

# 部分线性可加模型的随机约束岭估计

李 静<sup>1</sup>, 李雪艳<sup>2</sup>, 安佰玲<sup>3</sup>

<sup>1</sup>中国劳动关系学院应用技术学院, 北京

<sup>2</sup>中国民航大学理学院, 天津

<sup>3</sup>淮北师范大学数学科学学院, 安徽 淮北

收稿日期: 2022年11月22日; 录用日期: 2022年12月12日; 发布日期: 2022年12月26日

## 摘 要

本文研究了部分线性可加模型这类半参数模型在线性部分自变量存在多重共线性, 同时还附加有随机约束条件时的估计问题。基于针对半参数模型的Profile最小二乘技术, 处理多重共线性问题的岭估计方法以及针对随机约束的混合估计技术, 构造了参数分量的随机约束岭估计, 给出了估计量的偏与协方差, 并给出了所提估计量的渐近性质。最后通过数值模拟验证了所提估计方法的表现。

## 关键词

部分线性可加模型, 随机线性约束, Profile最小二乘估计, 岭估计

# A Stochastic Restricted Ridge Estimator in Partially Linear Additive Models

Jing Li<sup>1</sup>, Xueyan Li<sup>2</sup>, Bailing An<sup>3</sup>

<sup>1</sup>School of Applied Technology, China University of Labor Relations, Beijing

<sup>2</sup>School of Science, Civil Aviation University of China, Tianjin

<sup>3</sup>School of Mathematical Sciences, Huaibei Normal University, Huaibei Anhui

Received: Nov. 22<sup>nd</sup>, 2022; accepted: Dec. 12<sup>th</sup>, 2022; published: Dec. 26<sup>th</sup>, 2022

## Abstract

This paper studies the estimation of a partially linear additive model when there are stochastic linear restrictions on the parameter components and multicollinearity problem exists, simultaneously. Based on the profile least square method of semiparametric model, ridge estimation approach for multicollinearity problem and mixed estimation technique for stochastic restrictions, a stochastic ridge estimator for the parametric components was proposed and its bias and covariance

were provided. Meanwhile, the asymptotic distribution of this proposed estimator has also been proved. Finally, some simulations are carried out to study the finite properties of the proposed estimator.

## Keywords

Partially Linear Additive Model, Stochastic Linear Restriction, Profile Least Squares Approach, Ridge Estimation

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

作为可加模型的推广, 部分线性变系数模型近年来受到了广泛的关注, 在理论和应用上都得到了深入的研究。该模型一般记为如下形式

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + m_1(Z_{1i}) + \cdots + m_q(Z_{qi}) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

其中  $Y_i$  是因变量观测值,  $\mathbf{X}_i$  是  $p \times 1$  的自变量观测值,  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  是对应的未知回归系数,  $Z_{ki}$  是一元连续性变量观测值,  $m_1(\cdot), \dots, m_q(\cdot)$  是未知的非参数光滑参数, 模型误差  $\varepsilon_i$  相互独立且均值为零和方差为  $\sigma^2$ 。

对于半参数模型(1), 模型中回归系数和非参数函数的估计是一个重要的任务, 基于不同的估计思想, 已经有多种方法提出, 包括 Ospomer 和 Ruppert (1999) [1]提出的 backfitting 方法、Li (2000) [2]所使用的级数估计方法, Manzan 和 Zerom (2005) [3]提出的边际积分方法以及 Wei 和 Liu (2012) [4]使用的 Profile 最小二乘方法。

使用回归模型分析实际数据的过程中, 经常会遇到多重共线性问题, 原因是变量之间存在较强的线性关系。多重共线性的存在对回归模型的统计推断结果有直接的影响, 能将问题分析带入歧途, 比如在理论上最小二乘估计依然是最小方差线性无偏估计, 但此时估计量的方差很大, 即估计量的均方误差很大。在回归分析结果中的直接体现就是未知系数的估计值绝对值偏大或者其正负号与实际意义相反。为了解决上面存在的问题, 目前有多种处理思路, 比如包括对自变量的剔除, 构造系数的有偏估计, 以牺牲估计量的无偏性换取估计量均方误差的降低。使用较多的有偏估计方法包括常见的岭估计和主成分估计, 以及 Liu (1993) [5]提出的 Liu 估计。此外, 分析实际研究中有时还能从问题本身以及已有研究中得到一些额外的信息, 如果将这些额外信息纳入到模型推断中, 则有助于提高模型的推断结果。附加额外信息, 常常可以转化为对回归系数的约束问题, 从而有了回归模型的精确线性约束估计, 随机约束估计, 以及不等式约束估计。如何在多重共线性和附加额外信息同时存在的情况下对模型进行估计得到了关注。

对于模型(1), 其线性部分存在多重共线性时的估计问题得到了重视。Wei 和 Wang (2016) [6]给出了参数分量的 Liu 估计及其约束 Liu 估计的渐近性质, Wei (2016) [7]研究了参数分量的随机约束估计及其性质, 在此基础上, 王肖南和魏传华(2016) [8]构造了模型中参数分量的随机约束 Liu 估计。

相比于 Liu 估计, 模型(1.1)的岭估计研究相对较少, 蓝文英(2013) [9]基于 backfitting 方法给出了参数分量的岭估计和约束岭估计。为了更进一步研究模型(1)的岭估计, 本文在上述研究基础上将给出模型

(1.1)在线性部分存在多重共线性且存在随机约束下的估计问题, 提出参数分量的随机约束岭估计, 并给出估计量的渐近性质。研究将 Özkale (2009) [10]的结果从线性回归模型推广到了部分线性可加模型。与此类似的是魏传华等(2014) [11]和刘超等(2017) [12]分别研究了部分线性模型和部分线性变系数模型的随机约束岭估计。

同 Özkale (2009)等文献一样, 随机约束条件表述为如下形式

$$\mathbf{b} = \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\eta} \quad (2)$$

其中  $\mathbf{b}$  是  $k \times 1$  的已知向量,  $\mathbf{A}$  是一个已知的  $k \times p$  行满秩矩阵。  $\boldsymbol{\eta}$  为均值为  $\mathbf{0}$  协方差矩阵为  $\sigma^2 \boldsymbol{\Omega}$  的随机向量, 其中  $\boldsymbol{\Omega}$  为一已知正定矩阵。针对模型(1)~(2), 我们将重点研究参数分量  $\boldsymbol{\beta}$  的估计问题。因此本文是 Wei (2016) [7]和蓝文英(2013) [9]的推广。

第 2 节提出部分线性可加模型中参数部分的随机约束岭估计, 并研究其渐近性质。第 3 节利用数值模拟考察所提估计方法的有效性, 第 4 节给出结论。

## 2. 随机约束岭估计及其性质

为了构造模型的随机约束岭估计, 首先介绍 Profile 最小二乘技术。为了方便叙述, 假定模型(1.1)中  $q = 2$ , 设  $\{Y_i, \mathbf{X}_i, Z_{1i}, Z_{2i}\}_{i=1}^n$  来自于模型(1.1), 则有

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + m_1(Z_{1i}) + m_2(Z_{2i}) + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (3)$$

假定模型(3)中的参数分量  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  已知, 则模型(3)可以写成如下一个标准的可加模型:

$$Y_i - \mathbf{X}_i^T \boldsymbol{\beta} = m_1(Z_{1i}) + m_2(Z_{2i}) + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (4)$$

对  $k = 1, 2$ , 记

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \vdots \\ \mathbf{X}_n^T \end{bmatrix}, \mathbf{m}_k = \begin{bmatrix} m_k(Z_{k1}) \\ m_k(Z_{k2}) \\ \vdots \\ m_k(Z_{kn}) \end{bmatrix}, \mathbf{D}_{Z_k}^k = \begin{bmatrix} 1 & (Z_{k1} - Z_k)/h_k \\ 1 & (Z_{k2} - Z_k)/h_k \\ \vdots & \vdots \\ 1 & (Z_{kn} - Z_k)/h_k \end{bmatrix}.$$

由局部线性估计方法的原理, 关于第  $k$  个非参数函数的局部线性光滑矩阵可记为:

$$\mathbf{S}_k = \begin{bmatrix} \mathbf{e}_1^T \left\{ \mathbf{D}_{Z_{k1}}^{kT} \mathbf{K}_{Z_{k1}} \mathbf{D}_{Z_{k1}}^k \right\}^{-1} \mathbf{D}_{Z_{k1}}^T \mathbf{K}_{Z_{k1}} \\ \mathbf{e}_1^T \left\{ \mathbf{D}_{Z_{k2}}^{kT} \mathbf{K}_{Z_{k2}} \mathbf{D}_{Z_{k2}}^k \right\}^{-1} \mathbf{D}_{Z_{k2}}^T \mathbf{K}_{Z_{k2}} \\ \vdots \\ \mathbf{e}_1^T \left\{ \mathbf{D}_{Z_{kn}}^{kT} \mathbf{K}_{Z_{kn}} \mathbf{D}_{Z_{kn}}^k \right\}^{-1} \mathbf{D}_{Z_{kn}}^T \mathbf{K}_{Z_{kn}} \end{bmatrix},$$

其中  $\mathbf{e}_1 = (1, 0)^T$ ,  $\mathbf{K}_{Z_k} = \text{diag}\{K_{h_k}(Z_{k1} - Z_k), K_{h_k}(Z_{k2} - Z_k), \dots, K_{h_k}(Z_{kn} - Z_k)\}$ , 这里

$K_{h_k}(\cdot) = K(\cdot/h_k)/h_k$ ,  $K(\cdot)$  是核函数,  $h_k$  是窗宽。由 Opsomer 和 Ruppert (1997) [13]可知, 未知的非参数函数  $m_k$  可通过求解如下方程得到:

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{S}_1^* \\ \mathbf{S}_2^* & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1^* \\ \mathbf{S}_2^* \end{bmatrix} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \quad (5)$$

其中  $\mathbf{S}_k^* = (\mathbf{I}_n - \mathbf{1}\mathbf{1}^T) \mathbf{S}_k$ ,  $k = 1, 2$ 。进一步我们可以得到  $\mathbf{m}_1$  和  $\mathbf{m}_2$  的 backfitting 估计为:

$$\hat{\mathbf{m}}_1 = \mathbf{W}_1(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \quad \hat{\mathbf{m}}_2 = \mathbf{W}_2(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad (6)$$

$$\mathbf{W}_1 = \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} (\mathbf{I}_n - \mathbf{S}_1^*), \quad \mathbf{W}_2 = \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1} (\mathbf{I}_n - \mathbf{S}_2^*).$$

将  $\hat{m}_1$  和  $\hat{m}_2$  代入模型(3), 可得如下线性回归模型

$$\bar{Y}_i = \bar{X}_i^T \boldsymbol{\beta} + \varepsilon_i, i = 1, 2, \dots, n, \quad (7)$$

其中  $\bar{Y} = (\bar{Y}_1, \dots, \bar{Y}_n)^T = (\mathbf{I}_n - \mathbf{S})\mathbf{Y}$ ,  $\bar{X} = (\bar{X}_1, \dots, \bar{X}_n)^T = (\mathbf{I}_n - \mathbf{S})\mathbf{X}$  和  $\mathbf{S} = \mathbf{W}_1 + \mathbf{W}_2$ 。

使用最小二乘法估计线性模型(7), 从而得到线性部分回归系数  $\boldsymbol{\beta}$  的 profile 最小二乘估计

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left[ (\bar{Y} - \bar{X}\boldsymbol{\beta})^T (\bar{Y} - \bar{X}\boldsymbol{\beta}) \right] = (\bar{X}^T \bar{X})^{-1} \bar{X}^T \bar{Y}. \quad (8)$$

下面我们在模型(7)上考虑随机约束条件(2), 构造随机约束岭估计, 类似于 Özkale (2009) [10]、魏传华等(2014) [11]和刘超等(2017) [12], 构造如下的目标函数

$$L(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ (\bar{Y} - \bar{X}\boldsymbol{\beta})^T (\bar{Y} - \bar{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}^T \boldsymbol{\beta} + (\mathbf{b} - \mathbf{A}\boldsymbol{\beta})^T \boldsymbol{\Omega}^{-1} (\mathbf{b} - \mathbf{A}\boldsymbol{\beta}) \right\}$$

求取函数  $L(\boldsymbol{\beta})$  关于  $\boldsymbol{\beta}$  的偏导数, 并令其等于零, 有

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\bar{X}^T \bar{Y} + 2\bar{X}^T \bar{X}\boldsymbol{\beta} + 2k\boldsymbol{\beta} - 2\mathbf{A}^T \boldsymbol{\Omega}^{-1} (\mathbf{b} - \mathbf{A}\boldsymbol{\beta}) = \mathbf{0} \quad (9)$$

求解(9), 整理可得  $\boldsymbol{\beta}$  的随机约束岭估计为

$$\hat{\boldsymbol{\beta}}_R^S(k) = (\bar{X}^T \bar{X} + k\mathbf{I}_p + \mathbf{A}^T \boldsymbol{\Omega}^{-1} \mathbf{A})^{-1} (\bar{X}^T \bar{Y} + \mathbf{A}^T \boldsymbol{\Omega}^{-1} \mathbf{b}) \quad (10)$$

记  $\mathbf{E} = \bar{X}^T \bar{X} + k\mathbf{I}_p$ , 由 Rao 等(2008) [14]中的定理 A.18(iii), 可得

$$(\bar{X}^T \bar{X} + k\mathbf{I}_p + \mathbf{A}^T \boldsymbol{\Omega}^{-1} \mathbf{A})^{-1} = \mathbf{E}^{-1} - \mathbf{E}^{-1} \mathbf{A}^T (\boldsymbol{\Omega} + \mathbf{A} \mathbf{E}^{-1} \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{E}^{-1} \quad (11)$$

将(11)式代入(10)式整理可得  $\hat{\boldsymbol{\beta}}_R^S(k)$  的另外一种等价形式为:

$$\hat{\boldsymbol{\beta}}_R^S(k) = \hat{\boldsymbol{\beta}}_R(k) - (\bar{X}^T \bar{X} + k\mathbf{I}_p)^{-1} \mathbf{A}^T \left[ \boldsymbol{\Omega} + \mathbf{A} (\bar{X}^T \bar{X} + k\mathbf{I}_p)^{-1} \mathbf{A}^T \right]^{-1} [\mathbf{A}(k) - \hat{\boldsymbol{\beta}}_R(k)\mathbf{b}] \quad (12)$$

其中  $\hat{\boldsymbol{\beta}}_R(k) = (\bar{X}^T \bar{X} + k\mathbf{I}_p)^{-1} \bar{X}^T \bar{Y}$  是  $\boldsymbol{\beta}$  的岭估计。

下面给出  $\hat{\boldsymbol{\beta}}_R^S(k)$  的一些理论性质。

**定理 1.** 参数分量  $\boldsymbol{\beta}$  的 Profile 混合岭估计  $\hat{\boldsymbol{\beta}}_R^S(k)$  有如下性质

$$\text{Bias}(\hat{\boldsymbol{\beta}}_R^S(k)) = \boldsymbol{\beta} - kS_{sr}^{-1}\boldsymbol{\beta} + S_{sr}^{-1}\bar{X}^T (\mathbf{I}_n - \mathbf{S})\mathbf{M},$$

$$\text{Cov}[\hat{\boldsymbol{\beta}}_R^S(k)] = S_{sr}^{-1}\boldsymbol{\Psi}S_{sr}^{-1}$$

其中:  $S_{sr} = \bar{X}^T \bar{X} + k\mathbf{I}_p + \mathbf{A}^T \boldsymbol{\Omega}^{-1} \mathbf{A}$ ,  $\mathbf{M} = \mathbf{m}_1 + \mathbf{m}_2$ ,

$$\boldsymbol{\Psi} = \sigma^2 \left[ \bar{X}^T (\mathbf{I}_n - \mathbf{S})(\mathbf{I}_n - \mathbf{S})^T \bar{X} + \mathbf{A}^T \boldsymbol{\Omega}^{-1} \mathbf{A} \right]$$

**证明:** 根据  $\hat{\boldsymbol{\beta}}_R^S(k)$  估计的表达式(12)可知

$$\begin{aligned} \hat{\boldsymbol{\beta}}_R^S(k) &= (\bar{X}^T \bar{X} + k\mathbf{I}_p + \mathbf{A}^T \boldsymbol{\Omega}^{-1} \mathbf{A})^{-1} (\bar{X}^T \bar{Y} + \mathbf{A}^T \boldsymbol{\Omega}^{-1} \mathbf{b}) \\ &= S_{sr}^{-1} (\bar{X}^T \bar{Y} + \mathbf{A}^T \boldsymbol{\Omega}^{-1} \mathbf{b}) \\ &= S_{sr}^{-1} \left[ \bar{X}^T (\mathbf{I}_n - \mathbf{S})\mathbf{X}\boldsymbol{\beta} + \bar{X}^T (\mathbf{I}_n - \mathbf{S})\mathbf{M} + \bar{X}^T (\mathbf{I}_n - \mathbf{S})\boldsymbol{\varepsilon} + \mathbf{A}^T \boldsymbol{\Omega}^{-1} (\mathbf{A}\boldsymbol{\beta} + \boldsymbol{\eta}) \right] \\ &= \boldsymbol{\beta} + S_{sr}^{-1} \left[ -k\boldsymbol{\beta} + \bar{X}^T (\mathbf{I}_n - \mathbf{S})\mathbf{M} + \bar{X}^T (\mathbf{I}_n - \mathbf{S})\boldsymbol{\varepsilon} + \mathbf{A}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\eta} \right] \end{aligned}$$

由  $E\boldsymbol{\varepsilon} = \mathbf{0}, E\boldsymbol{\eta} = \mathbf{0}, \text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$  和  $\text{Cov}(\boldsymbol{\eta}) = \sigma^2 \boldsymbol{\Omega}$ , 我们有

$$\begin{aligned} \text{Bias}[\hat{\boldsymbol{\beta}}_R^S(k)] &= E[\hat{\boldsymbol{\beta}}_R^S(k)] - \boldsymbol{\beta} = S_{sr}^{-1}[-k\boldsymbol{\beta} + \bar{\mathbf{X}}^T(\mathbf{I}_n - S)\mathbf{M}], \\ \text{Cov}[\hat{\boldsymbol{\beta}}_R^S(k)] &= E\left[\left(\hat{\boldsymbol{\beta}}_R^S(k) - E\hat{\boldsymbol{\beta}}_R^S(k)\right)\left(\hat{\boldsymbol{\beta}}_R^S(k) - E\hat{\boldsymbol{\beta}}_R^S(k)\right)^T\right] = S_{sr}^{-1}\boldsymbol{\Psi}S_{sr}^{-1}. \end{aligned}$$

在给出定理 2 之前先给出一些假设条件:

**条件 1.** 核函数  $K(\cdot)$  为对称密度函数, 具有紧支撑。

**条件 2.**  $Z_k$  的密度函数  $f_k(Z_k)$  是 Lipschitz 连续且有有界支撑  $\Omega_k, k=1, 2$ 。

**条件 3.**  $m_k(\cdot), k=1, 2$  具有二阶连续导数。

**条件 4.**  $n \rightarrow \infty$  时,  $h_k \rightarrow 0$  时,  $nh_k/\log n \rightarrow \infty$  和  $nh_k^8 \rightarrow 0, k=1, 2$ 。

**定理 2.** 在满足以上条件 1~4 假设下, 随机约束岭估计  $\hat{\boldsymbol{\beta}}_R^S(k)$  是渐近正态的,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_R^S(k) - \boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}^{-1}).$$

$$\text{其中 } \boldsymbol{\Sigma} = E\left[\mathbf{X}_i - \sum_{k=1}^2 E(\mathbf{X}_i | Z_{ki})\right]\left[\mathbf{X}_i - \sum_{k=1}^2 E(\mathbf{X}_i | Z_{ki})\right]^T.$$

**证明:** 由定理 1 中的结论可知,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_R^S(k) - \boldsymbol{\beta}) = \left(\frac{S_{sr}}{n}\right)^{-1} \frac{1}{\sqrt{n}}[-k\boldsymbol{\beta} + \bar{\mathbf{X}}^T(\mathbf{I}_n - S)\mathbf{M} + \bar{\mathbf{X}}^T(\mathbf{I}_n - S)\boldsymbol{\varepsilon} + \mathbf{A}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta}]$$

根据 Wei 和 Liu (2012) [4] 中引理 6.2 知,

$$\frac{1}{n}(\bar{\mathbf{X}}^T\bar{\mathbf{X}}) \xrightarrow{p} \boldsymbol{\Sigma},$$

此外有  $\frac{1}{n}\mathbf{A}^T\boldsymbol{\Omega}^{-1}\mathbf{A} = o_p(1)$ , 结合以上式子易知:

$$\frac{1}{n}(\bar{\mathbf{X}}^T\bar{\mathbf{X}} + k\mathbf{I}_p + \mathbf{A}^T\boldsymbol{\Omega}^{-1}\mathbf{A}) \xrightarrow{p} \boldsymbol{\Sigma}. \quad (13)$$

由 Wei 和 Liu (2012) [4] 中的引理 6.3, 我们有

$$\frac{1}{\sqrt{n}}\bar{\mathbf{X}}^T(\mathbf{I}_n - S)\mathbf{M} = o_p(1), \quad \frac{1}{\sqrt{n}}\bar{\mathbf{X}}^T(\mathbf{I}_n - S)\boldsymbol{\varepsilon} \xrightarrow{D} N(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}) \quad (14)$$

事实上,  $E\mathbf{A}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta} = \mathbf{0}, \text{Cov}(\mathbf{A}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta}) = \mathbf{A}^T\boldsymbol{\Omega}^{-1}\mathbf{A}$ , 则有

$$\frac{1}{\sqrt{n}}\mathbf{A}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta} = o_p(1) \quad (15)$$

因此结合(13), (14)和(15), 根据 Slutsky 定理可得

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_R^{\text{PM}}(k) - \boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}^{-1}).$$

**注 2.1.** 当  $\boldsymbol{\Omega} = \mathbf{0}$  时, 随机约束条件(1.2)转化为精确的线性约束条件  $\mathbf{A}\boldsymbol{\beta} = \mathbf{b}$ , 此时由(11)式变为

$$\hat{\boldsymbol{\beta}}_R^S(k) = \hat{\boldsymbol{\beta}}_R(k) - (\bar{\mathbf{X}}^T\bar{\mathbf{X}} + k\mathbf{I}_p)^{-1}\mathbf{A}^T\left[\mathbf{A}(\bar{\mathbf{X}}^T\bar{\mathbf{X}} + k\mathbf{I}_p)^{-1}\mathbf{A}^T\right]^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}}_R(k) - \mathbf{b}).$$

该估计就是 Wei 和 Wang (2016) [6] 针对部分线性可加模型在精确约束条件下构造的 profile 最小二乘约束岭估计。

### 3. 数值模拟

本节将通过 Monte Carlo 数值模拟的方法对所提出的估计进行有效性检验。考虑如下的部分线性可加模型产生数据：

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + x_{4i}\beta_4 + x_{5i}\beta_5 + m_1(z_{1i}) + m_2(z_{2i}) + \varepsilon_i, i = 1, 2, \dots, n,$$

同 Wei (2016) [7] 中的设置一样，模型中自变量和非参数函数进行如下设定

$z_{1i} \sim U(0, 1), z_{2i} \sim U(-1, 1), m_1(z_{1i}) = \sin(2\pi z_{1i}), m_2(z_{2i}) = z_{2i}^3 + 3z_{2i}^2 - 5z_{2i} - 1$   $\beta = (2, 1, 1, 3)^T$  为了构造多重共线性，运用 McDonald 和 Galerneau (1975) [15] 中生成具有多重共线性自变量的方法生成解释变量

$x_1, x_2, \dots, x_4$ ，具体为：

$$x_{ij} = (1 - \rho^2)^{1/2} \omega_{ij} + \rho \omega_{i(p+1)}, \quad i = 1, 2, \dots, n; \quad j = 1, \dots, p$$

其中， $\omega_{ij}$  是独立的标准正态随机数， $\rho$  是一个具体的数值以确保任何四个解释变量在理论上是相关的，分别取  $\rho = 0.90, 0.99, 0.999$  刻画不同程度的复共线性问题。样本量  $n = 50, 80$  和  $120$  三种情形。为了考察模型误差的分布对估计结果的影响，选取模型误差  $\varepsilon_i$  服从如下的正态分布和均匀分布 1)  $\varepsilon_i \sim N(0, 0.5^2)$ ，2)  $\varepsilon_i \sim U\left(-\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}\right)$ 。模拟中核函数选用 Epanechnikov 核： $K(x) = 0.75(1-x^2)I_{|x| \leq 1}$ ，为了计算方便，窗宽设定为  $h_1 = h_2 = n^{-1/5}$ 。随机约束的设定同刘超等(2017) [12] 一样，参数分量  $(\beta_1, \beta_2, \beta_3, \beta_4) = (2, 1, 3, 3)$ ，随机约束设定为

$$E\eta_1 = E\eta_2 = 0, \text{Var}(\eta_1) = 0.04, \text{Var}(\eta_2) = 0.16, \text{Cov}(\eta_1, \eta_2) = 0.2,$$

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix}, \Omega = \begin{bmatrix} 0.04 & 0.016 \\ 0.016 & 0.16 \end{bmatrix}$$

针对上面的具体设定，每种情况都重复进行 500 次模拟，每次求取参数分量

$(\beta_1, \beta_2, \beta_3, \beta_4)$  的 profile 最小二乘估计(PLS)、随机约束估计(S)、岭估计(R)和提出的随机约束岭估计(SR)，对于  $\beta = \{\beta_1, \beta_2, \beta_3, \beta_4\}$  的估计  $\beta^* = \{\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*\}$ ，以其估计均方误差(EMSE)来衡量其表现，

$$\text{EMSE}(\beta^*) = \frac{1}{500} \sum_{k=1}^{500} \sum_{j=1}^4 (\beta_{kj}^* - \beta_j)^2$$

其中  $\beta_{kj}^*$  是参数  $\beta_j$  的第  $k$  次重复时的估计值，模拟结果见表 1。

**Table 1.** EMSEs of the estimators

**表 1.** 不同估计量的 EMSEs

$\rho$	$\beta$	$n = 50$		$n = 80$		$N = 120$	
		N	U	N	U	N	U
0.9	PLS	0.1196	0.1247	0.0645	0.0666	0.0404	0.0402
	S	0.0953	0.1024	0.0595	0.0595	0.0405	0.0396
	R	0.1217	0.1253	0.0656	0.0658	0.0404	0.0401
	SR	0.0940	0.1007	0.0596	0.0583	0.0405	0.0391
0.99	PLS	1.2161	1.1273	0.6010	0.6375	0.3882	0.3991
	S	0.5556	0.4787	0.3542	0.3562	0.2464	0.2359
	R	1.0613	1.0115	0.5935	0.6525	0.3909	0.4101
	SR	0.3732	0.3051	0.2837	0.2743	0.2140	0.2038

## Continued

	PLS	11.4675	11.1771	6.1310	5.9372	3.6720	3.5115
0.999	S	3.7561	3.6777	2.1531	2.1473	1.4106	1.3089
	R	3.8133	3.7323	2.5967	2.5143	2.1214	2.0966
	SR	0.8646	0.8844	0.5577	0.5948	0.4926	0.4501

从模拟结果不难看出,随着样本量的增加,四类估计的 EMSE 都在变小,同时模型误差的分布对结果影响很小。四类估计之间对比,随着共线性程度变高,随机约束岭估计的 EMSE 比其他三类估计显著的小,从而反映了所提方法的有效性。

#### 4. 总结

作为非参数回归和线性回归模型的推广,部分线性模型、可加模型、变系数和单指标模型等半参数模型近年来得到了广泛的关注。对于这些半参数模型的研究,较少讨论变量的多重共线性问题。本文针对部分线性可加模型这一半参数模型,研究了附加随机约束条件时的有偏估计问题,构造了线性部分回归系数的随机约束岭估计,并研究了估计量的性质。

半参数模型的模型结构对比线性回归模型更为复杂,在自变量存在多重共线性的情况下,如何构造有效的有偏估计量是实际分析中不可避免的问题。

#### 基金项目

中国劳动关系学院教育教学改革立项项目(JG1406);2020年度安徽高等学校自然科学基金项目(KJ2020A1200)。

#### 参考文献

- [1] Opsomer, J.D. and Ruppert, D. (1999) A Root- $n$  Consistent Backfitting Estimator for Semiparametric Additive Modeling. *Journal of Computational and Graphical Statistics*, **8**, 715-732. <https://doi.org/10.1080/10618600.1999.10474845>
- [2] Manzano, S. and Zeromb, D. (2005) Kernel Estimation of a Partially Linear Additive Model. *Statistics & Probability Letters*, **72**, 313-322. <https://doi.org/10.1016/j.spl.2005.02.005>
- [3] Liu, X., Wang, L. and Liang, H. (2011) Estimation and Variable Selection for Semiparametric Additive Partial Linear Model. *Statistica Sinica*, **21**, 1225-1248. <https://doi.org/10.5705/ss.2009.140>
- [4] Wei, C.H. and Liu, C.L. (2012) Statistical Inference on Semiparametric Partially Linear Additive Models. *Journal of Nonparametric Statistics*, **24**, 809-823. <https://doi.org/10.1080/10485252.2012.716155>
- [5] Liu, K.J. (2003) Using Liu-Type Estimator to Combat Collinearity. *Communication in Statistics-Theory and Methods*, **32**, 285-292. <https://doi.org/10.1081/STA-120019959>
- [6] Wei, C.H. and Wang, X.N. (2016) Liu-Type Estimator in Semiparametric Partially Linear Additive Models. *Journal of Nonparametric Statistics*, **28**, 459-468. <https://doi.org/10.1080/10485252.2016.1163357>
- [7] Wei, C.-H. (2016) A Stochastic Restricted Estimator in Partially Linear Additive Models. *Far East Journal of Theoretical Statistics*, **52**, 35-477. <https://doi.org/10.17654/TS052010035>
- [8] 王肖南, 魏传华. 部分线性可加模型的随机约束 Liu 估计[J]. 中央民族大学学报(自然科学版), 2016, 25(1): 80-85.
- [9] 蓝文英. 部分线性可加模型基于 backfitting 方法的岭估计[J]. 中央民族大学学报(自然科学版), 2013, 22(S1): 68-71.
- [10] Özkale, M.R. (2009) A Stochastic Restricted Ridge Regression Estimator. *Journal of Multivariate Analysis*, **100**, 1706-1716. <https://doi.org/10.1016/j.jmva.2009.02.005>
- [11] 魏传华, 郭双, 王肖南. 部分线性模型的随机约束岭估计[J]. 数学的实践与认识, 2014, 44(13): 249-254.
- [12] 刘超, 韦杰, 魏传华. 部分线性变系数模型的随机约束岭估计[J]. 应用数学, 2017, 30(4): 774-779.

- 
- [13] Opsomer, J.D. and Ruppert, D. (1997) Fitting a Bivariate Additive Model by Local Polynomial Regression. *The Annals of Statistics*, **25**, 186-211. <https://doi.org/10.1214/aos/1034276626>
- [14] Rao, C.R., Shalabh, Toutenburg, H. and Heumann, C. (2008) *Linear Models and Generalization: Least Squares and Alternatives*. Springer, Berlin.
- [15] McDonald, G.C. and Galarneau, D.I. (1975) A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of American Statistical Association*, **70**, 407-416. <https://doi.org/10.1080/01621459.1975.10479882>