

# 马氏转移对数正态模型参数置信区间的 Bootstrap 估计

陈明镜

(西华师范大学数学与信息学院, 四川 南充 637002)

**摘要:** 马氏状态转移对数正态模型参数的区间估计很难得到。文章提出的 Bootstrap 法不仅能在 Excel 环境下克服似然函数关于参数求导以及求导后对随机变量求期望的困难, 而且能弥补使用费希尔信息量的倒数  $I(\Theta)^{-1}$  低估估计量波动的缺陷, 同时能方便地得到估计量的分布和协方差阵, 为研究模型系数估计的稳定性提供了切实可行的统计方法。

**关键词:** RSLN 模型; Bootstrap 法; 置信区间; 费希尔信息

**中图分类号:** O212

**文献标识码:** A

## 1 马氏状态转移对数正态模型

马氏状态转移模型 (Markov Regime Switching) 是最近很热门的研究领域, 比如 Kim et al(2008)<sup>[1]</sup>考虑了内生马氏转移模型中内生变量的估计。Tak Kuen Siu 和 Hailiang Yang(2009)<sup>[2]</sup>通过马氏状态转移模型刻画风险资产收益均值和方差的变化进而对期权定价改进。严太华和陈明玉(2009)<sup>[3]</sup>利用马氏状态转移模型对我国上证股指的波动进行了阶段性诊断。马氏转移对数正态模型 (Regime Switching Log Normal RSLN) 可以追溯到 Bollen(1998)<sup>[4]</sup>, 他把对数正态分布引入到状态转移中, 从而既保持了独立对数正态模型的简单易行性, 也弥补了独立对数正态模型的许多缺陷。Hardy(2001)<sup>[5]</sup>继承了文献[4]的思路, 把对数正态转移模型用于具有独立帐户的保险精算研究和风险管理。

RSLN 模型假设  $F_i = \Phi(\mu_i, \sigma_i^2)$  在不同状态之间按照马尔科夫链的方式转移。此模型背后的理论很形象, 即认为市场会从一种相对平稳的状态转移到一种相对波动的状态, 这一观点在 2006 年到 2008 年的股票市场得到了验证。

假设收益率  $Y_t$  在时段  $[t, t+1)$  处于状态  $\pi_t$  之中,  $\pi_t = 1, \dots, K$  给定所处状态下, 收益率服从正态分布, 即:

$$Y_t | \pi_t \sim F(\mu_{\pi_t}, \sigma_{\pi_t}^2)$$

其中  $K$  是一个确定的数,  $\mu_{\pi_t}, \sigma_{\pi_t}^2$  是第  $q$  个状态的均值和方差。

记  $p_{ij} = \Pr(\pi_{t+1} = j | \pi_t = i), i, j = 1, \dots, K$  表示是从状态  $i$  转移到状态  $j$  的概率。

设模型的参数集  $\Theta = (\mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2, p_{11}, \dots, p_{KK})^T$  则随机向量  $Y_1, \dots, Y_n$  的联合密度函数

$$f(y_1, \dots, y_n) = f(y_1) \cdot f(y_2 | y_1) \cdots f(y_n | y_1, \dots, y_{n-1}) \tag{1}$$

其中式(1)右边第  $t$  项是

$$f(y_t | y_1, \dots, y_{t-1}) = \sum_{\pi_t} \left[ \sum_{\pi_{t-1}} P(\pi_{t-1} | y_{t-1}, \dots, y_1) P(\pi_t | \pi_{t-1}) \right] \cdot f(y_t | \pi_t, y_{t-1}, \dots, y_1) \tag{2}$$

其中  $f(y_t | \pi_t, y_{t-1}, \dots, y_1) = f(y_t | \pi_t)$

使用贝叶斯法则推导

$$p(\pi_{t-1} | y_{t-1}, \dots, y_1) = \sum_{\pi_{t-2}=1}^K \frac{f(\pi_{t-1}, \pi_{t-2}, y_{t-1} | y_{t-2}, \dots, y_1)}{f(y_{t-1} | y_{t-2}, \dots, y_1)} \tag{3}$$

起始分布

$$P_{00} = (P(\pi_0 = 1 | y_0), \dots, P(\pi_0 = K | y_0))^T$$

假设使用平稳分布, 也可以使用其它后验分布作为起始分布, 如果使用平稳分布, 即:

$$(\Pi_1, \dots, \Pi_K)Q = (\Pi_1, \dots, \Pi_K)$$

## 2 参数区间估计的困难

使用极大似然估计法(MLE)估计模型参数是一种广泛使用的方法。在样本是独立的条件下,极大似然估计量有很多优秀的性质:渐近无偏性,渐近有效性,渐近正态性和一致性。对模型的参数进行估计后,我们还总希望能得到估计量的置信区间。此时可以利用对数似然函数求得费希尔(Fisher)信息量  $I(\Theta)$ ,从而得到克拉美-罗(Cramer-Rao)下界,即估计量的方差不会小于这个下界。对含有  $s$  个元素的未知参数向量,  $I(\Theta)$  是一个  $s \times s$  的矩阵,其中第  $i$  行第  $j$  列的元素是:

$$I(\Theta)_{ij} = E \left[ - \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln l(\Theta) \right] \quad (4)$$

求期望是针对随机向量  $Y$ 。在式(1)和式(3)的形式下对式(4)的计算是相当复杂的:通过对数似然函数求关于参数的偏导会因为无解析式而变得困难;即便通过计算机程序能够求得偏导,但是对随机向量求期望又是一个相当困难的任务,而随机向量的分布又不是一个常规的概率密度函数形式,且随机变量之间存在相关性。

另外,尽管  $I(\Theta)^{-1}$  经常被用来作为极大似然估计量的渐近方差。但是使用  $I(\Theta)^{-1}$  作为渐近方差很可能低估估计量的方差。

还有一点也不能忽视,尽管极大似然估计法有上述的优秀性质,但是对于时间序列而言,除非时间序列是严格平稳的,极大似然估计量是不会具有上述的多种优秀渐近性质。由此可见,求极大似然估计量的区间估计不仅存在理论上的缺陷,而且实际操作起来也相当麻烦。本文针对RSLN模型参数的极大似然估计法作了修正和改良,主要针对参数的区间估计使用参数Bootstrap法,从而避免了对数似然函数对多个参数的求导和对随机向量  $Y$  的求期望这两项艰巨的任务,同时通过参数Bootstrap法可以得到极大似然估计量的分布和各个参数的估计量的相关性估计,更重要的一点是,参数Bootstrap法可以在Microsoft Excel环境下通过使用“规划求解”和VBA编程轻松实现。

## 3 参数 Bootstrap法的运用与结果对比

为了方便对比,本文使用Hardy(2001)的结果,见表1。Hardy(2001)曾使用加拿大多伦多300月指数(TSE300)1956-2000年的对数收益率拟合两个状态转换的RSLN模型。模型使用了极大似然估计法,并通过费希尔(Fisher)信息量给出了参数估计量的标准误。

表1 Hardy(2001)的极大似然估计和根据渐近正态性得到的90%的置信区间

估计值	置信区间
$\mu_1 = 0.0125$	(0.0092 0.0158)
$\mu_2 = -0.0163$	(-0.0328 0.0001)
$\sigma_1 = 0.0348$	(0.0332 0.0365)
$\sigma_2 = 0.0777$	(0.0662 0.0893)
$p_{12} = 0.0361$	(0.0148 0.0575)
$p_{21} = 0.2111$	(0.1059 0.3164)

采用Bootstrap法,共生成2000条样本,使用极大似然估计法对每个样本进行参数估计,从而得到90%的置信区间。生成样本的过程可以在Excel环境下方便地按照下面的过程完成:

- (1)生成均匀分布随机数  $u \sim U(0, 1)$ ;
- (2)如果  $u < P(\pi_1 = 1)$ , 则假设  $\pi_1 = 1$ , 否则  $\pi_1 = 2$ ;
- (3)生成随机数  $z \sim N(0, 1)$ ;
- (4)  $Y_1 = \mu_{\pi_1} + \sigma_{\pi_1} z$  则为第一时刻的收益率;
- (5)生成随机数  $u \sim U(0, 1)$ , 若  $u < p_{\pi_1, 1}$ , 则假设  $\pi_2 = 1$ , 否则  $\pi_2 = 2$ ;
- (6)重复第(3)步到第(5)步的过程,生成  $Y_i$ ;
- (7)重复第(1)步到第(6)步的过程,生成2000条样本。

Bootstrap法的90%置信区间与费希尔(Fisher)信息量法相比,区间两端点值相差不大,但是Bootstrap法的置信区间略宽于费希尔(Fisher)信息量法的置信区间长度(表2),这不仅因为Bootstrap法的随机性,更因为以  $I(\Theta)^{-1}$  作为渐近方差会低估估计量的方差,这一点在直方图里看得更清楚(图1),特别是  $p_{21}$  和  $p_{12}$  的右尾较厚,说明关于这两个估计量的方差会较大。但是使用费希尔(Fisher)信息量法和根据渐近正态性得到的置信区间却会掩盖估计量的波动性。另外,关于这六个估计量的峰度系数分别是3.214, 9.099, 3.599, 5.639, 11.923, 11.526, 说明与正态分布比较,这六个估计量均有不同程度的后尾特征,而对这六个参数的估计值分别进行Jarque-Bera正态性检验也均拒绝了正态分布的原假设。进一步说明参数的估计量波动会大于费希尔信息量法的波动,由此可见费希尔(Fisher)信息量法的劣势。

表2 Bootstrap法下90%置信区间

估计值	置信区间
$\mu_1 = 0.0125$	(0.0095 0.0158)
$\mu_2 = -0.0163$	(-0.0458 0.0026)
$\sigma_1 = 0.0348$	(0.0317 0.0373)
$\sigma_2 = 0.0777$	(0.0621 0.0922)
$p_{12} = 0.0361$	(0.0140 0.0811)
$p_{21} = 0.2111$	(0.1055 0.4763)

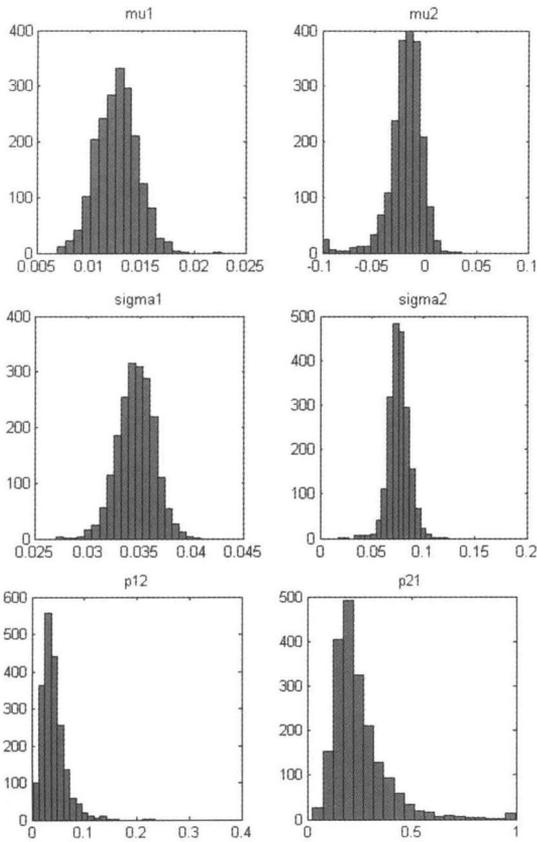


图 1 六个参数的直方图

估计量的协方差阵

$$\begin{pmatrix}
 1 & -0.0751 & -0.0879 & -0.1982 & 0.1066 & -0.0132 \\
 & 1 & -0.3121 & 0.0833 & 0.2716 & -0.3620 \\
 & & 1 & 0.2522 & -0.4620 & 0.0628 \\
 & & & 1 & -0.2104 & 0.0080 \\
 & & & & 1 & 0.3156 \\
 & & & & & 1
 \end{pmatrix}$$

说明估计量之间存在不同大小的相关性。 $\mu_1, \mu_2$  之间有负相关性, 这符合模型期望用两个不同的正态分布拟合数据的初衷;  $\sigma_1, \sigma_2$  存在正相关性, 说明当两个正态分布的均值向相反方向变化时, 两个分布的方差都会增大;  $p_{21}$  和  $p_{12}$  的正相关性说明两个状态之间的转换频率加剧。这些数值结果均符合模型的假设。

#### 4 结束语

本文提出简单易行的 Bootstrap 法, 这种方法能避开复杂的求偏导和求期望运算, 不仅能在 Excel 环境下轻松实现, 而且弥补了  $I(\Theta)^{-1}$  的缺陷, 同时能方便地得到估计量的分布特征, 也为研究模型系数估计的稳定性提供了切实可行的统计方法。

#### 参考文献:

[1] Kim C J, Piger J, Startz R. Estimation of markov regime switching regression models with endogenous switching [ J ]. Journal of Econometrics 2008, 143 ( 2 ): 263-273

[2] Tak Kuen Siu, Yang Hailiang. Option pricing when the regime switching risk is priced [ J ]. Acta Mathematicae Applicatae Sinica (English version), 2009, 25( 3 ): 369-388

[3] 严太华, 陈明玉. 基于马尔科夫切换模型的上证指数周收益率时间序列分析 [ J ]. 中国管理科学, 2009, 17( 6 ): 33-38.

[4] Bollen N P B. Valuing Options in Regime Switching Models [ J ]. Journal of Derivatives 1998, 6: 38-49

[5] Hardy M R. A Regime Switching Model of Long Term Stock Returns [ J ]. North American Actuarial Journal 2001, 5( 2 ): 41-53

## Confidence in Parameters' Interval of Markov Regime Switching Lognormal Model Based on Bootstrap Method

CHEN Ming-jing

( College of Mathematics and Information, China West Normal University, Nanchong 637002, China )

**Abstract** The estimation of Markov Regime Switching Lognormal Model parameters' confidence interval is difficult to be gained. This paper recommends Bootstrap method that can be easily realized in Excel and remedy the drawbacks of  $I(\Theta)^{-1}$ , moreover we can get confidence interval, estimators' distributions and variance-covariance matrix and provides an easy but reliable method.

**Key words** RSLN Model; Bootstrap method; confidence interval; Fisher information