应用概率统计　第二十五卷
第三期　2009年6月

Chinese Journal of Applied Probability
and Statistics Vol.25 No.3 Jun. 2009

具有线性趋势的回归信度模型中的估计和检验 *

唐　国　强

(华东师范大学金融与统计学院, 上海, 200241;　桂林工学院数理系, 桂林, 541004)

摘　要

研究具有线性趋势回归信度模型的参数估计和检验. 对该模型的回归系数和随机效应的方差, 利用正交变换法得到了它们的极大似然估计, 并得到了参数的无偏估计. 对随机效应和是否有线性趋势采用似然比检验, 得到了似然统计量较好的近似P值, 并对检验的功效进行了模拟研究.

关键词:　回归信度模型, 正交变换, 似然比检验.

学科分类号:　O212.1.

§1.　引　　言

保险精算学中的信度理论(Credibility Theory)萌芽于20世纪20年代, 至今已有80年的历史. 在非寿险精算理论与实务中信度理论具有重要地位(参看文[1]). 根据信度理论厘定信度保费, 是非寿险保费计算的一个重要方法. 信度建模就是一种费率拟订过程, 在这个过程中, 精算师根据过去的单个风险或者一组风险的经验, 调整未来的保险费. 在精算科学中, 很多问题需要数学模型去预测未来的保险成本, 尤其是短期, 所以, 信度理论越来越受到人们的重视.

为了适应各种情况下的信度保费, 人们建立了各种各样的信度模型. 如Buhlmann信度模型、Buhlmann-Straub信度模型、回归信度模型和分层信度模型等(参看文[2]). 随着现代统计学的发展, 1999年Frees, E.W.将纵向数据的分析方法用于信度模型的分析, 得到了上述四种信度模型的信度保费(参看文[3]). 对精算师来说, 在各种信度模型中选择合适的模型和参数结构是必要的, 否则就会出现大的偏差, 从统计角度说就是对模型进行检验(参看文[5]). 但以前对信度模型的研究主要集中在信度保费的计算上, 没有给出模型的检验, 本文研究了信度模型的参数估计, 并对模型类型进行了检验.

在文[3]中, 对方差采用极大似然估计和限制极大似然估计, 但由于没有显示解, 很多情况不得不用迭代计算. 本文对具有线性趋势的回归信度模型的回归系数和随机效应的方差, 利用正交变换法得到了具有显示表达式的极大似然估计, 并对随机效应采用似然比检验, 得到了似然比较好的近似P值, 最后利用似然比的思想得到了对是否有线性趋势的F检验.

*国家自然科学基金(10661006)、广西教育厅科研项目(200707MS133)资助.

本文2007年11月2日收到, 2008年4月14日收到修改稿.

具有线性趋势回归信度模型是: 将某险种分成n类风险(或组), 比如汽车保险, 可按不同的年龄、职业、地区等进行风险分类. 每类风险观测到T期索赔数据, 要预测每类风险$T+1$期的索赔, 也称信度保费. 即有观测数据

$$y_{ij} \qquad i=1,2,\cdots,n,\ j=1,2,\cdots,T,$$

y_{ij}满足下列模型

$$y_{ij}=\beta_1+j\beta_2+\alpha_i+\epsilon_{ij} \qquad i=1,2,\cdots,n,\ j=1,2,\cdots,T, \tag{1.1}$$

其中β_1是总平均, β_2是趋势项, 比如索赔随着通货膨胀会呈现一定的趋势. 从$j\beta_2$可以看出, 观察值与时间具有线性关系. β_1,β_2是固定效应, α_i是第i组的随机效应, ϵ_{ij}是误差项, 且$\mathsf{Var}\,(\alpha_i)=\sigma_\alpha^2$, $\mathsf{Var}\,(\epsilon_{ij})=\sigma_\epsilon^2$, $\alpha_i\sim N(0,\sigma_\alpha^2)$, $\epsilon_{ij}\sim N(0,\sigma_\epsilon^2)$, 所有$\alpha_i$与$\epsilon_{ij}$相互独立.

将第i组的模型写成矩阵形式为:

$$\boldsymbol{Y}_i=\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{U}_i\alpha_i+\boldsymbol{\epsilon}_i \qquad i=1,2,\cdots,n, \tag{1.2}$$

其中$\boldsymbol{Y}_i=(y_{i1},y_{i2},\cdots,y_{iT})'$, $\boldsymbol{X}_i=(\boldsymbol{X}_{i1},\boldsymbol{X}_{i2})=\begin{pmatrix}1&1\\1&2\\\vdots&\vdots\\1&T\end{pmatrix}$, $\boldsymbol{U}_i=\mathbf{1}_T$, $\mathbf{1}_T$表示元素为1的列向量, $\boldsymbol{\epsilon}_i=(\epsilon_{i1},\epsilon_{i2},\cdots,\epsilon_{iT})'$, $\boldsymbol{\beta}=(\beta_1,\beta_2)'$, 则$\boldsymbol{Y}_i$的协方差矩阵为

$$\mathsf{Cov}\,(\boldsymbol{Y}_i)=\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'+\sigma_\epsilon^2\boldsymbol{I}_T\,\widehat{=}\,\boldsymbol{V}_i. \tag{1.3}$$

下面我们先研究信度保费的计算.

§2. 信度保费的预测

为了得到具有线性趋势回归信度模型的信度保费的预测, 我们需要下列引理(参看文献[3]).

引理 2.1 Y是$N\times 1$随机向量, 均值为$\mathsf{E}(Y)=X\beta$, 协方差阵$\mathsf{Var}\,(Y)=V$. 需要预测随机变量W, 均值$\mathsf{E}(W)=\lambda^t\beta$, 方差$\mathsf{Var}\,(W)=\sigma_W^2$. 记$\mathsf{Cov}\,(W,Y)\,\widehat{=}\,\mathsf{Cov}_{\,WY}$, 则$Y$对$W$最佳线性预测为

$$W^*=\mathsf{E}(W)+\mathsf{Cov}_{\,WY}V^{-1}[Y-\mathsf{E}(Y)]=\lambda^t\beta+\mathsf{Cov}_{\,WY}V^{-1}(Y-X\beta).$$

假设参数已估计出来了, 我们来求信度保费的最佳线性预测.

将整个数据模型写成矩阵形式为:

$$\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{U}\boldsymbol{\alpha}+\boldsymbol{\epsilon}, \tag{2.1}$$

其中$\boldsymbol{Y}=(y_{11},y_{12},\cdots,y_{1T},y_{21},\cdots,y_{nT})'=(\boldsymbol{Y}'_1,\boldsymbol{Y}'_2,\cdots,\boldsymbol{Y}'_n)'$, $\boldsymbol{X}=\boldsymbol{1}_n\otimes\boldsymbol{X}_i$, $\boldsymbol{U}=\boldsymbol{I}_n\otimes\boldsymbol{1}_T$, $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_n)'$, $\boldsymbol{\epsilon}=(\epsilon_{11},\epsilon_{12},\cdots,\epsilon_{1T},\epsilon_{21},\cdots,\epsilon_{nT})'=(\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_2,\cdots,\boldsymbol{\epsilon}_n)'$.

我们要预测信度保费是每类风险的$T+1$期的索赔, 即

$$\mathsf{E}(y_{i,T+1}|\alpha_i)=\beta_1+(T+1)\beta_2+\alpha_i,\qquad i=1,2,\cdots,n. \tag{2.2}$$

令$W_i=\beta_1+(T+1)\beta_2+\alpha_i$, $i=1,2,\cdots,n$, 则$\mathsf{E}(W_i)=\beta_1+(T+1)\beta_2$, 当$j=i$时, $\mathsf{Cov}\,(W_i,\boldsymbol{Y}_j)=\sigma_\alpha^2\boldsymbol{1}'_T$, 当$j\neq i$时, $\mathsf{Cov}\,(W_i,\boldsymbol{Y}_j)=0$, 所以由文[3]第(3)式, 得$W_i$的最佳线性预测为

$$\begin{aligned} W_i^* &= \mathsf{E}(W_i)+\mathsf{Cov}_{W_i\boldsymbol{Y}}\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta}) \\ &= \Big(1-\frac{T\sigma_\alpha^2}{\sigma_\epsilon^2+T\sigma_\alpha^2}\Big)\beta_1+(T+1)\Big(1-\frac{T\sigma_\alpha^2}{2(\sigma_\epsilon^2+T\sigma_\alpha^2)}\Big)\beta_2+\frac{T\sigma_\alpha^2}{\sigma_\epsilon^2+T\sigma_\alpha^2}\overline{\boldsymbol{Y}}_i, \end{aligned} \tag{2.3}$$

其中$\overline{\boldsymbol{Y}}_i=T^{-1}\sum\limits_{j=1}^{T}y_{ij}$. 由上面的计算我们得下列定理.

定理 2.1 模型(1.1)的信度保费的估计为

$$\widehat{y}_{i,T+1}=(1-\xi)\widehat{\beta}_1+(T+1)(1-\xi/2)\widehat{\beta}_2+\xi\overline{\boldsymbol{Y}}_i,\qquad i=1,2,\cdots,n, \tag{2.4}$$

其中$\xi=T\widehat{\sigma_\alpha^2}/(\widehat{\sigma_\epsilon^2}+T\widehat{\sigma_\alpha^2})$, $\overline{\boldsymbol{Y}}_i=T^{-1}\sum\limits_{j=1}^{T}y_{ij}$.

从上面的信度保费的估计可以看出. 某类风险的信度保费为总平均, 趋势效应和本身均值的加权和, 这样信度保费是综合了总体风险和个体风险, 是比较合理的. 要计算信度保费, 我们要估计参数β_1, β_2, σ_α^2和σ_ϵ^2, 下面研究他们的极大似然估计.

§3. 参数的估计

我们先将数据进行正交变换, 然后求出模型参数的极大似然估计.

令$\boldsymbol{H}_*$是T阶正交矩阵, 形如$\boldsymbol{H}_*=\begin{pmatrix}\boldsymbol{1}'_T/\sqrt{T}\\ \boldsymbol{H}\end{pmatrix}$, 其中$\boldsymbol{H}$是$(T-1)\times T$阶行正交阵, 则$\boldsymbol{H}\boldsymbol{H}'=\boldsymbol{I}_{T-1}$, 并且有$\boldsymbol{H}\boldsymbol{1}_T=\boldsymbol{0}$. $\boldsymbol{H}$有很多种选择, 其中最为常用的是取

$$\begin{aligned} \boldsymbol{H} &= (\boldsymbol{H}'_2,\cdots,\boldsymbol{H}'_T)', \\ \boldsymbol{H}_2 &= (1/\sqrt{1\cdot 2},-1/\sqrt{1\cdot 2},0,\cdots,0), \\ \boldsymbol{H}_3 &= (1/\sqrt{2\cdot 3},1/\sqrt{2\cdot 3},-2\sqrt{2\cdot 3},0,\cdots,0), \end{aligned}$$

依此类推. 令$\boldsymbol{Z}_i=\boldsymbol{H}_*\boldsymbol{Y}_i$, $\boldsymbol{Z}_i=(z_{i1},z_{i2},\cdots,z_{iT})'$, $\boldsymbol{Z}_{i(2)}=(z_{i2},z_{i3},\cdots,z_{iT})'$. 我们有下面一些结果:

1) $\boldsymbol{Z}_i$的协方差矩阵为对角阵:

$$\mathsf{Cov}\,(\boldsymbol{Z}_i)=\mathrm{diag}(\sigma_\epsilon^2+T\sigma_\alpha^2,\sigma_\epsilon^2,\cdots,\sigma_\epsilon^2),$$

由此可见$\{z_{it}:i=1,\cdots,n,t=1,\cdots,T\}$相互独立.

2) 由于正交阵$\boldsymbol{H}_*$的第1个行向量为$\boldsymbol{1}_T'/\sqrt{T}$, 所以

$$z_{i1}=\Big(\sum_{t=1}^T y_{it}\Big)\Big/\sqrt{T},\qquad \mathsf{E}(z_{i1})=\sqrt{T}\beta_1+\frac{(1+T)\sqrt{T}}{2}\beta_2.$$

因此$\{z_{i1}:i=1,\cdots,n\}$独立, 分布分别为正态分布,

$$N\Big(\sqrt{T}\Big(\beta_1+\frac{1+T}{2}\beta_2\Big),\sigma_\epsilon^2+T\sigma_\alpha^2\Big)\mathrel{\widehat{=}}N(\mu,\sigma_\tau^2)\qquad i=1,2,\cdots,n.$$

3) 由于$\boldsymbol{H}\boldsymbol{1}_T=0$, 所以$\mathsf{E}(\boldsymbol{Z}_{i(2)})=\boldsymbol{H}\boldsymbol{X}_{i2}\beta_2$, 令$\boldsymbol{\eta}=(\eta_2,\eta_3,\cdots,\eta_T)'=\boldsymbol{H}\boldsymbol{X}_2\beta_2$, 因此对$t=2,3,\cdots,T$, $\{z_{it}:i=1,\cdots,n\}$独立同分布$N(\eta_t,\sigma_\epsilon^2)$正态分布.

利用$\{z_{i1}:i=1,\cdots,n\}$可以得到$\sqrt{T}\{\beta_1+[(1+T)/2]\beta_2\}$和$\sigma_\epsilon^2+T\sigma_\alpha^2$的极大似然估计分别为

$$\widehat{\mu}=\overline{z}_1=\frac{1}{n}\sum_{i=1}^n z_{i1},\qquad \widehat{\sigma}_\tau^2=\mathrm{SS1}=\frac{1}{n}\sum_{i=1}^n(z_{i1}-\overline{z}_1)^2.\tag{3.1}$$

利用当$t=2,3,\cdots,T$时, $\{z_{it}:i=1,\cdots,n\}$独立同分布$N(\eta_t,\sigma_\epsilon^2)$正态分布, 得$\beta_2,\sigma_\epsilon^2$的极大似然估计:

$$\widehat{\beta}_2=\Big[1\Big/\sum_{t=2}^T\sum_{i=1}^n(\boldsymbol{H}_t\boldsymbol{X}_{i2})^2\Big]\cdot\sum_{t=2}^T\sum_{i=1}^n\boldsymbol{H}_t\boldsymbol{X}_{i2}z_{it},\tag{3.2}$$

$$\widehat{\sigma}_\epsilon^2=\frac{1}{n(T-1)}\sum_{t=2}^T\sum_{i=1}^n(z_{it}-\boldsymbol{H}_t\boldsymbol{X}_{i2}\widehat{\beta}_2)^2.\tag{3.3}$$

综合上面的讨论得我们得下面定理,

定理 3.1 β_1, β_2, σ_α^2和σ_ϵ^2的极大似然估计

$$\widehat{\beta}_1=\frac{\overline{z}_1}{\sqrt{T}}-\frac{1+T}{2}\widehat{\beta}_2,\qquad \widehat{\beta}_2=\Big[1\Big/\sum_{t=2}^T\sum_{i=1}^n(\boldsymbol{H}_t\boldsymbol{X}_{i2})^2\Big]\cdot\sum_{t=2}^T\sum_{i=1}^n\boldsymbol{H}_t\boldsymbol{X}_{i2}z_{it},$$

$$\widehat{\sigma}_\alpha^2=\frac{1}{T}(\mathrm{SS1}-\widehat{\sigma}_\epsilon^2)=\frac{1}{T}\Big(\frac{1}{n}\sum_{i=1}^n(z_{i1}-\overline{z}_1)^2-\widehat{\sigma}_\epsilon^2\Big),$$

$$\widehat{\sigma}_\epsilon^2=\frac{1}{n(T-1)}\sum_{t=2}^T\sum_{i=1}^n(z_{it}-\boldsymbol{H}_t\boldsymbol{X}_{i2}\widehat{\beta}_2)^2,$$

其中

$$\mathrm{SS1}=\frac{1}{n}\sum_{i=1}^n(z_{i1}-\overline{z}_1)^2,\qquad \overline{z}_1=\frac{1}{n}\sum_{i=1}^n z_{i1},$$

$\boldsymbol{H}_t$是$\boldsymbol{H}$的行向量, 而$\boldsymbol{H}$是$(T-1)\times T$阶行正交阵, 且$\boldsymbol{H}\boldsymbol{H}'=\boldsymbol{I}_{T-1}$, $\boldsymbol{H}\boldsymbol{1}_T=0$.

定理 3.2 $\widehat{\beta}_1$, $\widehat{\beta}_2$ 分别是 β_1, β_2 的无偏估计; $\{n(T-1)/[n(T-1)-1]\}\widehat{\sigma}_\epsilon^2$, $(1/T)\cdot\{[n/(n-1)]\widehat{\sigma}_\tau^2-[n(T-1)/[n(T-1)-1]]\widehat{\sigma}_\epsilon^2\}$分别是$\widehat{\sigma}_\epsilon^2$, $\widehat{\sigma}_\alpha^2$的无偏估计.

证明: 前面的参数估计, 采用的是正交变换后的数据, 由于变换后仍然是正态分布型数据, 所以由参考文献[6]的定理4.1.4知$\widehat{\beta}_2$是β_2的无偏估计, $\widehat{\mu}$是μ的无偏估计, 而$\mu=\sqrt{T}\{\beta_1+[(1+T)/2]\beta_2\}$, 所以$\widehat{\beta}_1$也是$\beta_1$的无偏估计.

同样由上述定理的结果知$\{n(T-1)/[n(T-1)-1]\}\widehat{\sigma}_\epsilon^2$是$\widehat{\sigma}_\epsilon^2$的无偏估计, $[n/(n-1)]\widehat{\sigma}_\tau^2$是$\sigma_\tau^2$的无偏估计, 而$\sigma_\tau^2=\sigma_\epsilon^2+T\sigma_\alpha^2$, 所以$(1/T)\{[n/(n-1)]\widehat{\sigma}_\tau^2-[n(T-1)/[n(T-1)-1]]\widehat{\sigma}_\epsilon^2\}$是$\sigma_\alpha^2$的无偏估计. □

索赔数据有没有随机效应呢? 如没有, 我们就没必要将风险进行分类了, 下面对随机效应进行检验.

§4. 随机效应的存在性检验

我们利用似然比对随机效应进行检验, 要求似然比的渐近P值, 需要用到下面的引理(参考文献[4]).

引理 4.1 假设变量Z的h阶矩$\mathsf{E}(Z^h)$形如

$$\mathsf{E}(Z^h)=C\Big(\prod_{j=1}^{q}b_j^{b_j}\Big/\prod_{k=1}^{r}a_k^{b_k}\Big)^h\cdot\Big\{\prod_{k=1}^{r}\Gamma[a_k(1+h)+\xi_k]\Big/\prod_{j=1}^{q}\Gamma[b_j(1+h)+\eta_j]\Big\}, \tag{4.1}$$

其中C是一个常数, 它使得Z的0阶矩$\mathsf{E}(Z^0)=1$, 所以

$$C=\prod_{j=1}^{q}\Gamma[b_j+\eta_j]\Big/\prod_{k=1}^{r}\Gamma[a_k+\xi_k]. \tag{4.2}$$

除了$a_k=O(n)$ $(k=1,2,\cdots,r)$和$b_j=O(n)$ $(j=1,2,\cdots,q)$之外, 其余的ξ_k $(k=1,2,\cdots,r)$和η_j $(j=1,2,\cdots,q)$, 以及q和r都与n无关. 如果$\sum\limits_{k=1}^{r}a_k=\sum\limits_{j=1}^{q}b_j$, 若取

$$f=-2\Big[\sum_{k=1}^{r}\xi_k-\sum_{j=1}^{q}\eta_j-\frac{1}{2}(r-q)\Big], \tag{4.3}$$

$$\rho=1-\frac{1}{f}\Big[\sum_{k=1}^{r}a_{k=1}^{-1}B_2(\xi_k)-\sum_{j=1}^{q}b_{j=1}^{-1}B_2(\eta_j)\Big], \tag{4.4}$$

$$\omega_2=-\frac{1}{6\rho^2}\Big\{\sum_{k=1}^{r}a_k^{-2}B_3[(1-\rho)a_k+\xi_k]-\sum_{j=1}^{q}b_j^{-2}B_3[(1-\rho)b_j+\eta_j]\Big\}, \tag{4.5}$$

其中$B_2(h)$和$B_3(h)$分别是二次和三次Bernoulli多项式:

$$B_2(h)=h^2-h+\frac{1}{6},\qquad B_3(h)=h^3-\frac{3}{2}h^2+\frac{1}{2}h.$$

则我们有

$$\mathsf{P}(-2\ln Z \geq x) = \mathsf{P}(\chi^2(f) \geq x) + O(n^{-1}), \tag{4.6}$$

$$\mathsf{P}(-2\rho\ln Z \geq x) = \mathsf{P}(\chi^2(f) \geq x) + O(n^{-2}), \tag{4.7}$$

$$\mathsf{P}(-2\rho\ln Z \geq x) = \mathsf{P}(\chi^2(f) \geq x) + \omega_2[\mathsf{P}(\chi^2(f+4) \geq x) - \mathsf{P}(\chi^2(f) \geq x)] + O(n^{-3}). \tag{4.8}$$

引理的证明见参看文献[4]的8.2.4节.

对随机效应检验, 即检验问题为

$$H_0: \sigma_\alpha^2 = 0 \leftrightarrow H_1: \sigma_\alpha^2 \neq 0.$$

事实上, 我们所考虑解决的检验问题为

$$H_0: \sigma_\tau^2 = \sigma_\epsilon^2 \leftrightarrow H_1: \sigma_\tau^2 \neq \sigma_\epsilon^2,$$

其中$\sigma_\tau^2 = \sigma_\epsilon^2 + T\sigma_\alpha^2$就是§3中$z_{i1}$的方差.

当原假设$\sigma_\alpha^2 = 0$成立时, β_1, β_2和σ_ϵ^2的极大似然估计为

$$\widehat{\beta}_1 = \frac{\overline{z}_1}{\sqrt{T}} - \frac{1+T}{2}\widehat{\beta}_2,$$

$$\widehat{\beta}_2 = \Big[1\Big/\sum_{t=2}^{T}\sum_{i=1}^{n}(\boldsymbol{H}_t\boldsymbol{X}_{i2})^2\Big] \cdot \sum_{t=2}^{T}\sum_{i=1}^{n}\boldsymbol{H}_t\boldsymbol{X}_{i2}z_{it}, \tag{4.9}$$

$$\widehat{\sigma}_\epsilon^2 = \frac{1}{nT}\Big[\sum_{i=1}^{n}(z_{i1}-\overline{z}_1)^2 + \sum_{t=2}^{T}\sum_{i=1}^{n}(z_{i1} - \boldsymbol{H}_t\boldsymbol{X}_{i2}\widehat{\beta}_2)^2\Big], \tag{4.10}$$

其中$\overline{z}_1 = n^{-1}\sum\limits_{i=1}^{n} z_{i1}$. 备择假设下的极大似然估计见§3节.

则似然比检验统计量:

$$\begin{aligned}\Lambda &= \frac{\max\limits_{H_0} L(\beta_1,\beta_2,\sigma_\epsilon^2)}{\max\limits_{H_0\cup H_1} L(\beta_1,\beta_2,\sigma_\epsilon^2,\sigma_\alpha^2)} \\ &= \frac{\Big(\frac{1}{nT}\Big[\sum\limits_{i=1}^{n}(z_{i1}-\overline{z}_1)^2 + \sum\limits_{t=2}^{T}\sum\limits_{i=1}^{n}(z_{it} - \boldsymbol{H}_t\boldsymbol{X}_{i2}\widehat{\beta}_2)^2\Big]\Big)^{-nT/2}}{\Big(\frac{1}{n}\sum\limits_{i=1}^{n}(z_{i1}-\overline{z}_1)^2\Big)^{-n/2}\Big(\frac{1}{n(T-1)}\sum\limits_{t=2}^{T}\sum\limits_{i=1}^{n}(z_{it} - \boldsymbol{H}_t\boldsymbol{X}_{i2}\widehat{\beta}_2)^2\Big)^{-n(T-1)/2}}.\end{aligned} \tag{4.11}$$

在Λ较小时, 拒绝原假设. 全空间被估计的独立参数的个数为4, 而原假设成立时参数空间被估计的独立参数的个数为3, 所以渐近χ^2分布的自由度为1, 在原假设成立即没有随机效应时我们有

$$-2\ln\Lambda \xrightarrow{L} \chi^2(1), \qquad n\to\infty.$$

从而得到检验的渐近P值为$\mathsf{P}(\chi^2(1) \geq -2\ln\Lambda)$.

下面我们对检验的渐近P值进行改进:

令$P_1=\sum\limits_{i=1}^{n}(z_{i1}-\overline{z}_1)^2$, $P_2=\sum\limits_{t=2}^{T}\sum\limits_{i=1}^{n}(z_{it}-\boldsymbol{H}_t\boldsymbol{X}_{i2}\widehat{\beta}_2)^2$, 则似然比可化为

$$\Lambda=\frac{(nT)^{nT/2}}{n^{n/2}(n(T-1))^{n(T-1)/2}}\Big(\frac{P_1}{P_1+P_2}\Big)^{n/2}\Big(\frac{P_2}{P_1+P_2}\Big)^{n(T-1)/2}. \tag{4.12}$$

令$P=P_1/(P_1+P_2)$, 则似然比可化为

$$\Lambda=\frac{(nT)^{nT/2}}{n^{n/2}(n(T-1))^{n(T-1)/2}}(P)^{n/2}(1-P)^{n(T-1)/2}.$$

在原假设成立时, $P\sim\mathrm{Beta}\big((n-1)/2,[n(T-1)-1]/2\big)$分布. 下面计算似然比$\Lambda$的$h$阶矩,

$$\begin{aligned}\mathsf{E}(\Lambda^h)\ &=\ \frac{T^{nTh/2}}{(T-1)^{n(T-1)h/2}}\frac{\Gamma((nT-2)/2)}{\Gamma((n-1)/2)\Gamma([n(T-1)-1]/2)}\\ &\quad\cdot\frac{\Gamma([n(h+1)-1]/2)\Gamma([n(T-1)(h+1)-1]/2)}{\Gamma([nT(h+1)-2]/2)}.\end{aligned} \tag{4.13}$$

对比引理, 若取

$$r=2,\quad a_1=\frac{n}{2},\quad a_2=\frac{n(T-1)}{2},\quad \xi_1=\xi_2=-\frac{1}{2},\quad q=1,\quad b_1=\frac{nT}{2},\quad \eta_1=-1,$$

则可利用引理对P值进行改进. 下面计算f, ρ和ω_2

$$\begin{aligned}
f\ &=\ -2\Big[\sum_{k=1}^{2}\Big(-\frac{1}{2}\Big)-\sum_{j=1}^{1}(-1)-\frac{1}{2}(2-1)\Big]=1,\\
\rho\ &=\ 1-\frac{1}{f}\Big[\sum_{k=1}^{2}a_{k=1}^{-1}B_2(-1/2)-\sum_{j=1}^{1}b_j^{-1}B_2(-1)\Big]=1-\frac{11T^2-26T+26}{6n(T-1)T},\\
\omega_2\ &=\ -\frac{1}{6\rho^2}\Big\{\sum_{k=1}^{2}a_k^{-2}B_3[(1-\rho)a_k+\xi_k]-\sum_{j=1}^{1}b_{j=1}^{-2}B_3[(1-\rho)b_j+\eta_j]\Big\}\\
&=\ -\frac{4}{6n^2\rho^2}\Big\{B_3\Big[\frac{5T^2-20T+26}{12T(T-1)}\Big]+\frac{1}{(T-1)^2}B_3\Big[\frac{11T^2-32T+26}{12T}\Big]\Big\}\\
&\quad-\frac{1}{T^2}B_3\Big[\frac{11T^2-38T+38}{12(T-1)}\Big]\\
&=\ -\frac{4}{6n^2\rho^2}\Big\{\Big(h_1^3+\frac{1}{(T-1)^2}h_2^3-\frac{1}{T^2}h_3^3\Big)-\frac{3}{2}\Big(h_1^2+\frac{1}{(T-1)^2}h_2^2-\frac{1}{T^2}h_3^2\Big)\\
&\quad+\frac{1}{2}\Big(h_1+\frac{1}{(T-1)^2}h_2-\frac{1}{T^2}h_3\Big)\Big\},
\end{aligned}$$

其中,

$$h_1=\frac{5T^2-20T+26}{12T(T-1)},\qquad h_2=\frac{11T^2-32T+26}{12T},\qquad h_3=\frac{11T^2-38T+38}{12(T-1)}.$$

根据引理, 随机效应检验的似然比检验的精度为$O(n^{-1})$, $O(n^{-2})$和$O(n^{-3})$的渐近P值分别为$\mathsf{P}(\chi^2(1)\geq -2\ln\Lambda)+O(n^{-1})$, $\mathsf{P}(\chi^2(1)\geq -2\rho\ln\Lambda)+O(n^{-2})$, $\mathsf{P}(\chi^2(1)\geq -2\rho\ln\Lambda)+\omega_2[\mathsf{P}(\chi^2(5)\geq -2\rho\ln\Lambda)-\mathsf{P}(\chi^2(1)\geq -2\rho\ln\Lambda)]+O(n^{-3})$. 一般用精度为$O(n^{-3})$的渐近$P$值.

下面我们对随机效应存在性检验的势进行模拟, 取$\alpha=0.05$, $T=5$, $\sigma_\epsilon^2=1$, 模拟得到的势为

表1 随机效应存在性检验的势

	$n=10$	$n=30$	$n=50$
$\sigma_\alpha^2=0.5$	0.281	0.779	0.926
$\sigma_\alpha^2=0.7$	0.595	0.980	0.990
$\sigma_\alpha^2=0.9$	0.607	0.974	0.998

从模拟的结果可以看到, 随着样本容量的增加, 检验统计量的表现越来越好.

§5. 线性趋势的检验

检验线性趋势的检验问题是:

$$H_0:\beta_2=0\leftrightarrow H_1:\beta_2\neq 0.$$

当原假设是$H_0:\beta_2=0$时, 模型为

$$y_{ij}=\beta_1+\alpha_i+\epsilon_{ij}\qquad i=1,2,\cdots,n,\ j=1,2,\cdots,T. \tag{5.1}$$

得参数的极大似然估计:

$$\widehat{\beta}_1=\frac{1}{n\sqrt{T}}\sum_{i=1}^n z_{i1},\qquad \widehat{\sigma}_\epsilon^2=\frac{1}{n(T-1)}\sum_{t=2}^T\sum_{i=1}^n z_{it}^2,\qquad \widehat{\sigma}_\alpha=\frac{1}{T}(\widehat{\sigma}^2-\widehat{\sigma}_\epsilon^2),$$

其中$\widehat{\sigma}^2=n^{-1}\sum\limits_{i=1}^n(z_{i1}-\sqrt{T}\widehat{\beta}_1)$.

似然比检验统计量为:

$$\begin{aligned}\Lambda &= \frac{\max\limits_{H_0} L(\beta_1,\beta_2,\sigma_\epsilon^2)}{\max\limits_{H_0\cup H_1} L(\beta_1,\beta_2,\sigma_\epsilon^2,\sigma_\alpha^2)}\\ &= \frac{\Big(\frac{1}{n}\sum\limits_{i=1}^n(z_{i1}-\overline{z}_1)^2\Big)^{-n/2}\Big(\frac{1}{nT}\sum\limits_{t=2}^T\sum\limits_{i=1}^n z_{it}^2\Big)^{-n(T-1)/2}}{\Big(\frac{1}{n}\sum\limits_{i=1}^n(z_{i1}-\overline{z}_1)^2\Big)^{-n/2}\Big(\frac{1}{n(T-1)}\sum\limits_{t=2}^T\sum\limits_{i=1}^n(z_{it}-\boldsymbol{H}_t\boldsymbol{X}_{i2}\widehat{\beta}_2)^2\Big)^{-n(T-1)/2}}\\ &= \frac{T^{n(T-1)/2}\Big(\sum\limits_{t=2}^T\sum\limits_{i=1}^n z_{it}^2\Big)^{-n(T-1)/2}}{(T-1)^{n(T-1)/2}\Big(\sum\limits_{t=2}^T\sum\limits_{i=1}^n(z_{it}-\boldsymbol{H}_t\boldsymbol{X}_{i2}\widehat{\beta}_2)^2\Big)^{-n(T-1)/2}}.\end{aligned} \tag{5.2}$$

令$Q_1=\sum\limits_{t=2}^{T}\sum\limits_{i=1}^{n}z_{it}^2$, $Q_2=\sum\limits_{t=2}^{T}\sum\limits_{i=1}^{n}(z_{it}-\boldsymbol{H}_t\boldsymbol{X}_{i2}\widehat{\beta}_2)^2$, 则$Q_2\sim\sigma_\epsilon^2\chi^2(n(T-1)-1)$分布. 在原假设成立时, $Q_1\sim\sigma_\epsilon^2\chi^2(n(T-1))$分布.

似然比Λ作为检验统计量, 在Λ比较小时拒绝原假设, 等价与用Q_1/Q_2做检验统计量, 在它比较大时拒绝原假设. 又

$$\frac{Q_1}{Q_2}=\frac{Q_1-Q_2}{Q_2}+1. \tag{5.3}$$

用Q_1/Q_2做检验统计量, 等价于用$(Q_1-Q_2)/Q_2$做检验统计量, 在它比较大时拒绝原假设. 而$Q_1-Q_2\sim\sigma_\epsilon^2\chi^2(1)$分布, 且与$Q_2$独立, 所以取检验统计量

$$F=\frac{(Q_1-Q_2)/1}{Q_2/[n(T-1)-1]}. \tag{5.4}$$

在原假设成立时, $F\sim F(1,n(T-1)-1)$分布, 检验的P值为

$$P=\mathsf{P}(F(1,n(T-1)-1)\geq F). \tag{5.5}$$

下面我们对线性趋势检验的势进行模拟, 样本容量分别取为$n=10,30,50$, 取$\alpha=0.05$, $T=5$, $\beta_1=1$, 模拟得到的势为

表2 线性趋势检验的势

	$n=10$	$n=30$	$n=50$
$\beta_2=0.1$	0.074	0.128	0.202
$\beta_2=0.3$	0.289	0.721	0.916
$\beta_2=0.5$	0.692	0.988	0.997

从模拟的结果可以看到, 随着样本容量的增加, 检验统计量的表现越来越好.

§6. 例 子

下面是Hachemeister (1975)文[7]研究从1970年7月到1973年6月五个地区十二个季度中汽车第三者损失险的经验损失. 表3是损失索赔数据.

在[3]中, 对上述数据建立了带有线性趋势的信度模型来研究信度保费的计算, 即建立了模型:

$$y_{ij}=\beta_1+j\beta_2+\alpha_i+\epsilon_{ij}\qquad i=1,2,\cdots,5,\ j=1,2,\cdots,12.$$

要用此模型来研究上述数据, 首先应检验模型类型是否相符, 否则会出现较大的偏差. 下面我们来检验有没有线性趋势:

$$H_0:\beta_2=0\leftrightarrow H_1:\beta_2\neq0.$$

采用(5.5)式的检验统计量F, 得到P值:

$$P = \mathsf{P}(F(1,54) \geq F) < 0.001.$$

故拒绝原假设, 即有线性趋势.

表3 损失索赔数据

	1	2	3	4	5
1	1738	1364	1759	1223	1456
2	1642	1408	1685	1146	1499
3	1794	1597	1479	1010	1609
4	2051	1444	1763	1257	1741
5	2079	1342	1674	1426	1482
6	2234	1675	2103	1532	1572
7	2032	1470	1502	1953	1606
8	2035	1448	1622	1123	1735
9	2115	1464	1828	1343	1607
10	2262	1831	2155	1243	1573
11	2267	1612	2233	1762	1613
12	2517	1471	2059	1306	1690

再检验是否有随机效应,

$$H_0 : \sigma_\alpha^2 = 0 \leftrightarrow H_1 : \sigma_\alpha^2 \neq 0.$$

利用第4节的渐近P值的计算, 得

$$\mathsf{P}(\chi^2(1) \geq -2\ln(\Lambda)) < 0.001.$$

故拒绝原假设, 即有随机效应. 从上面的检验说明模型与数据类型是相符的, 这样就可以利用该模型求信度保费了.

利用第3节参数估计和第2节的信度保费的预测, 可以预测5个地区第13个季度的信度保费, 表示为: $\widehat{y}_{i,13}$, $i = 1, 2, \cdots, 5$, 结果在表4中, 第一行表地区, 第二行表13季度的信度保费估计.

表4 信度保费估计

	1	2	3	4	5
13	2257	1728	2025	1584	1812

感谢 衷心感谢王静龙教授的指导.

参 考 文 献

[1] 王静龙等, 非寿险精算, 北京: 中国人民大学出版社, 2004.

[2] Buhlmann, H. and Gisler, A., *A Course in Credibility Theory and Its Application*, Springer, 2005.

[3] Frees, E.W., Young, V.R. and Luo Y., A longitudinal date analysis interpretation of credibility models, *Insurance: Mathematics and Economics*, **24**(1999), 229–248.

[4] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis* (Second Edition), New York: John Wiley, 1984.

[5] 茆诗松, 王静龙等, 高等数理统计, 北京: 高等教育出版社, 施普林格出版社, 1998.

[6] 王松桂等, 线性模型引论, 北京: 科学出版社, 2004.

[7] Hachemeister, C.A., Credibility for regression models with application to trend in credibility theory and application, *Proceedings of the Berkeley Actuarial Research Conference on Credibility*, Academic Press, New York, 1975, 129–163.

[8] Lo, C.H., Fung, W.K. and Zhu, Z.Y., Generalized estimating equations for variance and covariance parameters in regression credibility models, *Insurance: Mathematics and Economics* **39**(2006), 99–113.

[9] Antonio, K. and Beirlant, J., Actuarial statistics with generalized linear mixed models, *Insurance: Mathematics and Economics*, **40**(2007), 58–76.

Estimating and Testing Parameter for Regression Credibility Model with Linear Trend

Tang Guoqiang

(*School of Finance and Statistics, East China Normal University, Shanghai, 200241*)

(*Department of Mathematics and Physics, Guilin University of Technology, Guilin, 541004*)

In this paper, the parameters of regression credibility model with linear trend are estimated and tested. Orthogonal transformation is used to estimate parameter and unbiased estimate of parameters are obtained. Likelihood ratio test is used to test randomness and linear trend. The better P-value of likelihood ratio test is got and Monte-Carlo simulation is performed.

Keywords: Regression credibility model, orthogonal transformation, likelihood ratio test.

AMS Subject Classification: 62P05.