应用概率统计　第二十三卷
第三期　2007 年 8 月

Chinese Journal of Applied Probability
and Statistics Vol. 23 No. 3 Aug. 2007

# 离散型半参数广义线性纵向数据模型的方差成分检验 *

曾林蕊　　朱仲义
(华东师范大学统计系，上海，200062)　(复旦大学统计系，上海，200433)
卢一强
(解放军信息工程大学电子技术学院，郑州，450004)



**摘　要**

在回归分析中，随机误差是否存在方差非齐性是大家十分关心的问题，本文根据 Laplace 展开原理针对随机效应的影响研究了基于纵向数据的离散型半参数广义线性模型的方差成分检验，得到了 Score 检验统计量，最后通过一个实例和计算机模拟验证了本文所提出的方法的有效性．

**关 键 词**：纵向数据，半参数模型，Score 检验，随机效应.
**学 科 分 类 号**：O212.1.

## §1. 引　　言

在回归分析中，随机误差是否存在方差非齐性是大家十分关心的问题，如果方差存在非齐性，将会对模型的统计分析带来很大的影响，虽然有一些数据变换方法可使变换后的数据具有方差齐性，但必须检验数据的异方差是否存在，因此，模型的方差成分检验是一个非常重要的问题．Zhang & Weiss (2000) 把产生方差非齐性的原因分为两种，一种是由随机效应产生的不可解释的非齐性；另一种是方差可表示为协变量函数时的可解释的非齐性．对于线性随机效应模型，Zhang & Weiss (2000) 已做了系统研究 [1]，而 [2] 中对于非线性随机效应模型分别讨论了群体内，群体间和多变量的异方差性的检验问题．对于广义回归模型，由于方差总是非齐的．因而不存在异方差检验问题，但其方差问题仍然存在．在该类模型中，方差变异的检验问题转化为偏离名义离差的检验问题．[3, 4] 中利用随机效应法研究广义线性模型的方差成分检验，[5] 中研究了广义非线性随机效应模型的方差成分检验，[6] 中同样利用随机效应法讨论了 Poisson 和二项回归模型中偏大离差的检验问题．而半参数回归模型作为一种既含有参数分量，又含有非参数分量的统计模型，它更能充分利用数据中提供的信息，是一类更具普遍性的模型．对该类模型的方差成分检验问题研究成果较少，[7] 中研究了半参数随机效应模型的异方差检验问题．并推导了检验统计量的极限分布．本文在以上研究的基础上，根据 Laplace 展开原理利用 Score 检验法讨论了离散型半参数广义线性纵向数据模型的方差成分检验问题．而在离散型情形，可通过研究随机效应的存在来研究模型的方差成分检验问题．

假设第 $i$ 个受试单元第 $j$ 次的观察值 $y_{ij}$ 关于随机效应 $\mathbf{u}_i$ 的条件密度为

$$p(y_{ij}|\mathbf{u}_i) = \exp\{y_{ij}\theta_{ij} - b(\theta_{ij}) - c(y_{ij})\}, \tag{1}$$

* 国家自然科学基金项目 (10371042, 10501053) 和上海市教委科研项目 (040B10) 资助.
本文 2005 年 12 月 8 日收到，2006 年 4 月 10 日收到修改稿.

$i=1,2,\cdots,m,\ j=1,2,\cdots,n_i$. 且假设

$$\sum_{i=1}^{m} n_i = n,$$

这时有

$$\mu_{ij}=\mathsf{E}(y_{ij}|\mathbf{u}_i)=\dot{b}(\theta_{ij}),\qquad \mathsf{Var}\,(y_{ij}|\mathbf{u}_i)=V_{ij}=\ddot{b}(\theta_{ij}).$$

从而可得基于纵向数据的半参数广义线性随机效应模型

$$\eta_{ij}=f(\mu_{ij})=\mathbf{x}_{ij}^T\boldsymbol{\beta}+g(t_{ij})+\mathbf{z}_{ij}^T\mathbf{u}_i, \tag{2}$$

其中 $f(\cdot)$ 是已知的单调联系函数， $g(\cdot)$ 为未知的光滑函数， $\boldsymbol{\beta}$ 为 $p\times 1$ 的未知参数， $\mathbf{u}_i$ 为 $q$ 维随机效应， $\mathbf{u}_i$ 独立同分布，假设 $\mathbf{u}_i\sim(0,D(\boldsymbol{\delta}))$, $\boldsymbol{\delta}$ 为 $s$ 维向量. $b,c$ 为已知函数. 当组间随机效应 $\mathbf{u}_i$ 不存在时， $\ddot{b}(\theta_{ij})$ 称为 $y_{ij}$ 的名义离差，而当随机效应存在时该模型就会存在着偏离名义离差的问题.

由上可知 $\mathbf{y}_i=(y_{i1},y_{i2},\cdots,y_{in_i})^T$ 关于 $\mathbf{u}_i$ 的条件密度为

$$p_i=p(\mathbf{y}_i|\mathbf{u}_i)=\exp\Big[\sum_{j=1}^{n_i}\{y_{ij}\theta_{ij}-b(\theta_{ij})-c(y_{ij})\}\Big], \tag{3}$$

其中 $\mathbf{u}_i$ 具有共同的分布 $T$, 期望为 0, 方差为 $D(\boldsymbol{\delta})$, $\boldsymbol{\delta}$ 为 $s$ 维向量， $\boldsymbol{\delta}=0$ 时， $D(\boldsymbol{\delta})=0$. 从而模型 (1)、(2) 的方差成分检验问题可化为假设检验

$$H_0:\boldsymbol{\delta}=0;\qquad H_1:\boldsymbol{\delta}\neq 0. \tag{4}$$

假设 $\mathbf{t}^0=(t_1^0,\cdots,t_r^0)$ 为对应于 $t_{ij}$ $(i=1,\cdots,m,\ j=1,\cdots,n_i)$ 的不相等递增序列， $N_i$ 为 $n_i\times r$ 的矩阵 $(i=1,\cdots,m)$ 且使得 $N_i\mathbf{g}=(g(t_{i1}),\cdots,g(t_{in_i}))^T$, 其中 $\mathbf{g}=(g(t_1^0),\cdots,g(t_r^0))^T$, 而 $X_i$, $Z_i$ 为第 $i$ 个单元内部的设计矩阵，它们分别为 $n_i\times p$, $n_i\times q$ 的矩阵， $X=(X_1^T,X_2^T,\cdots,X_m^T)^T$, $Z=\mathrm{diag}(Z_1,Z_2,\cdots,Z_m)$分别为 $n\times p$, $n\times mq$ 的矩阵. $\boldsymbol{\eta}_i=(\eta_{i1},\cdots,\eta_{in_i})^T$, $\boldsymbol{\eta}=(\boldsymbol{\eta}_1^T,\cdots,\boldsymbol{\eta}_m^T)^T$, $N=(N_1^T,N_2^T,\cdots,N_m^T)^T$, $\mathbf{u}=(\mathbf{u}_1^T,\cdots,\mathbf{u}_m^T)^T$, 则 (2) 可改写为

$$\boldsymbol{\eta}=X\boldsymbol{\beta}+N\mathbf{g}+Z\mathbf{u}.$$

又设 $\mathbf{y}_i$ 的边际密度为 $q_i$, 则有

$$q_i=\int p(\mathbf{y}_i|\mathbf{u}_i)\mathrm{d}T(\mathbf{u}_i), \tag{5}$$

其中 $T(\mathbf{u}_i)$ 为 $\mathbf{u}_i$ 的分布，将 $p(\mathbf{y}_i|\mathbf{u}_i)$ 在 $\mathbf{u}_i=0$ 处 Talor 展开可以得到

$$p(\mathbf{y}_i|\mathbf{u}_i)\approx p(\mathbf{y}_i|0)+\Big(\frac{\partial p_i}{\partial\mathbf{u}_i}\Big)\Big|_{\mathbf{u}_i=0}(\mathbf{u}_i-0)+\frac{1}{2}(\mathbf{u}_i-0)^T\frac{\partial^2 p_i}{\partial\mathbf{u}_i\partial\mathbf{u}_i^T}\Big|_{\mathbf{u}_i=0}(\mathbf{u}_i-0),$$

把上式代入 (5) 式可得

$$q_i\approx p_{i0}+\frac{1}{2}\mathrm{tr}\Big[\frac{\partial^2 p_i}{\partial\mathbf{u}_i\partial\mathbf{u}_i^T}\Big|_{\mathbf{u}_i=0}D(\boldsymbol{\delta})\Big], \tag{6}$$

其中 $p_{i0}=p(\mathbf{y}_i|0)$, 所以上述模型的惩罚对数似然函数可近似表示为

$$\begin{aligned} l(\boldsymbol{\xi}) &= \log\Big(\prod_{i=1}^{m} q_i\Big)-\frac{1}{2}\lambda\mathbf{g}^T K\mathbf{g} \\ &= \sum_{i=1}^{m}\log\Big[p_{i0}+\frac{1}{2}\mathrm{tr}\Big(\frac{\partial^2 p_i}{\partial\mathbf{u}_i\partial\mathbf{u}_i^T}\Big|_{\mathbf{u}_i=0}D(\boldsymbol{\delta})\Big)\Big]-\frac{1}{2}\lambda\mathbf{g}^T K\mathbf{g}. \end{aligned}$$

由于

$$\frac{\partial^2 p_i}{\partial\mathbf{u}_i\partial\mathbf{u}_i^T}=p_{i0}\sum_{j=1}^{n_i}\mathbf{z}_{ij}(-w_{ij}-\varepsilon_{ij}e_{ij})\mathbf{z}_{ij}^T+p_{i0}\Big(\sum_{j=1}^{n_i}\frac{\mathbf{z}_{ij}e_{ij}}{V_{ij}\dot f(\mu_{ij})}\Big)\Big(\sum_{j=1}^{n_i}\frac{\mathbf{z}_{ij}^T e_{ij}}{V_{ij}\dot f(\mu_{ij})}\Big),$$

$$\log(1+x)\approx x,$$

所以

$$\begin{aligned} l(\boldsymbol{\xi}) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\Big[y_{ij}\theta_{ij}-b(\theta_{ij})-c(y_{ij})\Big]+\frac{1}{2}\sum_{i=1}^{m}\Big[\Big(\sum_{j=1}^{n_i}\frac{\mathbf{z}_{ij}^T e_{ij}}{V_{ij}\dot f(\mu_{ij})}\Big)D(\boldsymbol{\delta})\Big(\sum_{j=1}^{n_i}\frac{\mathbf{z}_{ij}e_{ij}}{V_{ij}\dot f(\mu_{ij})}\Big) \\ &\quad+\sum_{j=1}^{n_i}\{-w_{ij}-\varepsilon_{ij}e_{ij}\}\mathbf{z}_{ij}^T D(\boldsymbol{\delta})\mathbf{z}_{ij}\Big]-\frac{1}{2}\lambda\mathbf{g}^T K\mathbf{g}, \end{aligned} \tag{7}$$

其中

$$\begin{aligned} &\boldsymbol{\xi}=(\boldsymbol{\delta}^T,\boldsymbol{\beta}^T,\mathbf{g}^T)^T, && e_{ij}=y_{ij}-\mu_{ij}, \\ &V_{ij}=\ddot b(\theta_{ij}), && w_{ij}=[V_{ij}\dot f^2(\mu_{ij})]^{-1}, \\ &\varepsilon_{ij}=\frac{\dot V_{ij}\dot f(\mu_{ij})+\ddot f(\mu_{ij})V_{ij}}{V_{ij}^2\dot f^3(\mu_{ij})}, && \dot V_{ij}=\frac{\partial V_{ij}}{\partial\mu_{ij}}, \end{aligned}$$

$K$ 是一个 $r\times r$ 的非负定矩阵, 且仅与节点 $\{t_i^0, i=1,2,\cdots,r\}$ 有关 (详见 Green, Silverman, 1994), $\lambda$ 为光滑参数, 假设 $\lambda$ 已用广义交叉核实法得到.

## §2. Score 检验统计量

为方便起见, 引入一些记号, 记

$$\begin{aligned} &W=\mathrm{diag}\{w_{ij}, i=1,2,\cdots,m, j=1,2,\cdots,n_i\}, \\ &W_1=\mathrm{diag}\Big\{\frac{\ddot f(\mu_{ij})}{V_{ij}\dot f^4(\mu_{ij})}, i=1,2,\cdots,m, j=1,2,\cdots,n_i\Big\}, \\ &\dot D_k=\frac{\partial D(\boldsymbol{\delta})}{\partial\delta_k}\Big|_{\delta=0}, \qquad k=1,\cdots,s, \end{aligned}$$

$Q=(Q_1^T,\cdots,Q_m^T)^T$ 为 $n\times s$ 的矩阵, $Q_i$ 为 $n_i\times s$ 的矩阵, 它的第 $j$ 行第 $k$ 列元素为 $\mathbf{z}_{ij}^T\dot D_k\mathbf{z}_{ij}$, 而

$$R_i=2T_i+\mathrm{diag}\Big\{\frac{\dddot V_{ij}}{V_{ij}\dot f^4(\mu_{ij})}+\frac{\ddot f^2(\mu_{ij})}{V_{ij}\dot f^6(\mu_{ij})}, j=1,2,\cdots,n_i\Big\}$$

为 $n_i\times n_i$ 的矩阵, 其中 $T_i$ 也为 $n_i\times n_i$ 的矩阵, 它的第 $j_1$ 行第 $j_2$ 列元素为 $w_{ij_1}w_{ij_2}$, $R=\mathrm{diag}\{R_i, i=1,2,\cdots,m\}$. 则对于假设检验 (4) 有如下定理.

**定理 1** $\widehat{\boldsymbol{\xi}}_0$ 表示 $H_0$ 成立时 $\boldsymbol{\xi}$ 的惩罚极大似然估计，则模型 (1)、(2) 的方差成分检验的 Score 检验统计量为

$$\mathrm{SC}=\Big\{\Big(\frac{\partial l(\boldsymbol{\xi})}{\partial \boldsymbol{\delta}}\Big)^T I_{\boldsymbol{\delta}}^{-1}\Big(\frac{\partial l(\boldsymbol{\xi})}{\partial \boldsymbol{\delta}}\Big)\Big\}_{\widehat{\boldsymbol{\xi}}_0}, \tag{8}$$

其中

$$\begin{aligned}\frac{\partial l(\boldsymbol{\xi})}{\partial \delta_k} &= \frac{1}{2}\sum_{i=1}^{m}\Big[\Big(\sum_{j=1}^{n_i}\frac{\mathbf{z}_{ij}^T e_{ij}}{V_{ij}\dot f(\mu_{ij})}\Big)\dot D_k\Big(\sum_{j=1}^{n_i}\frac{\mathbf{z}_{ij}e_{ij}}{V_{ij}\dot f(\mu_{ij})}\Big)\\ &\quad+\sum_{j=1}^{n_i}\{-w_{ij}-\varepsilon_{ij}e_{ij}\}\mathbf{z}_{ij}^T\dot D_k\mathbf{z}_{ij}\Big],\qquad k=1,2,\cdots,s,\end{aligned} \tag{9}$$

$$I_\delta=\left[\frac{1}{4}Q^TRQ-\frac{1}{4}Q^TW_1(X,N)\begin{pmatrix}X^TWX & X^TWN\\ N^TWX & N^TWN+\lambda K\end{pmatrix}^{-1}\begin{pmatrix}X^T\\ N^T\end{pmatrix}W_1Q\right]_{\widehat{\boldsymbol{\xi}}_0}. \tag{10}$$

**证明：** 当 $H_0$ 成立时，由 (7) 式直接对 $\delta_k$ 求偏导便知 (9) 式成立. 同样当 $H_0$ 成立时，经过计算有

$$\begin{aligned}\mathsf{E}\Big(-\frac{\partial^2 l(\boldsymbol{\xi})}{\partial\delta_k\partial\delta_l}\Big) &= \frac{1}{4}\sum_{i=1}^{m}\Big[\sum_{j_1\neq j_2}\frac{\mathbf{z}_{ij_1}^T\dot D_k\mathbf{z}_{ij_2}\mathbf{z}_{ij_1}^T\dot D_l\mathbf{z}_{ij_2}}{V_{ij_1}\dot f^2(\mu_{ij_1})V_{ij_2}\dot f^2(\mu_{ij_2})}\\ &\quad+\sum_{j=1}^{n_i}\Big(\frac{\ddot V_{ij}+2}{V_{ij}^2\dot f^4(\mu_{ij})}\frac{\ddot f^2(\mu_{ij})}{V_{ij}\dot f^6(\mu_{ij})}\Big)(\mathbf{z}_{ij}^T\dot D_k\mathbf{z}_{ij})(\mathbf{z}_{ij}^T\dot D_l\mathbf{z}_{ij})\Big]\\ &= \frac{1}{4}\sum_{i=1}^{m}\Big\{2\Big(\sum_{j=1}^{n_i}\frac{\mathbf{z}_{ij}^T\dot D_k\mathbf{z}_{ij}}{V_{ij}\dot f^2(\mu_{ij})}\Big)\Big(\sum_{j=1}^{n_i}\frac{\mathbf{z}_{ij}^T\dot D_l\mathbf{z}_{ij}}{V_{ij}\dot f^2(\mu_{ij})}\Big)\\ &\quad+\sum_{j=1}^{n_i}\Big(\frac{\ddot V_{ij}}{V_{ij}^2\dot f^4(\mu_{ij})}+\frac{\ddot f^2(\mu_{ij})}{V_{ij}\dot f^6(\mu_{ij})}\Big)(\mathbf{z}_{ij}^T\dot D_k\mathbf{z}_{ij})(\mathbf{z}_{ij}^T\dot D_l\mathbf{z}_{ij})\Big\},\\ \mathsf{E}\Big(-\frac{\partial^2 l(\boldsymbol{\xi})}{\partial\beta_k\partial\beta_l}\Big) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}x_{ijk}x_{ijl}w_{ij},\\ \mathsf{E}\Big(-\frac{\partial^2 l(\boldsymbol{\xi})}{\partial g_k\partial g_l}\Big) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}N_{ijk}N_{ijl}w_{ij}+\lambda K_{kl},\\ \mathsf{E}\Big(-\frac{\partial^2 l(\boldsymbol{\xi})}{\partial\delta_k\partial\beta_l}\Big) &= -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{\ddot f(\mu_{ij})}{V_{ij}\dot f^4(\mu_{ij})}x_{ijl}\mathbf{z}_{ij}^T\dot D_k\mathbf{z}_{ij},\\ \mathsf{E}\Big(-\frac{\partial^2 l(\boldsymbol{\xi})}{\partial\delta_k\partial g_l}\Big) &= -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{\ddot f(\mu_{ij})}{V_{ij}\dot f^4(\mu_{ij})}N_{ijl}\mathbf{z}_{ij}^T\dot D_k\mathbf{z}_{ij},\\ \mathsf{E}\Big(-\frac{\partial^2 l(\boldsymbol{\xi})}{\partial\beta_k\partial g_l}\Big) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}x_{ijk}N_{ijl}w_{ij}.\end{aligned}$$

其中 $N_{ijk}$ 是 $N_i$ 的第 $j$ 行第 $k$ 列元素，$x_{ijk}$ 是 $X_i$ 的第 $j$ 行第 $k$ 列元素. 所以，当 $H_0$ 成立时，$\boldsymbol{\xi}$ 的 Fisher 信息矩阵为

$$I(\boldsymbol{\xi})=\begin{pmatrix}I_{\boldsymbol{\delta\delta}} & I_{\boldsymbol{\delta\psi}}\\ I_{\boldsymbol{\psi\delta}} & I_{\boldsymbol{\psi\psi}}\end{pmatrix}=\begin{pmatrix}\frac{1}{4}Q^TRQ & -\frac{1}{2}Q^TW_1X & -\frac{1}{2}Q^TW_1N\\ -\frac{1}{2}X^TW_1Q & X^TWX & X^TWN\\ -\frac{1}{2}N^TW_1Q & N^TWX & N^TWN+\lambda K\end{pmatrix},$$

其中 $\psi=(\beta^T,\mathbf{g}^T)^T$. 把上述结果代入 $I_{\boldsymbol{\delta}}=I_{\boldsymbol{\delta}\boldsymbol{\delta}}-I_{\boldsymbol{\delta}\psi}I_{\psi\psi}^{-1}I_{\psi\boldsymbol{\delta}}$, 即得 (10) 式, 从而定理 1 结论成立. #

## §3. 随机模拟与实例

本节我们通过计算机模拟和一个实例来检验 Score 统计量的性质.

### 3.1 随机模拟

我们针对 logistic 回归模型进行了模拟研究. 考虑如下模型

$$\begin{cases}\eta_{ij}=\log\dfrac{\mu_{ij}}{1-\mu_{ij}}=x_{ij}\beta+g(t_{ij})+u_i,\\ y_{ij}\sim b(1,\mu_{ij}),\end{cases}\qquad (i=1,\cdots,m,\ j=1,\cdots,n_i).$$

其中

$$\beta=1,\qquad t_{ij}=\mathrm{trun}\{(i+(m/5-1))/(m/5)\}/50+0.1(j-1),\qquad x_{ij}=5t_{ij}^2+e_{ij},$$

$e_{ij}$ 独立同分布来自 $N(0,0.5)$, $g(t_{ij})=\cos(\pi t_{ij})$, 随机变量 $u_i$ $(i=1,2,\cdots,m)$ 独立同分布, 来自如下混合分布

$$F=pN(-(1-p)\eta,\tau^2)+(1-p)N(p\eta,\tau^2),$$

由 Titterington (1985) 知, 该分布的均值为 0, 方差为 $\theta=p(1-p)\eta^2+\tau^2$. 类似于 Lin (1997), Zhu (2004), 我们考虑下面四种情况.

1. $\eta=\theta=\tau=0$, 此时为原假设;

2. $\eta=0$, $\tau^2=\theta=0.2, 0.4, 0.6, 0.8, 1.0$, 此时为正态分布;

3. $p=0.25$, $\eta=0.3500, 0.5800, 0.7500, 0.8500, 1.0000$, $\tau^2$ 分别取值使得 $\theta=0.2, 0.4, 0.6, 0.8, 1.0$. 此时为单峰正态混合模型 (unimodal normal mixture);

4. $p=0.25$, $\eta=0.9000, 1.2000, 1.4500, 1.7000, 1.9500$, $\tau^2$ 分别取值使得 $\theta=0.2, 0.4, 0.6, 0.8, 1.0$. 此时为双峰正态混合模型 (bimodal normal mixture).

对 $m=50, 75$, $n_i=10$, 及每一个 $\theta$ 值分别进行了 1000 次模拟, 得到了 Score 检验统计量的经验分布函数及势函数 (理论水平为 0.05), 具体做法是将 1000 次模拟中拒绝的次数与 1000 的比值作为势函数的值, 结果见图 1、图 2、表 1. 光滑参数采用广义交叉核实法选取. 从表 1 可以看出, 随着方差 $\theta$ 的增加以及 $m$ 的增大, 势函数也在增加, 这与预期的完全一致. 另外, Score 检验统计量对于方差相等的单峰混合正态分布、双峰混合正态分布和正态分布都有相似的功效, 这是因为 Score 统计量推导过程中并没有假定随机效应的具体分布形式, 所以, 分布的形式对结果的影响不是很大.

从图 1、图 2 可以看出当原假设成立时, 无论是 $m=50$ 或 $m=75$, Score 检验统计量的经验分布函数与 $\chi^2(1)$ 吻合得都非常好.

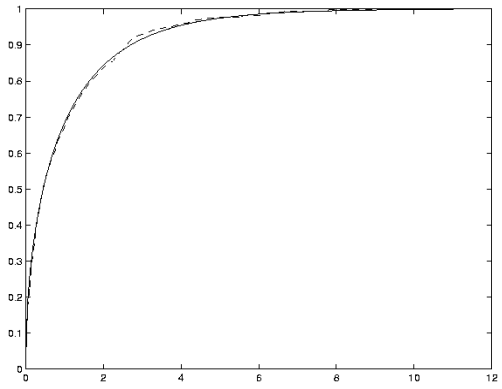

图 1 $m = 50$ 时 Score 检验统计量的经验分布函数 (对应点) 与卡方分布函数 (对应线) 的比较

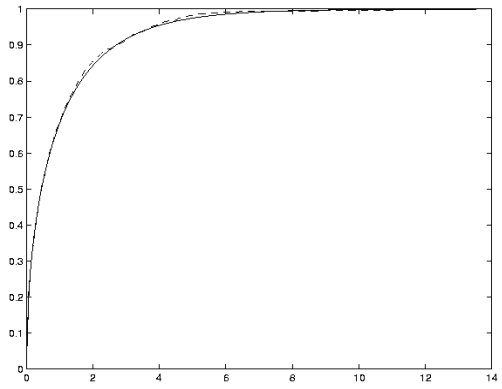

图 2 $m = 75$ 时 Score 检验统计量的经验分布函数 (对应点) 与卡方分布函数 (对应线) 的比较

表 1 Score 检验统计量的势函数

| | $\theta$ | Power | |
|---|---|---|---|
| | | $m = 50,\ n_i = 10$ | $m = 75,\ n_i = 10$ |
| Normal | 0 | 0.0550 | 0.0460 |
| | 0.2 | 0.1840 | 0.2560 |
| | 0.4 | 0.4650 | 0.6230 |
| | 0.6 | 0.7050 | 0.8730 |
| | 0.8 | 0.8580 | 0.9590 |
| | 1.0 | 0.9380 | 0.9870 |
| Unimodal Normal mixture | 0 | 0.0550 | 0.0460 |
| | 0.2 | 0.1880 | 0.2540 |
| | 0.4 | 0.4710 | 0.6050 |
| | 0.6 | 0.7130 | 0.8730 |
| | 0.8 | 0.8770 | 0.9690 |
| | 1.0 | 0.9280 | 0.9890 |
| Bimodal Normal mixture | 0 | 0.0550 | 0.0460 |
| | 0.2 | 0.2330 | 0.3080 |
| | 0.4 | 0.5150 | 0.6830 |
| | 0.6 | 0.8280 | 0.9030 |
| | 0.8 | 0.9160 | 0.9750 |
| | 1.0 | 0.9800 | 0.9970 |

### 3.2 实例

**例 1** 小孩呼吸感染数据

为了解儿童缺乏维生素 A 与呼吸感染之间的关系，对 275 个学龄前儿童每隔三月一次共 18 个月进行是否存在呼吸感染的检查，共得到 1200 个数据，$y_{ij} = 1$ 表示第 $i$ 个儿童第 $j$ 次检

查中存在呼吸感染，$y_{ij}=0$ 表示不存在，$t_{ij}$ 表示年龄，$x_{ij}$ 分别表示是否缺乏维生素 A (1 表示缺乏，0 表示不缺乏), 季节的余弦，季节的正弦，性别 (1 表示雌性，0 表示雄性), 身高，生长曲线是否有缺陷 (1 表示有，0 表示没有), Zeger & Karim (1991) 用广义线性随机效应模型对这组数据进行建模，Lin & Carroll (2001) 把年龄纳入非参数部分用半参数广义线性模型并利用核估计方法进行建模分析，Lin & Zhang (1999) 用半参数广义线性模型并利用拟似然估计方法对这组数据进行分析．现对这组数据进行方差齐性检验，由于数据是重复测量数据，我们采用下列随机效应模型：

$$\text{logit}\{\mathsf{P}(y_{ij}=1|u_i)\}=x_{ij}^T\beta+f(t_{ij})+u_i,$$

即考虑小孩之间有差异，这是比较合理的．设 $u_i$ 独立同分布 $N(0,\theta)$, 可用 Score 检验统计量来检查个体的差异．采用光滑样条的估计方法得到参数的估计及标准差如表 2 所示，非参数部分的估计如图 3 所示．至于光滑参数采用广义交叉核实法来确定，得到 $\widehat{\lambda}=1.259$, 另外，我们通过计算得到 Score 检验统计量的值为 $6.862>3.84=\chi^2_{0.05}(1)$, 检验的 $p$ 值为 0.007, 所以可以认为呼吸感染随着个体的变化差别较大．

表 2　参数的估计及标准差

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|---|---|---|
| 估计 | 0.6046 | -0.5886 | -0.1625 | -0.5071 | -0.0261 | -0.4276 |
| 标准差 | 0.4931 | 0.1983 | 0.1724 | 0.2647 | 0.0289 | 0.4689 |

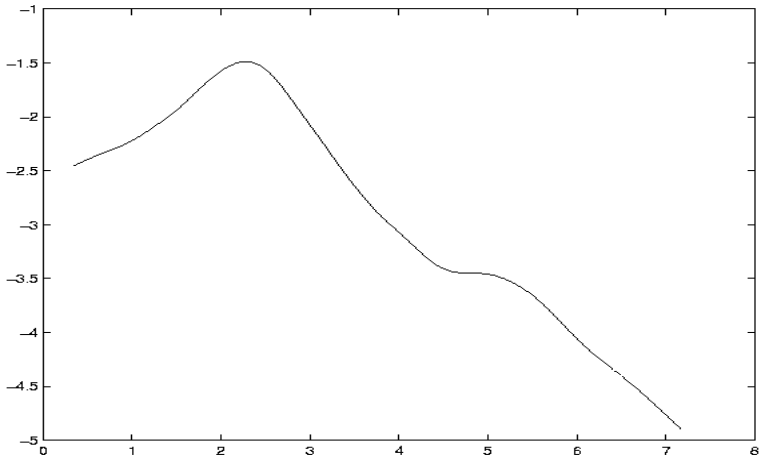


图 3　小孩呼吸感染数据非参数部分的估计

## 参　考　文　献

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

[2] 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.

[3] Jacqmin-Gadda, H. and Commenges, D., Testing of homogeneity for generalized linear models, *J. Am. Statist. Assoc.*, **90**(1995), 1237–1246.

[4] Lin, X.H., Variance component testing in generalized linear models with random effects, *Biometrika*, **84**(1997), 309–326.

[5] 韦博成，林金官，指数族广义非线性混合效应模型的变离差检验，东南大学学报（自然科学版），**32(3)** (2002), 528–535.

[6] Dean, C.B., Testing for overdispersion in Poisson and binomial regression models, *J. Am. Statist. Assoc.*, **87**(1992), 451–457.

[7] Zhu, Z.Y. and Fung, W.K., Variance component testing in semiparatric mixed models, *J. Multivariate Analy.*, **91**(2004), 107–118.

[8] Green, P.J. and Silverman, B.W., *Nonparametric Regression and Generalized Linear Models*, Chapman and Hall, London, 1994.

[9] Titterington, D.M., Smith, A.F.M. and Makov, U.E., *Statistical Analysis of Finite Mixture Distributions*, John Wiley and Sons, Chichester, New York, et al., 1985.

[10] Zeger, S.L. and Karim, M.R., Generalized linear models with random effects; a gibbs sampling approach, *JASA*, **86**(1991), 79–85.

[11] Lin, X.H. and Carroll, R.J., Semiparametric regression for clustered data using generalized estimation equations, *JASA*, **96**(2001), 1045–1056.

[12] Lin, X.H. and Zhang, D.W., Inference in generalized additive mixed models by using smoothing splines, *J.R. Statist. Soc. B*, **61(2)**(1999), 381–400.



## Variance Component Testing in Semiparametric Generalized Linear Mixed Model for Longitudinal Data

Zeng Linrui

(*Department of Statistics, East China Normal University, Shanghai, 200062*)

Zhu Zhongyi

(*Department of Statistics, Fudan University, Shanghai, 200433*)

Lu Yiqiang

(*Institute of Electronic Technology, the PLA Information Engineering University, Zhengzhou, 450004*)

The assumption of homoscedasticity is commonly concerned in regression analysis. According to the Laplace spread theory, we study the variance component testing in discrete semiparametric generalized linear model with random effects for longitudinal data, and the score statistic is obtained. An example and some simulation studies are implenented to verify the efficiency of our method.