国内半参数回归模型研究进展

王成勇
(襄樊学院 数学系，湖北 襄樊 441053)

摘要：对半参数回归模型的研究是近年来统计研究的热点之一，它结合线性回归模型和非参数回归模型，吸收了各自的优点，因此不论是在理论研究上还是实际应用中都具有重要意义. 文章总结了国内学者对半参数回归模型研究的贡献，理清了对该模型的理论研究演进的方向和脉络.

关键词：半参数回归模型；相合性；收敛速度
中图分类号：O212.1 **文献标志码**：A **文章编号**：1009-2854(2008)02-0008-06

半参数回归模型这一统计分支是近年来兴起的，无论在理论研究还是实际应用上，它都受到了许多统计学者的关注. 同其它回归模型问题一样，人们对此课题理论研究的兴趣主要集中在大样本性质上，并自八十年代初以来取得了丰硕的研究成果. 我国统计学者参与这方面的研究工作起步稍晚些，在 80 年代末期才开始这方面的系统研究，并在 90 年代中后期由高集体、洪圣岩、梁华、柴根象、薛留根、胡舒合、陈明华、钱伟民等学者带领掀起了一股研究热潮，论文遍布于国内数学类的重要期刊，因而国内学者对于半参数回归模型的研究达到了相当深入的程度.

对于半参数回归模型的研究主要集中于以下几个方面：一是在基本模型下，基于各种不同的估计方法或者基于误差的不同设定，探讨参数、非参数部分以及误差方差估计量的强相合性、弱相合性、p 阶平均相合性及其收敛速度，以及参数、误差方差估计量的渐近正态性等；二是把模型推广，探讨存在数据污染、存在数据截断或数据删失时，模型的估计及相应的估计量的性质；再就是实证与应用研究. 下面分别就各方面的研究进展与现状做一个比较详细的总结.

1 基本模型基于不同的估计方法或误差的不同设定下估计性质的研究

考虑半参数回归模型

$$Y_i = X_i'\beta + g(T_i) + e_i, \quad 1 \le i \le n \tag{1}$$

式(1)中，*i.i.d.*为独立同分布，e_i 为 *i.i.d.* 随机误差，$Ee_i = 0, Vare_i = \sigma^2$. $\{(X_i, T_i), 1 \le i \le n\}$ 为 *i.i.d.* 随机设计或固定非随机设计点列且 $X_i \in R^p$，β 是未知待估参数，$g(.)$ 是定义在 R 上的未知函数. 当 $\{(X_i, T_i), 1 \le i \le n\}$ 为 *i.i.d.* 随机设计时，(X_i, T_i) 与 e_i 相互独立.

模型(1)是 Engle, et al[1]在研究气候条件对电力需求影响这一实际问题时提出的. 随后，Heckman[2]研究了 $\{(X_i, T_i), 1 \le i \le n\}$ 为 *i.i.d.* 随机设计，$\{X_i\}$ 与 $\{T_i\}$ 相互独立，且 $g(.)$ 的估计取一类样条估计时，β 的加权最小二乘估计 $\hat{\beta}_n$ 的渐近正态性；同时，Rice[3]研究了 $\{(X_i, T_i), 1 \le i \le n\}$ 是固定设计，$g(.)$ 的估计取一类样条估计时，β 的估计的协方差函数的渐近性质；而 Chen[4]研究了当 $h_j(t) = E(X_{1j} \mid T = t)$ 关于 t 满足 $\alpha(0 < \alpha \le 1)$ 阶 Lipschitz 条件，$g(.)$ 的估计取一逐点多项式估计时，β 的加权最小二乘估计 $\hat{\beta}_n$ 的渐近正态性及 $g(.)$ 的估计的弱收敛速度. 一些学者还研究了当 $g(.)$ 的估计取一些样条估计时，β 的若干估计的性质.

其后，对模型(1)的研究的推进主要沿着三个方向：一是对加权最小二乘法(综合最小二乘法和非参数加权函数法)估计量大样本性质的深入细致的研究，这方面的文献相当多，并有许多深刻的结论；二是在估计方法上的改进和创新；三是将 *i.i.d.* 的误差 e_i 推广到各种相关情形，在此基础上讨论估计量的渐近性质.

1.1 加权最小二乘法(综合最小二乘法和非参数加权函数方法)估计性质的研究

在国内，高集体、洪圣岩、成平、梁华等学者是最先开始这方面研究的先驱者. 他们在 89 年到 92 年

收稿日期：2007-06-11
作者简介：王成勇(1973-)，男，湖北保康人，襄樊学院数学系讲师，华中科技大学经济学院在读博士，研究方向：计量经济学.

之间在这一课题研究的基础上先后在国内发表了近十篇文章，得到了许多深刻的结论，引起了国内学者对该模型理论研究的广泛关注. 其中，高集体和洪圣岩先后独立地研究了模型中当 $g(.)$ 的估计取 Parzen-Rosenblatt 核估计时，β 的加权最小二乘估计 $\hat{\beta}_n$ 的渐近正态性以及 $\hat{\beta}_n$ 和 $\hat{g}_n$ 的强弱收敛速度；高集体还进一步研究了当 $g(.)$ 的估计取一类核估计序列(包括常见的Parzen-Rosenblatt核和常见的离散核)时，β 的加权最小二乘估计 $\hat{\beta}_n$ 的渐近正态性及 $\hat{\beta}_n$ 和 $\hat{g}_n$ 的强弱收敛速度. 与此同时，高集体和洪圣岩还研究了模型中当 $g(.)$ 的估计取一类近邻估计时，$\hat{\beta}_n$ 的渐近正态性及 $\hat{\beta}_n$ 和 $\hat{g}_n$ 的强弱收敛速度. 高集体在其 92 年的博士论文中系统地研究了固定设计情形下参数 β 的加权最小二乘估计 $\hat{\beta}_n$ 的渐近正态性及其渐近正态的 Berry-Essen 界限，同时还给出了若干估计的强弱收敛速度，得到了许多有意义的结论. 随后，洪圣岩、成平先后讨论了模型(1)中，当 $\{(X_i,T_i),1\le i\le n\}$ 是 *i.i.d.* 随机设计情形且 $g(.)$ 的估计取核估计时，$\hat{\beta}_n$ 及其 t 统计量的渐近分布的 Berry-Essen 界限，洪圣岩[5]还研究了 $\hat{\beta}_n$ 和 $\sigma^2 = Ee_1^2$ 的估计的重对数律. 这些结果在文献[6]中有比较系统的总结.

高集体[7]在 $\{(X_i,T_i),1\le i\le n\}$ 是固定设计点列情形且 $g(.)$ 的估计取一类非参数权估计(包括常见的核估计和近邻估计)时得到了 β 和 $g(.)$ 加权最小二乘估计的强弱收敛速度；陈明华[8]针对文献[7]中的证明错误，在较基本的假设下得到了 $\hat{\beta}_n$ 和 $\sigma^2 = Ee_1^2$ 的估计 $\hat{\sigma}_n^2$ 的精确收敛速度—重对数律. 随后，陈明华[9]在文献[7]的基础上，进一步用 β 和 $\sigma^2 = Ee_1^2$ 的加权最小二乘估计 $\hat{\beta}_n$ 和 $\hat{\sigma}_n^2$ 构造了其 Bootstrap 统计量 $\hat{\beta}_n^*$ 和 $\hat{\sigma}_n^{*2}$，并证明了在给定原样本的条件下，$\sqrt{n}(\hat{\beta}_n^*-\hat{\beta}_n)$ 和 $\sqrt{n}(\hat{\sigma}_n^{*2}-\hat{\sigma}_n^2)$ 分别与 $\sqrt{n}(\hat{\beta}_n-\beta)$ 和 $\sqrt{n}(\hat{\sigma}_n^2-\sigma^2)$ 有相同的渐近分布. 陈明华[10]在 $\{X_i\}$ $\{T_i\}$ 都是固定设计点列且 $p=1$ 的情形证明了 β 和 $g(.)$ 以及误差方差 σ^2 估计的强相合性和 P-阶($P\ge 2$)平均相合性. 另外，胡舒合[11]将模型稍作修改，在 $X_i\in R, T_i\in R^p$ 为已知设计点列时，获得了 β 和 $g(.)$ 加权最小二乘估计的强相合、强一致相合、s 阶平均相合和 s 阶一致平均相合性.

由以上看出，对于基本模型(1)的理论研究细致而深入，特别是高集体、洪圣岩、陈平、梁华等学者得到了许多深刻而有意义的结论，对半参数回归模型的理论研究的推进作出了许多贡献，并开创了国内半参数回归模型研究的先河.

1.2 两阶段估计性质的研究

柴根象等[12]基于模型(1)的可加性，建立起两阶段估计方法. 文中假定 $\{X_i\}$ 为 p 维固定设计向量，$\{T_i\}$，$\{e_i\}$ 为 *i.i.d.* 点列且相互独立. 首先将模型(1)变换成一个标准的线性模型，利用最小二乘法得到 β 的第一次估计 β_n^*，并由标准线性模型的残差得出 $g(.)$ 的估计 $\hat{g}_n$，然后将 $\hat{g}_n$ 代回模型(1)，再次利用最小二乘法得出 β 的估计量 $\hat{\beta}_n$. 文章在比较弱的条件下证明了第一次估计 $\hat{\beta}_n^*$ 的强相合性和渐近正态性，以及当 $\hat{g}_n$ 为一类核估计时 $\hat{g}_n$ 的强相合性、一致强相合性，并且得到了其一致强收敛速度可达到非参数回归函数估计的最优一致强收敛速度 $(n^{-1}\log n)^{1/3}$ 的好结果；进一步得到了第二次估计 $\hat{\beta}_n$ 的强收敛速度以及 $\hat{\beta}_n$ 的渐近正态性，并且其渐近方差达到了 β 的所谓“正则估计”所能达到的最小渐近方差(文中首次提出模型(1)的两阶段估计，其中结论的条件及证明方法对后续的基于两阶段估计的文献有较大的影响)；随后，钱伟民、柴根象[13]将文献[12]中的核权函数改为最近邻权函数，重新讨论了 $\hat{\beta}_n$ 和 $\hat{g}_n$ 的大样本性质，得到了类似的结果. 钱伟民[14-15]基于两阶段估计构造了误差方差 σ^2 的估计量 $\hat{\sigma}_n^2$ 和 $Var(e_1^2)$ 的估计 $\hat{S}_n^2$，讨论了 $\hat{\sigma}_n^2$ 的 t 统计量的渐近正态性(因此可用于大样本检验和推断)及 $\hat{\sigma}_n^2$ 的 Bootstrap 逼近问题，在适当条件下证明了其 Bootstrap 逼近成立. 薛留根、韩建国[16]则在 $\{X_i\}$ 为随机设计点列下，研究最近邻权函数形式的估计量，获得了 $\hat{\beta}_n$、$\hat{g}_n$ 和 $\hat{\sigma}_n^2$ 的渐近正态性和最优收敛速度.

1.3 小波光滑法估计性质的研究

柴根象、徐克军[17]首次将一种成功应用的非参数估计方法—小波光滑法引入模型(1)的研究，给出了 β 和 $g(.)$ 的小波估计 $\hat{\beta}_n$ 和 $\hat{g}_n$，得到了 $\hat{\beta}_n$ 和 $\hat{g}_n$ 的弱相合性及其偏差和方差的渐近性质，以及 $\hat{\beta}_n$ 的渐近正态性. 钱伟民、柴根象[18]假设 $\{X_i\}$ 为随机设计，$\{T_i\}$ 为常数点列，得到了 β 的小波估计有强收敛速度 $O(n^{-1/3}\log n)$，$g(.)$ 的小波估计有一致强收敛速度 $O(n^{-1/3}(\log n)^{1/2}M_n)$，其中 M_n 以任意慢的速度趋于 ∞. 文

献[18]中与核估计和最近邻估计比较可知，对模型(1)应用小波估计是成功的；同时，陈敬雨等[19]还得到了 β 和 $g(.)$ 的小波估计的弱收敛速度. 钱伟民等[20]在文献[17]基础上进一步构造了误差方差 σ^2 的估计量 $\hat{\sigma}_n^2$，得到了 $\hat{\sigma}_n^2$ 的渐近正态性，构造出 $Var(e_1^2)$ 的估计 $\hat{D}_n^2$，讨论了 $\hat{D}_n^2$ 的弱相合性和 $\hat{\sigma}_n^2$ 的 t 统计量的渐近正态性，同时还构造了 $\hat{\beta}_n$ 的一个可用于大样本检验的 χ^2 统计量. 刘元金等[21]假定 $\{X_i\}$ $\{T_i\}$ 都是固定设计点列，$\{e_i\}$ 具有公共未知密度 $f(t)$，基于残差构造了 $f(t)$ 的小波估计，并证明了该估计的弱相合、强相合、渐近正态性和收敛速度. 薛留根[22]则在 $\{X_i\}$ $\{T_i\}$ 都是固定设计点列且 $p=1$ 的情形下，结合小波光滑方法和随机加权方法，用随机加权法构造了 β 的小波估计的随机加权统计量，并证明了其逼近精度可达到 $o(n^{-1/2})$. 姜玉英等[23]假定模型(1)中 $\{e_i\}$ 具有异方差 $\sigma_i^2 = f(u_i)$，$\{u_i\}$ 与 $\{X_i\}$ $\{T_i\}$ 都是固定设计点列，利用小波方法估计出异方差 $\hat{\sigma}_i^2$，用它来修正模型的异方差性，然后重新构造 β 的小波估计，并证明了估计量 $\hat{\beta}_n$ 的弱收敛速度及强相合性.

1.4 M-估计性质的研究

施沛德在其92年的博士论文中较为系统的研究了模型(1)中当 $g(.)$ 取逐点多项式逼近及B-样条逼近时参数分量和非参数分量的M-估计的若干渐近性质. 施沛德、滕新东[24]在此基础上进一步构造了一个 χ^2 统计量用于检验线性假设 $H_0: A^T\beta = \beta^*$. 其后，唐亚宁、赵选民[25]假定误差 $\{e_i\}$ *i.i.d.* 且具有公共未知密度 $f(u)$，$\{(X_i,T_i),1\le i\le n\}$ 与 $\{e_i\}$ 相互独立，用分段多项式逼近 $g(.)$，基于M-估计的残差构造出 $f(u)$ 的估计，然后证明了这个估计依概率收敛、几乎处处收敛、几乎一致收敛并得到了其收敛速度. 薛留根[26]则用随机加权法给出了半参数回归模型中参数的随机加权M估计，并在一般的条件下证明了用随机加权统计量的分布逼近原估计量误差的分布的强有效性，给出了M估计的最优强收敛速度.

1.5 误差序列推广后的估计性质的研究

陈明华[27]在 $\{(X_i,T_i),1\le i\le n\}$ 为固定设计且 $p=1$ 的条件下将 *i.i.d.* 的误差 e_i 推广为独立序列，讨论了在完全与截尾样本时 β 和 $g(.)$ 的估计的强相合性；胡舒合[28]在 $\{(X_i,T_i),1\le i\le n\}$ 为固定设计且 $p=1$ 的条件下将 *i.i.d.* 的误差 e_i 推广为 φ-混合序列和局部广义高斯序列，分别得到了 β 和 $g(.)$ 的估计 $\hat{\beta}_n$ 和 $\hat{g}_n$ 的强相合性以及 $\hat{g}_n$ 的强一致相合性，但所给的条件多而且强，不易验证；吴本忠[29]则在相同假设条件下，将误差序列推广为 ρ-混合和 φ-混合序列，证明了 β 和 $g(.)$ 的加权最小二乘估计的 $\hat{\beta}_n$ 和 $\hat{g}_n$ 的强相合性及 $\hat{g}_n$ 的一致强相合性. 任哲、陈明华[30]进一步讨论了 $\{e_i\}$ 为NA序列时 β、$g(.)$ 和 σ^2 的估计 $\hat{\beta}_n$、$\hat{g}_n$ 和 $\hat{\sigma}_n^2$ 的强相合性. 闫在在等[31]在 $\{e_i\}$ 为鞅差序列和模型的其它条件相当弱的情形下，以 k_n-近邻权函数构造出 β 和 $g(.)$ 的加权最小二乘估计，并证明了其强相合性和 $p(p>1)$ 阶平均相合性. 胡宏昌、胡迪鹤[32]则在 $\{X_i\}$ 为随机设计 $\{T_i\}$ 为常数点列情形，将 $\{e_i\}$ 推广为鞅差序列，得到了 β 和 $g(.)$ 的小波估计的强相合性. 柴根象等又把 $\{e_i\}$ 推广为 α-混合序列，构造了模型(1)中 β、$g(.)$ 和 σ^2 的局部多项式估计，得到了估计的渐近正态性和收敛速度. 潘光明、胡舒合等[33]又将 $\{e_i\}$ 推广为 L^q-Mixingale 情形，研究了估计量的 q-阶 $(q>1)$ 平均相合性.

1.6 两种对模型(1)的推广研究

胡舒合[34]在 Fraiman R[35]研究的一类非参数回归模型的基础上研究了如下半参数回归模型:

$$Y^{(j)}(x_{in},t_{in}) = \beta' t_{in} + g(x_{in}) + e^{(j)}(x_{in}),\quad 1\le j\le m, 1\le i\le n$$

其中 $\{e^{(j)}(x_{in}), j\ge 1\}$ 为随机误差且 $Ee^{(j)}(x_{in}) = 0$. $\{(x_{in},t_{in}),1\le i\le n\}$ 为已知设计点列且 $x_{in}\in R^p$，$t_{in}\in R$，β 是未知待估参数，$g(.)$ 是定义在 R 上的未知函数.

胡舒合[34]在 $\{e^{(j)}(x_{in}), j\ge 1\}$ 为两种常见的弱误差结构即 α-混合、φ-混合(不要求同分布)的情况下获得了参数分量 β 的综合最小二乘法与非参数权函数法估计出的 $\hat{\beta}_n$ 的强相合性；胡舒合接着又在文献[36]讨论了 β 和 g 的估计量的性质，在简洁的条件下证明了它们具有强相合性与 $r(r>2)$ 阶平均相合性.

孙孝前、尤进红[37]考察了综列数据半参数回归模型:

$$y_{ij} = x_{ij}'\beta + g(t_{ij}) + \varepsilon_{ij}, i = 1,2,\cdots,k,\ j = 1,2,\cdots n_i, \sum_{i=1}^k n_i = n$$

其中 ε_{ij} 是随机误差且 $E\varepsilon_{ij} = 0, Var\varepsilon_{ij} = \sigma_i^2$. 文中提出了 β 的一个迭代加权偏样条最小二乘估计，并建

立了估计的渐近正态性；随后，樊明智等[38]基于最小二乘法和局部线性拟合的方法建立了 β 、$g(.)$ 和 σ^2 (ε_{ij} 同方差时)的估计，证明了估计量的弱相合性并通过模拟研究说明了该方法在有限样本情况下具有良好的性质；田萍、薛留根[39]则采用最小二乘法结合非参数权函数法给出了它们的强相合性. 另外，钱伟民、李静茹[40]研究了一类纵向污染数据(模型(3))的半参数回归模型，证明了 β 、$g(.)$ 及污染参数 v 的两阶段估计量的强相合性.

当然，对模型(1)的估计方法绝不是仅限于上面提到的几种，除研究的比较多的这几种外，惩罚最小二乘法、局部多项式拟合结合最小二乘法、局部线性拟合结合最小二乘法、Bayes 估计和对数似然法等等许多估计方法都有学者进行研究，并不断的有新的估计方法提出. 对模型(1)的变形和推广也还有多种情形，由于这些模型自身以及应用上的局限性，对它们的研究不多. 另外，在对误差序列不同相关性假设的推广研究中，几乎涉及到了所有已知的误差序列相关的形式，这也从一个侧面说明误差相关的结构对一个模型是至关重要的，在设定模型时需要慎重处理.

2　关于污染数据半参数回归模型的研究

考虑如下形式的半参数 EV 模型：

$$\begin{cases} Y_i = x_i^T \beta + g(T_i) + \varepsilon_i \\ X_i = x_i + u_i \end{cases} \quad (i = 1,2,\cdots,n) \tag{2}$$

其中为 *i.i.d.* 随机误差向量，$E(\varepsilon_i, u_i^T)^T = 0$，$Cov(\varepsilon_i, u_i^T)^T = \sigma^2 I_{p+1}$，$\{x_i\}$ 为 p 维不可观测向量；$\{(X_i, T_i), 1 \le i \le n\}$ 为可观测 *i.i.d.* 随机样本，且与 $\{(\varepsilon_i, u_i^T)^T, 1 \le i \le n\}$ 相互独立.

崔恒建[41]结合非参数的核权函数法和广义最小二乘法分别给出了 β 、$g(.)$ 和 σ^2 的估计 $\hat{\beta}_n$ 、$\hat{g}_n$ 和 $\hat{\sigma}_n^2$，在一些基本的假设条件下获得了 $\hat{\beta}_n$ 和 $\hat{\sigma}_n^2$ 的强相合性和渐近正态性，并得到了 $g^*{}_n$ 的最优收敛速度；马俊玲等[42]采用不同的方法证得了 $\hat{\beta}_n$ 、$g^*{}_n$ 和 $\hat{\sigma}_n^2$ 的强相合性和渐近正态性. 田茂再[43]则研究了误差方差 σ^2 的大样本性质，得到了其渐近正态性和一致收敛性. 李范良、何灿芝[44]采用两阶段估计方法估计 $\hat{\beta}_n$ 、$g^*{}_n$，并证明了 $\hat{\beta}_n$ 的强相合性和渐近正态性. 刘强等[45]则运用小波光滑结合广义最小二乘法得出 $\hat{\beta}_n$ 、$g^*{}_n$ 和 $\hat{\sigma}_n^2$ 的强相合性，一致强相合性以及 $\hat{\sigma}_n^2$ 的强收敛速度.

另外，薛留根[46]将模型(2)推广为非线性半参数 EV 模型，分别探讨了 $\hat{\beta}_n$ 、$g^*{}_n$ 和 $\hat{\sigma}_n^2$ 的估计和 $m(\tilde{V}, \beta) = E[f(X, \beta) | \tilde{V}]$ 的经验似然推断，获得了良好的结论.

也有研究者考察与模型(2)不同的下述模型

$$Y_i = X_i' \beta + g(T_i) + e_i, \quad 1 \le i \le n$$

其中，e_i 为 *i.i.d.* 随机误差 $Ee_i = 0, Vare_i = \sigma_1^2$. $\{(X_i, T_i), 1 \le i \le n\}$ 为 *i.i.d.* 随机设计或固定非随机设计点列且 $X_i \in R^p$，β 是未知待估参数，$g(.)$ 是定义在 R 上的未知函数，但 $Y_1, \cdots, Y_n$ 受到另一独立同分布随机变量序列 $u_1, \cdots, u_n$ 的污染，Y_j, u_j 相互独立，Y 的分布为 $F(y)$. 我们仅能观测到

$$Y_j^* = (1 - v) Y_j + v u_j, \; 0 < v < 1 \tag{3}$$

或者

$$F_\alpha(y) = (1 - \alpha) F_1(y) + \alpha F_2(y) \tag{4}$$

郑祖康等[47]比较系统的给出了两类线性污染数据模型以及存在数据截断情形下的模型参数估计方法；潘建敏[48]推广上述结果到半参数回归模型情形，利用非参数权函数法结合矩估计方法给出了模型(3)、(4)下 β 、$g(.)$ 及污染参数 v 的估计；陈明华[49]进一步证明了线性污染数据模型和模型(3)的估计量的强相合性与渐近正态性. 任哲、陈明华[50]则给出了模型(3)下 β 、$g(.)$ 及污染参数 v 的最小一乘估计，进而证明了估计的相合性和渐近正态性. 随后，刘丽萍[51]采用[44]类似的方法证明了 $\hat{\beta}_n$ 、$g^*{}_n$ 和污染系数估计 $\hat{v}$ 的两阶段估计的强相合性.

存在数据污染或者是数据的观测存在误差，这在经济中是非常常见的情形，如果它们确实存在，但我们在建模时忽略了它们，将会导致比较严重的后果. 如由于此时变量的内生性导致我们对参数的估计连最基本的相合性都不能满足，那么我们的模型将可能会导致误导性的结论. 因此对于这一类模型的讨论是重

要的而且是有意义的.

3 存在数据截断(数据删失)时半参数回归模型的研究

考虑半参数回归模型

$$Y_i = X_i'\beta + g(T_i) + e_i,\quad 1 \le i \le n$$

其中 e_i 为 *i.i.d.* 随机误差，$Ee_i = 0, Vare_i = \sigma^2$. $\{(X_i, T_i), 1 \le i \le n\}$ 为 *i.i.d.* 随机设计或固定非随机设计点列且 $X_i \in R^p$，β 是未知待估参数，$g(.)$ 是定义在 R 上的未知函数. 与模型(1)不同的是，我们仅能观察

$$Z_i = \min(Y_i, C_i), \delta_i = I(Y_i \le C_i), i = 1,2,\cdots,n \tag{5}$$

$C_1,\cdots,C_n$ 表示截断的随机变量列且独立同分布有共同的连续分布函数 G.

王启华[52]分别就截断分布已知与未知两种情形，结合非参数的核权函数法与最小二乘法定义了 β 和 $g(.)$ 的估计，并证明了它们的强相合性与 $p(p \ge 2)$-阶平均相合性(本文为国内该类模型研究的基础性文献). 薛留根[53]则结合最近邻权函数法与最小二乘法定义了 β 和 $g(.)$ 的估计，并得到了它们的渐近正态性及弱收敛速度，结果与非截尾情形的结果基本一致. 邱瑾[54]引入文献[12]的两阶段估计方法，就截断分布已知的情形， 定义了 β 和 $g(.)$ 的估计，并证明了它们的强相合性.

另外，一些学者还对模型(5)进行了如下一些推广研究. 秦更生[55]在模型 $Y_i = g(X_i'\beta) + e_i, 1 \le i \le n$ 的数据存在随机右删失时，构造了指标系数 β 的方向估计 $\hat{\delta}_n$，证明了 $\hat{\delta}_n$ 的 $\sqrt{n}$-相合性及渐近正态性. 苟列红[56]则给出了右删失左截断半参数模型下风险率函数的极大似然估计，获得了这些估计的渐近正态性、对数率和重对数率.

事实上，在我们收集数据的过程中，由于人力、财力等各种条件的限制，数据存在截断或删失是一种常态，因而，对于这类模型的研究也是非常必要的. 近年来，国外研究(特别是在微观计量模型的研究)存在数据截断或删失的模型的文献似乎越来越多，模型的构造和估计方法越来越复杂.

4 实证与应用研究

到目前为止，国内用半参数回归模型作实证研究的文献尚不多见. 黄四民、梁华[57]采用文献[4]提出的逐点多项式结合最小二乘法的估计方法，把模型(1)引入居民消费结构研究，建立一种新的消费结构分析框架，分析了我国未来近四分之一世纪的居民消费结构变动趋势，并与线性回归模型分析的结果比较，说明半参数方法的优越性；梁华、熊健[58]又接着将采用两种半参数估计方法与线性模型结果相比较，进一步说明半参数方法的优越性. 这两篇文章是国内已知最早的该类文献. 王一兵[59]将模型(1)应用于商品房价格指数的研究，讨论了模型(1)的估计方法、Hausman 检验及参数标准误差的 Bootstrap 计算，实证表明半参数回归分析的效果优于普通最小二乘法. 还有一些学者如叶阿忠、戴丽娜、韦红梅、冯春山等、许允彬、赵卫亚、姜爱平等分别将模型(1)用于研究我国进出口对通货膨胀的影响、我国保险有效需求、市场风险度量、农村居民消费行为、我国人口预测等实证研究，得到了类似的结论.

此外，武汉大学测绘学院的孙海燕、张松林、王新洲、潘雄、丁士俊、陶本藻、张昆等把半参数方法应用于测量的平差模型研究，几乎涉及到了半参数模型的所有估计方法并有所创新，对测量的平差模型研究的推进起到了明显的促进作用.

从上面文中可以看出，半参数回归模型结合线性回归模型和非参数回归模型，吸收了各自的优点，参数分量用于对确定性影响因素进行分析，而非参数分量部分用于对随机干扰部分的刻画，能够更好的描述现实世界. 然而，目前国内的研究主要在于理论方面，还存在着理论上的结果过于复杂，难以在实际中应用等缺陷. 可以预料，对于该模型的研究将持续深入下去，并得到越来越广泛的应用.

参考文献：

[1] ENGKLE R F, RICE J. Semiparametric estimates of the relation between weather and electricity sales[J]. JASA, 1986, 81. 310-320.

[2] PAUL SPECKMAN. Kernel smoothing in partial linear models[J]. J. Roy. Statist. Soc. Ser B. 1988, 50 413-436.

[3] RICE J. Convergence rates for partially splined models[J]. Statistics and Probability Letters 1986, 4: 203-208.

[4] CHEN H. Convergence rates for parametric components in a partly linear model[J]. Ann. Statist. 1988, 16: 136-146.

[5] 洪圣岩, 成　平. 半参数回归模型参数估计的收敛速度[J]. 应用概率统计, 1994(1): 62-71.
[6] 高集体, 洪圣岩, 梁　华, 等. 半参数回归模型研究的若干进展[J]. 应用概率统计, 1994(1): 98-104.
[7] 高集体, 洪圣岩, 梁　华. 部分线性模型中估计的收敛速度[J]. 数学学报, 1995(5): 658-669.
[8] 陈明华. 固定设计下半参数回归模型参数估计的收敛速度[J]. 应用概率统计, 1998(2):149-158.
[9] 陈明华. 固定设计下半参数回归模型的参数估计的 Bootstrap 逼近[J]. 数学物理学报, 1999(2): 121-129.
[10] 陈明华. 固定设计下半参数回归模型估计的相合性[J]. 高校应用数学学报：A 辑, 1998(3). 301-310.
[11] 胡舒合. 一类半参数回归模型的估计问题[J]. 数学物理学报, 1999. S1: 541-549.
[12] 柴根象, 孙　平, 蒋泽云. 半参数回归模型的二阶段估计[J]. 应用数学学报, 1995(3). 353-363.
[13] 钱伟民, 柴根象. 一类半参数回归模型二阶段估计的渐近理论[J]. 同济大学学报：自然科学版, 1998(1): 77-82.
[14] 钱伟民. 半参数回归模型的误差方差估计的注记[J]. 同济大学学报：自然科学版, 1998(1): 83-86.
[15] 钱伟民. 半参数回归模型误差方差估计的 Bootstrap 逼近[J]. 数理统计与应用概率, 1998(4): 283-294.
[16] 薛留根, 韩建国. 半参数回归模型中二阶段估计的渐近性质[J]. 高校应用数学学报：A 辑, 2001, 16: 87-94.
[17] 柴根象, 徐克军. 半参数回归的线性小波光滑[J]. 应用概率统计, 1999, 15: 97-106.
[18] 钱伟民, 柴根象. 半参数回归模型小波估计的强逼近[J]. 中国科学：A 辑, 1999(3): 233-241.
[19] 陈敬雨, 钱伟民. 半参数回归模型小波估计的弱相合速度[J]. 同济大学学报：自然科学版, 1999, 27: 708-713.
[20] 钱伟民, 柴根象, 蒋凤瑛. 半参数回归模型的误差方差的小波估计[J]. 数学年刊：A 辑, 2000, 3: 341-350.
[21] 刘元金, 柴根象. 半参数模型误差分布小波估计的渐近理论[J]. 同济大学学报：自然科学版, 1999, 27: 463-468.
[22] 薛留根. 半参数回归模型中小波估计的随机加权逼近速度[J]. 应用数学学报, 2003, 26: 11-26.
[23] 姜玉英, 刘　强, 吴可法,. 半参数回归模型小波估计的相合性[J]. 福州大学学报：自然科学版, 2006, 34. 798-803.
[24] 施沛德, 滕新东. 固定设计点的半参数回归模型的 M 估计的渐进分布[J]. 数学进展, 1999, 28: 447-462.
[25] 唐亚宁, 赵选民. 半参数回归模型的误差分布的估计的大样本性质[J]. 纯粹数学与应用数学, 2000(1): 40-48.
[26] 薛留根. 半参数回归模型中随机加权 M 估计的强逼近[J]. 应用数学学报, 2002, 25: 591-604.
[27] 陈明华. 完全与截尾样本时半参数回归模型估计的强相合性[J]. 数学物理学报, 1999, 19: 501-506.
[28] 胡舒合. 固定设计下半参数回归模型估计的强相合性[J]. 数学学报, 1994(3): 393-401.
[29] 吴本忠. 混合误差半参数回归模型估计的强相合性[J]. 应用数学, 1998, 11(3): 27-31.
[30] 任　哲, 陈明华. NA 样本下半参数回归模型估计的强相合性[J]. 高校应用数学学报 A 辑, 2000, 15(4): 467-474.
[31] 闫在在, 吴伟志, 聂赞坎. 半参数回归模型的近邻估计—缺差误差序列情形[J].应用概率统计, 2001,17(1): 44-51.
[32] 胡宏昌, 胡迪鹤,. 半参数回归模型小波估计的强相合性[J]. 数学学报, 2006, 49(6):1417-1425.
[33] 潘光明, 胡舒合, 方利民, 等. 半参数回归模型估计的平均相合性[J]. 数学物理学报, 2003(5): 598-606.
[34] 胡舒合, 梅门昌. 弱误差结构下非参数、半参数回归模型[J]. 数学物理学报, 1996(1): 23-30.
[35] FRAIMAN R, INBARREN P. Nonparametric regression estimation in models with weak error's structure[J]. J Mult Anal, 1991, 37(2): 180-196.
[36] 胡舒合. 一类新的半参数回归模型中的相合估计[J]. 数学学报, 1997, 40(4): 527-537.
[37] 孙孝前, 尤进红. 纵向数据半参数进模中的迭代加权偏样条最小二乘估计[J]. 中国科学：A 辑, 2003, 33(5): 470-481.
[38] 樊明智, 王芬玲, 郭辉　纵向数据半参数回归模型的最小二乘局部线性估计[J]. 数理统计与管理, 2006, 25(2):170-174.
[39] 山　萍, 薛留根. 纵向数据半参数回归模型估计的强相合性[J]. 工程数学学报, 2006, 23(2): 369-373.
[40] 钱伟民, 李静茹. 纵向污染数据半参数回归模型中的强相合估计[J]. 同济大学学报：自然科学版, 2006, 34(8): 1105-1110.
[41] 崔恒建. 半参数 EV 模型的参数估计理论[J]. 科学通报, 1995, 16: 1444-1448.
[42] 马俊玲, 吴可法, 聂赞坎. 关于半参数函数关系模型的渐近正态性[J]. 数学年刊：A 辑, 2002(4): 475-482.
[43] 山茂再. 半参数模型中误差方差的大样本性质[J]. 湖南大学学报：自然科学版, 2000, 27(6): 4-10.
[44] 李范良, 何灿芝. 半参数 EV 模型参数的二阶段估计[J]. 经济数学, 2002(1): 50-54.
[45] 刘　强, 姜玉英, 吴可法. 半参数变址含误差函数关系模型的小波估计[J]. 应用数学学报, 2005, 28(2): 297-309.
[46] 薛留根. 核实数据下非线性半参数 EV 模型的经验似然推断[J]. 数学学报, 2006, 49(1): 146-156.
[47] 邓祖康. 关于两类污染数据回归分析的参数估计[J]. 高校应用数学学报, 1996(1): 31-40.
[48] 潘建敏. 污染数据半参数回归模型的估计方法[J]. 工程数学学报, 1997(3): 81-85.
[49] 陈明华. 污染数据回归分析中估计的强相合性[J]. 应用概率统计, 1998, 14: 73-78.
[50] 任　哲, 陈明华. 污染数据回归分析中参数的最小一乘估计[J]. 应用概率统计, 2000, 16: 262-268.
[51] 刘丽萍. 污染数据半参数回归模型中的强相合估计[J]. 同济大学学报：自然科学版, 2004, 32(6): 832-836.
[52] 王启华. 随机截断下半参数回归模型中的相合估计[J]. 中国科学：A 辑, 1995, 25(8): 819-833.
[53] 薛留根. 随机删失下半参数回归模型的估计理论[J]. 数学年刊：A 辑, 1999(6): 745-754.
[54] 邱　瑾. 删失场合半参数回归模型的二阶段估计[J]. 高校应用数学学报：A 辑, 1998, 13(3): 282-290.
[55] 秦更生. 随机删失下指标系数的半参数估计[J]. 四川大学学报：自然科学版, 1995, 32(1): 10-15.
[56] 苟列红. 左截断右删失数据下半参数模型风险率函数估计[J]. 应用数学学报, 2005, 28(4): 675-689.
[57] 黄四民, 梁　华. 用半参数部分线性模型分析居民消费结构[J]. 数量经济技术经济研究, 1994, 10: 33-39.
[58] 梁　华, 熊　健. 再论半参数部分线性模型在居民消费结构分析中的应用[J]. 数理统计与管理, 1995(6): 9-17.
[59] 王一兵. 半参数回归模型在商品房价格指数中的应用研究[J]. 统计研究, 2005(4): 25-30.

（下转第 17 页）

$$E\left\{\Sigma_{\bar{x}_k} A_k^T P_k \Psi(V_k)\Delta\bar{x}_k^T\right\} = \Sigma_{\bar{x}_k} A_k^T P_k \cdot E\Psi'(V_k) A_k \Sigma_{\bar{x}_k} \tag{36}$$

将(36)式代入(34)式得估值的验后协方差阵为：

$$\Sigma_{\hat{x}_k} = \Sigma_{\bar{x}_k} - 2\Sigma_{\bar{x}_k} A_k^T P_k E\Psi'(V_k) A_k \Sigma_{\bar{x}_k} + \Sigma_{\bar{x}_k} A_k^T P_k E\left\{\Psi(V_k)\Psi^T(V_k)\right\} \cdot P_k A_k \Sigma_{\bar{x}_k}$$

参考文献：

[1] 张志方，孙常胜. 线性控制系统教程[M]. 北京：科技出版社，1993

[2] 陈希儒，赵林城. 线性模型中的 M 方法[M]. 上海：上海科学技术出版社，1996.

[3] 庄常陵. 混合型正态分布的抗差最小二乘估计[J]. 襄樊学院学报，2007(2)：5-8.

[4] 庄常陵. 观测值对最小二乘估计的影响[J]. 绍兴文理学院学报，2006(9)：30-32.

[5] 周江文，黄幼才，杨元喜，等. 抗差最小二乘法[M]. 武汉：华中理工大学出版社，1997.

Robust Estimation on Observation Value Obeying Pollution Normal Distribution

ZHUANG Chang-ling

(Department of Mathematics, Xiangfan University, Xiangfan 441053, China)

Abstract: Robust estimation on observation value obeying pollution normal distribution is discussed according to the equivalent weight principle. And the author uses the influence function of the estimation to demonstrate that the given estimation can reduce the disturbance of gross error.

Key words: Robust Estimation; Polluted Normal Distribution; Influence Function.

(上接第13页)

Improvements of the Investigation for Semiparametric Regression Models

WANG Cheng-yong

(Department of Mathematics, Xiangfan University, Hubei Xiangfan 441053)

Abstract: The investigation for semiparametric regression models was one of the hot points in statistics research nowadays. Semiparametric regression models combined the advantages of linear regression models and the nonparametric models, it both has important meaning in theory research and applying research work. In this paper, we survey the recent research results of statisticians at home on semiparametric regression models, try to depict a picture of how the research work develop step by step.

Key words: Semiparametric regression models; Consistency; Convergence rate