变结构线性回归模型显著性检验的贝叶斯方法

代逸生①
(华东船舶工业学院管理工程系)

【摘要】以贝叶斯准则为基础,提出了检验变结构线性回归模型显著性的一种方法.在干扰项方差相等和不相等两种情形下,推导出其应用的一般公式.文中以实例说明了此方法的有效性.

关键词:突变理论,多元回归分析,经济数学
分类号:F224

0 引 言

在文献[1]中提出了检测有多次参数突变的变结构线性回归模型变化阀值的贝叶斯方法.该方法中假定模型参数变化的次数,也就是状态数是已知的.从文献[1]的结论中可以知道:在不同的状态数下,将得到不同的变化阀值.因此,需要解决状态数为多少时,所得到变结构线性回归模型最能反映实际规律这一问题.这就是模型的显著性检验.显著性检验的目的就是要在模型尽可能简单和拟合效果尽可能好之间进行权衡和折衷,选出实际经济背景明确,而本身规律性又强的模型.

作为模型显著性检验的准则有许多.本文将以 Bayes 准则为基础,建立变结构线性回归模型显著性检验的贝叶斯方法.有关文献见[2 ~ 5].

1 理论推导

一般地,设有多次参数突变的变结构线性回归模型为

$$Y_t=\begin{cases} X_{1t}\theta+\varepsilon_{1t} & Z_t\leqslant\pi_1 & \text{状态 1}\\ X_{2t}\theta+\varepsilon_{2t} & \pi<Z_t\leqslant\pi_2 & \text{状态 2}\\ \vdots & \vdots & \vdots\\ X_{it}'\theta_i+\varepsilon_{it} & \pi_{i-1}<Z_t\leqslant\pi_i & \text{状态 } i\\ \vdots & \vdots & \vdots\\ X_{kt}'\theta+\varepsilon_{kt} & \pi_{k-1}<Z_t & \text{状态 } k \end{cases} \tag{1}$$

这里 $X_{it}(i=1,2,\cdots,k)$ 是 m_i 维解释变量向量;

$\theta_i(i=1,2,\cdots,k)$ 是 m_i 维参数向量;

变量 Z_t 是引起参数突变的变量,称为研究变量;

$(\pi_1,\pi_2,\cdots,\pi_{k-1})$ 是研究变量 Z_t 在参数突变时的 $k-1$ 个取值,称为状态变化阀值,简称变化阀值;

k 是参数变化的次数,即状态数.

$$\varepsilon_{it}\overset{i.i.d}{\sim}N(0,\sigma_i^2)\quad i=1,2,\cdots,k,COV(\varepsilon_{t_1i},\varepsilon_{t_2i})=0\quad\forall\, t_1\neq t_2$$

---

① 代逸生:硕士,讲师.通讯地址:镇江华东船舶工业学院管理工程系,邮编:212003.
本文 1998 年 7 月 2 日收到.

令 $M_k(k=1,2,\cdots,L)$ 表示对已知的观测数据建立 k 个状态的模型，在 k 确定的情况下，可以利用文献[1]提出的方法，先确定 $k-1$ 个状态变化阈值，再对 k 个状态利用其相应的观测数据分别建立估计模型。因为变结构线性回归模型的状态数是有限的，因此，设定的最大状态数 L 一定存在。

由 Bayes 准则可知，在观测数据 D 下，对其建立有 k 个状态的模型 M_k 的后验概率密度 $P(M_K|D)$，正比于模型 M_K 出现的先验概率密度 $P(M_K)$ 和 M_K 条件下观测数据 D 的概率密度函数(似然函数)的乘积

$$P(M_K|D)\propto P(M_K)\cdot P(D|M_K) \tag{2}$$

根据文献[1,2]的推论，并沿用文献[1]的符号，得到在各状态干扰项方差相等时，M_K 的后验概率密度

$$P(M_K|D)\propto P(M_K)\cdot(2\pi\sigma^2)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}(Y-X\hat{\theta})^{\mathrm{T}}(Y-X\hat{\theta})\right\} \tag{3}$$

用方差 σ^2 的无偏估计

$$\hat{\sigma}^2=\frac{(Y-X\hat{\theta})^{\mathrm{T}}(Y-X\hat{\theta})}{N-\sum_{i=1}^{k}m_i}=\frac{1}{N-\sum_{i=1}^{k}m_i}\sum_{i=1}^{k}S_i \tag{4}$$

代入式(3)得

$$\begin{aligned}P(M_K|D)&\propto P(M_K)\cdot(2\pi\sigma^2)^{-N/2}\exp\left\{-\frac{N-\sum_{i=1}^{k}m_i}{2}\right\}\\&\propto P(M_K)\cdot\left(\frac{1}{N-\sum_{i=1}^{k}m_i}\sum_{i=1}^{k}S_i\right)^{-N/2}\cdot\exp\left\{-\frac{N-\sum_{i=1}^{k}m_i}{2}\right\}\end{aligned} \tag{5}$$

式(3)、(4)、(5)中　N—— 观测值 Y_i 的个数，即观测次数

S_i—— 第 i 个状态下模型的残差平方和

在各状态干扰项方差不相等的情况下，M_K 的后验概率密度为

$$\begin{aligned}P(M_K|D)\propto P(M_K)&\cdot(2\pi\sigma_1^2)^{-t_1/2}\exp\left\{-\frac{1}{2\sigma_1^2}(Y_1-X_1\hat{\theta}_1)^{\mathrm{T}}(Y_1-X_1\hat{\theta}_1)\right\}\\&\cdot(2\pi\sigma_2^2)^{-(t_2-t_1)/2}\exp\left\{-\frac{1}{2\sigma_2^2}(Y_2-X_2\hat{\theta}_2)^{\mathrm{T}}(Y_2-X_2\hat{\theta}_2)\right\}\\&\vdots\\&\cdot(2\pi\sigma_k^2)^{-(N-t_{k-1})/2}\exp\left\{-\frac{1}{2\sigma_k^2}(Y_k-X_k\hat{\theta}_k)^{\mathrm{T}}(Y_k-X_k\hat{\theta}_k)\right\}\end{aligned} \tag{6}$$

其中，σ_i^2 用它的无偏估计

$$\hat{\sigma}_i^2=\frac{(Y_i-X_i\hat{\theta}_i)^{\mathrm{T}}(Y_i-X_i\hat{\theta}_i)}{(t_i-t_{i-1})-m_i}=\frac{1}{(t_i-t_{i-1})-m_i}S_i\qquad i=1,2,\cdots,k \tag{7}$$

$$t_k=N,t_0=0$$

代入式(6)得

$$\begin{aligned}P(M_K|D)\propto P(M_K)&\cdot\left(\frac{1}{t_1-m_1}S_1\right)^{-t_1/2}\exp\left\{-\frac{t_1-m_1}{2}\right\}\\&\cdot\left(\frac{1}{t_2-t_1-m_2}S_2\right)^{-(t_2-t_1)/2}\exp\left\{-\frac{t_2-t_1-m_2}{2}\right\}\\&\vdots\\&\cdot\left(\frac{1}{N-t_{k-1}-m_k}S_k\right)^{-(N-t_{k-1})/2}\exp\left\{-\frac{N-t_{k-1}-m_k}{2}\right\}\end{aligned} \tag{8}$$

式(6)、(7)、(8)中，N—— 观测值 Y_i 的个数，即观测次数

S_i—— 第 i 个状态下模型的残差平方和

t_i-t_{i-1}—— 第 i 个状态下模型对应的观测数据个数($t_k=N,t_0=0$)

考察式(5)、(8)可以发现，$P(M_K)$ 取何种形式成为应用时问题的关键. 似然函数的值将随着状态数的增加而增加. 当状态数选择不合适，致使对实际数据的拟合有较大的偏差时，这种偏差应该由似然函数反映出来，此时似然函数的值较小. 当状态数选择合适，模型对实际数据有较好拟合时，似然函数值将较大，这时再增加状态数，似然函数值就不会有明显的提高. 另一方面，当状态数过于大时，每一状态中的数据个数必然减少，使得参数估计的精确度降低. 在实际问题中，甚至可能出现估计参数不合理、从而掩盖事物本质的现象. 可见，不可能无限制地增加状态数.

为了体现这两个方面的矛盾，限制状态数过大于，又使然函数在状态数小时起显著作用，同时克服实际模型先验概率密度 $P(M_K)$ 难于确定的困难，使用了 3 种形式的先验概率密度

(a) $P(M_k) \propto e^{-\lambda \cdot K}$

(b) $P(M_k) \propto e^{+\lambda \cdot \ln\left(n-\sum_{i=1}^{k} m_i\right)}$

(a) $P(M_k) \propto \left(n-\sum_{i=1}^{k} m_i\right)^{\wedge}$

(其中 $\lambda > 0$，称为控制参数) 进行了模拟计算，结果表明(b)形式的先验概率密度能正确地检验出状态数为多少时最合适. 从(b)的结构看，(b)形式的先验概率密度趋向于选取状态数少的模型，从而限制了状态数过于大. 将(b)代入式(5)、(8)分别得：

各状态干扰项方差相等时，模型 M_K 的后验概率密度为

$$P(M_K|D) \propto e^{+\lambda \ln\left(n-\sum_{i=1}^{k} m_i\right)} \cdot \left(\frac{1}{N-\sum_{i=1}^{k} m_i} \sum_{i=1}^{k} S_i\right)^{-N/2} \cdot \exp\left\{-\frac{N-\sum_{i=1}^{k} m_i}{2}\right\} \tag{9}$$

各状态干扰项方差不相等时，模型 M_k 的后验概率密度为

$$\begin{aligned} P(M_K|D) \propto e^{+\lambda \ln\left(n-\sum_{i=1}^{k} m_i\right)} & \cdot \left(\frac{1}{t_1-m_1} S_1\right)^{-t_1/2} \exp\left\{-\frac{t_1-m_1}{2}\right\} \\ & \cdot \left(\frac{1}{t_2-t_1-m_2} S_2\right)^{-(t_2-t_1)/2} \exp\left\{-\frac{t_2-t_1-m_2}{2}\right\} \\ & \vdots \\ & \cdot \left(\frac{1}{N-t_{k-1}-m_k} S_k\right)^{-(N-t_{k-1})/2} \exp\left\{-\frac{N-t_{k-1}-m_k}{2}\right\} \end{aligned} \tag{10}$$

2 举 例

例 1 仿真模型为

$$\begin{aligned} Y_t &= 5 + 0.7X_t + \varepsilon_t & t &= 1,\cdots,8 \\ Y_t &= 2 + 0.5X_t + \varepsilon_t & t &= 9,\cdots,16 \\ Y_t &= 2 + 0.8X_t - \varepsilon_t & t &= 17,\cdots,25 \end{aligned}$$

$X_t = 0.2t + 10\mathrm{RAND}(t)$ $t = 1,\cdots,25$ $\varepsilon_t \overset{i.i.d}{\sim} N(0,0.6)$ $t = 1,\cdots,25$

这是各状态干扰项方差相等的情况. 用式(9)计算得下表. 可见在 $\lambda = 4,5,6,7,8$ 时，显著性检验的 Bayes 方法式(9)都正确地选择了有 3 个状态的模型. 检验出状态数的同时，也检验出三状态模型的两个阀值：$\pi_1 = 8, \pi_2 = 16$.

	$\lambda=4$	$\lambda=5$	$\lambda=6$	$\lambda=7$	$\lambda=8$
$P(M_1\|.)$	$3.42D+26$	$9.06D+36$	$2.40D+47$	$6.36D+57$	$1.68D+68$
$P(M_2\|.)$	$5.18D+30$	$5.05D+40$	$4.92D+50$	$4.79D+60$	$4.67D+70$
$P(M_3\|.)$	$\underline{1.96D+35}$	$\underline{7.04D+44}$	$\underline{2.52D+54}$	$\underline{9.05D+63}$	$\underline{3.24D+73}$
$P(M_4\|.)$	$1.41D+35$	$1.86D+44$	$2.45D+53$	$3.23D+62$	$4.26D+71$
$P(M_5\|.)$	$4.58D+34$	$2.22D+43$	$1.08D+52$	$5.23D+60$	$2.54D+69$
$P(M_6\|.)$	$1.51D+34$	$2.69D+42$	$4.80D+50$	$8.57D+58$	$1.53D+67$

例 2　仿真模型为

$$Y_t = 0.45X_{1t} + 0.7X_{2t} + \varepsilon_{1t} \qquad t = 1,\cdots,9$$
$$Y_t = 0.2X_{1t} + 0.6X_{2t} + \varepsilon_{2t} \qquad t = 10,\cdots,18$$
$$Y_t = 0.35X_{1t} + 0.2X_{2t} + \varepsilon_{3t} \qquad t = 19,\cdots,27$$
$$X_{1t} = 0.25t + 4\mathrm{RAND}(t) \qquad t = 1,\cdots,27$$
$$X_{2t} = 0.5t + 6\mathrm{RAND}(t) \qquad t = 1,\cdots,27$$
$$\varepsilon_{1t} \overset{i.i.d}{\sim} N(0,0.8)$$
$$\varepsilon_{2t} \overset{i.i.d}{\sim} N(0,0.6)$$
$$\varepsilon_{3t} \overset{i.i.d}{\sim} N(0,0.4)$$

这是各个状态干扰项方差不相等的情况，利用式(10)计算得下表. 可见，除$\lambda=4$外，式(10)都正确地选择了 3 个状态的模型. 在检验出状态数的同时，也检验出三状态模型的两个阈值 $\pi_1=9,\pi_2=18$.

	$\lambda=7$	$\lambda=6$	$\lambda=5$	$\lambda=4$
$P(M_1\|.)$	$4.528D+03$	$3.261D+14$	$2.348D+25$	$\underline{1.690D+36}$
$P(M_2\|.)$	$1.844D+05$	$1.797D+15$	$1.751D+25$	$1.706D+35$
$P(M_3\|.)$	$\underline{2.420D+07}$	$\underline{3.192D+16}$	$\underline{4.210D+25}$	$5.552D+34$
$P(M_4\|.)$	$8.840D+06$	$1.158D+15$	$2.816D+24$	$5.026D+32$
$P(M_5\|.)$	$1.113D+07$	$1.234D+16$	$9.544D+23$	$2.305D+31$
$P(M_6\|.)$	$2.147D+07$	$2.351D+16$	$6.758D+24$	$2.209D+31$

3　结　论

本文提出的变结构线性回归模型显著性检验的贝叶斯方法，同文献[1]中提出的变化阈值的贝叶斯检测方法相结合，解决了变结构线性回归模型建模中的两个基本问题：模型显著性检验和模型变化阈值检测. 因此，利用本文及文献[1]的方法，可以较好地解决社会经济系统建模中具有一定普遍意义的变结构线性回归模型的建立问题.

参 考 文 献

1　代逸生. 贝叶斯方法在变结构线性回归模型建模中的应用. 镇江船舶学院学报，1992；(3)：1～6

2　陆　弘. 突变社会经济系统的建模及其预测：[学位论文]. 天津大学系统工程研究所，1986

3　Pole A M，Smith A F M. A Bayesian analysis of some threshold switching models. Journal of Econometrics，1985；(29)：97～119

（下转第 71 页）

A Researching On the Modelling Theory and Methodology for SCM

Chen Zhixiang, Ma Shihua, Chen Rongqiu
Management School, Huazhong University of Science & Technology

Abstract In this paper, the authors have built a model for supply chain designing based on the theory of "M" putted forward by the authors, which integrates the thoughts of MPSB, CE, OR. cybernetics, BPR. The characteristics of LD model are discussed.

Keywords: supply chain, model, concurrent engineering, thoughtsbase

(上接第 53 页)

4 Tsurumi H. A Bayesian test of the product life cycle hyposthsis as applied to U. S. demand for Color-TV sets. Int. Econ. Rev, 1980; (21): 1～25

5 Zeller A, Williams A D. Bayesian and non-Bayesian analysis of the regressions model with multivariate student-t error terms. JASS, 1976; (71): 400～405

Bayesian Method for Testing of Variable Structure Linear Regression Model

Dai Yisheng
Department of Management Engineering, East China Shipbuilding Institute

Abstract Based on Bayesian Rule, a method for testing of veriable structure linear regression model is presented. Then general formulas of the method applying to practical problem is given on the case of two different form of variances of disturbing item. The examples show effectiveness of the method.

Keywords: catastrophe theory, multiveriate regression analysis, economic mathematics