doi:10.16055/j.issn.1672-058X.2015.0009.003

部分线性回归模型中改进的差分估计及 SCAD

郭 雪 梅

(重庆大学 数学与统计学院,重庆 401331)

摘 要:考虑了部分线性回归模型中回归参数向量估计的问题,提出了具有更好性质的压缩差分估计,并且将 SCAD 惩罚函数运用到模型中得到 SCAD 估计,然后通过 Monte Carlo 模拟了压缩差分估计和 SCAD 估计的相关结果,并对它们之间的优劣进行了比较.

关键词:差分估计;局部线性回归模型;SCAD;压缩估计

中图分类号:O212 **文献标志码**:A **文章编号**:1672-058X(2015)09-0010-06

1 模型简介

考虑半参数回归模型:

$$Y_i = X_i^{\mathrm{T}} = \beta + f(T_i) + \varepsilon_i, i = 1,2,\cdots,n \tag{1}$$

其中 Y_i 是响应变量,受到(X_i,T_i)的影响,$X_i=(X_{i1},\cdots,X_{ip}{}^{\mathrm{T}})$是 p 维随机变量向量,T_i 是一个随机变量,$\beta=(\beta_1,\beta_2,\cdots,\beta_p)^{\mathrm{T}}$ 是未知的 p 维参数变量向量,$f(\cdot)$是一个未知的平滑函数,ε_i 是随机误差,且有 $E[\varepsilon_i|X_i,T_i]=0$,$\mathrm{Var}[\varepsilon_i|X_i,T_i]=\sigma^2$.

模型(1)是由 Engle 等提出的,主要是为了研究气候对电力需求的影响.半参数模型中参数部分带宽的选择不是一件容易的事,运用差分估计的方法是由 Hall 等(1990)[1]提出来的.压缩估计最初是由 C.Stein[2]在 1956 年提出的.关于压缩估计的方法有很多,其中 LAASO,SCAD 等是压缩估计中标志性的方法.虽然基于差分的估计和压缩估计得到了很大的发展,但是此文是第一个在部分回归模型中运用差分来提高压缩估计值的性质,并且和 SCAD 惩罚估计进行比较.

2 符号及假设

为了更好地介绍差分估计,列举了一些必要的符号以及假设,如下:

设 $h(t)=E(X_i|T_i=t)$,$s(t)=E(X_i^{\mathrm{T}}X_i|T_i=t)$,$\Lambda^{\alpha}(M)$是 Lipschitz 球,定义如下:对所有的$\{0\leqslant x,y\leqslant 1,k=0,\cdots,[a]-1\}$,有

$$\Lambda^{\alpha}(M)=\{g:|g^{(k)}(x)|\leqslant M,|g^{[\alpha]}(x)-g^{[\alpha]}(y)|\leqslant M|x-y|^{\alpha-[\alpha]},0\leqslant x,y\leqslant 1,k=0,\cdots,[\alpha]-1\}$$

其中$[\alpha]$是取整,$g^{[\alpha]}(x)$表示 $g(x)$的$[\alpha]$次导.

假设 1 ε_i 和(X_i,T_i)是相互独立的,对于,$\alpha>0$,$\gamma>0$,有

$$f(t)\in\boldsymbol{\Lambda}^{\alpha}(M_f)h(t)\in\boldsymbol{\Lambda}^{\gamma}(M_h)$$

收稿日期:2015-01-11;**修回日期**:2015-03-20.

作者简介:郭雪梅(1990-),女,重庆人,硕士研究生,从事参数估计和变量选择研究.

假设 2　$m\to\infty$,$m/n\to 0$,同时有 $\sum_{k=1}^{m} c_k^2 = O(m^{-1})$ 成立,其中 $C_K = \sum_{i=1}^{m+1-k} d_i d_{i+k}$.

假设 3　T_i 的边际密度函数是 Lipschitz 连续的,并且不为 0.

设 $T_{(1)} \leqslant T_{(2)} \leqslant \cdots \leqslant T_{(n)}$ 是 T_i 的顺序统计量,$(X_{(1)}, Y_{(1)}), \cdots, (X_{(n)}, Y_{(n)})$ 是 $(X_{(i)}, Y_{(i)})$ 所对应的值,注意 $(X_{(i)}, Y_{(i)})$ 不是顺序统计量,但是 $(X_{(i)}, Y_{(i)})$ 与 T_i 是一致的,那么模型(1)这可以重新写成如下的形式:

$$Y_{(i)} = X_{(i)}^{\mathrm{T}}\beta + f(T_{(i)}) + \varepsilon_{(i)}, i = 1, 2, \cdots, n \tag{2}$$

设向量 $d = (d_1, \cdots, d_{m+1})^{\mathrm{T}}$,$m$ 是差分的阶数,$d_1, \cdots, d_{m+1}$ 满足如下的条件:

$$\sum_{J=1}^{m+1} d_j = 0, \sum_{j=1}^{m+1} d_j^2 = 1 \tag{3}$$

定义如下 $(n-m)\times n$ 维的差分矩阵:

$$\boldsymbol{D} = \begin{pmatrix} d_0 & d_1 & \cdots & d_m & 0 & 0 & \cdots & 0 \\ 0 & d_0 & d_1 & \vdots & d_m & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & d_0 & d_1 & \cdots & d_m \end{pmatrix}$$

矩阵内的元素满足式(3).把这个差分序列运用到式(2)中,得到

$$\sum_{r=1}^{m+1} d_r Y_{(i+r-1)} = \sum_{r=1}^{m+1} d_r X_{(i+r-1)}^{\mathrm{T}}\beta + \sum_{r=1}^{m+1} d_r f(T_{(i+r-1)}) + \sum_{r=1}^{m+1} d_r \varepsilon_{(i+r-1)}$$
$$i = 1, 2, \cdots, n-m \tag{4}$$

通过假设 2 和假设 3,得到 $\sum_{r=1}^{m+1} d_r f(T_{(i+r-1)}) = o_p(1)$,从而式(4)可以表示为

$$\sum_{r=1}^{m+1} d_r Y_{(i+r-1)} \approx \sum_{r=1}^{m+1} d_r X_{(i+r-1)}^{\mathrm{T}}\beta + \sum_{r=1}^{m+1} d_r \varepsilon_{(i+r-1)}$$
$$i = 1, 2, \cdots, n-m \tag{5}$$

设 $\tilde{Y}_i = \sum_{r=1}^{m+1} d_r Y_{(i+r-1)}$, $\tilde{X}_i = \sum_{r=1}^{m+1} d_r X_{(i+r-1)}$, $\tilde{\varepsilon}_i = \sum_{r=1}^{m+1} d_r \varepsilon_{(i+r-1)}$,那么式(5)就可以表示成如下的形式:

$$\tilde{Y}_i \approx \tilde{X}_i^{\mathrm{T}}\beta + \tilde{\varepsilon}_i, i = 1, 2, \cdots, n-m \tag{6}$$

由于式(6)类似于线性模型,所以式(6)的估计值可以表示为

$$\tilde{\beta} = (\tilde{X}^{\mathrm{T}}\tilde{X})^{-1}\tilde{X}^{\mathrm{T}}\tilde{Y}_i^{\mathrm{T}} \tag{7}$$

其中 $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_{n-m})$, $\tilde{Y} = (\tilde{Y}_1, \tilde{Y}_2, \cdots, \tilde{Y}_{n-m})$.

3　差分压缩估计的构造

对于差分估计 $\tilde{\beta}$,Wang 等(2011)[3] 证明了在假设 1-3 成立,且有 $\alpha+\gamma>1/2$, $s(t)>0$ 时,$\tilde{\beta}$ 具有渐进正态性,也就是说

$$\sqrt{n}(\tilde{\beta} - \beta) \to N(0, \sigma^2 \textstyle\sum^{-1}) \tag{8}$$

其中, $\hat{\sum}_n = \frac{1}{n} X^{\boldsymbol{D}\tau} X^{\boldsymbol{D}} \xrightarrow{\text{a.s.}} \sum = E\{(X_i - E(X_i \mid T_i))(X_i - E(X_i \mid T_i))^{\mathrm{T}}\}$,且 $\sum$ 是可逆的.

假设有 β 的先验信息,使得 $\boldsymbol{R}\beta = r$ 成立,其中 $\boldsymbol{R}$ 是 $q\times p$ 的矩阵,其秩为 q,r 是 $q\times 1$ 的向量,$q \leqslant p$.设 $\hat{\beta}$ 是 β

的约束差分估计,当 $\boldsymbol{R}\beta=r$ 成立时,$\hat{\beta}$ 可以写成如下的形式:

$$\hat{\beta}=\tilde{\beta}-(\hat{\sum}_n)^{-1}\boldsymbol{R}^{\mathrm{T}}[\boldsymbol{R}(\hat{\sum}_n)^{-1}\boldsymbol{R}^{\mathrm{T}}]^{-1}(\boldsymbol{R}\tilde{\beta}-r) \tag{9}$$

设原假设 $\mathrm{H}_0:\boldsymbol{R}\beta=r$(备择假设:$\mathrm{H}_1:\boldsymbol{R}\beta\neq r$),提出如下的检验统计量:

$$L_n=\frac{n}{\hat{\sigma}^2}(\boldsymbol{R}\tilde{\beta}-r)^{\mathrm{T}}(\boldsymbol{R}(\hat{\sum}_n)^{-1}\boldsymbol{R}^{\mathrm{T}})^{-1}(\boldsymbol{R}\tilde{\beta}-r) \tag{10}$$

其中 $\hat{\sigma}^2=\frac{1}{n-m}\sum_{i=1}^{n-m}(\tilde{Y}-\tilde{X}_i^{\mathrm{T}}\tilde{\beta})^2$,并且在原假设 H_0 成立的情况下有 $L_n\equiv x_q^2$.通过参考 Ahmed(2012)[4] 和 Saleh(2006)[5] 这两篇文章,定义压缩差分估计(SDE)如下:

$$\hat{\hat{\beta}}=\tilde{\beta}-(q-2)(\tilde{\beta}-\hat{\beta})L_n^{-1},q\geqslant 3 \tag{11}$$

上述估计值是由无约束估计和约束差分估计的权重组成的,这个权重是原假设 H_0 中的检验统计量的函数.但是该估计值存在一些问题,当 $L_n\leqslant q-2$,$\hat{\hat{\beta}}$ 可能会使得 $\tilde{\beta}$ 朝着 $\hat{\beta}$,从而被过分地压缩,使得 $\tilde{\beta}$ 的符号出现一些问题.为了克服这个过度压缩的可能性,正压缩差分估计可以表示如下:

$$\hat{\hat{\beta}}^{+}=\hat{\beta}-[1-(q-2)L_n^{-1}]I(L_n>q-2)(\tilde{\beta}-\hat{\beta}),q\geqslant 3 \tag{12}$$

其中 I 是集合的示性函数.这个正准则的估计可以有效地控制 $\hat{\hat{\beta}}$ 过度压缩的情况.

4 LASSO 思想在模型(6)中的运用

给定惩罚函数 $p_\lambda(.)$ 和正则化参数 λ,那么模型(6)运用变量选择方法对参数求解的过程可以表示如下:

$$\varphi(\beta)=\frac{1}{2}(\tilde{Y}-\tilde{X}\beta)^{\mathrm{T}}(\tilde{Y}-\tilde{X}\beta)+np_\lambda(|\beta|)$$

LASSO 估计值可以表示为如下:

$$\hat{\beta}^{\mathrm{lasso}}=\arg\min_{\beta}\left\{\frac{1}{2}\sum_{i=1}^{n-m}(\tilde{Y}_i-\sum_{j=1}^{p}\tilde{X}_j^{\mathrm{T}}\beta_j)^2+n\lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{13}$$

LASSO 估计没有哲人性质,于是考虑 SCAD 估计.SCAD 估计是在 LASSO 估计的基础上发展的.具体可以参考文献 Fan 和 Li[6].将 SCAD 的惩罚函数运用到模型(6)中,可以得到 SCAD 估计如下:

$$\hat{\beta}^{\mathrm{SCAD}}=\arg\min_{\beta}\left\{\frac{1}{2}\sum_{i=1}^{n-m}(\tilde{Y}_i-\sum_{j=1}^{p}\tilde{X}_j^{\mathrm{T}}\beta_j)^2+n\sum_{j=1}^{p}p_\lambda(|\beta_j|)\right\} \tag{14}$$

其中 $p_\lambda(.)$ 是 SCAD 的惩罚函数,a 和 λ 是参数.

$$p_\lambda(|\beta|)=\begin{cases}\lambda|\beta|,|\beta|\leqslant\lambda\\ \dfrac{(a^2-1)\lambda^2-(|\beta|-a\lambda)^2}{2(a-1)},\lambda<|\beta|\leqslant a\lambda\\ \dfrac{(a+1)\lambda^2}{2},|\beta|>a\lambda\end{cases} \tag{15}$$

Fan 和 Li 证明了在各种情况下,当 $a\equiv 3.7$ 时,估计值有较好的性质,在第 5 部分的模拟中,也选择 $a\equiv 3.7$,与此同时,运用 GCV 准则来选择正则化参数参数 λ.

5　模　拟

在这一部分,将通过 Monte Carlo 模拟来比较压缩估计和 SCAD 惩罚函数估计.数据都是从式(14)线性回归模型中得到的.

$$Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + \cdots + X_{ip}\beta_p + f(T_i) + \varepsilon_i \tag{16}$$

其中 $T_i \sim$ i. i. d. $U(0,1)$, $X_{ij} \sim$ i. i. d. $N(0,1)$, $j=1,2,\cdots,p$, $\varepsilon_i \sim$ i. i. d. $N(0,1)$. 对于非参数部分,选择 $f(T_i)=\sin(3\pi T_i)$.

考虑在约束条件 $\boldsymbol{R}\beta=0$, $\boldsymbol{R}=(0_{p_2\times p_1}, I_{p_2})$ 下的压缩估计.那么假设检验的原假设为 $H_0:\beta_j=0, j=p_1+1, p_1+2,\cdots,p_1+p_2$, $p=p_1+p_2$.将回归系数分为两部分 $\beta=(\beta_1,\beta_2)=(\beta_1,0)$,其中 $\beta_1=(0.5,1,1.5,2)$,并且设权重矩阵 $\boldsymbol{W}=\boldsymbol{I}$.通过均方误差 MSE 来衡量 β 的估计值的风险.在计算 MSE 之后,会计算 $\hat{\beta}$, $\widehat{\beta}$, $\widehat{\beta}^{+}$ 与 $\tilde{\beta}$ 的相对有效性,公式为

$$\mathrm{EFF}(\hat{\beta}^{*}) = \frac{\mathrm{MSE}(\tilde{\beta})}{\mathrm{MSE}(\hat{\beta}^{*})} \tag{17}$$

其中 $\hat{\beta}^{*}$ 是 $\hat{\beta}$, $\widehat{\beta}$, $\widehat{\beta}^{+}$ 中的一个.

模拟过程中,考虑 $\delta=0$ 和 $\delta\neq0$ 两种情况.在 $\delta\neq0$ 的情况下,选择 $\delta=0,0.2,0.4,0.6,0.8,1.0,1.2,1.4,1.6$,SCAD 的正则化参数 λ 通过 GCV 准则来选择.$\hat{\beta}^{\mathrm{SCAD}}$ 的相对有效性可以通过数值结果来表示.为了求解这些估计的相对有效性,考虑 $n=30,50,100,200$, $p_1=4$, $p_2=3,5,7,9,11$.重复这样的模拟 3 000 次.

情况 1　$H_0:\boldsymbol{R}\beta=0$ 或者 $\delta=0$.

在不同的 n 和 $\delta=0$ 下,给出 SCAD 惩罚估计和压缩差分估计与 $\tilde{\beta}$ 的相对有效性,且 $(p_1,p_2)=\{(4,3),(4,5),(4,7),(4,9),(4,11)\}$,具体的结果会在表 1 中展示出来.

从表 1 中可以得出一般情况下,所有的估计值都优于 $\tilde{\beta}$,并且 $\hat{\beta}$ 是最好的.当 p_2 的维数增加,所有估计值的相对有效性就会增加,这就意味着当 β 的维数越来越大时,$\tilde{\beta}$ 的结果越来越差.当 n 比较小的时候,压缩差分估计的效果比 SCAD 惩罚估计的效果好.比如,从表 1 可以发现,当 $n=30$ 时,压缩估计的相对有效性高于 SCAD 惩罚估计.但另一方面,当 n 逐渐变大时,除了压缩差分估计,SCAD 估计的相对有效性是最大的.这可能是由于 SCAD 估计的哲人性质引起的.

表 1　$\hat{\beta}$, $\widehat{\beta}$, $\widehat{\beta}^{+}$, $\hat{\beta}^{\mathrm{SCAD}}$ 与 $\tilde{\beta}$ 的相对有效性($p_1=4$, $\delta=0$, $m=3$)

n	p_2	$\hat{\beta}$	$\widehat{\beta}$	$\widehat{\beta}^{+}$	$\hat{\beta}^{\mathrm{SCAD}}$
30	3	2.047	1.191	1.270	1.228
	5	2.965	1.613	1.761	1.550
	7	4.231	2.100	2.311	2.067
	9	5.921	2.778	2.998	2.706
	11	8.207	3.488	3.809	3.548

续表 1

n	p_2	$\hat{\beta}$	$\widehat{\beta}$	$\widehat{\beta}^{+}$	$\hat{\beta}^{\mathrm{SCAD}}$
	3	1.897	1.196	1.247	1.313
	5	2.606	1.580	1.701	1.606
50	7	3.362	1.966	2.137	1.987
	9	4.217	2.381	2.614	2.446
	11	5.148	2.816	3.089	2.861
	3	1.819	1.181	1.239	1.484
	5	2.429	1.507	1.632	1.807
100	7	2.980	1.865	2.001	2.112
	9	3.643	2.225	2.407	2.521
	11	4.309	2.592	2.817	2.862
	3	1.753	1.122	1.223	1.639
	5	2.265	1.501	1.601	2.042
200	7	2.871	1.822	1.956	2.468
	9	3.454	2.186	2.351	2.967
	11	4.291	2.508	2.691	3.378

情况 2　原假设 $H_0:\boldsymbol{R}\beta=0$ 不成立,或者 $\delta\neq 0$.表 2 为 $n=50,m=30$ 时,SCAD 惩罚估计和压缩差分别估计与 $\tilde{\beta}$ 的相对有效性.从表 2 中,得结论:当 $\delta=0$ 或者在 δ 在 0 附近时,$\hat{\beta}$ 的相对有效性是最好的.但随着 δ 越远离 0,$\hat{\beta}$ 的相对有效性越来越小,SCAD 惩罚估计的相对有效性是最大的.在 p_2 一定时,压缩估计的相对有效性随着 δ 的增大而减小,SCAD 惩罚估计的相对有效性有稍许的变动.当 δ 不变时,压缩估计的相对有效性和 SCAD 惩罚估计的有效性是 p_2 的增函数.同时也发现 $\widehat{\beta}^{+}$ 的相对有效性总是优于 $\widehat{\beta}$ 的相对有效性.

表 2　$\hat{\beta},\widehat{\beta},\widehat{\beta}^{+},\hat{\beta}^{\mathrm{SCAD}}$ 与 $\tilde{\beta}$ 的相对有效性($n=50,m=3$)

p_2	δ	$\hat{\beta}$	$\widehat{\beta}$	$\widehat{\beta}^{+}$	$\hat{\beta}^{\mathrm{SCAD}}$
	0.0	1.870	1.176	1.249	1.297
	0.2	1.860	1.186	1.237	1.305
	0.4	1.732	1.146	1.215	1.291
	0.6	1.584	1.162	1.189	1.330
3	0.8	1.377	1.126	1.154	1.312
	1.0	1.190	1.079	1.145	1.304
	1.2	1.018	1.078	1.086	1.306
	1.4	0.880	1.058	1.067	1.301
	1.6	0.765	1.055	1.058	1.295
	0.0	3.330	1.962	2.136	1.977
	0.2	3.170	1.941	2.097	1.998
	0.4	2.440	1.787	1.909	1.978
	0.6	2.225	1.664	1.741	1.981
7	0.8	1.813	1.542	1.594	2.036
	1.0	1.421	1.413	1.444	1.965
	1.2	1.120	1.331	1.346	2.010
	1.4	0.906	1.265	1.272	2.019
	1.6	0.744	1.206	1.208	2.003

续表 2

p_2	δ	$\hat{\beta}$	$\hat{\beta}$	$\hat{\beta}^{+}$	$\hat{\beta}^{SCAD}$
	0.0	5.173	2.830	3.093	2.873
	0.2	4.724	2.718	2.960	2.935
	0.4	3.860	2.465	2.651	2.902
	0.6	2.962	2.180	2.263	2.893
11	0.8	2.215	1.912	1.956	2.912
	1.0	1.683	1.692	1.710	2.958
	1.2	1.265	1.525	1.534	2.885
	1.4	0.995	1.415	1.418	2.898
	1.6	0.791	1.335	1.335	2.895

6 结　论

基于部分线性回归模型,提出了压缩差分估计,并将改进的压缩差分估计与 SCAD 惩罚估计进行了比较,从第 5 部分的模拟中可以看到,压缩差分估计和 SCAD 惩罚估计都具有较好的性质.与传统的差分估计相比,压缩差分估计具有明显的优越性,并且 SCAD 惩罚估计在样本量越大的时候体现出的优越性也越明显.对于模型(1),如果选择差分技术,则可以有效避免带宽的选择.因此,对于部分线性回归模型,此处提出的差分压缩估计是一个较为有用并且简单的方法.

参考文献:

[1] HALL P, KAY J W, TITTERINGTON D M. Asymptotically Optimal Difference-based Estimation of Variance in Nonparametric Regression[J]. Biometrika, 1990(77):521-528

[2] STEIN C. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution[J]. Proceedings of the Third Berkeley Symposium on Mathematics and Statistical Probability, 1956(1):197-206

[3] WANG L, BROWN L D, CAI T. A Difference Based Approach to the Semi-parametric Partial Linear Model[J]. Electronic Journal of Statistics, 2011(5):619-641

[4] AHMED S E, HOSSAIN S, DOKSUM K A. Lasso and Shrinkage Estimation in Weibull Censored Regression Models[J]. Journal of Statistical Panning and Inference, 2012(142):1273-1284

[5] SALEH A K, MD E. Theory of Preliminary Test and Stein-type Estimation with Applications[M]. New York: John Wiley & Sons, 2006

[6] FAN J, LI R. Variable Selection via Non-concave Penalized Likelihood and its Oracle Properties[J]. Journal of the American Statistical Association, 2001(96):1348-1360

SCAD and Improved Difference Estimation of Partially Linear Regression Model

GUO Xue-mei

(School of Mathematics and Statistics, Chongqing University, Chongqing 401331, China)

Abstract: This paper considers the problem of estimating a vector of regression parameters in partially regression model. Differential shrinkage estimator with better properties is proposed. The SCAD penalties are applied to the partially linear regression model. Results of the differential shrinkage estimators and SCAD estimators are simulated by Monte Carlo and compared in strength and weakness.

Key words: difference estimation; partially linear regression model; SCAD; shrinkage estimators