

部分线性变系数测量误差模型的约束Liu估计

李 静¹, 安佰玲²

¹中国劳动关系学院应用技术学院, 北京

²淮北师范大学数学科学学院, 安徽 淮北

收稿日期: 2022年12月9日; 录用日期: 2023年1月2日; 发布日期: 2023年1月11日

摘 要

本文考虑部分线性变系数测量误差模型的估计问题, 同时考虑线性部分自变量存在多重共线性和线性部分存在约束条件时两种情形, 基于校正的profile最小二乘技术与Liu估计方法, 分别构造了未知参数分量的Liu估计和约束Liu估计, 并研究了所提估计量的渐进性质。

关键词

部分线性变系数模型, 测量误差模型, Liu估计, Profile最小二乘估计

Restricted Liu Estimation in Partially Linear Varying Coefficient Measurement Error Model

Jing Li¹, Bailing An²

¹School of Applied Technology, China University of Labor Relations, Beijing

²School of Mathematical Sciences, Huaibei Normal University, Huaibei Anhui

Received: Dec. 9th, 2022; accepted: Jan. 2nd, 2023; published: Jan. 11th, 2023

Abstract

This paper considers estimation of semiparametric partially linear varying coefficient measurement error model when the problem of multicollinearity exists and linear restrictions on the parameter components are available. Based on the corrected profile least-squares approach and Liu estimation method, the Liu estimator and the corresponding restricted Liu estimator for the parametric component are constructed, and their statistical properties are given.

Keywords

Partially Linear Varying Coefficient Model, Measurement Error Model, Liu Estimator, Profile Least-Squares Approach

Copyright © 2023 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

测量误差模型是统计学和计量经济学领域的重点研究内容之一, 近二十年来, 半参数测量误差模型的研究得到了越来越多的关注。本文研究如下的带有测量误差的半参数部分线性变系数模型

$$\begin{cases} Y = X^T \boldsymbol{\beta} + Z^T \boldsymbol{\alpha}(U) + \varepsilon \\ V = X + \boldsymbol{\xi} \end{cases} \quad (1)$$

其中, Y 为因变量观测值, Z, X 和 U 分别为自变量观测值, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ 为模型中线性部分的未知待估系数, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \alpha_2(\cdot), \dots, \alpha_q(\cdot))^T$ 为模型中变系数部分的未知系数函数。 ε 是均值为零的模型误差, 测量误差 $\boldsymbol{\xi}$ 与 (Y, X, Z, U) 独立, 有 $E\boldsymbol{\xi} = \mathbf{0}, \text{Cov}(\boldsymbol{\xi}) = \boldsymbol{\Sigma}_{\boldsymbol{\xi}}$, 一般假定协方差矩阵 $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ 已知。

对于模型(1), You和Chen (2006) [1]基于部分线性变系数模型的profile最小二乘估计提出了一种校正估计方法, Wei (2012) [2]在这个估计方法的基础上对模型(1)考虑了未知参数分量 $\boldsymbol{\beta}$ 的约束估计, 并讨论了对应的约束检验问题。

使用多元回归模型分析实际数据时, 不同自变量之间往往存在较强的线性关系, 我们将这种现象称之为多重共线性。多重共线性会导致模型推断的错误, 比如最小二乘估计虽然在理论上还有无偏估计, 但方差很大, 基于实际数据得到的估计结果表现为回归系数的正负号与实际问题相反, 回归系数估计值的绝大致异常大等。为了解决多重共线性问题, 一方面可以在变量选择上进行深入分析, 采用删除一些自变量的方式。另一方面, 构造有偏估计用以降低估计量的均方误差, 目前得到研究较多的有偏估计有岭估计、主成分估计和Liu估计。对于测量误差模型, 由于自变量不能精确观测, 其有偏估计的研究也相对复杂, 因为目前研究结果相对较少, 有关的结果可参考文[3] [4] [5]。对于半参数测量误差模型的有偏估计, 文[6] [7] [8]针对部分线性测量误差模型分别研究了岭估计和Liu估计。对于模型(1), 文[9]构造了模型(1)的岭估计。基于以上研究, 本文主要给出模型(1)的Liu估计和对应的约束Liu估计, 并给出估计量的渐近性质。

论文第2节将构造参数分量的Liu估计, 并给出所提估计量的渐近性质。第3节将构造参数分量的约束Liu估计并研究其性质。定理的证明在第4节给出。结论将在第5节给出。

2. 参数分量的 Liu 估计

为了构造模型的Liu估计和约束Liu估计, 首先介绍You和Chen (2006) [1]针对模型(1)提出校正Profile最下二乘法估计, 我们要在这个估计方法的基础构造有偏估计。记 $\{Y_i, X_i, Z_i, U_i\}_{i=1}^n$ 为来自模型(1)的样本数据, 这里先假设自变量 X 可以被精确观测, 则有如下的模型

$$Y_i = X_i^T \boldsymbol{\beta} + Z_i^T \boldsymbol{\alpha}(U_i) + \varepsilon_i, i = 1, 2, \dots, n. \quad (2)$$

假定参数分量 β 已知, 则模型(2)可等价写成如下形式的变系数模型

$$(Y_i - \mathbf{X}_i^T \beta) = \alpha_1(U_i)Z_{i1} + \cdots + \alpha_q(U_i)Z_{iq} + \varepsilon_i \quad (3)$$

采用局部线性光滑方法对模型(43)进行估计。假设系数函数 $\{\alpha_j(\cdot), j=1, 2, \dots, q\}$ 二阶连续可导, 对于 u_0 附近的点 u , 由Taylor展开可得

$$\alpha_j(u) \approx \alpha_j(u_0) + \alpha'_j(u_0)(u - u_0), j=1, 2, \dots, q \quad (4)$$

其中 $\alpha'_j(u) = \partial \alpha_j(u) / \partial u$, $\{\alpha_j(u_0), \alpha'_j(u_0)\}, j=1, 2, \dots, q\}$ 的局部加权最小二乘估计可通过极小化下式得到

$$\sum_{i=1}^n \left((Y_i - \mathbf{X}_i^T \beta) - \sum_{j=1}^q (\alpha_j(u_0) + \alpha'_j(u_0)(U_i - u_0)) Z_{ij} \right)^2 K_h(U_i - u_0) \quad (5)$$

其中核函数 $K_h(\cdot) = K(\cdot/h)/h$, h 是窗宽。

为了方便叙述, 记

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \vdots \\ \mathbf{X}_n^T \end{bmatrix}, D_{u_0} = \begin{bmatrix} \mathbf{Z}_1^T & \frac{U_1 - u_0}{h} \mathbf{Z}_1^T \\ \mathbf{Z}_2^T & \frac{U_2 - u_0}{h} \mathbf{Z}_2^T \\ \vdots & \vdots \\ \mathbf{Z}_n^T & \frac{U_n - u_0}{h} \mathbf{Z}_n^T \end{bmatrix};$$

以及 $\mathbf{W}_{u_0} = \text{diag}(K_h(U_1 - u_0), K_h(U_2 - u_0), \dots, K_h(U_n - u_0))$, 那么可得系数函数 $\alpha(u_0)$ 的局部线性估计为

$$\hat{\alpha}(u_0) = (\mathbf{I}_q \quad \mathbf{0}_q) \{D_{u_0}^T \mathbf{W}_{u_0} D_{u_0}\}^{-1} D_{u_0}^T \mathbf{W}_{u_0} (Y - X\beta) \quad (6)$$

其中矩阵 \mathbf{I}_q 和 $\mathbf{0}_q$ 分别为 q 维的单位阵和元素全为零的矩阵。用得到的 $\hat{\alpha}(U_i)$ 代替(2)中系数函数 $\alpha(U_i)$, 经过简单整理可得如下的线性模型

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

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

$$\mathbf{S} = \begin{bmatrix} (\mathbf{Z}_1^T \quad \mathbf{0}) [D_{U_1}^T \mathbf{W}_{U_1} D_{U_1}]^{-1} D_{U_1}^T \mathbf{W}_{U_1} \\ (\mathbf{Z}_2^T \quad \mathbf{0}) [D_{U_2}^T \mathbf{W}_{U_2} D_{U_2}]^{-1} D_{U_2}^T \mathbf{W}_{U_2} \\ \vdots \\ (\mathbf{Z}_n^T \quad \mathbf{0}) [D_{U_n}^T \mathbf{W}_{U_n} D_{U_n}]^{-1} D_{U_n}^T \mathbf{W}_{U_n} \end{bmatrix}$$

如果自变量 \mathbf{X}_i 能被精确观测, 那么对上面得到的模型(7)使用最小二乘法可以直接得到 β 的估计。然而, 模型(1)中得不到自变量 \mathbf{X}_i 的观测值, 得到的观测数据实际是 V_i , 直接用 V_i 来直接代替 \mathbf{X}_i 对模型(7)进行估计, 由于测量误差的存在, 所构造的估计不是相合估计量。对于由于测量误差带来的这一问题, You和Chen (2006) [1]对Profile最小二乘估计进行校正, 得到了如下的校正估计

$$\hat{\beta} = \arg \min_{\beta \in R^p} \left[(\bar{Y} - \bar{V}\beta)^T (\bar{Y} - \bar{V}\beta) - n\beta^T \Sigma_{\xi} \beta \right] = (\bar{V}^T \bar{V} - n\Sigma_{\xi})^{-1} \bar{V}^T \bar{Y} \quad (8)$$

其中 $V = (V_1, \dots, V_n)^T$, $\bar{V} = (\mathbf{I}_n - \mathbf{S})V$ 。

类似于普通线性回归模型中 Liu 估计量的构造, 构造如下的辅助函数

$$F_1(\beta) = (\bar{Y} - \bar{V}\beta)^T (\bar{Y} - \bar{V}\beta) - n\beta^T \Sigma_\xi \beta + (d\hat{\beta} - \beta)^T (d\hat{\beta} - \beta) \quad (9)$$

$F_1(\beta)$ 关于 β 求导, 并令其偏导数等于 0, 有

$$\frac{\partial F_1(\beta)}{\partial \beta} = -2\bar{V}^T (\bar{Y} - \bar{V}\beta) - 2n\Sigma_\xi \beta - 2(d\hat{\beta} - \beta) = 0 \quad (10)$$

整理可得参数分量 β 的 Liu 估计

$$\hat{\beta}_{Liu} = (\bar{V}^T \bar{V} - n\Sigma_\xi + I_p)^{-1} (\bar{V}^T \bar{Y} + d\hat{\beta}) = (\bar{V}^T \bar{V} - n\Sigma_\xi + I_p)^{-1} (\bar{V}^T \bar{V} - n\Sigma_\xi + dI_p) \hat{\beta} \quad (11)$$

下面给出所提估计量 $\hat{\beta}_{Liu}$ 的渐近性质. 定义 $\Gamma(u) = E(\mathbf{Z}\mathbf{Z}^T | U = u)$, $\Phi(u) = E(\mathbf{Z}\mathbf{X}^T | U = u)$,

$\Sigma_1 = E\left\{ \left[\mathbf{X}_1 - \Phi^T(U_1)\Gamma^{-1}(U_1)\mathbf{Z}_1 \right]^{\otimes 2} \right\}$ 以及 $\Sigma_2 = E\left\{ \left[\mathbf{X}_1 - \Phi^T(U_1)\Gamma^{-1}(U_1)\mathbf{Z}_1 + \xi_1 \right] (\varepsilon_1 - \xi_1^T \beta) + \Sigma_\xi \beta \right\}^{\otimes 2}$, $A^{\otimes 2}$

表示 AA^T .

定理 1. 若第 4 节中的假设 A.1~A.6 成立, 参数分量 β 的 Liu 估计有如下性质

$$\sqrt{n}(\hat{\beta}_{Liu} - \beta) \xrightarrow{D} N(\mathbf{0}, \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}).$$

该结论表明基于校正 Profile 最小二乘估计方法的 $\hat{\beta}_{Liu}$ 的渐近性质与 Profile 最小二乘估计 $\hat{\beta}$ 的相同。

3. 参数分量的约束 Liu 估计

本节考虑考虑如下的线性约束条件

$$A\beta = b \quad (12)$$

其中 A 是 $k \times p$ 维的已知矩阵, 且 $\text{rank}(A) = k$, b 是 $k \times 1$ 维的已知向量. 下面构造部分线性变系数测量误差模型(1)在约束条件(12)下的约束 Liu 估计。

构造如下的辅助函数

$$F_2(\beta, \lambda) = (\bar{Y} - \bar{V}\beta)^T (\bar{Y} - \bar{V}\beta) - n\beta^T \Sigma_\xi \beta + (d\hat{\beta} - \beta)^T (d\hat{\beta} - \beta) + 2\lambda^T (A\beta - b) \quad (13)$$

其中 λ 是 k 维 Lagrange 乘子. 针对函数 $F_2(\beta, \lambda)$ 关于 β , λ 分别求导, 并令偏导数等于 0, 有

$$\frac{\partial F_2(\beta, \lambda)}{\partial \beta} = -2\bar{V}^T (\bar{Y} - \bar{V}\beta) - 2n\Sigma_\xi \beta - 2(d\hat{\beta} - \beta) + 2A^T \lambda = 0 \quad (14)$$

和

$$\frac{\partial F_2(\beta, \lambda)}{\partial \lambda} = A\beta - b = 0 \quad (15)$$

(14)式整理可得

$$\beta = \hat{\beta}_{Liu} - (\bar{V}^T \bar{V} - n\Sigma_\xi + I_p)^{-1} A^T \lambda \quad (16)$$

将(16)式代入(15), 有下式成立

$$b = A\hat{\beta}_{Liu} - A(\bar{V}^T \bar{V} - n\Sigma_\xi + I_p)^{-1} A^T \lambda \quad (17)$$

如果 $A(\bar{V}^T \bar{V} - n\Sigma_\xi + I_p)^{-1} A^T$ 可逆, 则有 λ 的估计为

$$\hat{\lambda} = \left[A(\bar{V}^T \bar{V} - n \Sigma_{\xi} + I_p)^{-1} A^T \right]^{-1} (A \hat{\beta}_{Liu} - b) \quad (18)$$

将上式代入(16), 可得 β 的约束 Liu 估计为

$$\hat{\beta}_{Liu}^R = \hat{\beta}_{Liu} - (\bar{V}^T \bar{V} - n \Sigma_{\xi} + I_p)^{-1} A^T \left[A(\bar{V}^T \bar{V} - n \Sigma_{\xi} + I_p)^{-1} A^T \right]^{-1} (A \hat{\beta}_{Liu} - b) \quad (19)$$

对于 $\hat{\beta}_{Liu}^R$, 显然有 $A \hat{\beta}_{Liu}^R = b$ 。

下面给出所提估计量 $\hat{\beta}_{Liu}^R$ 的渐近性质。

定理2. 若第4节中的假设A.1~A.6成立, $\hat{\beta}_{Liu}^R$ 有如下性质

$$\sqrt{n}(\hat{\beta}_{Liu}^R - \beta) \xrightarrow{D} N\left[\mathbf{0}, (I_p - D) \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1} (I_p - D)^T\right].$$

$$D = \Sigma_1^{-1} A^T \left[A \Sigma_1^{-1} A^T \right]^{-1} A$$

该结论表明基于校正Profile最小二乘估计方法的约束Liu估计 $\hat{\beta}_{Liu}^R$ 与Wei (2012) [2]中提出的约束估计的渐近分布相同。

4. 定理的证明

下面给出定理1和定理2成立需要的条件, 这些条件是 You 和 Chen (2006) [1]采用过的。

(A.1) 随机变量 U 具有有界支撑 Π , 其密度函数 $f(\cdot)$ 在其支撑上满足 Lipschitz 连续, 且不为 0。

(A.2) 对于任一 $U \in \Omega$, 矩阵 $E(\mathbf{z}\mathbf{z}^T | U)$ 为非奇异, $E(\mathbf{z}\mathbf{z}^T | U)$, $E(\mathbf{z}\mathbf{z}^T | U)^{-1}$ 和 $E(\mathbf{z}\mathbf{x}^T | U)$ 都是 Lipschitz 连续的。

(A.3) 存在 $s > 2$ 使得 $E\|\mathbf{x}\|^{2s} < \infty$ 和 $E\|\mathbf{z}\|^{2s} < \infty$, 对于 $\varepsilon < 2 - s^{-1}$ 使得 $n^{2\varepsilon-1}h \rightarrow \infty$ 。

(A.4) $\{\alpha_j(\cdot), j = 1, \dots, q\}$ 二阶连续可导。

(A.5) 函数 $K(\cdot)$ 为对称密度函数, 具有紧支撑。

(A.6) $nh^8 \rightarrow 0$ 和 $nh^2/(\log n)^2 \rightarrow \infty$ 。

引理1 若假设A.1~A.6成立, 参数分量 β 的校正最小二乘估计有如下性质

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(\mathbf{0}, \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}).$$

该引理是 You 和 Chen (2006) [1] 的定理 3.1。

定理1的证明。

由 You 和 Chen (2006) [1] 中的引理 A.3 和 A.4, 可得

$$\frac{1}{n}(\bar{V}^T \bar{V} - n \Sigma_{\xi} + I_p) \xrightarrow{p} \Sigma_1, \frac{1}{n}(\bar{V}^T \bar{V} - n \Sigma_{\xi} + dI_p) \xrightarrow{p} \Sigma_1$$

再由引理1, 利用 Slutsky 定理, 可得

$$\sqrt{n}(\hat{\beta}_{Liu} - \beta) \xrightarrow{D} N(\mathbf{0}, \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}).$$

定理2的证明。

由 You 和 Chen (2006) [1] 中的引理 A.3 和 A.4, 可得

$$\frac{1}{n}(\bar{V}^T \bar{V} - n \Sigma_{\xi} + I_p) \xrightarrow{p} \Sigma_1$$

从而有

$$(\bar{V}^T \bar{V} - n \Sigma_{\xi} + I_p)^{-1} A^T \left[A (\bar{V}^T \bar{V} - n \Sigma_{\xi} + I_p)^{-1} A^T \right]^{-1} A \xrightarrow{p} D$$

利用Slutsky定理, 可得

$$\sqrt{n}(\hat{\beta}_{Liu}^R - \beta) \xrightarrow{D} N \left[\mathbf{0}, (I_p - D) \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1} (I_p - D)^T \right].$$

5. 总结

本文主要研究了部分线性变系数测量误差模型的 Liu 估计问题, 针对线性部分自变量存在多重共线性这一问题构造了有偏估计方法, 同时也考虑了线性约束下的 Liu 估计。本文只考虑了线性部分的自变量的多重共线性问题, 没有考虑变系数部分存在多重共线性的问题, 这在实际问题分析中同样重要, 将是以后研究的方向。

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] You, J.H. and Chen, G.M. (2006) Estimation of a Semiparametric Varying-Coefficient Partially Linear Errors-in-Variables Model. *Journal of Multivariate Analysis*, **97**, 324-341. <https://doi.org/10.1016/j.jmva.2005.03.002>
- [2] Wei, C.H. (2012) Statistical Inference for Restricted Partially Linear Varying Coefficient Errors-in-Variables Models. *Journal of Statistical Planning and Inference*, **142**, 2464-2472. <https://doi.org/10.1016/j.jspi.2012.02.041>
- [3] Shalabh, G.G. and Misra, N. (2007) Restricted Regression Estimation in Measurement Error Models. *Computational Statistics & Data Analysis*, **52**, 1149-1166. <https://doi.org/10.1016/j.csda.2007.05.011>
- [4] Saleh, A.K.Md.E. and Shalabh (2014) A Ridge Regression Estimation Approach to the Measurement Error Model. *Journal of Multivariate Analysis*, **123**, 68-84. <https://doi.org/10.1016/j.jmva.2013.08.014>
- [5] Ghapani, F., Rasekh, A.R. and Babadi, B. (2016) The Weighted Ridge Estimator in Stochastic Restricted Linear Measurement Error Models. *Statistical Papers*, **59**, 709-723. <https://doi.org/10.1007/s00362-016-0786-3>
- [6] 李静, 李雪艳. 半参数 EV 模型的岭估计[J]. 统计与管理, 2016(2): 15-16.
- [7] Emami, H. (2018) Ridge Estimation in Semiparametric Linear Measurement Error Models. *Linear Algebra and Its Applications*, **552**, 127-146. <https://doi.org/10.1016/j.laa.2018.04.016>
- [8] 张婷婷. 部分线性测量误差模型的 Liu 估计[D]: [硕士学位论文]. 重庆: 重庆大学, 2020.
- [9] 曹连英, 毕琳. 变系数部分线性误差变量模型的岭估计[J]. 统计与决策, 2020(24): 25-27.