相对较少；参考 Wong and Severini (1991), Shen (1997), Chen and Shen (1998)。这个方法通过在筛参数空间 $\Theta_n=B\times\mathcal{H}_n$ 上最大化样本目标函数 $\widehat{Q}_n(\beta,h)$ 来同时估计 $\beta$ 和 $h$；这里 $\widehat{Q}_n(\beta,h)$ 是样本平均 $\frac{1}{n}\sum_{i=1}^{n} l(\beta,h,Z_i)$。Wong and Severini (1991) 证明了当参数空间为 $\Theta_n\equiv\Theta=B\times\mathcal{H}$ 时非参数 MLE 估计量的平滑泛函的 $\sqrt{n}-$ 渐近正态性和有效性。Shen (1997) 将他们的结果推广到筛 MLE，并且考虑了高度弯曲的（非线性）最不利方向这种情形。Chen and Shen (1998) 将 Shen (1997) 的结果拓展到平稳弱相关数据的一般筛 M-回归上。

### 4.2.1 筛 M-估计量的平滑泛函的渐近正态性

用 $\widehat{\theta}_n=(\widehat{\beta}_n,\widehat{h}_n)=\arg\max_{(\beta,h)\in B\times\mathcal{H}_n}\frac{1}{n}\sum_{i=1}^{n} l(\beta,h,Z_i)$ 表示 $\theta_o=(\beta_o,h_o)$ 的筛 M-估计。在本小节中我们给出一个简单的适用于 $\theta_o$ 的平滑泛函的代入法估计量的 $\sqrt{n}-$ 渐近正态性理论。更具有一般性的版本可以参考 Shen (1997) 和 Chen and Shen (1998)。

假设 $\Theta=B\times\mathcal{H}$ 在 $\theta_o$ 上为凸，使得对所有的小 $\tau\in[0,1]$ 和所有的固定 $\theta\in\Theta$ 有 $\theta_o+\tau[\theta-\theta_o]\in\Theta$。假设方向导数

$$\frac{\partial l(\theta_o,z)}{\partial\theta}[\theta-\theta_o]\equiv\lim_{\tau\to0}\frac{l(\theta_o+\tau[\theta-\theta_o],z)-l(\theta_o,z)}{\tau}$$

对几乎所有在支集 $Z$ 中的 $z$ 都是适定的。

赋予 $\Theta=B\times\mathcal{H}$ 范数 $\|\cdot\|$。假设我们感兴趣的泛函，$f:\Theta\to\mathcal{R}$，是平滑的（定义如下）：

$$\frac{\partial f(\theta_o)}{\partial\theta}[\theta-\theta_o]\equiv\lim_{\tau\to0}\frac{f(\theta_o+\tau[\theta-\theta_o])-f(\theta_o)}{\tau}$$

是适定的，同时

$$\|\frac{\partial f(\theta_o)}{\partial \theta}\| \equiv \sup_{\{\theta\in\Theta:\|\theta-\theta_o\|>0\}} \frac{|\frac{\partial f(\theta_o)}{\partial \theta}[\theta-\theta_o]|}{\|\theta-\theta_o\|} < \infty$$

下面，我们假设范数 $\|\cdot\|$ 导出一个定义在由 $\Theta-\theta_o$ 生成的完备化空间（用 $\bar{V}$ 表示）上的内积 $\langle\cdot,\cdot\rangle$。根据 Riesz 表示定理 (Riesz representation theorem)，存在 $v^*\in\bar{V}$ 使得对于任何 $\theta\in\Theta$ 都有

$$\frac{\partial f(\theta_o)}{\partial \theta}[\theta-\theta_o] = \langle\theta-\theta_o, v^*\rangle \quad \text{当且仅当} \quad \|\frac{\partial f(\theta_o)}{\partial \theta}\| < \infty$$

假设筛 M-估计 $\hat{\theta}_n$ 以快于 $\delta_n$ 的速度收敛到 $\theta_o$ (即 $\|\hat{\theta}_n-\theta_o\| = o_P(\delta_n)$)。用 $\varepsilon_n$ 表示满足 $\varepsilon_n = o(n^{-1/2})$ 的任意序列，用 $\mu_n(g(Z)) = \frac{1}{n}\Sigma_{t=1}^n\{g(Z_t)-E(g(Z_t))\}$ 表示以函数 $g$ 索引的经验过程。之前我们定义过 $K(\theta_o,\theta)\equiv E[l(\theta_o,Z_i)-l(\theta,Z_i)]$。

**条件 4.1.** *(i) 存在一个正数 $\omega>0$ 使得对满足 $\|\theta-\theta_o\|=o(1)$ 的所有 $\theta\in\Theta_n$ 都有一致地 $|f(\theta)-f(\theta_o)-\frac{\partial f(\theta_o)}{\partial\theta}[\theta-\theta_o]| = O(\|\theta-\theta_o\|^\omega)$; (ii) $\|\frac{\partial f(\theta_o)}{\partial\theta}\|<\infty$; (iii) 存在一个 $\pi_n v^*\in\Theta_n$ 使得 $||\pi_n v^*-v^*||\times\|\hat{\theta}_n-\theta_o\| = o_P(n^{-1/2})$。*

**条件 4.2.** $\sup_{\{\theta\in\Theta_n:\|\theta-\theta_o\|\le\delta_n\}}\mu_n(l(\theta,Z)-l(\theta\pm\varepsilon_n\pi_n v^*,Z)-\frac{\partial l(\theta_o,Z)}{\partial\theta}[\pm\varepsilon_n\pi_n v^*]) = O_P(\varepsilon_n^2)$。

**条件 4.3.** $K(\theta_o,\hat{\theta}_n)-K(\theta_o,\hat{\theta}_n\pm\varepsilon_n\pi_n v^*) = \pm\varepsilon_n\langle\hat{\theta}_n-\theta_o,\pi_n v^*\rangle + o(n^{-1})$。

**条件 4.4.** *(i)* $\mu_n(\frac{\partial l(\theta_o,Z)}{\partial\theta}[\pi_n v^*-v^*]) = o_P(n^{-1/2})$; *(ii)* $E\{\frac{\partial l(\theta_o,Z)}{\partial\theta}[\pi_n v^*]\} = o(n^{-1/2})$。

**条件 4.5.** $n^{1/2}\mu_n(\frac{\partial l(\theta_o,Z)}{\partial\theta}[v^*]) \xrightarrow{d} \mathcal{N}(0,\sigma_{v^*}^2)$, 其中 $\sigma_{v^*}^2>0$。

我们注意到对经典非线性 M-估计（例如在 Newey and McFadden (1994) 中总结过的情形）而言，仍然需要条件 4.1(i)(ii), 4.2, 4.3 和 4.5 (虽然使用略微不同的表达式)，然而由于对标准非线性 M-估计有 $\pi_n v^*=v^*$，所以条件 4.1(iii) 和 4.4自动得到满足。注意对 i.i.d. 数据，只要 $\sigma_{v^*}^2 = Var\left(\frac{\partial l(\theta_o,Z)}{\partial\theta}[v^*]\right)>0$，条件 4.5 就成立。如果 $l(\theta,Z)$ 也同时在 $\theta\in\Theta_n$,$\|\theta-\theta_o\|=o(1)$, 上依路径可导，那么条件 4.2和 4.3可以分别从条件 4.2' 和 4.3' 推出，其中

**条件 4.2'.** $\sup_{\{\bar{\theta}\in\Theta_n:\|\bar{\theta}-\theta_o\|\le\delta_n\}}\mu_n\left(\frac{\partial l(\bar{\theta},Z)}{\partial\theta}[\pi_n v^*]-\frac{\partial l(\theta_o,Z)}{\partial\theta}[\pi_n v^*]\right) = o_P(n^{-1/2})$。

**条件 4.3'.** $E\{\frac{\partial l(\hat{\theta}_n,Z)}{\partial\theta}[\pi_n v^*]\} = \langle\hat{\theta}_n-\theta_o,\pi_n v^*\rangle + o(n^{-1/2})$。

条件 4.2 (或 4.2') 可以通过应用引理 4.2 来验证。条件 4.3 (或 4.3') 可以在研究者选择 Hilbert 范数 $\|\theta-\theta_o\|$ 时得到验证。

当参数空间 $\Theta$ 不是凸集时，可能需要对条件 4.2, 4.3 和 4.4 做出一些调整；这方面的具体讨论可以参考 Shen (1997) 和 Chen and Shen (1998)。

**定理 4.2.** *假设条件 4.1–4.5 成立，同时 $\|\hat{\theta}_n-\theta_o\|^\omega = o_P(n^{-1/2})$。那么对筛 M-估计 $\hat{\theta}_n$, 有 $n^{1/2}(f(\hat{\theta}_n)-f(\theta_o)) \xrightarrow{d}\mathcal{N}(0,\sigma_{v^*}^2)$。*

定理 4.2 的证明可以很容易地从 Shen (1997) 和 Ai and Chen (1999) 的证明得到。在具体应用中，我们需要指定一个 Hilbert 范数 $\|\theta-\theta_o\|$ 来计算表达式 $v^*$。Wong and Severini (1991) 和 Shen (1997) 针对筛 MLE 方法采用了 Fisher 范数 $\|\theta-\theta_o\|^2 = E\{\frac{\partial l(\theta_o,Z_i)}{\partial\theta}[\theta-\theta_o]\}^2$。Ai and Chen (1999, 2003) 为他们的筛 MD 和筛广义最小二乘（Sieve GLS）估计引入了一个类-Fisher 范数。在下一个小节中我们详细说明如何在筛 GLS 问题中应用定理 4.2来得到有限维参数部分的 $\sqrt{n}$-渐近正态性。

### 4.2.2 筛 GLS 估计的渐近正态性

我们介绍过对所有属于第一子类的条件矩约束（$E\{\rho(Z,\theta_o)|X\}=0$，其中 $\rho(Z,\theta)-\rho(Z,\theta_o)$ 并不取决于内生变量 $Y$）模型 (2.8)，可以使用筛 GLS 方法来估计未知参数 $\theta_o=(\beta_o,h_o)\in B\times\mathcal{H}$：

$$\widehat{\theta}_n=(\widehat{\beta}_n,\widehat{h}_n)=\arg\min_{(\beta,h)\in B\times\mathcal{H}_n}\frac{1}{n}\sum_{i=1}^n\rho(Z_i,\beta,h)'\Sigma(X_i)^{-1}\rho(Z_i,\beta,h),$$

上式中 $\Sigma(X_i)$ 是一个正定加权矩阵。当 $\Sigma(X_i)$ 已知（比如单位矩阵）时，这类模型属于筛 M-估计，其中 $l(\theta,Z_i)=-\rho(Z_i,\theta)'\Sigma(X_i)^{-1}\rho(Z_i,\theta)/2$。关于最优加权矩阵 $\Sigma_o(X_i)\equiv Var\{\rho(Z_i,\theta_o)|X_i\}$ 的估计可以参考 4.3 节和备注 4.3。

我们现在应用定理 4.2来得到筛 GLS 估计量 $\widehat{\beta}_n$ 的 $\sqrt{n}$ 渐近正态性。定义范数 $\|\theta-\theta_o\|^2=E\{(\frac{\partial\rho(Z_i,\theta_o)}{\partial\theta}[\theta-\theta_o])\Sigma(X_i)^{-1}(\frac{\partial\rho(Z_i,\theta_o)}{\partial\theta}[\theta-\theta_o])\}$。对 $j=1,...,d_\beta$，令

$$D_{w_j}(X)=\frac{\partial\rho(Z,\beta,h_o(\cdot))}{\partial\beta_j}|_{\beta=\beta_o}-\frac{\partial\rho(X,\beta_o,h_o(\cdot)+\tau w_j(\cdot))}{\partial\tau}|_{\tau=0}=\frac{\partial\rho(Z,\theta_o)}{\partial\beta_j}-\frac{\partial\rho(Z,\theta_o)}{\partial h}[w_j],$$

$w=(w_1,...,w_{d_\beta})$，$D_w(X)=\left(D_{w_1}(X),...,D_{w_{d_\beta}}(X)\right)=\frac{\partial\rho(Z,\theta_o)}{\partial\beta'}-\frac{\partial\rho(Z,\theta_o)}{\partial h}[w]$ 是一个 $X$ 的 $d_\rho\times d_\beta-$ 矩阵值可测函数。令 $w^*=(w_1^*,...,w_{d_\beta}^*)$, 其中对于 $j=1,...,d_\beta$, $w_j^*$ 是下列问题的解：

$$E\{D_{w_j^*}(X)'\Sigma(X)^{-1}D_{w_j^*}(X)\}=\inf_{w_j}E\{D_{w_j}(X)'\Sigma(X)^{-1}D_{w_j}(X)\}$$

令 $D_{w^*}(X)=\frac{\partial\rho(Z,\theta_o)}{\partial\beta'}-\frac{\partial\rho(Z,\theta_o)}{\partial h}[w^*]$。令 $v_\beta^*=\left(E\{D_{w^*}(X)'\Sigma(X)^{-1}D_{w^*}(X)\}\right)^{-1}\lambda$, $v_h^*=-w^*v_\beta^*$ 和 $v^*=(v_\beta^*,v_h^*)$。

**假设 4.1.** *(i)* $\beta_o\in int(B)$; *(ii)* $E[D_{w^*}(X)'\Sigma(X)^{-1}D_{w^*}(X)]$ 是正定矩阵；*(iii)* 存在一个 $\pi_nv^*\in\Theta_n$ 满足 $||\pi_nv^*-v^*||\times\|\hat{\theta}_n-\theta_o\|=o_P(n^{-1/2})$。

**假设 4.2.** *(i)* $\Sigma(X)$ 和 $\Sigma_o(X)\equiv Var\{\rho(Z_i,\theta_o)|X\}$ 都一致地在 $X$ 上正定有界；*(ii)* $\rho(Z,\theta)$ 是 $\theta\in\Theta(\|\theta-\theta_o\|=o(1))$ 的两次连续依路径可导函数；*(iii)* 条件 *4.2'* 和 *4.3'* 在 $\frac{\partial l(\overline{\theta},Z)}{\partial\theta}[\pi_nv^*]=-\rho(Z,\overline{\theta})'\Sigma(X)^{-1}\{\frac{\partial\rho(Z,\overline{\theta})}{\partial\theta}[\pi_nv^*]\}$ 下对所有的 $\overline{\theta}\in\Theta_n$ $(\|\overline{\theta}-\theta_o\|=o(1))$ 成立；*(iv)* $\{Z_i\}_{i=1}^n$ 是 *i.i.d.* 数据，$E\{\rho(Z,\theta_o)|X\}=0$, 以及对所有 $\theta\in\Theta$ 有 $E\{\rho(Z,\theta)-\rho(Z,\theta_o)|X\}=\rho(Z,\theta)-\rho(Z,\theta_o)$。

**命题 4.1.** 令 $\hat{\theta}_n$ 表示筛 *GLS* 估计。如果假设 *4.1 - 4.2* 成立，那么 $n^{1/2}(\widehat{\beta}_n-\beta_o)\xrightarrow{d}\mathcal{N}(0,V_1^{-1}V_2V_1^{-1})$ 其中

$$V_1=E[D_{w^*}(X)'\Sigma(X)^{-1}D_{w^*}(X)],\ V_2=E[D_{w^*}(X)'\Sigma(X)^{-1}\Sigma_o(X)\Sigma(X)^{-1}D_{w^*}(X)].$$

证明. 令 $f(\theta)=\lambda'\beta$, 这里 $\lambda$ 是 $\mathcal{R}^{d_\beta}$ 中的任意单位向量。很明显，条件 4.1(i) 在 $\frac{\partial f(\theta_o)}{\partial\theta}[\theta-\theta_o]=(\beta-\beta_o)'\lambda$ 和 $\omega=\infty$ 下成立。除此之外，在假设 4.1(i)(ii) 之下，我们有 $v^*=(v_\beta^*,v_h^*)$ 和

$$\|v^*\|^2=\sup_{\{\theta\in\Theta:\|\theta-\theta_o\|>0\}}\frac{\{(\beta-\beta_o)'\lambda\}^2}{\|\theta-\theta_o\|^2}=\lambda'\left(E\{D_{w^*}(X)'\Sigma(X)^{-1}D_{w^*}(X)\}\right)^{-1}\lambda<\infty,$$

因此条件 4.1可以由假设 4.1 得到。注意到

$$\frac{\partial l(\theta_o,Z)}{\partial\theta}[\theta-\theta_o]=-\rho(Z,\theta_o)'\Sigma(X)^{-1}\left(\frac{\partial\rho(Z,\theta_o)}{\partial\beta'}(\beta-\beta_o)+\frac{\partial\rho(Z,\theta_o)}{\partial h}[h-h_o]\right),$$

我们有

$$E\{\frac{\partial l(\theta_o,Z)}{\partial\theta}[\pi_n v^*]\}=-E\left\{\rho(Z,\theta_o)'\Sigma(X)^{-1}\left(\frac{\partial\rho(Z,\theta_o)}{\partial\beta'}(v_\beta^*)+\frac{\partial\rho(Z,\theta_o)}{\partial h}[\pi_n v_h^*]\right)\right\}=0,$$

因此条件 4.4(ii) 自动成立。由于

$$\frac{1}{n}\sum_{t=1}^n\frac{\partial l(\theta_o,Z_t)}{\partial\theta}[\pi_n v^*-v^*]=\frac{-1}{n}\sum_{t=1}^n\rho(Z_t,\theta_o)'\Sigma(X_t)^{-1}\left(\frac{\partial\rho(Z_t,\theta_o)}{\partial h}[\pi_n v_h^*-v_h^*]\right),$$

根据 Chebyshev 不等式和假设 4.1(iii) 和 4.2(i)，我们得到

$$\frac{1}{n}\sum_{i=1}^n\frac{\partial l(\theta_o,Z_i)}{\partial\theta}[\pi_n v^*-v^*]=o_P(n^{-1/2}),$$

因此条件 4.4(i) 被满足。由于数据是 i.i.d. 和在假设 4.1(ii) 和 4.2(i) 下

$$\begin{aligned}\sigma_{v^*}^2 &= Var\left\{\frac{\partial l(\theta_o,Z)}{\partial\theta}[v^*]\right\}=Var\left\{\rho(Z,\theta_o)'\Sigma(X)^{-1}\left(\frac{\partial\rho(Z,\theta_o)}{\partial\beta'}-\frac{\partial\rho(Z,\theta_o)}{\partial h}[w^*]\right)(v_\beta^*)\right\}\\ &= (v_\beta^*)'E\left\{D_{w^*}(X)'\Sigma(X)^{-1}\Sigma_o(X)\Sigma(X)^{-1}D_{w^*}(X)\right\}(v_\beta^*)=\lambda'V_1^{-1}V_2V_1^{-1}\lambda>0,\end{aligned}$$

满足条件 4.5。根据定理 4.2, 我们得到，对任意单位向量 $\lambda\in\mathcal{R}^{d_\beta}$, 有 $n^{1/2}\lambda'(\widehat{\beta}_n-\beta_o)\xrightarrow{d}\mathcal{N}(0,\sigma_{v^*}^2)$。所以 $\sqrt{n}(\widehat{\beta}_n-\beta_o)\xrightarrow{d}\mathcal{N}(0,V_1^{-1}V_2V_1^{-1})$。 □

**备注 4.2.** 筛 GLS 估计量 $\widehat{\beta}_n$ 的渐近方差 $V_1^{-1}V_2V_1^{-1}$ 可以用 $\widehat{V}_1^{-1}\widehat{V}_2\widehat{V}_1^{-1}$ 一致地估计，其中

$$\widehat{V}_1=\frac{1}{n}\sum_{i=1}^n\left(\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial\beta'}-\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial h}[\widehat{w}]\right)'\Sigma(X_i)^{-1}\left(\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial\beta'}-\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial h}[\widehat{w}]\right),$$

$$\widehat{V}_2=\frac{1}{n}\sum_{i=1}^n\left(\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial\beta'}-\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial h}[\widehat{w}]\right)'\Sigma(X_i)^{-1}\widehat{\Sigma}_o(X_i)\Sigma(X_i)^{-1}\left(\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial\beta'}-\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial h}[\widehat{w}]\right),$$

$\widehat{w}=(\widehat{w}_1,...,\widehat{w}_{d_\beta})$ 解决下列筛最小化问题：对 $j=1,...,d_\beta$,

$$\min_{w_j\in\mathcal{H}_n}\sum_{i=1}^n\left(\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial\beta_j}-\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial h}[w_j]\right)'[\Sigma(X_i)]^{-1}\left(\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial\beta_j}-\frac{\partial\rho(Z_i,\widehat{\theta}_n)}{\partial h}[w_j]\right),$$

同时 $\widehat{\Sigma}_o(X_i)$ 可以是 $\Sigma_o(X_i)$ 的任意一致非参估计量；关于核估计量可以参考 Ai and Chen (1999)，关于 $\Sigma_o(X_i)$ 的级数 LS 估计量可以参考 Ai and Chen (2003, 2004a)。

### 4.2.3 例子：具有单调性约束的部分可加均值回归问题

假设 i.i.d. 数据 $\{Y_t, X_t' = (X_{0t}', X_{1t}, ..., X_{qt})\}_{t=1}^n$ 由以下模型生成：

$$Y_i = X_{0i}'\beta_o + h_{o1}(X_{1i}) + ... + h_{oq}(X_{qi}) + e_i, \quad E[e_i|X_i] = 0$$

令 $\theta_o = (\beta_o', h_{o1}, ..., h_{oq})' \in \Theta = B \times \mathcal{H}$ 表示我们关注的参数，其中 $B$ 是 $\mathcal{R}^{d_\beta}$ 的一个紧凑子集，$\mathcal{H}$ 同 3.2.1小节中的定义一样。由于 $h_{o1}(\cdot)$ 可以有常数项，我们假设 $X_0$ 并不包含常数自变量，$\dim(X_0) = d_\beta$, $\dim(X_j) = 1$, $j = 1, ..., q$, $\dim(X) = d_\beta + q$, $\dim(Y) = 1$。我们通过在 $\Theta_n = B \times \mathcal{H}_n$ 上最大化目标函数 $\widehat{Q}_n(\theta) = n^{-1}\sum_{t=1}^n l(\theta, Y_t, X_t)$ 来估计回归函数 $\theta_o(X) = X_{0t}'\beta_o + \sum_{j=1}^q h_{oj}(X_{jt})$，其中 $l(\theta, Y_t, X_t) = -\frac{1}{2}[Y_t - X_{0t}'\beta - \sum_{j=1}^q h_j(X_{jt})]^2$。令 $\|\theta - \theta_o\|^2 = E\{X_{0t}'(\beta - \beta_o) + \sum_{j=1}^q [h_j(X_{jt}) - h_{oj}(X_{jt})]\}^2$。

注意到 $D_{w^*}(X)' = X_0 - \sum_{k=1}^q w^{*k}(X_k)$，其中 $w^{*k}(X_k)$, $k = 1, ..., q$ 解决

$$\inf_{w^k, k=1,...,q\,:\, E[|X_0 - \sum_{k=1}^q w^k(X_k)|_e^2] > 0} E[(X_0 - \sum_{k=1}^q w^k(X_k))(X_0 - \sum_{k=1}^q w^k(X_k))']$$

**命题 4.2.** *如果假设 3.1 和下列条件成立：(i) $\beta_o \in int(B)$; (ii) $\Sigma_o(X)$ 是正有界矩阵；(iii) $E[X_0X_0']$ 是正定矩阵；$E[D_{w^*}(X)'D_{w^*}(X)]$ 是正定矩阵；(iv) $w^{*j}$ 的每个元素属于 Hölder 空间 $\Lambda^{m_j}$, $m_j > 1/2$, $j = 1, ..., q$。令 $k_{jn} = O(n^{1/(2p_j+1)})$, $j = 1, ..., q$。那么 $n^{1/2}(\widehat{\beta}_n - \beta_o) \xrightarrow{d} \mathcal{N}(0, V_1^{-1}V_2V_1^{-1})$ 其中 $V_1 = E[D_{w^*}(X)'D_{w^*}(X)]$, $V_2 = E[D_{w^*}(X)'\Sigma_o(X)D_{w^*}(X)]$。*

*证明.* 我们用命题 4.4 来证明上述结果。令 $\Theta_n = B \times \mathcal{H}_n$ 和 $\mathcal{H}_n = \mathcal{H}_n^1 \times \cdots \times \mathcal{H}_n^q$，其中 $\mathcal{H}_n^j$, $j = 1, 2, ..., q$, 同 3.2.1 小节中的定义一致。根据命题 3.3 的证明，我们知道只要 $p = \min\{p_1, \cdots, p_q\} > 0.5$，就有 $\|\widehat{\theta}_n - \theta_o\| = O_P(n^{-p/(2p+1)})$。这和假设 (iv) 推出假设 4.1(iii) 成立。根据度量 $\|\cdot\|$ 的定义条件 4.3' 自然地成立。现在只剩下验证条件 4.2'：

$$\mu_n\left(\{X_0'[v_\beta^*] + \sum_{j=1}^q [\pi_n v_{h_j}^*(X_j)]\}\{X_0'[\beta - \beta_o] + \sum_{j=1}^q [h_j(X_j) - h_{oj}(X_j)]\}\right) = o_P(n^{-1/2}),$$

一致地在满足 $\|\theta - \theta_o\| \le \delta_n = O(n^{-p/(2p+1)})$ 的 $\theta \in \Theta_n$ 上。应用 Chen et al. (2003) 中的定理 3 (或适用于 i.i.d. 情形的引理 4.2), 假设 (i)-(iv) 和假设 3.1 ($h_j \in \mathcal{H}^j = \Lambda_c^{m_j}$ , $m_j > 1/2$, 对所有 $j = 1, ..., q$) 可以推出 4.2'；还可以参考 van der Vaart and Wellner (1996)。 □

注意对众所周知的部分线性回归模型 $Y_i = X_{0i}'\beta_o + h_{o1}(X_{1i}) + e_i$, $E[e_i|X_i] = 0$, 我们可以计算得到 $D_{w^*}(X)' \equiv X_0 - w^{*1}(X_1)$ 的显性解，其中 $w^{*1}(X_1) = E\{X_0|X_1\}$。因此如果 $E\{X_0|X_1\}$ 足够平滑，假设 (iv) 将被满足。关于 $\beta_o$ 的半参有效估计，可以参考备注 4.3。

### 4.2.4 筛 MLE 估计的效率

Wong (1992) 和 Wong and Severini (1991) 证明了参数估计平滑泛函的代入非参 MLE 估计的渐近有效性。Shen (1997) 将他们的结果拓展到筛 MLE 中。这里我们总结 Wong (1992) 和 Shen (1997) 的结果。其他相关研究可以参考 Begun et al. (1983), Ibragimov and Has'minskii (1991), Bickel et al. (1993)。

这里目标函数是 $\widehat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^n l(Z_i,\theta)$, 其中 $l(Z_i,\theta) = \log p(Z_i,\theta)$ 是在单点观测 $Z_i$ 处的对数似然函数。我们用 Fisher 范数: $\|\theta-\theta_o\|^2 = E\{\frac{\partial \log p(Z_i,\theta_o)}{\partial\theta}[\theta-\theta_o]\}^2$。之前介绍过如下定义: 如果下列条件 (1)(2) 都满足, 则称一个概率族 $\{P_\theta : \theta\in\Theta\}$ 在 $\theta_o$ 处 "局部渐近正态" (LAN): (1) 对任意在 $\Theta-\theta_o$ 的线性生成空间中的 $g$, 对所有小的 $t\geq 0$ 有 $\theta_o + tn^{-1/2}g \in \Theta$; (2)

$$\frac{dP_{\theta_o+n^{-1/2}g}}{dP_{\theta_o}}(Z_1,\cdots,Z_n) = \exp\left\{\Sigma_n(g) - \frac{1}{2}\|g\|^2 + R_n(\theta_o,g)\right\},$$

其中 $\Sigma_n(g)$ 是 $g$ 的线性函数, $\Sigma_n(g) \xrightarrow{d} \mathcal{N}(0,\|g\|^2)$ 以及 $\mathrm{plim}_{n\to\infty} R_n(\theta_o,g) = 0$ (两个极限都是在真实概率测度 $P_o = P_{\theta_o}$ 下); 参考如 LeCam (1960)。

为了避免 "超-效率"现象, 我们需要为估计量添加一定的正则性条件。在无限维情况下估计某平滑泛函时, Wong (1992, p.58) 定义了一个在 Bahadur (1964) 意义上的依路径正则的估计类别。如果满足下列条件, 我们称 $f(\theta_o)$ 的估计量 $T_n(Z_1,\cdots,Z_n)$ 是依路径正则的: 对任意正实数 $\tau>0$ 和任意在 $\Theta-\theta_o$ 的线性生成空间中的 $g$, 我们有

$$\limsup_{n\to\infty} \mathrm{P}_{\theta_{n,\tau}}(T_n < f(\theta_{n,\tau})) \leq \liminf_{n\to\infty} \mathrm{P}_{\theta_{n,-\tau}}(T_n < f(\theta_{n,-\tau})),$$

其中 $\theta_{n,\tau} = \theta_o + n^{-1/2}\tau g$。

**定理 4.3.** *[Wong (1992), Shen (1997)]* 条件 *LAN* 成立; 假设泛函 $f:\Theta\to\mathcal{R}$ 在 $\theta_o$ 处 *Frechet*-可导并满足 $0 < \|\frac{\partial f(\theta_o)}{\partial\theta}\| < \infty$。则在上述条件下对任意依路径正则的 $f(\theta_o)$ 的估计 $T_n$, 以及对任意正常数 $\tau>0$, 有

$$\limsup_{n\to\infty} \mathrm{P}_o\left(\sqrt{n}|T_n - f(\theta_o)| \leq \tau\right) \leq \mathrm{P}_o\left(\left|\mathcal{N}(0,\|\frac{\partial f(\theta_o)}{\partial\theta}\|^2)\right| \leq \tau\right),$$

其中 $\mathcal{N}(0,\|\frac{\partial f(\theta_o)}{\partial\theta}\|^2)$ 是服从均值为 *0*, 方差为 $\|\frac{\partial f(\theta_o)}{\partial\theta}\|^2$ 的正态分布的标量随机变量。

**定理 4.4.** *[Shen 1997]* 假设保证 $n^{1/2}(f(\hat{\theta}_n) - f(\theta_o)) \xrightarrow{\mathrm{P}_{\theta_o}} \mathcal{N}(0,\sigma^2_{v^*})$, 其中 $\sigma^2_{v^*} = \|\frac{\partial f(\theta_o)}{\partial\theta}\|^2$ 成立的条件满足, 同时如果 *LAN* 成立, 那么对 $f(\theta)$ 的代入法筛 *MLE* 估计, 任意正的常数 $\tau>0$, 以及任意在 $\Theta-\theta_o$ 的线性生成空间中的 $g$ 都有

$$n^{1/2}(f(\hat{\theta}_n) - f(\theta_{n,\tau})) \xrightarrow{\mathrm{P}_{\theta_{n,\tau}}} \mathcal{N}(0,\sigma^2_{v^*}),$$

其中 $\theta_{n,\tau} = \theta_o + n^{-1/2}\tau g$。这里 $\xrightarrow{\mathrm{P}_\theta}$ 表示在概率测度 $P_\theta$ 下依分布收敛。

### 4.3 筛联立 MD 估计: 正态性和有效性

正如我们在 2.1 节中提到过的, 绝大多数结构化计量模型都属于半参条件矩框架: $E[\rho(Z,\beta_o,h_o(\cdot))|X] = 0$, 其中差异 $\rho(Z,\beta,h(\cdot)) - \rho(Z,\beta_o,h_o(\cdot))$ 取决于内生变量 $Y$。关于这类模型的 $\beta_o$ 和 $h_o$ 的筛联立 MD 估计的相关理论, 现有研究还并不太多。这方面讨论可以参考 Newey and Powell (1989, 2003) 和 Ai and Chen (1999, 2003)。筛联立 MD 方法通过在筛参数空间 $\Theta_n = B\times\mathcal{H}_n$ 上求解最小化样本二次型 $\frac{1}{n}\sum_{i=1}^n \widehat{m}(X_i,\beta,h)'[\widehat{\Sigma}(X_i)]^{-1}\widehat{m}(X_i,\beta,h)$ 问题来同时估计 $\beta_o$ 和 $h_o$; 其中 $\widehat{m}(X_i,\beta,h)$ 是函数 $m(X,\beta,h) \equiv E[\rho(Z,\beta,h(\cdot))|X]$

的任意非参估计量，依概率 $\widehat{\Sigma}(X) \to \Sigma(X)$，$\Sigma(X)$ 是一个正定的加权矩阵。Ai and Chen (1999, 2003) 证明了 $\beta_o$ 的筛 MD 估计量 $\widehat{\beta}$ 的 $\sqrt{n}-$ 渐近正态性。

为了实现 $\beta_o$ 的半参数有效估计，Ai and Chen (1999) 首先提出了以下最优加权筛 MD 估计三步法：

**第 1 步**. 通过求解下面的最小化问题获得初始的一致筛 MD 估计量 $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{h}_n)$：

$$\min_{\theta=(\beta,h)\in B\times\mathcal{H}_n} \frac{1}{n}\sum_{i=1}^{n} \widehat{m}(X_i,\theta)'\widehat{m}(X_i,\theta),$$

其中 $\widehat{m}(X_i,\theta)$ 是条件均值函数 $m(X,\theta) \equiv E[\rho(Z,\beta,h(\cdot))|X]$ 的任意非参估计量。

**第 2 步**. 通过任意非参方法（如核、最近邻域或级数 LS 估计）用第一步中得到 $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{h}_n)$ 取得最优加权矩阵 $\Sigma_o(X) \equiv Var[\rho(Z,\beta_o,h_o(\cdot))|X]$ 的一致估计量 $\widehat{\Sigma}_o(X)$。

**第 3 步**. 通过解决下列最小化问题，计算最优加权估计量 $\widetilde{\theta}_n = (\widetilde{\beta}_n, \widetilde{h}_n)$:

$$\min_{\theta=(\beta,h)\in B\times\mathcal{H}_n} \frac{1}{n}\sum_{i=1}^{n} \widehat{m}(X_i,\theta)'[\widehat{\Sigma}_o(X_i)]^{-1}\widehat{m}(X_i,\theta);$$

作为另一种估计 $\beta_o$ 的有效方式，Ai and Chen (2003) 提出了局域连续更新筛 MD 方法：

**第 1 步**. 通过求解下列最小化问题取得初步的一致筛 MD 估计量 $\widehat{\theta}_n$

$$\min_{\theta\in B\times\mathcal{H}_n} \sum_{i=1}^{n} \widehat{m}(X_i,\theta)'\widehat{m}(X_i,\theta)$$

其中 $\widehat{m}(X_i,\theta)$ 是 $m(X,\theta) \equiv E[\rho(Z,\beta,h(\cdot))|X]$ 的级数 LS 估计量 (2.15)。

**第 2 步**. 通过求解下列最小化问题计算最优加权筛 MD 估计量 $\widetilde{\theta}_n = (\widetilde{\beta}_n, \widetilde{h}_n)$：

$$\min_{\theta=(\beta,h)\in N_{on}} \frac{1}{n}\sum_{i=1}^{n} \widehat{m}(X_i,\theta)'[\widehat{\Sigma}_o(X_i,\theta)]^{-1}\widehat{m}(X_i,\theta),$$

其中 $N_{on}$ 是在筛空间 $B\times\mathcal{H}_n$ 内的 $\theta_o = (\beta_o, h_o)$ 的递缩邻域；$\widehat{\Sigma}_o(X_i,\theta)$ 是条件方差函数 $\Sigma_o(X,\theta) \equiv Var[\rho(Z,\beta,h(\cdot))|X]$ 的任意非参估计量。为了计算上述第 2 步，我们可以使用从第 1 步中得到的 $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{h}_n)$ 作为起始点。

Ai and Chen (1999) 考虑条件均值 $m(\cdot,\theta)$ 与条件方差 $\Sigma_o(\cdot,\theta)$ 的核估计；Ai and Chen (2003) 则提出了 $m(\cdot,\theta)$ 和 $\Sigma_o(\cdot,\theta)$ 的级数 LS 估计法。用 $\{p_{0j}(X), j = 1,2,...,k_{m,n}\}$ 表示一系列已知基函数，随着 $k_{m,n}\to\infty$，令 $p^{k_{m,n}}(X) = (p_{01}(X),...,p_{0k_{m,n}}(X))'$ 以及 $P = (p^{k_{m,n}}(X_1),...,p^{k_{m,n}}(X_n))'$，则这些基函数可以用来很好地近似任意 $X$ 的实值平方可积函数。这时，条件方差 $\Sigma_o(X,\theta) \equiv Var[\rho(Z,\theta)|X]$ 的一个级数 LS 估计量为：

$$\widehat{\Sigma}_o(X,\theta) \equiv \sum_{i=1}^{n} \rho(Z_i,\theta)\rho(Z_i,\theta)'p^{k_{m,n}}(X_i)'(P'P)^{-1}p^{k_{m,n}}(X)$$

同样的，可以用 $\widehat{\Sigma}_o(X) \equiv \widehat{\Sigma}_o(X,\widehat{\theta}_n)$ 很简单地估计 $\Sigma_o(X) = Var[\rho(Z,\theta_o)|X]$。

我们陈述下列适用于条件矩约束 $E[\rho(Z,\beta_o,h_o(\cdot))|X] = 0$ 类模型的有关 $\beta_o$ 的半参数有效估计的相关结果。具体细节可以参考 Ai and Chen (2003)。对 $j = 1,...,d_\beta$，令

$$
\begin{aligned}
D_{w_j}(X) &\equiv \frac{\partial E\{\rho(Z,\beta,h_o(\cdot))|X\}}{\partial \beta_j}|_{\beta=\beta_o} - \frac{\partial E\{\rho(X,\beta_o,h_o(\cdot)+\tau w_j(\cdot))|X\}}{\partial \tau}|_{\tau=0} \\
&\equiv \frac{\partial m(X,\theta_o)}{\partial \beta_j} - \frac{\partial m(X,\theta_o)}{\partial h}[w_j]\ ,
\end{aligned}
$$

$$
E\left\{D_{w_{oj}}(X)'\Sigma_o(X)^{-1}D_{w_{oj}}(X)\right\} = \inf_{w_j} E\left\{D_{w_j}(X)'\Sigma_o(X)^{-1}D_{w_j}(X)\right\},
$$

$w_o = (w_{o1},...,w_{od_\beta})$；用 $D_{w_o}(X) \equiv \left(D_{w_{o1}}(X),...,D_{w_{od_\beta}}(X)\right)$ 表示一个 $d_\rho \times d_\beta-$ 矩阵取值的 $X$ 的可测函数。

**定理 4.5.** *用 $\widetilde{\beta}_n$ 表示三步法最优加权筛 MD 估计量或两步法局域连续更新筛 MD 估计量。在 Ai and Chen (2003, 定理 6.1 和 6.2) 中用到的条件下，估计量 $\widetilde{\beta}_n$ 是半参数有效的并满足 $\sqrt{n}(\widetilde{\beta}_n - \beta_o) \xrightarrow{d} \mathcal{N}(0, V_o^{-1})$，其中*

$$
V_o = E\left[D_{w_o}(X)'[\Sigma_o(X)]^{-1}D_{w_o}(X)\right]
$$

Ai and Chen (2003) 还为 $\widetilde{\beta}_n$ 的渐近方差 $V_o^{-1}$ 提供了一个简单的一致估计量 $\widehat{V}_o^{-1}$，其中

$$
\widehat{V}_o = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \widehat{m}(X_i,\widetilde{\theta}_n)}{\partial \beta'} - \frac{\partial \widehat{m}(X_i,\widetilde{\theta}_n)}{\partial h}[\widehat{w}_o]\right)'\{\widehat{\Sigma}_o(X_i)\}^{-1}\left(\frac{\partial \widehat{m}(X_i,\widetilde{\theta}_n)}{\partial \beta'} - \frac{\partial \widehat{m}(X_i,\widetilde{\theta}_n)}{\partial h}[\widehat{w}_o]\right),
$$

其中 $\widehat{w}_o = (\widehat{w}_{o1},...,\widehat{w}_{od_\beta})$ 解决如下筛最小化问题：

$$
\min_{w_j \in \mathcal{H}_n}\sum_{i=1}^{n}\left(\frac{\partial \widehat{m}(X_i,\widetilde{\theta}_n)}{\partial \beta_j} - \frac{\partial \widehat{m}(X_i,\widetilde{\theta}_n)}{\partial h}[w_j]\right)'[\widehat{\Sigma}_o(X_i)]^{-1}\left(\frac{\partial \widehat{m}(X_i,\widetilde{\theta}_n)}{\partial \beta_j} - \frac{\partial \widehat{m}(X_i,\widetilde{\theta}_n)}{\partial h}[w_j]\right)
$$

对所有 $j = 1,...,d_\beta$；并且

$$
\frac{\partial \widehat{m}(X,\theta)}{\partial \beta_j} - \frac{\partial \widehat{m}(X,\theta)}{\partial h}[w_j] \equiv \sum_{i=1}^{n}\left(\frac{\partial \rho(Z_i,\theta)}{\partial \beta_j} - \frac{\partial \rho(Z_i,\theta)}{\partial h}[w_j]\right)p^{k_{m,n}}(X_i)'(P'P)^{-1}p^{k_{m,n}}(X)
$$

**备注 4.3: (1)** 最近 Chen and Pouzo (2006) 已经将 Ai and Chen (2003) 中得到的 $\sqrt{n}$ 渐近正态性和有效性的结论拓展到当广义残差函数 $\rho(Z,\beta,h(\cdot))$ 并不在 $\theta = (\beta,h)$ 上逐点连续的情形。

**(2)** 无论 $\rho(Z,\beta,h(\cdot)) - \rho(Z,\beta_o,h_o(\cdot))$ 是否取决于内生变量 $Y$，在模型 $E[\rho(Z,\beta_o,h_o(\cdot))|X] = 0$ 下三步最优加权筛 MD 估计法都可以达到 $\beta_o$ 估计的最优半参数效率上界。然而当 $\rho(Z,\beta,h(\cdot)) - \rho(Z,\beta_o,h_o(\cdot))$ 并不取决于内生变量 $Y$ 时，可以应用下列更简单的三部筛 GLS 估计方法来得到 $\beta_o$ 的有效估计（该方法在 Ai and Chen (1999) 中首次提出）：

**第 1 步**. 通过求解下列最小化问题，得到初始一致筛 GLS 估计量 $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{h}_n)$：

$$
\min_{(\beta,h)\in B\times\mathcal{H}_n}\frac{1}{n}\sum_{i=1}^{n}\rho(Z_i,\beta,h(\cdot))'\rho(Z_i,\beta,h(\cdot))
$$

**第 2 步**. 基于第一步得到的初始点估计结果 $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{h}_n)$, 应用任意半参方法计算一个一致的 $\Sigma_o(X) = Var[\rho(Z, \theta_o)|X]$ 的估计量 $\widehat{\Sigma}_o(X)$（例如 $\widehat{\Sigma}_o(X) = \widehat{\Sigma}_o(X, \widehat{\theta}_n)$）。

**第 3 步**. 通过求解下列最小化问题得到最优加权 GLS 估计量 $\widetilde{\theta}_n = (\widetilde{\beta}_n, \widetilde{h}_n)$

$$\min_{(\beta,h)\in B\times\mathcal{H}_n} \frac{1}{n}\sum_{i=1}^{n} \rho(Z_i, \beta, h(\cdot))'[\widehat{\Sigma}_o(X_i)]^{-1}\rho(Z_i, \beta, h(\cdot))$$

也就是说，对所有属于条件矩约束的第一子类 (2.8) 的模型，$E\{\rho(Z, \beta_o, h_o)|X\} = 0$, 其中 $\rho(Z, \theta) - \rho(Z, \theta_o)$ 并不取决于内生变量 $Y$, 我们有简单的三步筛 GLS 估计量 $\widetilde{\beta}_n$ 也满足 $\sqrt{n}(\widetilde{\beta}_n - \beta_o) \xrightarrow{d} \mathcal{N}(0, V_o^{-1})$。很明显，下面的连续更新筛 GLS 方法也可以得到 $\beta_o$ 的半参数有效一致估计：

$$(\widetilde{\beta}_{cgls}, \widetilde{h}_{cgls}) = \arg\min_{(\beta,h)\in B\times\mathcal{H}_n} \frac{1}{n}\sum_{i=1}^{n} \rho(Z_i, \beta, h(\cdot))'[\widehat{\Sigma}_o(X_i, \beta, h(\cdot))]^{-1}\rho(Z_i, \beta, h(\cdot))$$

对于条件矩约束模型（不存在未知函数 $h_o$），$E[\rho(Z, \beta_o)|X] = 0$，现在已经有许多不同的 $\beta_o$ 的有效估计方法，包括 Donald et al. (2003) 提出的经验似然估计，Newey and Smith (2004) 提出的广义经验似然估计 (GEL)，Kitamura et al. (2004) 提出的基于核方法的经验似然估计，Antoine, Bonnal and Renault (2006) 提出的连续更新最小距离方法或欧几里得条件经验似然估计法等等。直觉上我们可以将上述方法直接应用在更一般的条件矩框架 $E[\rho(Z, \beta_o, h_o())|X] = 0$ 下，其中可以用筛分方法去近似未知函数 $h_o()$。事实上，Zhang and Gijbels (2003) 已经在特例 $E[\rho(Z, \beta_o, h_o(X))|X] = 0$ 下考虑了筛经验似然方法，其中 $h_o$ 只是条件变量 $X$ 的函数；更一般的情况可以参考 Otsu (2005)。

近来，Ai and Chen (2004a, b) 已经考虑了半参数条件矩框架 $E[\rho_j(Z, \beta_o, h_o())|X_j] = 0$，$j = 1, ..., J$ 其中 $J$ 的大小是有限的，而且每个条件矩都有自己的条件变量集合 $X_j$，该集合在不同方程中可以包含不同的元素。他们的这一拓展将对估计具有不完全信息的半参数结构模型有所帮助。

# 5 结语

在这篇文章中，我们分析和总结了一些近来应用筛分方法对计量经济学模型进行非参数和半参数估计的相关大样本结果。我们的关注点在于未知函数的筛估计的一般一致性和收敛速度，以及平滑泛函的筛估计的 $\sqrt{n}-$ 渐近正态性。我们通过一些实例来展示一般筛估计的相关理论。我们希望这些例子可以充分地反映出一般筛极值估计方法的多样性和灵活性。由于篇幅所限，我们无法在一篇文章中讨论所有关于筛估计理论的研究进展。因此，我们在本节中将简要介绍其他有关筛分方法的课题供读者参考。

首先，尽管目前仍然缺乏有关使用筛分方法来进行假设检验的一般理论，我们已经看到一些使用筛分方法来实现一致设定检验的研究。例如 Hong and White (1995) 使用级数 LS 估计量检验了（有限维）参数回归模型；Hart (1997) 则使用级数估计量给出了许多一致的检验；Stinchcombe and White (1998) 使用了神经网络筛来检验（有限维）参数条件矩约束 $E[\rho(Z, \beta_o)|X] = 0$；Li et al. (2003) 应用了样条级数估计量来检验半参数/非参数回归模型。一个较新的进展是 Song (2005)，他提出通过条件鞅变换来一致地检验半非参数回归模型，其中未知函数可以使用筛分方法来估计。其他参考文献包括 Wooldridge (1992), Bierens (1990),

Bierens and Ploberger (1997) 以及 de Jong (1996)。此外，原则上所有现有的基于核方法或局部线性回归方法的假设检验都可以使用筛分方法来实现，例如 Robinson (1989), Fan and Li (1996), Lavergne and Vuong (1996), Chen and Fan (1999), Fan and Linton (1999), Ait-Sahalia et al. (2001), Horowitz and Spokoiny (2001) 以及 Fan et al. (2001)。

其次，本文并没有具体介绍数据导向的筛空间选择问题。在具体实践中，许多现有的模型选择方法（例如交叉验证 (CV)、广义 CV 以及 AIC 等）都直接被用作筛空间选择方法；这是由于参数模型与筛分方法有许多相关的地方；具体关于如何使用包括级数估计量在内的半非参数估计量的细节可以参考 Ichimura and Todd (2006) 中的综述部分，以及 Stone et al. (1997) 中的总结和 Ruppert et al. (2003) 中关于在拓展线性模型中使用样条筛实现模型选择的讨论。现在统计学中已经有一些论文（例如 Barron et al. (1999) 和 Shen and Ye (2002)）研究了基于数据的筛基选择问题。关于给定一种筛基，如何通过数据得到最优的筛基项数，学术界已经有很多结果。例如 Li (1987), Andrews (1991a), Hurvich et al. (1998), Donald and Newey (2001), Coppejans and Gallant (2002), Phillips and Ploberger (2003), Fan and Peng (2004) 以及 Imbens et al. (2005)。特别指出 Andrews (1991a) 证明了 CV 方法作为适用于具有异方差残差项的非参数最小二乘回归的选择级数项的方法是渐近最优的。Imbens et al. (2005) 则为平均处理效应参数的有效半参数估计证明了类似的结果；其中第一步对条件均值使用级数估计法估计。如果可以拓展上述研究结果来处理更广泛的通过筛分方法估计的半非参数模型，这将极大地促进对这类问题的理解。

第三，目前几乎没有关于半参数有效筛估计量的大样本性质的高阶细化的研究。包括 Linton (1995) 与 Heckman et al. (1998) 在内的许多作者已经指出半参数方法的一阶渐近结果可能会误导研究者。在核估计量方面，包括 Robinson (1995), Linton (1995, 2001), Nishiyama and Robinson (2000, 2005), Xiao and Linton (2001) and Ichimura 和 Linton (2002) 在内的许多文章都已经研究过核估计量的高阶细化问题。如果可以将上述结果拓展到使用筛分方法的半参数有效估计量上，这会对促进筛分方法的广泛应用起到积极的作用。

最后，考虑到在具体实施的层面上筛分方法相对易于应用，但获得其大样本性质却比较困难，我们也许可以通过将筛分方法同核估计法或局部线性回归方法（参考 Fan and Gijbels, 1996）结合起来取各家之所长来处理不同的问题。最近，包括 Horowitz and Mammen (2004) 和 Horowitz and Lee (2005) 在内的几篇文章已经展示了这种结合的实用性。

# 参考文献


[1] Ai, C. (1997) "A Semiparametric Maximum Likelihood Estimator", *Econometrica,* 65, 933-964.

[2] Ai, C., and X. Chen (2003) "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions", *Econometrica*, 71, 1795-1843. Working paper version, 1999.

[3] Ai, C., and X. Chen (2004a) "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables", Working paper, New York University.

[4] Ai, C., and X. Chen (2004b) "On Efficient Sequential Estimation of Semi-nonparametric Moment Models", Working paper, New York University.

[5] Ait-Sahalia, Y., P. Bickel and T. Stoker (2001) "Goodness-of-fit Tests for Kernel Regression with an Application to Option Implied Volatilities", *Journal of Econometrics*, 105, 363-412.

[6] Amemiya, T. (1985) *Advanced Econometrics*. Cambridge: Harvard University Press.

[7] Anastassiou, G. and X. Yu (1992a) "Monotone and Probabilistic Wavelet Approximation", *Stochastic Analysis and Applications*, 10, 251-264.

[8] Anastassiou, G. and X. Yu (1992b) "Convex and Convex-Probabilistic Wavelet Approximation", *Stochastic Analysis and Applications*, 10, 507-521.

[9] Andrews, D. (1991a) "Asymptotic Optimality of Generalized $C_L$, Cross-validation, and Generalized Cross-validation in Regression with Heteroskedastic Errors", *Journal of Econometrics,* 47, 359-377.

[10] Andrews, D. (1991b) "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models", *Econometrica,* 59, 307-345.

[11] Andrews, D. (1992) "Generic Uniform Convergence", *Econometric Theory,* 241-257.

[12] Andrews, D. (1994a) "Empirical process method in econometrics", in R.F. Engle III and D.F. McFadden (eds.), *The Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.

[13] Andrews, D. (1994b) "Asymptotics for Semi-parametric Econometric Models via Stochastic Equicontinuity", *Econometrica,* 62, 43-72.

[14] Andrews, D. and M. Schafgans (1998) "Semiparametric Estimation of the Intercept of a Sample Selection Model", *Review of Economic Studies*, 65, 497-517.

[15] Andrews, D. and Y. Whang (1990) "Additive Interactive Regression Models: Circumvention of the Curse of Dimensionality", *Econometric Theory*, 6, 466-479.

[16] Antoine, B., H. Bonnal and E. Renault (2006) "On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood", forthcoming in *Journal of Econometrics*.

[17] Bahadur, R.R. (1964) "On Fisher's bound for asymptotic variances", *Ann. Math. Statist.* 35, 1545-1552.

[18] Bansal, R., D. Hsieh and S. Viswanathan (1993) "A New Approach to International Arbitrage Pricing", *The Journal of Finance*, 48, 1719-1747.

[19] Bansal, R. and S. Viswanathan (1993) "No Arbitrage and Arbitrage Pricing: A New Approach", *The Journal of Finance*, 48(4), 1231-1262.

[20] Barnett, W.A., J. Powell and G. Tauchen (1991) *Non-parametric and Semi-parametric Methods in Econometrics and Statistics*. Cambridge University Press, New York.

[21] Barron, A.R. (1993) "Universal Approximation Bounds for Superpositions of a Sigmoidal Function", *IEEE Trans. Information Theory*, 39, 930-945.

[22] Barron, A., L. Birgé, P. Massart (1999) "Risk bounds for model selection via penalization", *Probab. Theory Related Fields*, 113, 301-413.

[23] Begun, J., W. Hall, W. Huang and J.A. Wellner (1983) "Information and asymptotic efficiency in parametric-nonparametric models", *The Annals of Statistics*, 11, 432-452.

[24] Bierens, H. (1990) "A Consistent Conditional Moment Test of Functional Form", *Econometrica,* 58, 1443-1458.

[25] Bierens, H. (2006) "Semi-Nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results", forthcoming in *Econometric Theory*.

[26] Bierens, H. and J. Carvalho (2006) "Semi-nonparametric competing risks analysis of recidivism", forthcoming in *Journal of Applied Econometrics*.

[27] Bierens, H. and W. Ploberger (1997) "Asymptotic Theory of Integrated Conditional Moment Tests", *Econometrica*, 65, 1129-1151.

[28] Bickel, P.J., C.A.J. Klaassen, Y. Ritov and J.A. Wellner (1993) *Efficient and adaptive estimation for semiparametric models*. The John Hopkins University Press, Baltimore.

[29] Birgé, L., and P. Massart (1998) "Minimum contrast estimators on sieves: Exponential bounds and rates of convergence", *Bernoulli*, 4 , 329-375

[30] Birman, M. and M. Solomjak (1967) "Piece-wise Polynomial Approximations of Functions in the Class $W_p^\alpha$", *Mathematics of the USSR Sbornik* 73 295-317.

[31] Blundell, R. and J. Powell (2003) "Endogeneity in Nonparametric and Semiparametric Regression Models", in M. Dewatripont, L.P. Hansen, and S. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications*, 2, 312-357, Cambridge: Cambridge University Press.

[32] Blundell, R., M. Browning and I. Crawford (2003) "Non-parametric Engel Curves and Revealed Preference", *Econometrica*, 71, 205-240.

[33] Blundell, R., X. Chen and D. Kristensen (2001) "Semiparametric Engel Curves with Endogenous Expenditure", manuscript, University of College London and New York University.

[34] Blundell, R., A. Duncan and K. Pendakur (1998) "Semiparametric Estimation and Consumer Demand", *Journal of Applied Econometrics*, 13, 435-461.

[35] Brendstrup, B. and H. Paarsch (2004) "Identification and Estimation in Sequential, Asymmetric, English Auctions", manuscript, University of Iowa.

[36] Cai, Z., J. Fan and Q. Yao (2000) "Functional-coefficient Regression Models for Nonlinear Time Series", *Journal of American Statistical Association*, 95, 941-956.

[37] Cameron, S. and J. Heckman (1998) "Life Cycle Schooling and Dynamic Selection Bias", *Journal of Political Economy,* 106, 262-333.

[38] Campbell, J. and J. Cochrane (1999) "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior", *Journal of Political Economy*, 107, 205-251.

[39] Carrasco, M., J.-P. Florens and E. Renault (2006) "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization", in J.J. Heckman and E.E. Leamer (eds.), *The Handbook of Econometrics*, vol. 6. North-Holland, Amsterdam.

[40] Chamberlain, G. (1992) "Efficiency Bounds for Semiparametric Regression", *Econometrica*, 60, 567-596.

[41] Chapman, D. (1997) "Approximating the Asset Pricing Kernel", *Journal of Finance,* 52(4), 1383-1410.

[42] Chen, R. and R. Tsay (1993) "Functional-coefficient Autoregressive Models", *Journal of American Statistical Association*, 88, 298-308.

[43] Chen, X. and T. Conley (2001) "A New Semiparametric Spatial Model for Panel Time Series", *Journal of Econometrics,* 105, 59-83.

[44] Chen, X. and Y. Fan (1999) "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics*, 91, 373-401

[45] Chen, X. and S. Ludvigson (2003) "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models", manuscript, New York University.

[46] Chen, X. and D. Pouzo (2006) "Efficient Estimation of Semi-nonparametric Conditional Moment Models With Possibly Nonsmooth Moments", manuscript, New York University.

[47] Chen, X. and X. Shen (1996) "Asymptotic Properties of Sieve Extremum Estimates for Weakly Dependent Data with Applications", manuscript, University of Chicago.

[48] Chen, X. and X. Shen (1998) "Sieve Extremum Estimates for Weakly Dependent Data", *Econometrica*, 66, 289-314.

[49] Chen, X. and H. White (1998) "Nonparametric Adaptive Learning with Feedback", *Journal of Economic Theory*, 82, 190-222.

[50] Chen, X. and H. White (1999) "Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators", *IEEE Tran. Information Theory,* 45, 682-691.

[51] Chen, X. and H. White (2002) "Asymptotic Properties of Some Projection-based Robbins-Monro Procedures in a Hilbert Space," *Studies in Nonlinear Dynamics and Econometrics*, vol. 6, issue 1, article 1.

[52] Chen, X., Y. Fan and V. Tsyrennikov (2004a) "Efficient Estimation of Semiparametric Multivariate Copula Models", forthcoming in *Journal of the American Statistical Association*.

[53] Chen, X., L.P. Hansen and J. Scheinkman (1998) "Shape-preserving Estimation of Diffusions", manuscript, University of Chicago.

[54] Chen, X., H. Hong and E. Tamer (2005) "Measurement Error Models with Auxiliary Data", *Review of Economic Studies*, 72, 343-366.

[55] Chen, X., H. Hong and A. Tarozzi (2004b) "Semiparametric Efficiency in GMM Models of Nonclassical Measurement Errors, Missing Data and Treatment Effects", manuscript, New York University and Duke University.

[56] Chen, X., O. Linton and I. van Keilegom (2003) "Estimation of Semiparametric Models when the Criterion Function is not Smooth", *Econometrica*, 71, 1591-1608.

[57] Chen, X., J. Racine and N. Swanson (2001) "Semiparametric ARX Neural Network Models with an Application to Forecasting Inflation", *IEEE Tran. Neural Networks,* 12, 674-683.

[58] Chernozhukov, V., G. Imbens and W. Newey (2006) "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions", forthcoming in *Journal of Econometrics*.

[59] Chui, C. (1992) *An Introduction to Wavelets*. San Diego: Academic Press, Inc.

[60] Cochrane, J. (2001) *Asset Pricing*. Princeton University Press, Princeton, NJ.

[61] Constantinides, G. (1990) "Habit-formation: A Resolution of the Equity Premium Puzzle", *Journal of Political Economy*, 98, 519-543.

[62] Coppejans, M. (2001) "Estimation of the Binary Response Model using a Mixture of Distributions Estimator (MOD)", *Journal of Econometrics,* 102, 231-261.

[63] Coppejans, M. and A.R. Gallant (2002) "Cross-validated SNP density estimates", *Journal of Econometrics,* 110, 27-65.

[64] Cosslett, S. (1983) "Distribution-Free Maximum Likelihood Estimation of the Binary Choice Model", *Econometrica*, 51, 765-782.

[65] Cybenko, G. (1990) "Approximation by Superpositions of a Sigmoid Function", *Mathematics of Control, Signals and Systems*, 2, 303-314.

[66] Darolles, S., J.-P. Florens and E. Renault (2002): "Nonparametric Instrumental Regression," mimeo, GREMAQ, University of Toulouse.

[67] Das, M., W.K. Newey and F. Vella (2003) "Nonparametric Estimation of Sample Selection Models", *Review of Economic Studies*, 70, 33-58.

[68] Daubechies, I. (1992) *Ten Lectures on Wavelets*, Philadelphia, SIAM.

[69] de Boor, C. (1978) *A Practical Guide to Splines*. Springer-Verlag, New York.

[70] Dechevsky, L. and S. Penev (1997) "On Shape-Preserving Probabilistic Wavelet Approximators", *Stochastic Analysis and Applications*, 15, 187-215.

[71] de Jong, R. (1996) "The Bierens Test Under Data Dependence", *Journal of Econometrics* 72 1-32.

[72] de Jong, R. (2002) "A Note on 'Convergence Rates and Asymptotic Normality for Series Estimators,': Uniform Convergence Rates", *Journal of Econometrics* 111, 1-9.

[73] DeVore, R.A. (1977a) "Monotone Approximation by Splines", *SIAM Journal on Mathematical Analysis*, 8, 891-905.

[74] DeVore, R.A. (1977b) "Monotone Approximation by Polynomials", *SIAM Journal on Mathematical Analysis*, 8, 906-921.

[75] DeVore, R.A. and G. G. Lorentz (1993) *Constructive Approximation*. Springer-Verlag, Berlin.

[76] Donald, S. and W. Newey (2001) "Choosing the Number of Instruments", *Econometrica*, 69, 1161-1191.

[77] Donald, S., G. Imbens and W. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55–93.

[78] Donoho, D. L., I. M. Johnstone, G. Kerkyacharian and D. Picard (1995) "Wavelet Shrinkage: Asymptopia?" *Journal of the Royal Statistical Society, Series B*, 57, 301-369.

[79] Doukhan, P., P. Massart and E. Rio (1995) "Invariance Principles for Absolutely Regular Empirical Processes," *Ann. Inst. Henri Poincaré - Probabilités et Statistiques*, 31, 393-427.

[80] Duncan, G.M. (1986) "A Semiparametric Censored Regression Estimator", *Journal of Econometrics*, 32, 5-34.

[81] Eastwood, B. and A. Gallant (1991) "Adaptive Rules for Seminonparametric Estimators that Achieve Asymptotic Normality", *Econometric Theory*, 7, 307-340.

[82] Eggermont, P. and V. LaRiccia (2001) *Maximum Penalized Likelihood Estimation, Volume I: Density Estimation*. Springer, New York

[83] Eichenbaum, M. and L.P. Hansen (1990) "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data", *Journal of Business and Economic Statistics*, 8, 53-69.

[84] Elbadawi, I., A.R. Gallant and G. Souza (1983) "An Elasticity Can Be Estimated Consistently Without A Prior Knowledge of Functional Form", *Econometrica*, 51, 1731-1751

[85] Engle, R. and G. Gonzalez-Rivera (1991) "Semiparametric ARCH Models", *Journal of Business and Economic Statistics*, 9, 345-359.

[86] Engle, R., C. Granger, J. Rice and A. Weiss (1986) "Semiparametric Estimates of the Relation between Weather and Electricity Sales", *Journal of the American Statistical Association*, 81, 310-320.

[87] Engle, R.F. and D.F. McFadden (1994) *The Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.

[88] Engle, R. and G. Rangel (2004) "The Spline GARCH Model for Unconditional Volatility and its Global Macroeconomic Causes", Working paper, New York University.

[89] Fan, J. and I. Gijbels (1996) *Local Polynomial Modelling and Its Applications* Chapman and Hall, London.

[90] Fan, J. and H. Peng (2004) "On Non-concave Penalized Likelihood with Diverging Number of Parameters", *The Annals of Statistics*, 32, 928-961.

[91] Fan, J. and Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods* Springer-Verlag, New York.

[92] Fan, J., C. Zhang and J. Zhang (2001) "Generalized Likelihood Ratio Statistics and Wilks Phenomenon", *The Annals of Statistics*, 29, 153-193.

[93] Fan, Y. and Q. Li (1996) "Consistent Model Specification Tests: Omitted Variables, Parametric and Semiparametric Functional Forms", *Econometrica*, 64, 865-890

[94] Fan, Y. and O. Linton (1999) "Some Higher Order Theory for a Consistent Nonparametric Model Specification Test", working paper LSE.

[95] Flinn, C. and J. Heckman (1982) "New Methods for Analyzing Structural Models of Labor Force Dynamics", *Journal of Econometrics*, 18, 115-168.

[96] Florens, J.P. (2003) "Inverse Problems and Structural Econometrics: the Example of Instrumental Variables", in M. Dewatripont, L.P. Hansen, and S. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications*, 2, 284-311, Cambridge: Cambridge University Press.

[97] Gabushin, (1967) "Inequalities for Norms of Functions and their Derivatives in the $L_p$ Metric", *Matematicheskie Zametki*, 1, 291-298.

[98] Gallant, A.R. (1987) "Identification and Consistency in Seminonparametric Regression", in T. F. Bewley (ed.), *Advances in Econometrics*, vol. I, 145-170, Cambridge University Press.

[99] Gallant, A.R. and D. Nychka (1987) "Semi-non-parametric maximum likelihood estimation", *Econometrica*, 55, 363-390.

[100] Gallant, A.R. and G. Souza (1991) "On the Asymptotic Normality of Fourier Flexible Form Estimates", *Journal of Econometrics*, 50, 329-353.

[101] Gallant, A.R. and G. Tauchen (1989) "Semiparametric Estimation of Conditional Constrained Heterogenous Processes: Asset Pricing Applications", *Econometrica*, 57, 1091-1120.

[102] Gallant, A.R. and G. Tauchen (1996) "Which Moments to Match?" *Econometric Theory*, 12, 657-681.

[103] Gallant, A.R. and G. Tauchen (2004) "EMM: A Program for Efficient Method of Moments Estimation, Version 2.0 User's Guide", Working paper, Duke University.

[104] Gallant, A.R. and H. White (1988a) "There Exists a Neural Network That does not Make Avoidable Mistakes", in *Proceedings of the IEEE 1988 International Conference on Neural Networks*, Vol. 1, IEEE, New York, pp. 657-664.

[105] Gallant, A.R. and H. White (1988b) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.

[106] Gallant, A.R. and H. White (1992) "On Learning the Derivatives of an Unknown Mapping with Multilayer Feedforward Networks", *Neural Networks* 5, 129-138.

[107] Gallant, A.R., D. Hsieh and G. Tauchen (1991) "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 1974-83", in Barnett, W.A., J. Powell and G. Tauchen (eds.), *Non-parametric and Semi-parametric Methods in Econometrics and Statistics,* 199-240, Cambridge: Cambridge University Press.

[108] Geman, S. and C. Hwang (1982) "Nonparametric Maximum Likelihood Estimation by the Method of Sieves", *The Annals of Statistics*, 10, 401-414.

[109] Girosi, F. (1994) "Regularization theory, radial basis functions and networks", In *From Statistics to Neural Networks. Theory and Pattern Recognition Applications*, V. Cherkassky, J.H. Friedman, and H. Wechsler, eds. Springer-Verlag, Berlin.

[110] Granger, C.W.J., and T. Terasvirta (1993) *Modelling nonlinear economic relationships*, Oxford, New York.

[111] Grenander, U. (1981) *Abstract Inference*, New York: Wiley Series.

[112] Haerdle, W. and O. Linton (1994) "Applied Nonparametric Methods", in R.F. Engle III and D.F. McFadden (eds.), *The Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.

[113] Haerdle, W., M. Mueller, S. Sperlich and A. Werwatz (2004) *Nonparametric and Semiparametric Models,* New York: Springer

[114] Hahn, J. (1998) "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects", *Econometrica*, 66, 315-332.

[115] Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables", *Annals of Statistics*, 33, 2904-2929.

[116] Hansen, L.P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

[117] Hansen, L.P. (1985) "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators", *Journal of Econometrics*, 30, 203-238.

[118] Hansen, L.P. and K. Singleton (1982) "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", *Econometrica*, 50, 1269-86.

[119] Hansen, L.P. and S. Richard (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models", *Econometrica*, 55, 587-613.

[120] Hansen, M.H. (1994) *Extended Linear Models, Multivariate Splines, and ANOVA*. Ph. D. Dissertation, Department of Statistics, University of California at Berkeley.

[121] Hart, J. (1997) *Nonparametric Smoothing and Lack-of-Fit Tests,* New York: Springer-Verlag.

[122] Hausman, J. and W. Newey (1995) "Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss", *Econometrica*, 63, 1445-1467.

[123] Heckman, J. (1979) "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-161.

[124] Heckman, J. and B. Singer (1984) "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, 68, 839-874.

[125] Heckman, J. and R. Willis (1977) "A Beta Logistic Model for the Analysis of Sequential Labor Force Participation of Marries Women", *Journal of Political Economy*, 85, 27-58.

[126] Heckman, J., H. Ichimura, J. Smith and P. Todd (1998) "Characterization of Selection Bias Using Experimental Data", *Econometrica*, 66, 1017-1098.

[127] Hirano, K., G. Imbens and G. Ridder (2003) "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score", *Econometrica*, 71, 1161-1189.

[128] Hong, Y. and H. White (1995) "Consistent Specification Testing via Nonparametric Series Regression", *Econometrica*, 63, 1133-1159.

[129] Honore, B. (1990) "Simple Estimation of a Duration Model with Unobserved Heterogeneity," *Econometrica* 58, 453-473.

[130] Honore, B. (1994) "A Note on the Rate of Convergence of Estimators of Mixtures of Weibulls", manuscript, Northwestern University.

[131] Honore, B. and E. Kyriazidou (2000) "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica*, 68, 839-874.

[132] Hornik, K., M. Stinchcombe and H. White (1989) "Multilayer feedforward networks are universal approximators", *Neural Networks*, 2, 359-366.

[133] Hornik, K., M. Stinchcombe, H. White and P. Auer (1994) "Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives", *Neural Computation*, 6, 1262-75.

[134] Horowitz, J. (1992) "A smoothed Maximum Score Estimator for the Binary Response Model", *Econometrica*, 60, 505-531.

[135] Horowitz, J. (1998) *Semiparametric Methods in Econometrics*. New York Springer-Verlag.

[136] Horowitz, J. and S. Lee (2005) "Nonparametric Estimation of An Additive Quantile Regression Model", *Journal of the American Statistical Association*, 100, 1238-1249.

[137] Horowitz, J. and E. Mammen (2004) "Nonparametric Estimation of An Additive Model with A Link Function", *Annals of Statistics*, 32, 2412-2443.

[138] Horowitz, J. and V. Spokoiny (2001) "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative", *Econometrica*, 69, 599-631.

[139] Hu, Y. and S. Schennach (2006) "Identification and Estimation of Nonclassical Nonlinear Errors-in-Variables Models With Continuous Distributions Using Instruments," Working paper, University of Texas, Austin.

[140] Huang, J.Z. (1998a) "Projection estimation in multiple regression with application to functional ANOVA models", *The Annals of Statistics*, 26, 242-272.

[141] Huang, J.Z. (1998b) "Functional ANOVA models for generalized regression", *Journal of Multivariate Analysis*, 67, 49-71.

[142] Huang, J.Z. (2001) "Concave extended linear modeling: a theoretical synthesis", *Statistica Sinica*, 11, 173-197.

[143] Huang, J.Z. (2003) "Local asymptotics for polynomial spline regression", *The Annals of Statistics*, 31, 1600-1635.

[144] Huang, J.Z., C. Kooperberg, C.J. Stone and Y.K. Truong (2000) "Functional ANOVA modeling for proportional hazards regression", *The Annals of Statistics*, 28, 960-999.

[145] Hurvich, C., J. Simonoff and C. Tsai (1998) "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion", *Journal of the Royal Statistical Society, Series B*, 60, 271-293.

[146] Hutchinson, J., A. Lo and T. Poggio (1994) "A non-parametric approach to pricing and hedging derivative securities via learning networks", *J. of Finance*, 3, 851-889.

[147] Ibragimov, I.A. and R.Z. Has'minskii (1991) "Asymptotically normal families of distributions and efficient estimation", *The Annals of Statistics*, 19, 1681-1724.

[148] Ichimura, H. (1993) "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models", *Journal of Econometrics,* 58, 71-120.

[149] Ichimura, H. and S. Lee (2006) "Characterization of the Asymptotic Distribution of Semiparametric M-Estimators", manuscript, UCL.

[150] Ichimura, H. and O. Linton (2002) "Asymptotic Expansions for Some Semiparametric Program Evaluation Estimators", working paper IFS and LSE.

[151] Ichimura, H. and P. Todd (2006) "Implementing Nonparametric and Semiparametric Estimators", in J.J. Heckman and E.E. Leamer (eds.), *The Handbook of Econometrics*, vol. 6. North-Holland, Amsterdam.

[152] Imbens, G., W. Newey and G. Ridder (2005) "Mean-squared-error Calculations for Average Treatment Effects", manuscript, UC Berkeley.

[153] Ishwaran, H. (1996a) "Identification and Rates of Estimation for Scale Parameters in Location Mixture Models", *The Annals of Statistics*, 24, 1560-1571.

[154] Ishwaran, H. (1996b) "Uniform Rates of Estimation in the Semiparametric Weibull Mixture Models", *The Annals of Statistics*, 24, 1572-1585.

[155] Jovanovic, B. (1979) "Job Matching and the Theory of Turnover", *Journal of Political Economy*, 87, 972-990.

[156] Judd, K. (1998) *Numerical Method in Economics*, MIT University Press.

[157] Kim, J. and D. Pollard (1990) "Cube Root Asymptotics", *The Annals of Statistics*, 18, 191-219.

[158] Kitamura, Y., G. Tripathi and H. Ahn (2004) "Empirical Likelihood-based Inference in Conditional Moment Restriction Models", *Econometrica*, 72, 1667-1714.

[159] Khan, S. (2005) "An Alternative Approach to Semiparametric Estimation of Heteroskedastic Binary Response Models", manuscript, University of Rochester.

[160] Klein, R. and R. Spady (1993) "An Efficient Semiparametric Estimator for Binary Response Models", *Econometrica,* 61, 387-421.

[161] Koenker, R. and G. Bassett (1978) "Regression quantiles", *Econometrica*, 46, 33-50.

[162] Koenker, R. and I. Mizera (2003) "Penalized Triograms: Total Variation Regularization for Bivariate Smoothing", *J. Royal Stat. Soc., B,* 66, 145-163.

[163] Koenker, R., P. Ng and S. Portnoy (1994) "Quantile Smoothing Splines", *Biometrika*, 81, 673-680.

[164] Kooperberg, C., C.J. Stone and Y.K. Truong (1995a) "Hazard regression", *Journal of the American Statistical Association*, 90, 78-94.

[165] Kooperberg, C., C.J. Stone and Y. K. Truong (1995b) "Rate of convergence for logspline spectral density estimation", *Journal of Time Series Analysis*, 16, 389-401.

[166] Lavergne, P. and Q. Vuong (1996) "Nonparametric Selection of Regressors: the Nonnested Case", *Econometrica*, 64, 207-219.

[167] LeCam, L. (1960) "Local asymptotically normal families of distributions", Univ. California Publications in Statist. **3**, 37-98.

[168] Lee, S. (2003) "Efficient Semiparametric Estimation of a Partially Linear Quantile Regression Model", *Econometric Theory*, 19, 1-31.

[169] Li, K. (1987) "Asymptotic Optimality for $C_p$, $C_L$, Cross-validation, and Generalized Cross-validation: Discrete Index Set", *Annals of Statistics* 15, 958-975.

[170] Li, Q. and J. Racine (2006) *Nonparametric Econometrics Theory and Practice*, forthcoming in Princeton University Press.

[171] Li, Q., C. Hsiao and J. Zinn (2003) "Consistent Specification Tests for Semiparametric/Nonparametric Models Based on Series Estiamtion Methods", *Journal of Econometrics*, 112, 295-325.

[172] Linton, O. (1995) "Second Order Approximation in the Partially Linear Regression Model", *Econometrica,* 63, 1079-1112.

[173] Linton, O. (2001) "Edgeworth Approximations for Semiparametric Instrumental Variable Estimators and Test Statistics", *Journal of Econometrics,* 106, 325-368.

[174] Linton, O. and E. Mammen (2005) "Estimating Semiparametric ARCH($\infty$) Models by Kernel Smoothing Methods", *Econometrica,* 73, 771-836.

[175] Lorentz, G. (1966) *Approximation of functions*, New York: Holt.

[176] Mahajan, A. (2004) "Identification and Estimation of Single Index Models with Misclassified Regressors", manuscript, Stanford University.

[177] Makovoz, Y. (1996) "Random approximants and neural networks", *J. Approximation Theory*, **85**, 98-109.

[178] Manski, C. (1985) "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics,* 27, 313-334.

[179] Manski, C. (1994) "Analog Estimation of Econometric Models", in R.F. Engle III and D.F. McFadden (eds.), *The Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.

[180] Matzkin, R.L. (1994) "Restrictions of Economic Theory in Nonparametric Methods", in R.F. Engle III and D.F. McFadden (eds.), *The Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.

[181] McCaffrey, D., S. Ellner, A. Gallant and D. Nychka (1992) "Estimating the Lyapunov exponent of a chaotic system with nonparametric regression," *Journal of the American Statistical Association* 87, 682-695.

[182] Meyer, Y. (1992) *Ondelettes et operateurs I: Ondelettes*, Paris: Hermann.

[183] Murphy, S. and A. van der Vaart (2000) "On Profile Likelihood", *Journal of the American Statistical Association*, 95, 449-465.

[184] Newey, W.K. (1988) "Two Step Series Estimation of Sample Selection Models", manuscript, MIT Department of Economics.

[185] Newey, W.K. (1990a) "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99-135.

[186] Newey, W.K. (1990b) "Efficient Instrumental Variables Estimation of Nonlinear Models", *Econometrica*, 58, 809-837.

[187] Newey, W.K. (1991) "Uniform Convergence in Probability and Stochastic Equicontinuity", *Econometrica*, 59, 1161-1167.

[188] Newey, W.K. (1993) "Efficient Estimation of Models with Conditional Moment Restrictions, in *Handbook of Statistics*, Vol. 11, G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., Amsterdam: North-Holland.

[189] Newey, W.K. (1994a) "The Asymptotic Variance of Semiparametric Estimators", *Econometrica,* 62, 1349-1382.

[190] Newey, W.K. (1994b) "Series Estimation of Regression Functionals", *Econometric Theory,* 10, 1-28.

[191] Newey, W.K. (1997) "Convergence Rates and Asymptotic Normality for Series Estimators", *Journal of Econometrics,* 79, 147-168.

[192] Newey, W.K. (2001) "Flexible Simulated Moment Estimation of Nonlinear Errors in Variables Models," *Review of Economics and Statistics*, 83, 616-627.

[193] Newey, W.K. and D. F. McFadden (1994) "Large sample estimation and hypothesis testing", in R.F. Engle III and D.F. McFadden (eds.), *The Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.

[194] Newey, W.K. and J.L Powell (2003) "Instrumental Variable Estimation of Nonparametric Models", *Econometrica,* 71, 1565-1578. Working paper version, 1989.

[195] Newey, W.K. and R. Smith (2004) "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", *Econometrica*, 72, 219-256.

[196] Newey, W.K., J.L. Powell and F. Vella (1999) "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica,* 67, 565-603.

[197] Nishiyama, Y. and P.M. Robinson (2000) "Edgeworth Expansions for Semiparametric Averaged Derivatives", *Econometrica* 68, 931-980.

[198] Nishiyama, Y. and P.M. Robinson (2005) "The Bootstrap and the Edgeworth Correction for Semiparametric Averaged Derivatives", *Econometrica* 73, 903-980.

[199] Ossiander, M. (1987) "A central limit theorem under metric entropy with $L_2$ bracketing", *The Annals of Probability,* 15, 897-919.

[200] Otsu, T. (2005) "Sieve Conditional Empirical Likelihood Estimation of Semiparametric Models", manuscript, Yale University.

[201] Pagan, A. and A. Ullah (1999) *Nonparametric Econometrics*, Cambridge University Press.

[202] Pakes, A. and S. Olley (1995) "A Limit Theorem for A Smooth Class of Semiparametric Estimators", *Journal of Econometrics*, 65, 295-332.

[203] Pastorello, S., V. Patilea and E. Renault (2003) "Iterative and recursive estimation in structural non-adaptive models", *Journal of Business & Economic Statistics*, 21, 449-509.

[204] Phillips, P.C.B. (1998) "New Tools for Understanding Spurious Regressions", *Econometrica,* 66, 1299-1325.

[205] Phillips, P.C.B. and W. Ploberger (2003) "An Introduction to Best Empirical Models when the Parameter Space is Infinite Dimensional", *Oxford Bulletin of Economics and Statistics*, 65, 877-890.

[206] Pinkse, J. (2000) "Nonparametric Two-step Regression Estimation When Regressors and Errors are Dependent", *Canadian Journal of Statistics*, 28, 289-300.

[207] Pollard, D. (1984) *Convergence of Statistical Processes*. Springer-Verlag, New York.

[208] Polk, C., S. Thompson and T. Vuolteenaho (2003) "New Forecasts of the Equity Premium," manuscript, Harvard University.

[209] Portnoy, S. (1997) "Local Asymptotics for Quantile Smoothing Splines," *The Annals of Statistics*, 25, 387-413.

[210] Powell, J., J. Stock and T. Stoker (1989) "Semiparametric Estimation of Index Coefficients", *Econometrica,* 57, 1403-1430.

[211] Powell, J. (1994) "Estimation of Semiparametric Models", in R.F. Engle III and D.F. McFadden (eds.), *The Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.

[212] Robinson, P. (1988) "Root-N-Consistent Semiparametric Regression", *Econometrica,* 56, 931-954.

[213] Robinson, P. (1989) "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series", *Review of Economic Studies*, 56, 511-534.

[214] Robinson, P. (1995) "The Normal Approximation for Semiparametric Averaged Derivatives", *Econometrica,* 63, 667-680.

[215] Ruppert, D., M. Wand and R. Carroll (2003) *Semiparametric Regression*, Cambridge: Cambridge University Press.

[216] Schumaker, L. (1981) *Spline Functions: Basic Theory.* New York: John Wiley & Sons.

[217] Severini, T. and H.W. Wong (1992) "Profile Likelihood and Conditionally Parametric Models", *The Annals of Statistics*, 20, 1768-1802.

[218] Shen, X. (1997) "On Methods of Sieves and Penalization", *The Annals of Statistics,* 25, 2555-2591.

[219] Shen, X. and W. Wong (1994) "Convergence Rate of Sieve Estimates", *The Annals of Statistics,* 22, 580-615.

[220] Shen, X. and J. Ye (2002) "Adaptive Model Selection", *Journal of American Statistical Association* 97, 210-221.

[221] Shintani, M. and O. Linton (2003) "Nonparametric Neural Network Estimation of Lyapunov Exponents and a Direct Test for Chaos," *Journal of Econometrics*, forthcoming.

[222] Song, K. (2005) "Testing Semiparametric Conditional Moment Restrictions Using Conditional Martingale Transforms," manuscript, Yale University, Dept. of Economics.

[223] Stinchcombe, M. (2002) "Some Genericity Analyses in Nonparametric Econometrics", manuscript, University of Texas, Austin, Dept. of Economics.

[224] Stinchcombe, M. and H. White (1998) "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative", *Econometric Theory,* 14, 295-325.

[225] Stone, C.J. (1982) "Optimal Global Rates of Convergence for Nonparametric Regression", *The Annals of Statistics,* 10, 1040-1053.

[226] Stone, C.J. (1985) "Additive regression and other nonparametric models", *The Annals of Statistics,* 13, 689-705.

[227] Stone, C.J. (1986) "The dimensionality reduction principle for generalized additive models", *The Annals of Statistics,* 14, 590-606.

[228] Stone, C.J. (1990) "Large-sample inference for log-spline models", *The Annals of Statistics,* 18, 717-741.

[229] Stone, C.J. (1994) "The use of polynomial splines and their tensor products in multivariate function estimation" (with discussion), *The Annals of Statistics,* 22, 118-184.

[230] Stone, C.J., M. Hansen, C. Kooperberg and Y.K. Truong (1997) "Polynomial splines and their tensor products in extended linear modeling" (with discussion), *The Annals of Statistics,* 25, 1371-1470.

[231] Strawderman, R.L. and A. A. Tsiatis (1996) "On the asymptotic properties of a flexible hazard estimator," *The Annals of Statistics,* 24 , 41-63.

[232] Timan, A.F. (1963) *Theory of Approximation of Functions of a Real Variable*, MacMillan, New York.

[233] Van de Geer, S. (1993) "Hellinger-consistency of certain nonparametric maximum likelihood estimators", *The Annals of Statistics,* 21, 14-44.

[234] Van de Geer, S. (1995) "The method of sieves and minimum contrast estimators", *Mathematical Methods of Statistics,* 4 , 20-38.

[235] Van de Geer, S. (2000) *Empirical Processes in M-estimation*, Cambridge University Press.

[236] Van der Vaart, A. (1991) "On Differentiable Functionals", *The Annals of Statistics,* 19, 178-204.

[237] Van der Vaart, A. and J. Wellner (1996) *Weak Convergence and Empirical Processes: with Applications to Statistics,* New York: Springer-Verlag.

[238] Vapnik, V. (1998) *Statistical Learning Theory,* New York: Wiley Interscience.

[239] Wahba, G. (1990) *Spline Models for Observational Data,* CBMS-NSF Regional Conference Series, Philadelphia.

[240] White, H. (1984) *Asymptotic Theory for Econometricians.* Academic Press.

[241] White, H. (1990) "Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings", *Neural Networks*, 3, 535-550.

[242] White, H. (1994) *Estimation, Inference and Specification Analysis*, Cambridge University Press.

[243] White, H. and J. Wooldridge (1991) "Some results on sieve estimation with dependent observations", in Barnett, W.A., J. Powell and G. Tauchen (eds.), *Non-parametric and Semi-parametric Methods in Econometrics and Statistics*, 459-493, Cambridge: Cambridge University Press.

[244] Wong, W.H. (1992) "On Asymptotic Efficiency in Estimation Theory", *Statistica Sinica*, 2, 47-68.

[245] Wong, W.H. and T. Severini (1991) "On Maximum Likelihood Estimation in Infinite Dimensional Parameter Spaces", *The Annals of Statistics,* 19, 603-632.

[246] Wong, W.H. and X. Shen (1995) "Probability inequalities for likelihood ratios and convergence rates for sieve MLE's", *The Annals of Statistics,* 23, 339-362.

[247] Wooldridge, J. (1992) "A Test for Functional Form Against Nonparametric Alternatives", *Econometric Theory* 8, 452–475.

[248] Wooldridge, J. (1994) "Estimation and Inference for Dependent Processes", in R.F. Engle III and D.F. McFadden (eds.), *The Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.

[249] Xiao, Z. and O. Linton (2001) "Second Order Approximation for an Adaptive Estimator in a Linear Regression", *Econometric Theory* 17, 984-1024.

[250] Zhang, J. and I. Gijbels (2003) "Sieve Empirical Likelihood and Extensions of the Generalized Least Squares", *Scandinavian Journal of Statistics*, 30, 1-24.

[251] Zhou, S., X. Shen and D. A. Wolfe (1998) "Local Asymptotics for Regression Splines and Confidence Regions", *The Annals of Statistics,* 26 1760-1782.