

可交换条件下的结构回归模型^{*}

金 华, 何 春, 方积乾

(中山大学公共卫生学院, 广东 广州 510080)

摘 要: 在可交换条件下, 当响应变量为一维时, 利用结构回归模型 $Y_{ij} = \alpha_i + \chi_{ij}\beta_i^T + \epsilon_{ij}$, $(\chi_{ij}, \epsilon_{ij})^T \sim N[(\mu_{\chi}, 0)^T, A]$, $j = 1, \dots, n_i, i = 0, 1$ 研究总体平均处理效应, 通过采用极大似然的方法, 给出模型各参数的估计: $\alpha_i, \beta_i, \sigma_{\epsilon}^2, \mu_{\chi}, \Sigma_{\chi}$, 从而给出总体平均处理效应的极大似然估计 ATE。

关键词: 可交换条件; 总体平均处理效应; 结构回归模型; 极大似然估计

中图分类号: O212.1 **文献标识码:** A **文章编号:** 0529-6579(2003)03-0004-04

Rosenbaum 等^[1]得到, 当处理分配 S 是强可忽略时, 对暴露组和对照组中的协变量匹配求均值之差可以得到总体平均处理效应的无偏估计。但是, 上述方法计算过程复杂, 在协变量过多的情况下还会面临稀疏数据问题^[2]。解决办法之一是只对估计倾向分数进行分层, 不过计算还是繁琐, 在数据较少时也难以检验模型真伪^[3]。其它方法也存在一些问题^[4]。本文在可交换条件下, 当响应变量为一维时, 利用结构回归模型估计总体平均处理效应 ATE。

以 U 表示总体, 假设其中每个个体接受一种处理。为简便记, 假设处理只有 2 种水平: 一是暴露, 一是对照。令 S 为处理分配变量, 即 $S = 1$ 表示个体接受处理, 而 $S = 0$ 表示个体用作对照。对每一个体 u , 它都可能接受任何一种处理, 因而相应的响应变量值可能有所不同。假设个体 u 有 2 种可能不同的响应变量 $Y_S(u)$, 它表示个体 u 在处理分配 S 下的响应变量值。

一个处理对于某个体的处理效应是该个体暴露与对照下的响应变量值之差。对单位 u 而言, 暴露 $S = 1$ (相对于 $S = 0$) 的处理效应是 $Y_1(u)$ 与 $Y_0(u)$ 之差, 即

$$Y_1(u) - Y_0(u) \quad (1)$$

我们感兴趣的因果参数是个体效应, 即式 (1) 中的处理效应值。然而, 对于个体 u 而言, 不可能同时观察到 $Y_1(u)$ 与 $Y_0(u)$ 之值, 因而也就不可能观测到 u 的个体效应。通常的解决方法是估计总体的平均处理效应

$$ATE = E(Y_1 - Y_0)$$

即是个体效应 $Y_1(u) - Y_0(u)$ 在总体 U 上的期望值, 其中 Y_S 表示总体在处理分配 S 下的响应变量值。

1 有关引理

引理 1 设 $Z = (z_{ij})_{n \times n}$, 其中 z_{ij} 为独立未知数。对于任一实函数 z , 定义

$$\frac{\partial z}{\partial Z} = \left[\frac{\partial z}{\partial z_{ij}} \right]_{n \times n}$$

则

(1) 对于任一 n 维常数行向量 t ,

$$\frac{\partial}{\partial Z}(tZt') = t't;$$

(2) 若 $|Z| > 0$, 则 $\frac{\partial}{\partial Z}(\ln |Z|) = (Z^{-1})'$

引理 2 若模型 (3) 成立, 则 $\chi, M_{\chi}\chi_i$ 都与

$$\frac{1}{n_i} \sum_{j=1}^{n_i} (\chi_{ij} - \chi_i)^T (\epsilon_{ij} - \bar{\epsilon}) \text{ 相互独立。}$$

2 结构回归模型

定义 1 对于给定的协变量 χ , 称暴露组与对照组可交换, 如果

$$E(Y_1 | S = 1, \chi) = E(Y_1 | S = 0, \chi)$$

$$E(Y_0 | S = 1, \chi) = E(Y_0 | S = 0, \chi) \quad (2)$$

(2) 式要求, 对于给定的协变量 χ , 暴露组与对照组在暴露状态相同的假设下响应变量的平均值相等。

* 收稿日期: 2002-10-23

基金项目: 国家自然科学基金重点课题资助项目 (39930160)

作者简介: 金华 (1967 年生), 男, 博士, 讲师, 通讯联系人: 方积乾; E-mail: fangjq@gzsums.edu.cn

定义 2 可交换条件下多维结构回归模型定义为:

$$\begin{cases} Y_{ij} = \alpha_t + \lambda_{ij}\beta_t^T + \epsilon_{ij} \\ (\lambda_{ij}, \epsilon_{ij})^T \sim \text{NI}[(\mu_\lambda, 0)^T, \mathbf{A}] \\ j = 1, \dots, n_t, t = 0, 1 \end{cases} \quad (3)$$

其中 $\mathbf{A} = \begin{bmatrix} \Sigma_{\lambda\lambda} & 0 \\ 0 & \sigma_\epsilon^2 \end{bmatrix}$, $\sim \text{NI}$ 表示“服从正态分布且相互独立”, $(\lambda_{ij}, \epsilon_{ij})^T$ 为 $(\lambda_{ij}, \epsilon_{ij})$ 的转置, $\lambda_{ij}, \mu_\lambda, \beta_t$ 均为 k 维行向量, β_t^T 是 β_t 的转置; Y_{ij} 是在第 t 种处理下取得的响应变量的第 j 个观察值, λ_{ij} 是对应于 Y_{ij} 的在协变量上的度量. 这里假设 Y_{ij} 与 λ_{ij} 之间的真实关系是线性的, 线性回归系数为 β_t , ϵ_{ij} 是随机误差分量, λ_{ij} 的协方差矩阵 $\Sigma_{\lambda\lambda}$ 是正定的, 且 $\sigma_\epsilon > 0$. 对于模型 (3), $\alpha_1 - \alpha_0$ 不再是处理 $S = 1$ (相对于 $S = 0$) 的效应向量. 如果对于给定的协变量 λ , 暴露组与对照组可交换, 即 (2) 式成立, 那么处理 $S = 1$ (相对于 $S = 0$) 的平均处理效应为:

$$\begin{aligned} ATE &= E(Y_1) - E(Y_0) = \\ &E_\lambda(E(Y_1 | \lambda) - E(Y_0 | \lambda)) = \\ &E_\lambda(E(Y_1 | S = 1, \lambda) - E(Y_0 | S = 0, \lambda)) = \\ &E_\lambda((\alpha_1 + \lambda\beta_1^T) - (\alpha_0 + \lambda\beta_0^T)) = \\ &(\alpha_1 - \alpha_0) + \mu_\lambda(\beta_1^T - \beta_0^T) \end{aligned} \quad (4)$$

其中 $E_\lambda(\cdot)$ 表示对总体求数学期望.

注: 当 $\beta_1 = \beta_0$ 时, 处理 $S = 1$ (相对于 $S = 0$) 的平均处理效应为:

$$ATE = \alpha_1 - \alpha_0$$

3 模型参数的极大似然估计

定理 1 假设模型 (3) 成立, 则模型参数的极大似然估计分别是

$$\begin{aligned} \hat{\beta}_t &= M_{\lambda Y_t} M_{\lambda\lambda}^{-1}, \\ \hat{\alpha}_t &= \mathbf{Y}_t - \lambda_t \hat{\beta}_t^T = Y_t - \lambda_t M_{\lambda Y_t}^{-1} M_{\lambda\lambda}^T, \\ \hat{\sigma}_t^2 &= M_{Y_t Y_t} - M_{\lambda Y_t}^T M_{\lambda\lambda}^{-1} M_{\lambda Y_t}, \quad t = 0, 1 \\ \hat{\mu}_\lambda &= \bar{\lambda}, \quad \hat{\Sigma}_{\lambda\lambda} = M_{\lambda\lambda} \end{aligned}$$

其中, $\bar{\lambda} = \frac{1}{n_t} \sum_{j=1}^{n_t} \lambda_{ij}$,

$$M_{\lambda\lambda} = \frac{1}{n_t} \sum_{j=1}^{n_t} (\lambda_{ij} - \bar{\lambda})(\lambda_{ij} - \bar{\lambda}),$$

$$Y_t = \frac{1}{n_t} \sum_{j=1}^{n_t} Y_{ij}, \quad \bar{\lambda} = \frac{1}{n_1 + n_0} \sum_{t=0}^1 n_t \bar{\lambda}_t,$$

$$M_{\lambda\lambda} = \frac{1}{n_1 + n_0} \sum_{t=0}^1 \left[\sum_{j=1}^{n_t} (\lambda_{ij} - \bar{\lambda})(\lambda_{ij} - \bar{\lambda}) \right]$$

$$M_{\lambda Y_t} = \frac{1}{n_t} \sum_{j=1}^{n_t} (\lambda_{ij} - \bar{\lambda})(Y_{ij} - Y_t),$$

$$M_{Y_t Y_t} = \frac{1}{n_t} \sum_{j=1}^{n_t} (Y_{ij} - Y_t)^2$$

证明 记 $\mu_\lambda = \mu$, $\Sigma_{\lambda\lambda} = \Sigma$ 并将 σ_ϵ^2 简写为 σ_t , 由模型 (3) 可知

$$(\lambda_{ij}, Y_{ij})^T \sim \text{NI}[(\mu_\lambda, \alpha_t + \mu_\lambda \beta_t^T)^T, \mathbf{A}], \quad j = 1, 2, \dots, n_t; \quad t = 0, 1$$

其中协方差矩阵为

$$\mathbf{A}_t = \begin{bmatrix} \Sigma & \Sigma \beta_t^T \\ \beta_t \Sigma & \beta_t \Sigma \beta_t^T + \sigma_t \end{bmatrix}$$

于是

$$\mathbf{A}_t^{-1} = \begin{bmatrix} \Sigma^{-1} + \beta_t^T \sigma_t^{-1} \beta_t & -\beta_t^T \sigma_t^{-1} \\ -\sigma_t^{-1} \beta_t & \sigma_t^{-1} \end{bmatrix}$$

令

$$\mathbf{B}_t = \begin{bmatrix} \mathbf{B}_{1t} & \mathbf{B}_{0t} \\ \mathbf{B}_{0t}^T & \mathbf{B}_{2t} \end{bmatrix} = \mathbf{A}_t^{-1}$$

则子样 $\{(\lambda_{ij}, Y_{ij}), j = 1, \dots, n_t; t = 0, 1\}$ 的似然函数为:

$$\begin{aligned} L &= \prod_{t=0}^1 (2\pi)^{-\frac{1}{2}|\mathbf{A}_t|} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{n_t} (\lambda_{ij} - \mu, Y_{ij} - (\alpha_t + \mu \beta_t^T)) \mathbf{B}_t \begin{bmatrix} (\lambda_{ij} - \mu)^T \\ (Y_{ij} - (\alpha_t + \mu \beta_t^T))^T \end{bmatrix} \right\} \end{aligned} \quad (5)$$

于是

$$-2 \ln L = \sum_{t=0}^1 [\Delta_t + n_t \ln((2\pi)^2 |\mathbf{A}_t|)]$$

其中

$$\begin{aligned} \Delta_t &= \sum_{j=1}^{n_t} (\lambda_{ij} - \mu) \mathbf{B}_{1t} (\lambda_{ij} - \mu)^T + \\ &2 \sum_{j=1}^{n_t} (\lambda_{ij} - \mu) \mathbf{B}_{0t} (Y_{ij} - (\alpha_t + \mu \beta_t^T)) + \\ &\sum_{j=1}^{n_t} (Y_{ij} - (\alpha_t + \mu \beta_t^T))^2 \mathbf{B}_{2t} \end{aligned}$$

$$\frac{\partial(-2 \ln L)}{\partial \alpha_t} = 0, \text{ 得}$$

$$Y_t = \lambda_t \beta_t^T + \alpha_t \quad (6)$$

从而,

$$Y_t - (\alpha_t + \mu \beta_t^T) = (\lambda_t - \mu) \beta_t^T \quad (7)$$

$$\text{令 } \frac{\partial(-2 \ln L)}{\partial \beta_t} = 0, \text{ 由 (6) 式得}$$

$$\beta_t M_{\lambda\lambda} = M_{\lambda Y_t}^T$$

$$\text{即} \quad \beta_t = M_{\chi_t Y_t}^T M_{\chi_t \chi_t}^{-1} \quad (8)$$

$$\text{令} \frac{\partial(-2\ln L)}{\partial \sigma_t} = 0, \text{联合(7), (8) 式得}$$

$$\sigma_t = M_{Y_t Y_t} - M_{\chi_t Y_t}^T M_{\chi_t \chi_t}^{-1} M_{\chi_t Y_t}$$

$$\text{令} \frac{\partial(-2\ln L)}{\partial(\Sigma^{-1})} = 0, \text{由引理 1 得}$$

$$\Sigma = \frac{1}{n_1 + n_0} \sum_{i=0}^1 \left[\sum_{j=1}^{n_i} (\chi_{ij} - \mu)^T (\chi_{ij} - \mu) \right] \quad (9)$$

$$\text{令} \frac{\partial(-2\ln L)}{\partial \mu} = 0, \text{由(7) 式得}$$

$$\sum_{i=0}^1 n_i (\chi_{ij} - \mu) \Sigma^{-1} = 0,$$

从而

$$\mu = \frac{1}{n_1 + n_0} \sum_{i=0}^1 n_i \chi_i = \chi \quad (10)$$

(10) 式代入 (9) 式得

$$\Sigma = \frac{1}{n_1 + n_0} \sum_{i=0}^1 \left[\sum_{j=1}^{n_i} (\chi_{ij} - \chi)^T (\chi_{ij} - \chi) \right] \quad (11)$$

(8) 式代入 (6) 式得

$$\hat{\alpha}_t = Y_t - \chi_t \beta_t^T = Y_t - \chi_t M_{\chi_t Y_t}^{-1} M_{\chi_t Y_t}^T \quad (12)$$

4 总体平均处理效应极大似然估计

下面给出在可交换条件下, 总体平均处理效应 ATE 的极大似然估计。

定理 2 假设模型 (3) 成立, 如果对于给定的协变量 χ , 暴露组与对照组可交换, 即 (2) 式成立, 那么:

(1) 总体平均处理效应 ATE 的极大似然估计是:

$$E_L = (Y_1 - Y_0) - (\chi_1 - \chi_0) \frac{1}{n_1 + n_0} \sum_{i=0}^1 n_{1-i} \beta_i^T$$

(2) 总体平均处理效应 ATE 的极大似然估计是它的一致无偏估计;

(3) 在大样本下, E_L 渐近服从正态分布 $N(ATE, \sigma^2)$, 其中

$$\sigma^2 = \sum_{i=0}^1 \frac{\sigma_i^2}{n_i} + \frac{1}{n_1 + n_0} (\beta_1 - \beta_0) \Sigma_{\chi\chi} (\beta_1 - \beta_0)^T$$

其极大似然估计为:

$$\sigma^2 = \sum_{i=0}^1 \frac{\sigma_i^2}{n_i} + \frac{1}{n_1 + n_0} (\beta_1 - \beta_0) \hat{\Sigma}_{\chi\chi} (\beta_1 - \beta_0)^T$$

证明

(1) 总体平均处理效应 ATE 的极大似然估计

(4) 式为:

$$ATE = (\alpha_1 - \alpha_0) + \mu_\chi (\beta_1^T - \beta_0^T)$$

整理得:

$$ATE = (\alpha_1 + \mu_\chi \beta_1^T) - (\alpha_0 + \mu_\chi \beta_0^T)$$

由定理 1 知:

$$\hat{\alpha}_t = Y_t - \chi_t \beta_t^T, t = 0, 1,$$

$$\mu_\chi = \chi = \frac{1}{n_1 + n_0} \sum_{i=0}^1 n_i \chi_i$$

代入上式得:

$$E_L = (\hat{\alpha}_1 + \mu_\chi \beta_1^T) - (\hat{\alpha}_0 + \mu_\chi \beta_0^T) =$$

$$(Y_1 - \chi_1 \beta_1^T + \frac{1}{n_1 + n_0} \sum_{i=0}^1 n_i \chi_i \beta_1^T) -$$

$$(Y_0 - \chi_0 \beta_0^T + \frac{1}{n_1 + n_0} \sum_{i=0}^1 n_i \chi_i \beta_0^T) =$$

$$(Y_1 - Y_0) - \frac{n_1 + n_0}{n_1 + n_0} \chi_1 \beta_1^T + \frac{n_1}{n_1 + n_0} \chi_1 \beta_1^T +$$

$$\frac{n_0}{n_1 + n_0} \chi_0 \beta_1^T + \frac{n_1 + n_0}{n_1 + n_0} \chi_0 \beta_0^T -$$

$$\frac{n_1}{n_1 + n_0} \chi_1 \beta_0^T - \frac{n_0}{n_1 + n_0} \chi_1 \beta_0^T = (Y_1 - Y_0) -$$

$$\frac{n_0 \chi_1 \beta_1^T - n_0 \chi_0 \beta_1^T - n_1 \chi_0 \beta_0^T + n_1 \chi_0 \beta_0^T}{n_1 + n_0} =$$

$$(Y_1 - Y_0) - (\chi_1 - \chi_0) \frac{1}{n_1 + n_0} \sum_{i=0}^1 n_{1-i} \beta_i^T$$

所以总体平均处理效应 ATE 的极大似然估计是:

$$E_L = (Y_1 - Y_0) -$$

$$(\chi_1 - \chi_0) \frac{1}{n_1 + n_0} \sum_{i=0}^1 n_{1-i} \beta_i^T$$

(2) 由模型 (3) 及 (8) 式即可得

$$\beta^T = M_{\chi\chi}^{-1} M_{\chi Y_t} =$$

$$\left[M_{\chi\chi}^{-1} \frac{1}{n_t} \sum_{j=1}^{n_t} ((\chi_{tj} - \chi_t)^T (Y_{tj} - Y_t))^T \right]^T =$$

$$\beta_t^T + \frac{1}{n_t} M_{\chi\chi}^{-1} \sum_{j=1}^{n_t} (\chi_{tj} - \chi_t)^T (\epsilon_{tj} - \bar{\epsilon}_t)$$

所以, 由大数定理容易知道, ATE 的极大似然估计 E_L 是它的一致估计, 又利用引理 2, 可知 χ_t 与 β_t^T 相互独立, 而

$$E(Y_t) = \alpha_t + \mu_\chi \beta_t^T$$

$$E(\chi_t) = \mu_\chi$$

$$E(\beta_t^T) = \beta_t^T, t = 0, 1$$

因此,

$$E(E_L) = E(Y_1 - Y_0) -$$

$$E(\chi_1 - \chi_0) \left[\frac{1}{n_1 + n_0} \sum_{i=0}^1 n_{1-i} E(\beta_i^T) \right] = ATE$$

即 E_L 是总体平均处理效应 ATE 的无偏估计。

(3) 注意到

$$E_L \approx (Y_1 - Y_0) -$$

$$(\chi_1 - \chi_0) \left[\frac{1}{n_1 + n_0} \sum_{i=0}^1 n_{1-i} \beta_i^T \right] =$$

$$(\alpha_1 + \chi_1 \frac{n_1}{n_1 + n_0} (\beta_1^T - \beta_0^T) + \bar{\epsilon}_1) -$$

$$(\alpha_0 + \gamma_0 \frac{n_0}{n_1 + n_0} (\beta_0^T - \beta_1^T) + \epsilon_0)$$

其方差为：

$$\begin{aligned} D(E_L) &\approx \\ \sum_{t=0}^1 &\left[\frac{\sigma_t^2}{n_t} + \left(\frac{n_t}{n_1 + n_0} \right)^2 (\beta_1 - \beta_0) D(\gamma) (\beta_1^T - \beta_0^T) \right] = \\ \sum_{t=0}^1 &\left[\frac{\sigma_t^2}{n_t} + \frac{1}{n_1 + n_0} (\beta_1 - \beta_0) \Sigma_{\gamma\gamma} (\beta_1 - \beta_0)^T \right] = \sigma^2 \end{aligned}$$

由中心极限定理知， E_L 渐近服从正态分布 $N(ATE, \sigma^2)$

其中， σ^2 的极大似然估计为：

$$\sigma^2 = \sum_{t=0}^1 \frac{\hat{\sigma}_t^2}{n_t} + \frac{1}{n_1 + n_0} (\beta_1 - \beta_0) \hat{\Sigma}_{\gamma\gamma} (\beta_1 - \beta_0)^T.$$

参考文献：

[1] ROSENBAUM P R, RUBIN D B. The central role of the propensity score in observational studies for causal effects [J] . Biometrika, 1983, 70: 41— 55.

[2] ROBINS J M, GREENLAND S. The role of model selection in causal inference from nonexperimental data[J] . American Journal of Epidemiology, 1986, 123: 393— 402.

[3] ROBINS J M, GREENLAND S. Adjusting for differential rates of prophylaxis therapy for PCP in high versus low dose AZT treatment arms in an AIDS randomized trial[J] . J American Statistical Association, 1994, 89: 737— 749.

[4] ROBINS J M, RITOV Y. Toward a curse-of-dimensionality appropriate (CODA) asymptotic theory for semiparametric models[J] . Statistics in Medicine, 1997, 16: 285— 319.

A Structural Regression Model under the Exchangeability Conditions

JIN Hua, HE Chun, FANG Ji-qian

(Department of Medical Statistics School of Public Health, Sun Yat-sen University, Guangzhou 510080, China)

Abstract: The structural regression: $Y_{ij} = a_t + \gamma_{jt}\beta_t^T + \epsilon_{jt}, (\gamma_{jt}, \epsilon_{jt})^T \sim N[(\mu_\gamma, 0)^T, A], j = 1, \cdots, n_t; t = 0, 1$, is used to study the population-averaged treatment effects under exchangeability conditions when the response variable is a one-dimensional vector. The estimators of the model parameters: $\alpha_t, \beta_t, \sigma_t^2, \mu_\gamma, \Sigma_{\gamma\gamma}$ and the estimation of the population-averaged treatment effects are given by using the maximum likelihood method.

Key words: exchangeability conditions; population-averaged treatment effects; structural regression models; maximum likelihood estimators