应用概率统计 第二十五卷
第一期 2009年2月

Chinese Journal of Applied Probability
and Statistics Vol.25 No.1 Feb. 2009

具有一致相关的纵向数据模型中方差和相关系数的齐性检验 *

范俊花[1,2] 林金官[1⋆] 韦博成[1]

([1]东南大学数学系, 南京, 210096; [2]东南大学成贤学院, 南京, 210088)

摘 要

在纵向数据分析中, 模型方差的齐性是一个基本假定, 但是该假定未必正确. 林金官、韦博成[1]讨论了具有AR(1)误差的非线性纵向数据模型中方差和相关系数的齐性检验. 本文对具有一致相关协方差结构的纵向数据模型, 研究了方差齐性和相关系数齐性的检验, 得到了检验的score统计量, 并应用于葡萄糖数据. 最后, 本文还给出了模拟结果.

关键词: 纵向数据, 一致相关模型, 异方差, 相关系数, score检验.

学科分类号: O212.2.

§1. 引言与模型

纵向数据模型分析是近年来统计学的热点课题之一, 主要用于探索各组受试单元在不同时间或空间上的重复观测数据的统计性质. Diggle et al[2]系统论述了纵向数据分析, 讨论了基于线性和广义线性模型的纵向数据的统计分析. 为了用线性或非线性模型拟合纵向数据, 需要确定一个合理的协方差结构, 协方差结构是否合理直接影响到统计分析. 刻画纵向数据协方差结构有三种可能因素, 即序列相关、随机效应和常规的随机误差. Diggle et al[2], Pinheiro and Bates[3], Laird and Ware[4]等用随机效应和随机误差刻画了线性纵向数据模型, 并在随机效应和随机误差的方差齐性假设下对模型进行了统计推断. 林金官和韦博成[5][6][1]分别研究了非线性纵向数据模型中随机效应的存在性和相关性检验问题, 随机效应和随机误差独立时组间、组内方差的齐性检验问题, 具有一阶自相关误差的非线性纵向数据模型的方差齐性以及相关系数齐性的检验问题等.

一致相关是另一种重要的相关形式, Diggle et al[2], Wolfinger[7]对该相关形式进行了讨论. 本文讨论了具有一致相关协方差结构的纵向数据的相关系数齐性检验和异方差检验的问题, 得到比较便于计算的score检验统计量, 利用所得到的score检验统计量可检验所选择的协方差结构是否合理.

---

*国家自然科学基金(10671032)、江苏省自然科学基金(BK2008284)和东南大学科技基金(基地类)资助.

⋆通讯作者, E-mail: jglin@seu.edu.cn.

本文2005年9月22日收到, 2007年4月1日收到修改稿.

设有m个个体, $(\boldsymbol{Y}_i, X_i)$为第i个个体的观测, $\boldsymbol{Y}_i$为$n_i \times 1$的响应变量, $X_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{in_i})^T$为$n_i \times p$的固定效应设计矩阵, 其观测时间为$t_{i1}, t_{i2}, \cdots, t_{in_i}$, 设$N = \sum\limits_{i=1}^{m} n_i$, $\boldsymbol{\beta}$为p维向量. 基本模型为:

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \varepsilon_{ij}, \qquad i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n_i. \tag{1.1}$$

一致相关性是指同一受试单元中任何两次测量之间具有相同的相关系数ϕ, 即corr$(\varepsilon_{ij}, \varepsilon_{ik}) = \phi_i$ $(j \neq k)$, 而$\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{in_i})^T$, 所以$\mathsf{Var}(\boldsymbol{\varepsilon}_i) = \sigma^2[I_{n_i} + \phi_i(J_{n_i} - I_{n_i})]$, 其中$I_{n_i}$为$n_i \times n_i$的单位阵, J_{n_i}为$n_i \times n_i$阶元素全为1的矩阵. 类似于文献[1], 纵向数据组内方差σ^2可能变异, 与Cook and Weisberg[8]类似, 设$\sigma_{ij}^2 = \sigma^2 m_{ij} = \sigma^2 m(\boldsymbol{z}_{ij}, \boldsymbol{\gamma})$, 且存在$\boldsymbol{\gamma}_0$使得$m(\boldsymbol{z}_{ij}, \boldsymbol{\gamma}_0) = 1$, 记$M_i = \mathrm{diag}(m_{i1}, m_{i2}, \cdots, m_{in_i})$为方差的权函数阵, $V_i = I_{n_i} + \phi_i(J_{n_i} - I_{n_i})$为一致相关结构, 则$\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2 M_i V_i)$. 同样各组一致相关系数$\phi$也可能变异, 参照Núñez-Antón and Zimmerman[9], 可将其参数化为$\phi_i = \phi w(\boldsymbol{v}_i, \boldsymbol{\gamma})$, 存在$\boldsymbol{\gamma}_0$使得相关系数权函数$w(\boldsymbol{v}_i, \boldsymbol{\gamma}_0) = c \neq 0$, 此时$V_i = I_{n_i} + \phi w(\boldsymbol{v}_i, \boldsymbol{\gamma})(J_{n_i} - I_{n_i})$.

本文第2节讨论方差齐性时一致相关系数的存在性检验; 第3节讨论方差齐性时一致相关系数的齐性检验, 一致相关系数齐性时方差齐性检验, 一致相关系数和方差齐性的联合检验. 第4节讨论相关性和方差齐性的联合检验. 第5节将各种情形应用到分析葡萄糖数据. 第6节进行功效模拟.

§2. 一致相关系数的存在性检验

本节首先在方差齐性时(即$M_i = I_{n_i}$), 研究相关系数的存在性检验. 令$\boldsymbol{\theta} = (\phi, \boldsymbol{\beta}^T, \sigma^2)^T$, $\boldsymbol{\theta}$的对数似然函数可表示为:

$$l(\boldsymbol{\theta}) = C - \frac{N}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{m}\log|V_i| - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\boldsymbol{\varepsilon}_i^T V_i^{-1}\boldsymbol{\varepsilon}_i. \tag{2.1}$$

其中C为常数, ϕ为有兴趣参数, 其它为多余参数; $V_i = I_{n_i} + \phi(J_{n_i} - I_{n_i})$, $\boldsymbol{\varepsilon}_i = \boldsymbol{Y}_i - X_i\boldsymbol{\beta}$. 于是, 一致相关系数的存在性检验化为假设检验问题:

$$H_0: \phi = 0; \qquad H_1: \phi \neq 0. \tag{2.2}$$

定理 2.1　对模型(1.1), 检验(2.2)的score统计量为:

$$\mathrm{SC}_\phi = \frac{1}{2\widehat{\sigma}^4}\{\boldsymbol{u}_1^T \mathbf{1}_m (\mathbf{1}_m^T U_2 \mathbf{1}_m)^{-1} \mathbf{1}_m^T \boldsymbol{u}_1\}_{\widehat{\boldsymbol{\theta}}}. \tag{2.3}$$

其中$U_2 = \mathrm{diag}\{\mathrm{tr}((J_{n_i} - I_{n_i})^2)\}$, $\mathbf{1}_m = \underbrace{(1, 1, \cdots, 1)}_{m}{}^T$, $\boldsymbol{u}_1 = (u_{11}, u_{12}, \cdots, u_{1m})^T$, $u_{1i} = \sum\limits_{k,l=1;k\neq l}^{n_i} \widehat{\varepsilon}_{ik}\widehat{\varepsilon}_{il}$.

证明: 为了推导score检验统计量, 现在先求$l(\boldsymbol{\theta})$关于各参数的一阶导数:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \phi}=-\frac{1}{2}\sum_{i=1}^{m}\mathrm{tr}(V_i^{-1}\dot{V}_i)+\frac{1}{2\sigma^2}\sum_{i=1}^{m}\boldsymbol{\varepsilon}_i^TV_i^{-1}\dot{V}_iV_i^{-1}\boldsymbol{\varepsilon}_i,\qquad \dot{V}_i=\frac{\partial V_i}{\partial \phi},$$
$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k}=\frac{1}{\sigma^2}\sum_{i=1}^{m}\boldsymbol{\varepsilon}_i^TV_i^{-1}X_{ik},\qquad \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma^2}=-\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{m}\boldsymbol{\varepsilon}_i^TV_i^{-1}\boldsymbol{\varepsilon}_i.$$

再对其求二阶导数, 根据$I_{\boldsymbol{Y}}(\boldsymbol{\theta})=\mathsf{E}_{\boldsymbol{\theta}}[-\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T]$, 经过较复杂计算并整理成矩阵形式, 并记$U_1=\mathrm{diag}\{\mathrm{tr}(V_i^{-1}(J_{n_i}-I_{n_i}))\}$, $U_2=\mathrm{diag}\{\mathrm{tr}(V_i^{-2}(J_{n_i}-I_{n_i})^2)\}$, $V=\mathrm{diag}\{V_i\}$, $X=(X_1^T,\cdots,X_m^T)^T$, 可得参数$\boldsymbol{\theta}$的Fisher信息阵为:

$$I_{\boldsymbol{Y}}(\boldsymbol{\theta})=\begin{bmatrix}\frac{1}{2}\mathbf{1}_m^TU_2\mathbf{1}_m & \mathbf{0} & \frac{1}{2\sigma^2}\mathbf{1}_m^TU_1\mathbf{1}_m\\ \mathbf{0} & \frac{1}{\sigma^2}X^TV^{-1}X & \mathbf{0}\\ \frac{1}{2\sigma^2}\mathbf{1}_m^TU_1\mathbf{1}_m & \mathbf{0} & \frac{N}{2\sigma^4}\end{bmatrix},$$

由$\partial l(\boldsymbol{\theta})/\partial\phi$, 经过计算与整理可得原假设$H_0:\phi=0$成立时的score函数为: $\partial l(\widehat{\boldsymbol{\theta}})/\partial\phi=[1/(2\widehat{\sigma}^2)]\cdot\mathbf{1}_m^T\boldsymbol{u}_1$. 另一方面, H_0成立时, 在$I_{\boldsymbol{Y}}^{-1}(\boldsymbol{\theta})$中, 对应于$\phi$的子阵为: $I^{\phi\phi}=2(\mathbf{1}_m^TU_2\mathbf{1}_m)^{-1}$. 由Cox and Hinkley[10], (2.2)的score检验统计量:

$$\mathrm{SC}_\phi=\Big[\Big(\frac{\partial l(\boldsymbol{\theta})}{\partial\phi}\Big)^TI^{\phi\phi}\Big(\frac{\partial l(\boldsymbol{\theta})}{\partial\phi}\Big)\Big]\Big|_{\widehat{\boldsymbol{\theta}}}.\tag{2.4}$$

其中$\widehat{\boldsymbol{\theta}}$为参数$\boldsymbol{\theta}$在零假设$H_0$下的最大似然估计, 将$\partial l(\widehat{\boldsymbol{\theta}})/\partial\phi$, $I^{\phi\phi}$代入(2.4)式即得检验$H_0:\phi=0$的score检验统计量为:

$$\mathrm{SC}_\phi=\frac{1}{2\widehat{\sigma}^4}\{\boldsymbol{u}_1^T\mathbf{1}_m(\mathbf{1}_m^TU_2\mathbf{1}_m)^{-1}\mathbf{1}_m^T\boldsymbol{u}_1\}_{\widehat{\boldsymbol{\theta}}}.$$

□

Score检验统计量(score test statistic)是广义似然比检验统计量的一种特殊形式. 它主要用于含有多余参数的复合检验问题. Cox and Hinkley[10], 韦博成等[11]有较详细论述. Score检验与似然比检验相比, 其最大优点是不需要计算备择假设下参数的极大似然估计, 而这在实际计算中往往是很复杂的. 另一方面, 在给定的正则条件下score检验统计量与似然比检验统计量的渐近分布相同, 检验功效相当. 所以近年来, 该统计量在文献中被广泛引用.

为利用(2.3)式检验(2.2), 需要得到SC_ϕ的精确分布或渐近分布, 下面讨论该问题. 假定$\boldsymbol{\theta}$在R^s (s为$\boldsymbol{\theta}$的维数)的紧子集Θ中, 令$\boldsymbol{\xi}=(\phi,\sigma^2)^T$, $\widehat{\boldsymbol{\theta}}=(0,\widehat{\boldsymbol{\beta}}^T,\widehat{\sigma}^2)^T$是参数$\boldsymbol{\theta}$在零假设$H_0$下的最大似然估计. 内点$\boldsymbol{\theta}_0=(0,\boldsymbol{\beta}_0^T,\sigma_0^2)^T$是$\Theta$中在原假设$\phi=0$下$\boldsymbol{\theta}$的真值. 参数$\boldsymbol{\beta}$与Vonesh and Carter[12]有同样假设. 另外假定: 对于$\boldsymbol{\theta}_0$邻域中的任何点$\boldsymbol{\theta}$, 当$m\longrightarrow\infty$时,

(i) $m^{-1}\sum\limits_{i=1}^{m} X_i^T(I_{n_i}+\phi(J_{n_i}-I_{n_i}))^{-1}X_i \longrightarrow A>0$, 则$m^{-1}I_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) \longrightarrow J_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta})>0$;

(ii) $\{n_i\}$一致有界, 则$m^{-1}I_{\sigma^2\sigma^2}(\boldsymbol{\theta}) \longrightarrow J_{\sigma^2\sigma^2}(\boldsymbol{\theta})>0$;

(iii) $m^{-1}I_{\boldsymbol{\xi}\boldsymbol{\xi}}(\boldsymbol{\theta}) \longrightarrow J_{\boldsymbol{\xi}\boldsymbol{\xi}}(\boldsymbol{\theta})>0$;

(iv) $m^{-1}[I_{\phi\phi}(\boldsymbol{\theta})-I_{\phi\sigma^2}(\boldsymbol{\theta})I^{-1}_{\sigma^2\sigma^2}(\boldsymbol{\theta})I_{\phi\sigma^2}(\boldsymbol{\theta})^T] \longrightarrow J_{\phi\phi}(\boldsymbol{\theta})-J_{\phi\sigma^2}(\boldsymbol{\theta})J^{-1}_{\sigma^2\sigma^2}(\boldsymbol{\theta})J_{\phi\sigma^2}(\boldsymbol{\theta})^T>0$,

其中, $I_{ab}(\boldsymbol{\theta})$为$I_{\boldsymbol{Y}}(\boldsymbol{\theta})$中参数(参数向量) (a,b)对应的子块.

下面在此正则条件下给出关于其渐近分布的定理.

定理 2.2　在上述正则条件假定下, 当H_0成立时SC_ϕ的渐近分布为$\chi^2(1)$.

证明:　在上述正则条件下, 据标准的渐近理论, 有: 当$\phi=0$时, (A1) $\widehat{\boldsymbol{\theta}} \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$; (A2) $m^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \xrightarrow{\mathcal{L}} N(0, J^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}_0))$; (A3) $m^{1/2}(\widehat{\sigma}^2-\sigma_0^2) \xrightarrow{\mathcal{L}} N(0, J^{-1}_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0))$; (A4) $m^{-1/2}\partial l(\boldsymbol{\theta})/\partial\boldsymbol{\xi}|_{\boldsymbol{\theta}_0} \xrightarrow{\mathcal{L}} N(0, J^{-1}_{\boldsymbol{\xi}\boldsymbol{\xi}}(\boldsymbol{\theta}_0))$; 则$\partial l(\boldsymbol{\theta})/\partial\phi|_{\widehat{\boldsymbol{\theta}}}$在$(0,\widehat{\boldsymbol{\beta}}^T,\sigma_0^2)$处的泰勒展式为:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial\phi}\Big|_{\widehat{\boldsymbol{\theta}}} = \frac{\partial l(\boldsymbol{\theta})}{\partial\phi}\Big|_{(0,\widehat{\boldsymbol{\beta}}^T,\sigma_0^2)^T} + \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\phi\partial\sigma^2}\Big|_{(0,\widehat{\boldsymbol{\beta}}^T,\sigma^{*2})^T}(\widehat{\sigma}^2-\sigma_0^2). \tag{2.5}$$

其中$\sigma^{*2}=\alpha\widehat{\sigma}^2+(1-\alpha)\sigma_0^2$, $0<\alpha<1$, 因在$\phi=0$时, $\widehat{\sigma}^2 \xrightarrow{\mathcal{P}} \sigma_0^2$, 所以$(0,\widehat{\boldsymbol{\beta}}^T,\sigma_0^{*2})^T \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$, 且有$-m^{-1}\partial^2 l(\boldsymbol{\theta})/(\partial\phi\partial\sigma^2)|_{(0,\widehat{\boldsymbol{\beta}}^T,\sigma_0^{*2})^T} \xrightarrow{\mathcal{P}} J_{\phi\sigma^2}(\boldsymbol{\theta}_0)$, 注意到

$$\frac{\partial l(\boldsymbol{\theta})}{\partial\sigma^2}\Big|_{\widehat{\boldsymbol{\theta}}} = \frac{\partial l(\boldsymbol{\theta})}{\partial\sigma^2}\Big|_{(0,\widehat{\boldsymbol{\beta}}^T,\sigma_0^2)^T} + \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\sigma^2\partial\sigma^2}\Big|_{(0,\widehat{\boldsymbol{\beta}}^T,\widetilde{\sigma}^2)^T}(\widehat{\sigma}^2-\sigma_0^2)=0.$$

其中$\widetilde{\sigma}^2=\alpha'\widehat{\sigma}^2+(1-\alpha')\sigma_0^2$, $0<\alpha'<1$. 显然有$(0,\widehat{\boldsymbol{\beta}}^T,\widetilde{\sigma}^2)^T \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$, 且$-m^{-1}\partial^2 l(\boldsymbol{\theta})/(\partial\sigma^2 \cdot\partial\sigma^2)|_{(0,\widehat{\boldsymbol{\beta}}^T,\widetilde{\sigma}^2)^T} \xrightarrow{\mathcal{P}} J_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)$, 故$m^{1/2}(\widehat{\sigma}^2-\sigma_0^2)$渐近等价于$J^{-1}_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)m^{-1/2}\partial l/\partial\sigma^2|_{(0,\widehat{\boldsymbol{\beta}}^T,\sigma_0^2)^T}$, 将其正态化, 表达式(2.5)可写为下面渐近等式:

$$\begin{aligned} & m^{-1/2}\frac{\partial l(\boldsymbol{\theta})}{\partial\phi}\Big|_{(0,\boldsymbol{\beta}_0^T,\sigma_0^2)^T} - J_{\phi\sigma^2}(\boldsymbol{\theta}_0)J^{-1}_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)m^{-1/2}\frac{\partial l(\boldsymbol{\theta})}{\partial\sigma^2}\Big|_{(0,\boldsymbol{\beta}_0^T,\sigma_0^2)^T} \\ = \; & [1,-J_{\phi\sigma^2}(\boldsymbol{\theta}_0)J^{-1}_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)]\Big(m^{-1/2}\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\xi}}\Big|_{\boldsymbol{\theta}_0}\Big). \end{aligned}$$

由(A4)可知在$\phi=0$下, $m^{-1/2}\partial l(\boldsymbol{\theta})/\partial\phi|_{\widehat{\boldsymbol{\theta}}} \xrightarrow{\mathcal{L}} N(0,(J^{\phi\phi})^{-1}(\boldsymbol{\theta}_0))$, 其中$(J^{\phi\phi})^{-1}(\boldsymbol{\theta}_0) = J_{\phi\phi}(\boldsymbol{\theta}_0)-J_{\phi\sigma^2}(\boldsymbol{\theta}_0)J^{-1}_{\sigma^2\sigma^2}(\boldsymbol{\theta}_0)J^T_{\phi\sigma^2}(\boldsymbol{\theta}_0)$. 而$mI^{\phi\phi} \xrightarrow{\mathcal{P}} J^{\phi\phi}(\boldsymbol{\theta}_0)$, 利用$\widehat{\boldsymbol{\theta}}$的收敛性, 当$m\longrightarrow\infty$时,

$$\mathrm{SC}_\phi = \Big[\Big(m^{-1/2}\frac{\partial l(\boldsymbol{\theta})}{\partial\phi}\Big)^T(mI^{\phi\phi})\Big(m^{-1/2}\frac{\partial l(\boldsymbol{\theta})}{\partial\phi}\Big)\Big]\Big|_{\widehat{\boldsymbol{\theta}}} \xrightarrow{\mathcal{L}} \chi^2(1).$$

□

§3. 一致相关系数和方差的齐性检验

3.1 相关系数的齐性检验

在实际问题中, 相关系数可能有变异, 此时为各受试组一致相关. 参照Núñez-Antón and Zimmerman[19], 可将ϕ_i参数化为

$$\phi_i = \phi w(\boldsymbol{v}_i, \boldsymbol{\gamma}), \tag{3.1}$$

其中, $\boldsymbol{v}_i$是协变量, 对ϕ_i, $i = 1, 2, \cdots, n$, 限定它们的值域为-1到1, 且假设存在$\boldsymbol{\gamma}_0$, 对所有的i, 使得$w(\boldsymbol{v}_i, \boldsymbol{\gamma}_0) = c \neq 0$, 其中$c$是一个与$i$无关的常数. 则一致相关的齐性检验化为如下的参数检验问题:

$$H_0: \boldsymbol{\gamma} = \boldsymbol{\gamma}_0; \qquad H_1: \boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0. \tag{3.2}$$

令$\boldsymbol{\theta} = (\boldsymbol{\gamma}^T, \boldsymbol{\beta}^T, \phi, \sigma^2)^T$且$V_i = I_{n_i} + \phi w(\boldsymbol{v}_i, \boldsymbol{\gamma})(J_{n_i} - I_{n_i})$, 其中$\boldsymbol{\gamma}$为$q$维有兴趣参数向量, 其余为多余参数. $\boldsymbol{\theta}$的对数似然函数为(2.1)式.

定理 3.1 对模型(1.1), 检验(3.2)的score统计量为:

$$\begin{aligned} \mathrm{SC}_{\boldsymbol{\gamma}1} = & \frac{1}{2\widehat{\sigma}^4}\{(\boldsymbol{u}_2 - \sigma^2 U_3 \mathbf{1}_m)^T \dot{W}[\dot{W}^T(A - A\mathbf{1}_m(\mathbf{1}_m{}^T A \mathbf{1}_m)^{-1}\mathbf{1}_m{}^T A)\dot{W}]^{-1} \\ & \cdot \dot{W}^T(\boldsymbol{u}_2 - \sigma^2 U_3 \mathbf{1}_m)\}_{\widehat{\boldsymbol{\theta}}}, \end{aligned} \tag{3.3}$$

其中$U_3 = \mathrm{diag}\{\mathrm{tr}(V_i^{-1}(J_{n_i} - I_{n_i}))\}$, $U_4 = \mathrm{diag}\{\mathrm{tr}(V_i^{-2}(J_{n_i} - I_{n_i})^2)\}$; $\dot{W} = [\partial w(\boldsymbol{v}_i, \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}_k]_{m\times q}$, $\boldsymbol{u}_2 = (u_{21}, u_{22}, \cdots, u_{2m})^T$, $u_{2i} = \widehat{\boldsymbol{\varepsilon}}_i^T \widehat{V}_i^{-1}(J_{n_i} - I_{n_i})\widehat{V}_i^{-1}\widehat{\boldsymbol{\varepsilon}}_i$, $A = U_4 - U_3 \cdot (\mathbf{1}_m \mathbf{1}_m^T / N) \cdot U_3$.

证明: 为了推导score检验统计量, 与定理2.1类似, 先求$l(\boldsymbol{\theta})$关于各参数的一阶导数, 再对其求二阶导, 根据$I_{\boldsymbol{Y}}(\boldsymbol{\theta}) = \mathsf{E}_{\boldsymbol{\theta}}[-\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T]$, 记$W = (w(\boldsymbol{v}_i, \boldsymbol{\gamma}))_{m\times 1}$, 经过较复杂计算并整理成矩阵形式, 可得参数$\boldsymbol{\theta}$的Fisher信息阵为:

$$I_{\boldsymbol{Y}}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\phi^2}{2}\dot{W}^T U_4 \dot{W} & \mathbf{0} & \frac{\phi}{2}\dot{W}^T U_4 W & \frac{\phi}{2\sigma^2}\dot{W}^T U_3 \mathbf{1}_m \\ \mathbf{0} & \frac{1}{\sigma^2}X^T V^{-1} X & \mathbf{0} & \mathbf{0} \\ \frac{\phi}{2}W^T U_4 \dot{W} & \mathbf{0} & \frac{1}{2}\mathbf{1}_m^T W^T U_4 W & \frac{1}{2\sigma^2}W^T U_3 \mathbf{1}_m \\ \frac{\phi}{2\sigma^2}\mathbf{1}_m^T U_3 \dot{W} & \mathbf{0} & \frac{1}{2\sigma^2}\mathbf{1}_m^T U_3 W & \frac{N}{2\sigma^4} \end{bmatrix},$$

由$\partial l(\boldsymbol{\theta})/\partial\boldsymbol{\gamma}$, 经过计算与整理可得原假设$H_0: \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$成立时的score函数为:

$$\frac{\partial l(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma}} = -\frac{\widehat{\phi}}{2}\dot{W}^T U_3 \mathbf{1}_m + \frac{\widehat{\phi}}{2\widehat{\sigma}^2}\dot{W}^T \boldsymbol{u}_2 = \frac{\widehat{\phi}}{2\widehat{\sigma}^2}[\dot{W}^T(\boldsymbol{u}_2 - \sigma^2 U_3 \mathbf{1}_m)]_{\widehat{\boldsymbol{\theta}}}.$$

另一方面, H_0成立时, 在$I_{\boldsymbol{Y}}^{-1}(\boldsymbol{\theta})$中, 对应于$\boldsymbol{\gamma}$的子阵为:

$$I^{\boldsymbol{\gamma\gamma}} = \Big\{\frac{\phi^2}{2}\dot{W}^T[A - A\mathbf{1}_m(\mathbf{1}_m^T A \mathbf{1}_m)^{-1}\mathbf{1}_m^T A]\dot{W}\Big\}^{-1}.$$

由Cox and Hinkley[10], (3.2)的score检验统计量为:

$$\mathrm{SC}_{\gamma 1}=\Big[\Big(\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}\Big)^T I^{\gamma\gamma}\Big(\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}\Big)\Big]\Big|_{\widehat{\boldsymbol{\theta}}}. \tag{3.4}$$

将$\partial l(\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\gamma}$, $I^{\gamma\gamma}$代入可得检验$H_0:\boldsymbol{\gamma}=\boldsymbol{\gamma}_0$的score检验统计量为(3.3)式. □

为检验$H_0:\boldsymbol{\gamma}=\boldsymbol{\gamma}_0$, 还需知道$\mathrm{SC}_{\boldsymbol{\gamma}}$的渐近分布. 为此假定$\boldsymbol{\theta}$在$R^s$ (s为$\boldsymbol{\theta}$的维数)的紧子集Θ中, 令$\boldsymbol{\xi}=(\boldsymbol{\gamma}^T,\phi,\sigma^2)^T$, $\boldsymbol{\eta}=(\phi,\sigma^2)^T$, $\widehat{\boldsymbol{\theta}}=(\boldsymbol{\gamma}_0^T,\widehat{\boldsymbol{\beta}}^T,\widehat{\phi},\widehat{\sigma}^2)^T$是参数$\boldsymbol{\theta}$在零假设$H_0$下的最大似然估计. 内点$\boldsymbol{\theta}_0=(\boldsymbol{\gamma}_0^T,\boldsymbol{\beta}_0^T,\phi_0,\sigma_0^2)^T$是$\Theta$中在原假设$\boldsymbol{\gamma}=\boldsymbol{\gamma}_0$下$\boldsymbol{\theta}$的真值, 另外我们有以下假定: 对于$\boldsymbol{\theta}_0$邻域中的任何点$\boldsymbol{\theta}$, 当$m\longrightarrow\infty$时, 一致的有

(i) $m^{-1}\sum\limits_{i=1}^m X_i^T(I_{n_i}+\phi w(\boldsymbol{v}_i,\boldsymbol{\gamma})(J_{n_i}-I_{n_i}))^{-1}X_i\longrightarrow A(\boldsymbol{\theta})>0$, 且$m^{-1}I_{\boldsymbol{\beta\beta}}(\boldsymbol{\theta})\longrightarrow J_{\boldsymbol{\beta\beta}}(\boldsymbol{\theta})>0$;

(ii) $m^{-1}I_{\boldsymbol{\eta\eta}}(\boldsymbol{\theta})\longrightarrow J_{\boldsymbol{\eta\eta}}(\boldsymbol{\theta})>0$;

(iii) $m^{-1}I_{\boldsymbol{\xi\xi}}(\boldsymbol{\theta})\longrightarrow I_{\boldsymbol{\xi\xi}}(\boldsymbol{\theta})>0$;

(iv) $m^{-1}[I_{\boldsymbol{\gamma\gamma}}(\boldsymbol{\theta})-I_{\boldsymbol{\gamma\eta}}(\boldsymbol{\theta})I_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\theta})I_{\boldsymbol{\gamma\eta}}(\boldsymbol{\theta})^T]\longrightarrow J_{\boldsymbol{\gamma\gamma}}(\boldsymbol{\theta})-J_{\boldsymbol{\gamma\eta}}(\boldsymbol{\theta})J_{\boldsymbol{\eta\eta}}^{-1}(\boldsymbol{\theta})J_{\boldsymbol{\gamma\eta}}(\boldsymbol{\theta})^T>0$;

其中, $I_{ab}(\boldsymbol{\theta})$为$I_{\boldsymbol{Y}}(\boldsymbol{\theta})$中参数(参数向量) (a,b)对应的子块.

与定理2.2类似, 在上述正则条件假定下, 当H_0成立时$\mathrm{SC}_{\boldsymbol{\gamma}1}$的渐近分布为$\chi^2(q)$. 因此当$H_0$成立时, 统计量$\mathrm{SC}_{\boldsymbol{\gamma}1}$渐近服从$\chi^2(q)$分布.

3.2 异方差检验

若受试单元之间的相关系数是齐性的, 即$\phi_i\equiv\phi$, 但方差σ_{ij}^2可能不一样, 将其参数化(Cook and Weisberg[8]), 并表示为以下参数化形式:

$$\sigma_{ij}^2=\sigma^2 m_{ij}=\sigma^2 m(\boldsymbol{z}_{ij},\boldsymbol{\gamma}). \tag{3.5}$$

其中$\boldsymbol{z}_{ij}$为协变量, $m(\cdot,\cdot)$关于参数$\boldsymbol{\gamma}$二阶可微且存在$\boldsymbol{\gamma}_0$, 对所有的$i=1,2,\cdots,m$, $j=1,2,\cdots,n_i$, 有$m(\boldsymbol{z}_{ij},\boldsymbol{\gamma}_0)=1$, 于是模型的异方差检验化为$H_0:\boldsymbol{\gamma}=\boldsymbol{\gamma}_0$.

记$\boldsymbol{\theta}=(\boldsymbol{\gamma}^T,\boldsymbol{\beta}^T,\phi,\sigma^2)^T$, $\boldsymbol{\theta}$的对数似然函数可由(2.1)式化为:

$$l_h(\boldsymbol{\theta})=C-\frac{N}{2}\log\sigma^2-\frac{1}{2}\sum_{i=1}^m\log|M_iV_i|-\frac{1}{2\sigma^2}\sum_{i=1}^m\boldsymbol{\varepsilon}_i^TV_i^{-1}M_i^{-1}\boldsymbol{\varepsilon}_i. \tag{3.6}$$

通过对$l_h(\boldsymbol{\theta})$求导即可得到所需的统计量. 下面首先对$l_h(\boldsymbol{\theta})$求一阶偏导数有:

$$\frac{\partial l_h(\boldsymbol{\theta})}{\partial\gamma_k}=-\frac{1}{2}\sum_{i=1}^m\mathrm{tr}(M_i^{-1}\dot{M}_{ik})+\frac{1}{2\sigma^2}\sum_{i=1}^m\boldsymbol{\varepsilon}_i^TV_i^{-1}M_i^{-1}\dot{M}_{ik}M_i^{-1}\boldsymbol{\varepsilon}_i,\qquad \dot{M}_{ik}=\frac{\partial M_i}{\partial\gamma_k},$$

$$\frac{\partial l_h(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}_k}=\frac{1}{\sigma^2}\sum_{i=1}^m\boldsymbol{\varepsilon}_i^TV_i^{-1}M_i^{-1}X_{ik},$$

$$\frac{\partial l_h(\boldsymbol{\theta})}{\partial \phi} = -\frac{1}{2}\sum_{i=1}^m \mathrm{tr}(V_i^{-1}\dot{V}_{i\phi}) + \frac{1}{2\sigma^2}\sum_{i=1}^m \boldsymbol{\varepsilon}_i^T V_i^{-1}\dot{V}_{i\phi}V_i^{-1}M_i^{-1}\boldsymbol{\varepsilon}_i, \qquad \dot{V}_{i\phi} = \frac{\partial V_i}{\partial \phi},$$

$$\frac{\partial l_h(\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^m \boldsymbol{\varepsilon}_i^T V_i^{-1}M_i^{-1}\boldsymbol{\varepsilon}_i.$$

再对其求二阶导及负期望, 经过计算整理可得$\boldsymbol{\theta}$的Fisher信息阵$I_{\mathbf{Y}}(\boldsymbol{\theta})$的非零子块分别为:

$$I_{\boldsymbol{\gamma\gamma}}^{(k,l)} = \frac{1}{2}\sum_{i=1}^m \mathrm{tr}(\dot{M}_{ik}\dot{M}_{il}), \qquad I_{\boldsymbol{\gamma}\phi}^{(k)} = \frac{1}{2}\sum_{i=1}^m \mathrm{tr}(\dot{V}_{i\phi}V_i^{-1}\dot{M}_{ik}) = I_{\phi\boldsymbol{\gamma}}^{(k)},$$

$$I_{\boldsymbol{\gamma}\sigma^2}^{(k)} = \frac{1}{2\sigma^2}\sum_{i=1}^m \mathrm{tr}(\dot{M}_{ik}) = I_{\sigma^2\boldsymbol{\gamma}}^{(k)}, \qquad I_{\boldsymbol{\beta\beta}} = \frac{1}{\sigma^2}X^TV^{-1}X,$$

$$I_{\phi\phi} = \frac{1}{2}\sum_{i=1}^m \mathrm{tr}(\dot{V}_{i\phi}V_i^{-1}\dot{V}_{i\phi}V_i^{-1}), \qquad I_{\sigma^2\sigma^2} = \frac{N}{2\sigma^4},$$

$$I_{\phi\sigma^2} = \frac{1}{2\sigma^2}\sum_{i=1}^m \mathrm{tr}(\dot{V}_{i\phi}V_i^{-1}) = I_{\sigma^2\phi},$$

其中$\dot{M}_{ik}, \dot{M}_{il}$在$\boldsymbol{\gamma}_0$处记值. 在假设$H_0: \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$成立的条件下, $\boldsymbol{\theta}$的Fisher信息阵的形式可表示为:

$$I_{\mathbf{Y}}(\boldsymbol{\theta}) = \begin{bmatrix} I_{\boldsymbol{\gamma\gamma}} & 0 & I_{\boldsymbol{\gamma}\phi} & I_{\boldsymbol{\gamma}\sigma^2} \\ 0 & I_{\boldsymbol{\beta\beta}} & 0 & 0 \\ I_{\phi\boldsymbol{\gamma}} & 0 & I_{\phi\phi} & I_{\phi\sigma^2} \\ I_{\sigma^2\boldsymbol{\gamma}} & 0 & I_{\sigma^2\phi} & I_{\sigma^2\sigma^2} \end{bmatrix}.$$

因此可知

$$I^{\boldsymbol{\gamma\gamma}} = (I_{\boldsymbol{\gamma\gamma}} - I_1I_2^{-1}I_1^T)^{-1}, \tag{3.7}$$

其中$I^{\boldsymbol{\gamma\gamma}}$表示$I_{\mathbf{Y}}(\boldsymbol{\theta})$的逆阵$I_{\mathbf{Y}}^{-1}(\boldsymbol{\theta})$中与$I_{\boldsymbol{\gamma\gamma}}$所对应的子块,

$$I_1 = [I_{\boldsymbol{\gamma}\phi}, I_{\boldsymbol{\gamma}\sigma^2}], \qquad I_2 = \begin{pmatrix} I_{\phi\phi} & I_{\phi\sigma^2} \\ I_{\sigma^2\phi} & I_{\sigma^2\sigma^2} \end{pmatrix}.$$

由$\partial l_h(\boldsymbol{\theta})/\partial \boldsymbol{\gamma}_k$可知score函数为:

$$\frac{\partial l_h(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma}} = \Big\{ -\frac{1}{2}\sum_{i=1}^m \mathrm{tr}(\dot{M}_{ik}) + \frac{1}{2\sigma^2}\sum_{i=1}^m \widehat{\boldsymbol{\varepsilon}}_i^T V_i^{-1}\dot{M}_{ik}\widehat{\boldsymbol{\varepsilon}}_i \Big\}_{q\times 1}.$$

将$\partial l_h(\widehat{\boldsymbol{\theta}})/\partial \boldsymbol{\gamma}$, $I^{\boldsymbol{\gamma\gamma}}$代入(3.4)式即可求出此检验的score统计量$\mathrm{SC}_{\boldsymbol{\gamma}2}$, 且其服从$\chi^2(q)$分布.

3.3 相关系数和方差齐性的联合检验

现在考虑一般情形, 一致相关系数和方差都有可能是变异的, 因此可将它们联合参数化为(3.1)和(3.5). 正如Zhang and Weiss[19]所述, 为避免参数繁冗, 我们假定$\boldsymbol{\gamma}$作为所有异

方差或非齐相关系数的共同参数, 即在(3.1)和(3.5)中采用同样的参数$\boldsymbol{\gamma}$. 于是, 与前两小节情形一样, 方差齐性和相关系数齐性的联合检验亦可化为假设检验问题(3.2).

记参数$\boldsymbol{\theta}=(\boldsymbol{\gamma}^T,\boldsymbol{\beta}^T,\phi,\sigma^2)^T$, 对以上假设检验问题, $\boldsymbol{\gamma}$为有兴趣参数, 设其为q维未知向量, 其余为多余参数. 这时$\boldsymbol{\theta}$的对数似然函数同为(3.6). 沿用前述记号, 当H_0成立时, 检验的score函数为

$$\frac{\partial l_h(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\gamma}}=\Big\{-\frac{1}{2}\sum_{i=1}^m\mathrm{tr}(\dot{M}_{ik}+V_i^{-1}\dot{V}_{ik})+\frac{1}{2\sigma^2}\sum_{i=1}^m\boldsymbol{\varepsilon}_i^T(V_i^{-1}\dot{V}_{ik}V_i^{-1}+V_i^{-1}\dot{M}_{ik})\boldsymbol{\varepsilon}_i\Big\}_{q\times1}.$$

与前面方法类似, 当H_0成立时, 可得到

$$I_{\boldsymbol{\gamma\gamma}}^{(k,l)}=\frac{1}{2}\sum_{i=1}^m\mathrm{tr}(V_i^{-1}\dot{V}_{ik}V_i^{-1}\dot{V}_{il}+2V_i^{-1}\dot{V}_{ik}\dot{M}_{il}+\dot{M}_{ik}\dot{M}_{il}),$$
$$I_{\boldsymbol{\gamma}\phi}^{(k)}=-\frac{1}{2}\sum_{i=1}^m\mathrm{tr}(V_i^{-1}\dot{V}_{ik}V_i^{-1}\dot{V}_{i\phi}+V_i^{-1}\dot{V}_{i\phi}\dot{M}_{ik})=I_{\phi\boldsymbol{\gamma}}^{(k)},$$
$$I_{\phi\phi}=-\frac{1}{2}\sum_{i=1}^m\mathrm{tr}(V_i^{-1}\dot{V}_{ik}+\dot{M}_{ik}).$$

其余子块与第3.2节的相同. 将其代入(3.4)式即可求得联合检验的score统计量$\mathrm{SC}_{\boldsymbol{\gamma}3}$, 且其服从$\chi^2(q)$分布.

§4. 相关性和方差齐性的联合检验

正如Tsai[13]指出的, 在时间序列数据中, 相关性和异方差性可能同时发生, 因此有必要同时检验模型的一致相关和异方差的存在性, 即检验是否有$\phi=0$, $\sigma_{ij}^2\equiv\sigma^2$. Tsai[13]讨论了非重复测量数据的线性回归模型的异方差和自相关的存在性检验; 韦博成和胡跃清[14]将Tsai[13]的结论推广到非线性回归模型; 林金官和韦博成[1]将其推广到误差为一阶自相关的非线性纵向数据模型. 本节将其推广到具有一致相关协方差结构的纵向数据模型中. 记$\boldsymbol{\psi}=(\boldsymbol{\gamma}^T,\phi)^T$, $\boldsymbol{\psi}_0=(\boldsymbol{\gamma}_0^T,0)^T$, 于是模型的相关性和异方差存在性的同时检验为$H_0:\boldsymbol{\psi}=\boldsymbol{\psi}_0$. 其对数似然函数同第3节, 记为$l_c(\boldsymbol{\theta})$.

由第3节中的$\partial l/\partial\boldsymbol{\gamma}_k$和$\partial l/\partial\phi$, 可得检验的score函数为:

$$\frac{\partial l_c(\boldsymbol{\theta})}{\partial\boldsymbol{\psi}}=\begin{bmatrix}\Big(-\frac{1}{2}\sum_{i=1}^m\mathrm{tr}(\dot{M}_{ik})+\frac{1}{2\sigma^2}\sum_{i=1}^m\boldsymbol{\varepsilon}_i^T\dot{M}_{ik}\boldsymbol{\varepsilon}_i\Big)_{q\times1}\\ -\frac{1}{2}\sum_{i=1}^m\mathrm{tr}(\dot{V}_{i\phi})+\frac{1}{2\sigma^2}\sum_{i=1}^m\boldsymbol{\varepsilon}_i^T\dot{V}_{i\phi}\boldsymbol{\varepsilon}_i\end{bmatrix},$$

再由$I_{\boldsymbol{Y}}(\boldsymbol{\theta})$, 当$H_0$成立时, 有

$$I^{\boldsymbol{\psi\psi}}=[I_{\boldsymbol{\psi\psi}}(\boldsymbol{\theta})-I_{\boldsymbol{\psi}\sigma^2}I_{\sigma^2\sigma^2}^{-1}I_{\boldsymbol{\psi}\sigma^2}^T]^{-1}=\begin{pmatrix}I_{\boldsymbol{\gamma\gamma}}-I_{\boldsymbol{\gamma}\sigma^2}I_{\sigma^2\sigma^2}^{-1}I_{\boldsymbol{\gamma}\sigma^2}^T & I_{\boldsymbol{\gamma}\phi}\\ I_{\phi\boldsymbol{\gamma}} & I_{\phi\phi}^{-1}\end{pmatrix}$$

($I^{\psi\psi}$为$\boldsymbol{\theta}$的Fisher阵$I(\boldsymbol{\theta})=\mathsf{E}[-\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T]$的逆矩阵$I^{-1}(\boldsymbol{\theta})$对应于$\boldsymbol{\psi}$的子块), 其中,

$$I_{\gamma\gamma}^{(k,l)}=\frac{1}{2}\sum_{i=1}^m \mathrm{tr}(\dot{M}_{ik}\dot{M}_{il}),$$
$$I_{\gamma\phi}^{(k)}=-\frac{1}{2}\sum_{i=1}^m \mathrm{tr}(\dot{V}_{i\phi}\dot{M}_{ik})=I_{\phi\gamma}^{(k)},$$
$$I_{\phi\phi}=-\frac{1}{2}\sum_{i=1}^m \mathrm{tr}(\dot{V}_{i\phi}\dot{V}_{i\phi}).$$

其中$\dot{M}_{ik},\dot{M}_{il},\dot{V}_{i\phi}$在$\boldsymbol{\psi}_0$处记值, 其余子块与前同.

由Cox and Hinkley[10], $H_0:\boldsymbol{\psi}=\boldsymbol{\psi}_0$的score检验统计量为:

$$\mathrm{SC}_{\boldsymbol{\psi}}=\Big[\Big(\frac{\partial l_c(\boldsymbol{\theta})}{\partial\boldsymbol{\psi}}\Big)^T I^{\psi\psi}\Big(\frac{\partial l_c(\boldsymbol{\theta})}{\partial\boldsymbol{\psi}}\Big)\Big]\Big|_{\widehat{\boldsymbol{\theta}}}, \tag{4.1}$$

其中$\widehat{\boldsymbol{\theta}}=(\boldsymbol{\psi}_0^T,\widehat{\boldsymbol{\beta}},\widehat{\sigma}^2)^T$是参数$\boldsymbol{\theta}$在零假设$H_0$下的最大似然估计. 将以上各式代入(4.1)可得检验的score统计量$\mathrm{SC}_{\boldsymbol{\psi}}$. 在类似前述正则条件假定下, 当$H_0$成立时$\mathrm{SC}_{\boldsymbol{\psi}}$的渐近分布为$\chi^2(q+1)$.

§5. 应用实例

葡萄糖数据 葡萄糖数据是由美国科罗拉多州医疗中心大学小儿科临床研究病房提供, Zerbe[15]将其作为成长曲线数据讨论其随机化检验, Chi and Reisel[16]为此数据建立具有AR(1)误差的随机效应纵向数据模型, Pan and Fang[17]将数据作为成长曲线数据建立模型并进行了诊断分析. 该数据通过对13个控制病人和20个肥胖病人测试其标准葡萄糖忍耐力. 实验过程为: 让这33个病人服用葡萄糖, 分别在0, 0.5, 1, 1.5, 2, 3, 4, 5小时后测试其血样. 这是一个典型的与时间有关的纵向数据, 实验目的是为研究比较控制组的病人和肥胖组的病人是否有显著区别. 我们采用其13个控制病人的数据来进行研究, 对于此数据集, 应用下列线性纵向数据模型进行拟合与分析:

$$Y_i=X_i^T\boldsymbol{\beta}+\boldsymbol{\varepsilon}_i,\qquad \boldsymbol{\varepsilon}_i\sim N(0,\sigma^2M_iV_i),\quad i=1,2,\cdots,13; \tag{5.1}$$

此处$Y_i=(y_{ij})$, 其中$j=1,2,\cdots,8$; y_{ij}为第i个控制病人第j次测试结果, Pan and Fang[17]采用的设计阵为

$$X_i=\begin{bmatrix}1&1&1&1&1&1&1&1\\0&0.5&1&1.5&2&3&4&5\\0&0.25&1&2.25&4&9&16&25\\0&0.125&1&3.375&8&27&64&125\end{bmatrix}.$$

我们将讨论以下问题: (a)方差齐性时的一致相关系数的存在性检验; (b)方差齐性时的一致相关系数的齐性检验; (c)一致相关系数齐性时方差齐性的检验; (d)方差齐性和一致相关系

数存在性的联合检验; (e)方差和一致相关系数齐性的联合检验. 我们利用$m_{ij}=\exp(z_{ij}\gamma)$ (取$\boldsymbol{z}_i=(z_{ij})^T=(0,0.5,1,1.5,2,3,4,5)^T$) 作为控制病人标准葡萄糖忍耐力的方差权函数; 取$w_i=\exp(v_i\gamma)/[1+\exp(v_i\gamma)]$, 其中$v_i$为各控制病人的属性量. 于是, 当$\gamma=0$时, $m_{ij}=1$及$w_i=1/2$, 两者都与i无关. 因而, 异方差或自相关的齐性检验为$H_0:\gamma=0$; $H_1:\gamma\neq 0$. 对情形(a), (d), 在方差齐性和$\phi=0$时, 参数$\boldsymbol{\beta},\sigma^2$的估计列于表1, 对情形(b), (c)和(e), 参数$\boldsymbol{\beta},\sigma^2$和$\phi$的估计列于表2.

表1 情形(a)和(d)的参数估计及其标准误

参数	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_3$	$\boldsymbol{\beta}_4$	σ^2
估计	4.0202	-1.8746	0.7854	-0.0834	0.3395
标准误	0.1452	0.2759	0.1368	0.0182	0.0471

表2 情形(b), (c)和(e)的参数估计及其标准误

参数	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_3$	$\boldsymbol{\beta}_4$	σ^2	ϕ
估计	4.0040	-1.7702	0.7370	-0.0777	0.2849	0.3782
标准误	0.1389	0.1993	0.0988	0.0132	0.0395	0.1681

利用前述结果, 计算出各检验的score统计量的值, 列于表3.

表3 异方差或自相关系数的齐性检验

检验	(a)	(b)	(c)	(d)	(e)
score统计量	**31.8252**	1.5152	0.0068	**52.0851**	0.4204
df	1	1	1	2	1
p-值	0.00	0.2183	0.9342	0.00	0.5168

由表3可知, 葡萄糖数据具有明显的一致相关结构, 且方差和一致相关系数具有显著齐性.

根据Vonesh and Chinchilli[18], 在正态假设下, 模型的选择可用AIC (Akaike's information criterion)和SBC (Schwarz's Bayesian information criterion)来判断. 我们各自计算出它们的AIC和SBC, 其结果列在表4中. 从该表中, 可以看出情形(a)和(d)协方差结构的AIC和SBC是最小的, 因此, 情形(a)和(d)之协方差结构是五种协方差结构中较好的协方差结构, 这与前面的诊断结果是相当一致的.

表4 葡萄糖数据的AIC和SBC

协方差结构	(a)	(b)	(c)	(d)	(e)
df	1	1	1	2	1
AIC	**-191.9698**	-170.6458	-170.6458	**-182.8549**	-170.6458
SBC	**-198.5807**	-178.5789	-178.5789	**-189.4659**	-178.5789

§6. 功效模拟

本节研究下列情形的模拟功效: (i)方差齐性时的一致相关系数的存在性检验; (ii)一致相关系数齐性时方差齐性的检验; (iii)方差齐性和一致相关系数存在性的联合检验; (iv)方差和一致相关系数齐性的联合检验. 为此我们考虑下列具有一致相关协方差结构的线性纵向数据模型:

$$y_{ij}=\boldsymbol{\beta}_1+x_{ij}\boldsymbol{\beta}_2+x_{ij}^2\boldsymbol{\beta}_3+x_{ij}^3\boldsymbol{\beta}_4+\boldsymbol{\varepsilon}_{ij},\qquad i=1,2,\cdots,m,\ j=1,2,\cdots,n,$$

其中m为测试单元个数, n为给定的重复测量次数, 且$\boldsymbol{\varepsilon}_i\sim N(0,\sigma^2M_iV_i)$, 其中$V_i=I_n+\phi_i(J_n-I_n)$, $M_i=\mathrm{diag}(m_{i1},m_{i2},\cdots,m_{in})$. 对于一致相关系数$\phi_i$和方差权函数$m_{ij}$我们假定它们具有如下形式:

$$m_{ij}=\exp\{z_{ij}\gamma\},\qquad \phi_i=\phi\frac{\exp\{v_i\gamma\}}{1+\exp\{v_i\gamma\}};\qquad i=1,2,\cdots,m,\ j=1,2,\cdots,n.$$

对前述讨论的几种检验问题, 选择若干m和n, 重复模拟1000次. 在每次模拟中计算相应的统计量SC的值, 并与0.05的显著性水平下的临界值相比, 若该值大于临界值, 则拒绝原假设H_0, 最后计算出1000次模拟中, 拒绝的次数百分比, 即为检验的近似功效.

1. 情形(i)具体模拟过程:

(1) 首先产生协变量$\{x_{ij},i=1,2,\cdots,m,j=1,2,\cdots,n\}$, 它们均为取自于[0,5]上的均匀随机数. 再对给定的ϕ和$\sigma^2=0.3$, 自$N(0,\sigma^2)$内产生具有相关性的误差序列$\{\varepsilon_{ij},i=1,2,\cdots,m,j=1,2,\cdots,n\}$.

(2) 对给定的$\boldsymbol{\beta}_1=4$, $\boldsymbol{\beta}_2=-2$, $\boldsymbol{\beta}_3=0.7$, $\boldsymbol{\beta}_4=-0.1$及得到的随机误差序列产生具有自相关误差的线性纵向数据序列$\{y_{ij},i=1,2,\cdots,m,j=1,2,\cdots,n\}$.

(3) 选择若干m和n, 对每一个ϕ值, 重复模拟1000次.

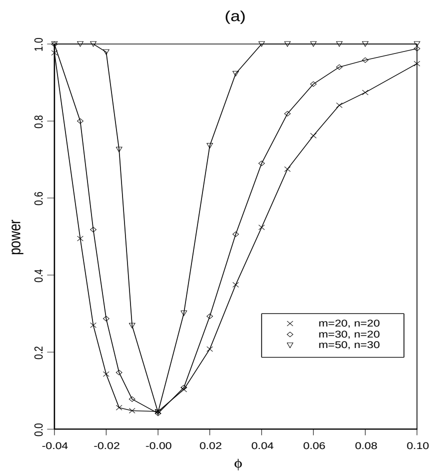

图1 方差齐性时一致相关系数存在性检验的功效模拟图

对于此检验问题, (m,n)分别取(20,20), (30,20), (50,30)作模拟, 其功效模拟图见图1. 由图可知它们在$\alpha=0.05$的条件下的水平均与0.05非常靠近(0.048,0.051). 当(m,n)取(20, 20)时, 检验统计量明显取得了较好的功效. 在原假设$H_0:\phi=0$附近, 当$|\phi|$增加时, 检验功效均增加, 且样本量愈大, 功效增加愈快. 从图形还可看出, 对设定的模拟模型来说, 当(m,n)取到(20,20)和(30,20)时, 功效图是左偏态的, 即检验统计量对$\phi<0$的检验较对$\phi>0$的检验更敏感, 但当(m,n)取到(50,30)时, 功效图趋于对称, 检验统计量对$\phi<0$和$\phi>0$的检验的敏感性接近.

2. 情形(ii)和(iv)具体模拟过程:

(1) 与1类似产生协变量$\{x_{ij}\}$, 再对给定的γ和$\sigma^2=0.3$产生具有相关性的误差序列$\{\varepsilon_{ij}\}$.

(2) 产生方差协变量$\{z_{ij}\}$, 它们是取自于[0,5]上的均匀随机数. 再对给定的γ, 据$m_{ij}=\exp\{z_{ij}\gamma\}$产生序列$\{m_{ij}\}$.

(3) 产生相关系数协变量$\{v_i, i=1,2,\cdots,m\}$, 它们是取自于[0,13]上的均匀随机数. 再对给定的γ, 产生一致相关系数序列ϕ_i $(i=1,2,\cdots,m)$.

(4) 对给定的$\boldsymbol{\beta}_1=4$, $\boldsymbol{\beta}_2=-2$, $\boldsymbol{\beta}_3=0.7$, $\boldsymbol{\beta}_4=-0.1$及得到的随机误差序列产生具有一致相关误差的线性纵向数据序列$\{y_{ij}\}$.

(5) 选择若干m和n, 对每一个γ值, 重复模拟1000次.

图2为一致相关系数齐性时异方差检验的功效模拟图; 图3为一致相关系数和方差齐性的联合检验功效模拟图. 从图中可看出, 在原假设$H_0:\gamma=0$成立时, 检验的功效在0.05附近, 说明检验是好的. 当$|\gamma|$增加时, 检验功效迅速增加.图形关于γ几乎是对称的. 还可以看出检验(ii)和(iv)在中小样本时, 即可得到理想功效.

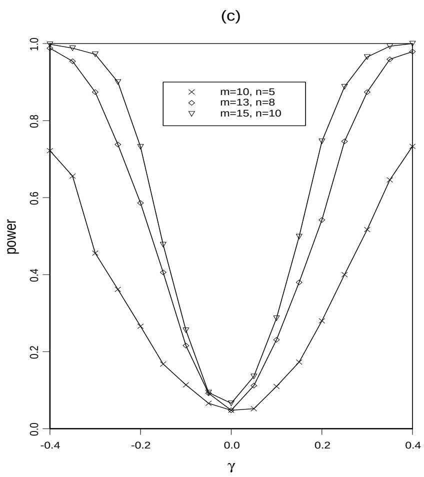

图2　一致相关系数齐性时异方差检验的功效模拟图

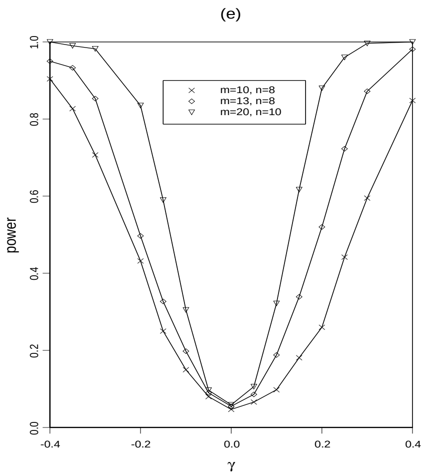

图3　一致相关系数和方差齐性检验的功效模拟图

3. 情形(iii)具体模拟过程:

(1) 同前面方法产生协变量$\{x_{ij}\}$, 再对给定的ϕ,γ和$\sigma^2=0.3$产生具有相关性的误差序列$\{\varepsilon_{ij}\}$, 并且产生序列$\{m_{ij}\}$.

(2) 对给定的$\boldsymbol{\beta}_1=4$, $\boldsymbol{\beta}_2=-2$, $\boldsymbol{\beta}_3=1$, $\boldsymbol{\beta}_4=-0.1$及得到的随机误差序列产生具有一致相关误差的线性纵向数据序列$\{y_{ij}\}$.

(3) 选择若干m和n, 对每一个ϕ和γ的值, 重复模拟1000次.

对于此检验问题, (m,n)分别取(13,8), (20,10)作模拟, 发现在原假设$H_0:\boldsymbol{\psi}=\boldsymbol{\psi}_0$条件下, 检验功效在0.05附近, 说明检验是好的. 且在原假设$H_0:\boldsymbol{\psi}=\boldsymbol{\psi}_0$附近, 当$\boldsymbol{\psi}$中$|\phi|$和$|\boldsymbol{\gamma}|$增加时, 检验功效均增加, 且样本量愈大, 功效增加愈快. 下表5即为检验的功效模拟结果.

表5　自相关性和异方差同时检验的模拟结果

(m,n)	$\boldsymbol{\gamma}\backslash\phi$	-0.05	-0.03	-0.02	-0.01	0	0.01	0.05	0.1	0.15	0.2	0.3
(13,8)	0					**0.049**						
	0.05	0.070	0.057	0.044	0.060	0.065	0.069	0.157	0.360	0.583	0.755	0.936
	0.1	0.163	0.116	0.107	0.108	0.114	0.148	0.221	0.592	0.615	0.772	0.937
	0.2	0.423	0.408	0.386	0.401	0.404	0.400	0.461	0.592	0.764	0.850	0.968
	0.3	0.768	0.737	0.708	0.729	0.726	0.714	0.768	0.827	0.893	0.942	0.988
(20,10)	0					**0.044**						
	0.01	0.154	0.067	0.045	0.039	0.039	0.079	0.271	0.606	0.834	0.942	0.996
	0.05	0.186	0.077	0.087	0.069	0.074	0.115	0.317	0.649	0.856	0.944	0.995
	0.1	0.385	0.246	0.225	0.213	0.227	0.260	0.431	0.714	0.901	0.964	0.998
	0.15	0.633	0.462	0.437	0.446	0.460	0.463	0.609	0.807	0.920	0.980	1
	0.2	0.830	0.729	0.725	0.711	0.685	0.725	0.776	0.899	0.969	0.993	0.997

§7. 结论及进一步的问题

对于一致相关系数和异方差的检验问题, 由于score检验只需计算原假设条件下的估计, 所以被广泛使用. 在前面几节, 关于一致相关协方差结构我们推导出检验一致相关系数和异方差的五个score检验统计量. 通过实例说明这些检验与以前的其它方法(比如AIC和SBC)可保持统一. 随机模拟又说明检验与样本量大小密切相关, 我们所作模拟均在中样本时就有比较好的模拟功效, 到大样本时效果更加好. 本文理论均可推广到非线性情形.

致谢 感谢审稿专家提出的宝贵的修改意见.

参 考 文 献

[1] 林金官, 韦博成, 非线性纵向数据模型中方差和自相关系数的齐性检验, *应用数学学报*, **27(3)**(2004), 466–480.

[2] Diggle, P.J., Heagerty, P., Liang, K.-Y. and Zeger, S.L., *Analysis of Longitudinal data*, New York, Oxford University Press, 2002.

[3] Pinheiro, J.C. and Bates, D.M., *Mixed-Effects Models in S and S-PLUS*, New York, Springer-Verlag, 2000.

[4] Laird, N.M. and Ware, J.H., Random-effects models for longitudinal data, *Biometrics*, **38(4)**(1982), 963–974.

[5] 林金官, 韦博成, 非线性纵向数据模型中自相关性和随机效应的存在性检验, *应用数学*, **17(1)**(2004), 42–48.

[6] 林金官, 韦博成, 非线性随机效应模型的异方差性检验, *系统科学与数学*, **22(2)**(2002), 245–256.

[7] Wolfinger, R.D., Heterogeneous variance-covariance structures for repeated measures, *Journal of Agricultural, Biological and Enviromental Statistics*, **1(2)**(1995), 205–230.

[8] Cook, R.D. and Weisberg, S., Diagnostic for heteroscedasticity in regression, *Biometrika*, **70**(1983), 1–10.

[9] Núñez-Antón, V. and Zimmerman, D.L., Modelling nonstationary longitudinal data, *Biometrics*, **56**(2000), 699–705.

[10] Cox, D.R. and Hinkley, D.V., *Theoretical Statistics*, London, Chapman and Hall, 1974.

[11] 韦博成, 鲁国斌, 史建清, *统计诊断引论*, 南京, 东南大学出版社, 1991.

[12] Vonesh, E.F. and Carter, R.L., Mixed-effects nonlinear models regression for unbalanced repeated measures, *Biometrics*, **48**(1992), 1–17.

[13] Tsai, C.L., Score test for the first-order autoregressive model with heteroscedasticity, *Biometrika*, **73**(1986), 455–460.

[14] 韦博成, 胡跃清, 非线性回归模型相关性和异方差性的检验, *工程数学学报*, **4**(1994), 1–12.

[15] Zerbe, G.O., Randomization analysis of the completely randomized design extended to growth and response curves, *Journal of the American Statistical Association*, **74**(1979), 215–221.

[16] Chi, E.M. and Reinsel, G.C., Models for longitudinal data with random effects and AR(1) errors, *Journal of the American Statistical Association*, **84**(1989), 452–459.

[17] Pan, J.X. and Fang, K.T., *Growth Curve Models and Statistical Diagnostics*, New York, Springer-Verlag, 2002.

[18] Vonesh, E.F. and Chinchilli, V.M., *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York: Marcel Dekker, Inc., 1997, 262–264.

[19] Zhang, F. and Weiss, R.E., Diagnosing explainable heterogeneity of variance in random-effects models, *Canad. J. Statust.*, **28**(2000), 3–18.

Testing for Homogeneity of Variance and Correlation Coefficients in Uniform Correlation Models Based on Longitudinal Data

Fan Junhua[1,2] Lin Jinguan[1] Wei Bocheng[1]

([1]*Department of Mathematics, Southeast University, Nanjing, 210096*)

([2]*School of Chengxian, Southeast University, Nanjing, 210088*)

In longitudinal data analysis, homogeneity of variance is a basic assumption. However, this assumption is not necessarily appropriate. Lin and Wei[1] considered the tests for homogeneity of within-individual variances and between-individual autocorrelation coefficients in nonlinear models with AR(1) errors based on longitudinal data. This paper discusses the tests for homogeneity of variances and correlation coefficients in longitudinal data model with uniform correlation covariance structure and obtains several score test statistics. The glucose data is used to illustrate our results. Power simulations of the proposed tests are given in this paper.

Keywords: Longitudinal data, uniform correlation model, heteroscedasticity, correlation coefficient, score test.

AMS Subject Classification: 62J25.