第24卷第5期 管理科学 ISSN 1672-0334 Vol. 24 No. 5 111-119
2011年10月 Journal of Management Science October, 2011

基于马尔科夫状态转移模型的股指收益率研究

瞿 慧，肖斌卿
南京大学 工程管理学院，南京 210093

摘要：金融资产收益率的分布是金融资产投资和风险管理等应用中的重要决定因素。针对经济和金融的潜在状态改变可能引起金融资产收益率分布结构性变化的现实情况，提出考虑收益率分布的时变性，将马尔科夫状态转移结构应用于中国股票指数对数收益率分布的建模，并提出使用混合正态分布模型刻画股指收益率在各状态的分布，建立隐马尔科夫状态转移-混合正态分布（HMS-MND）模型，使用期望最大化算法（E-M 算法）和 Baum-Welch 算法给出模型的参数估计，采用2002年7月1日至2010年10月29日沪深两市11种主要股票指数的对数日、周收益率作为实证数据。模型参数估计和似然比检验结果表明，大部分股票指数收益率的分布中都存在显著的马尔科夫状态转移结构，且 HMS-MND 模型可以较好地刻画对数收益率分布的高阶矩统计特征。因此，引入马尔科夫状态转移结构对股指收益率的相关研究具有重要意义。

关键词：收益率；马尔科夫状态转移；混合正态分布；E-M 算法；似然比检验
中图分类号：F830.91 **文献标识码**：A **文章编号**：1672-0334(2011)05-0111-09

1 引言

金融资产收益率的分布对金融资产投资和风险管理等具有重要意义，长期以来吸引了众多学者的研究兴趣。大量文献实证表明，传统的正态分布假设不能很好地刻画现实金融资产对数收益率的尖峰、厚尾和不对称性，这一局限性对于风险价值（value at risk，VaR）的计算等实际应用有重大影响。为此，学者们转向考察 Pareto 分布、t 分布、稳定分布、混合分布等多种统计分布形式对现实金融数据的刻画能力，以探寻更为合理的对数收益率分布模型，至今尚未达到一致共识。

本研究提出将马尔科夫状态转移结构应用于中国股票指数对数收益率分布的建模，并假设股指的对数收益率在马尔科夫状态转移结构的任一状态均符合混合正态分布，为对数收益率建立隐马尔科夫状态转移-混合正态分布模型（hidden Markov switching-mixed normal distribution model，HMS-MND model）。混合正态分布能在保持正态分布易操作性和有限高阶矩特征的同时，较好地刻画对数收益率的尖峰、厚尾特征，且其各组成成分自然对应于造成价格波动的不同质的新信息。马尔科夫状态转移结构的引入，旨在摆脱所有观测数据源于同一分布的严格假设，将观测数据分布的时间异质性纳入考虑，提高模型刻画时变收益率分布的能力。本研究通过实证对将马尔科夫状态转移结构引入中国股票指数收益率分布模型的意义做出评价。

2 相关研究评述

近年来，中国学者对股票和股票指数收益率的分布进行了大量研究，对正态性假设进行了实证否定，

收稿日期：2010-11-08 **修返日期**：2011-07-05
基金项目：国家自然科学基金（70932003，70901037）；中央高校基本科研业务费专项资金（1107011810，1118011804）；江苏省自然科学基金面上项目（BK2011561）；南通大学江苏沿海沿江发展研究院项目（Y201002）；教育部人文社会科学研究项目（09YJCZH061）；教育部留学回国人员科研启动基金
作者简介：瞿慧（1981-），女，江苏南通人，毕业于美国康奈尔大学，获博士学位，现任南京大学工程管理学院副教授，研究方向：计算金融等。E-mail：linda59qu@nju.edu.cn

并对国际上提出的多种分布模型拟合中国金融数据的能力进行了实证检验。封建强等[1]发现 t 分布能更好地拟合1996 年12 月16 日涨跌停板制度推出后的沪深股市收益率分布;楼小飞等[2]实证发现,1997 年 1 月 6 日至2002 年 6 月21 日沪深两市 A 股随机组合的周收益率不呈现正态分布;曹志广等[3]实证表明,广义双曲线分布和正态逆高斯分布可以较正态分布更好地拟合1997 年 1 月 2 日至2003 年 9 月19 日的上证综指日收益率分布;徐绪松等[4]实证表明,非正态稳定分布对上证指数和深证成指在1997 年 1 月 1 日至2004 年12 月31 日期间的日、周收益率分布拟合效果较好;赵秀娟等[5]发现非对称 Laplace 分布可以较好地拟合2003 年至2006 年中国17 支开放式基金的日收益率序列;宋丽娟等[6]实证了 Laplace 分布对1997 年 6 月10 日至2002 年 2 月28 日沪深两市 6 种主要指数日收益率分布特征的刻画能力;郑玉华等[7]以1996 年12 月16 日至2008 年11 月11 日的上证综指日数据为研究对象,实证表明混合正态分布模型可以更好地刻画过滤了自相关和异方差效应的收益率序列,并指出应当考虑混合正态分布参数的可变性。

为了使模型参数具有一定的可变性,提高模型刻画时间序列时变分布的能力,学者们常常选择引入马尔科夫状态转移结构。Cheung 等[8]和 Dueker 等[9]对外汇汇率模型的研究、Kim[10-11]对美联储利率和美国人均消费的建模、Lin 等[12]对道琼斯工业平均指数成分股票收益率移动平均模型的研究、Erlwein 等[13]对电价预测模型的研究、Dias 等[14]对亚洲股票指数收益率模型的研究、Geweke 等[15]对 S&P 500 指数日收益率分布模型的研究中都引入了马尔科夫状态转移结构。近年来,中国学者也开始关注马尔科夫状态转移结构在金融时间序列建模中的应用。刘金全等[16]和吴吉林等[17]都在对中国短期利率的研究中引入马尔科夫状态转移结构以刻画潜在的时变分布;高金余等[18]和严太华等[19]分别使用马尔科夫状态转移模型研究深证成指和上证综指的收益率。

大量研究表明,经济和金融的潜在状态改变可能引起金融时间序列分布的结构性变化,而在分布模型中引入马尔科夫状态转移结构可以有效捕捉这种内生性的结构转变。因此,本研究将马尔科夫状态转移结构引入中国股票指数收益率分布模型。对于对数收益率在马尔科夫状态转移结构任一状态的具体分布,不同于现有文献[18-19],本研究选择混合正态分布模型,主要出于以下两点考虑。一方面,市场价格和收益率的波动通常是新信息进入市场产生冲击造成的结果,正态分布随机变量是对新信息的自然刻画,而混合正态分布则自然对应于对不同质新信息的联合刻画[20]。另一方面,混合正态分布刻画对数收益率尖峰厚尾特征的能力早已得到大量文献的实证[20-22]。因此,本研究为中国股票指数的对数收益率建立隐马尔科夫状态转移-混合正态分布(HMS-MND)模型,模型名称中的“隐”表述了造成马尔科夫状态转移的策动变量以及模型的马尔科夫状态本身均不可观测的特性。

由于 HMS-MND 模型具有无法直接观测的隐状态随机变量,其完全数据的似然函数也具有随机性,本研究使用期望最大化算法(expectation-maximization algorithm,EM 算法)[23-24]对模型的参数进行估计,并使用 Baum-Welch 算法[24-25]提高 E-M 算法进行参数估计的效率。

3 隐马尔科夫状态转移-混合正态分布模型

本研究将马尔科夫状态转移结构引入中国股票指数对数收益率的分布模型,以摆脱所有观测数据源于同一分布的严格假设,提高模型刻画时变收益率分布的能力。假设观测到的对数收益率在 M 种可能的分布模型中切换,其中各种分布模型都是 K 个正态分布的加权混合(混合正态分布模型)。若当前观测到的收益率数据对应于编号为 $i(i=1,2,\cdots,M)$ 的混合正态分布模型,则称当前处于状态 i;由于与状态相对应的分布模型不可观测,因此我们称之为状态隐变量。假设各状态之间的转移概率仅取决于当前状态,而与状态转移的历史路径无关。这样,可以为股票指数的对数收益率建立如下的隐马尔科夫状态转移-混合正态分布模型。

$$p(r_n|Q_n=i)=\sum_{k=1}^{K}c_{ik}\varphi_{ik}(r_n|\mu_{ik},\sigma_{ik}) \quad i\in\Omega=\{1,2,\cdots,M\} \tag{1}$$

$$\forall i\in\Omega,\sum_{k=1}^{K}c_{ik}=1 \tag{2}$$

$$\forall n\leqslant N,P(Q_n|Q_{n-1},R_{n-1},\cdots,Q_1,R_1)=P(Q_n|Q_{n-1}) \tag{3}$$

$$\forall n\leqslant N,P(R_n|Q_N,R_N,Q_{N-1},R_{N-1},\cdots,Q_{n+1},R_{n+1},Q_n,Q_{n-1},R_{n-1},\cdots,Q_1,R_1)=P(R_n|Q_n) \tag{4}$$

其中,$p(r_n|Q_n=i)$为状态为 i 时对数收益率的概率密度函数;R_n为收盘价为S_n的股票指数在 n 日的对数收益率,r_n为 R_n 的取值,$R_n=\ln S_n-\ln S_{n-1}$;Q_n 为对数收益率在 n 日的状态隐变量,其取值范围是该隐马尔科夫链的状态空间 Ω,$\Omega=\{1,2,\cdots,M\}$;c_{ik} 为状态为 i 时的混合正态分布第 k 个组成成分的权重系数;$\varphi_{ik}(\cdot|\mu_{ik},\sigma_{ik})$为状态为 i 时的混合正态分布第 k 个组成成分(均值为 μ_{ik},标准差为 σ_{ik})的概率密度函数。(3) 式和(4) 式是模型的两个条件独立性假设,$P(\cdot)$ 为概率测度,N 为数据长度。

进一步假设隐马尔科夫链的状态转移概率与时间无关,a_{ij} 为状态 i 到状态 j 的一步转移概率,即 $P(Q_n=j|Q_{n-1}=i)=a_{ij},\forall 1<n\leqslant N$;$A$ 为隐马尔科夫链的一步转移概率矩阵,$A=\{a_{ij}\}_{M\times M}$;π_i 为状态 i 的初始概率,$\pi_i=P(Q_1=i)$;I 为隐马尔科夫链的初始概率分布,$I=\{\pi_i\}_{1\times M}$;令 B 为各状态混合正态分布组成成分的均值、标准差和相应权重,$B=\{\mu_{ik},\sigma_{ik},c_{ik}\}_{M\times K}$;$\Theta$ 为 HMS-MND 模型的完整参数集,$\Theta=\{A,I,B\}$。

将观测到的长度为 N 的对数收益率序列记作 R,$R=(R_1=r_1,\cdots,R_N=r_N)$,其伴随的隐状态序列记作 Q,$Q=(Q_1=q_1,\cdots,Q_N=q_N)$,$q_n$ 为状态隐变量 Q_n 的取值,

$n = 1,2,\cdots,N$。根据 HMS-MND 模型假设,完全数据的对数似然函数为

$$L(\Theta \mid R,Q) = \log p(R,Q \mid \Theta)$$
$$= \log[\pi_{q_1} p(r_1 \mid Q_1 = q_1)\prod_{n=2}^{N} a_{q_{n-1}q_n} p(r_n \mid Q_n = q_n)] \quad (5)$$

对模型的参数估计,即寻找最大化对数似然函数值的参数集 $\Theta^* = \{A^*, I^*, B^*\}$

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}}\, L(\Theta \mid R,Q) = \underset{\Theta}{\operatorname{argmax}}\, \log p(R,Q \mid \Theta) \quad (6)$$

4 模型参数估计和检验

4.1 期望最大化算法

考虑到隐状态序列 Q 无法观测,(5) 式中的对数似然函数作为随机变量 Q 的函数也具有不确定性。因此,采用期望最大化算法(E-M 算法),通过 E-步骤和 M-步骤的交替进行来迭代优化模型参数。

4.1.1 E-步骤

EM 算法的第一个步骤即建立关于对数似然函数的期望,以消除最优化目标函数的随机性。定义对数似然函数 $L(\Theta \mid R,Q)$ 的条件期望函数 $H(\Theta,\Theta^g)$,并以之为后继 M-步骤的优化对象,即

$$H(\Theta,\Theta^g) = E_{Q \mid R,\Theta^g}[\log p(R,Q \mid \Theta)]$$
$$= \sum_{Q \in \Omega^N} \log p(R,Q \mid \Theta) p(Q \mid R,\Theta^g) \quad (7)$$

其中,Θ^g 为当前的参数集估计值(或初始参数集),Ω^N 为 N 维隐状态序列 Q 的取值空间,$E_{Q \mid R,\Theta^g}$ 为已知股票收益率序列 R 和当前参数集 Θ^g 时对隐状态序列 Q 取条件期望。

将(5) 式代入(7) 式,则 $H(\Theta,\Theta^g)$ 函数可转化为

$$H(\Theta,\Theta^g) = \sum_{Q \in \Omega^N} \log p(R,Q \mid \Theta) p(Q \mid R,\Theta^g)$$
$$= \underbrace{\sum_{Q \in \Omega^N} \log \pi_{q_1} p(Q \mid R,\Theta^g)}_{(a)} +$$
$$\underbrace{\sum_{Q \in \Omega^N} (\sum_{n=2}^{N} \log a_{q_{n-1}q_n}) p(Q \mid R,\Theta^g)}_{(b)} +$$
$$\underbrace{\sum_{Q \in \Omega^N} [\sum_{n=1}^{N} \log p(r_n \mid Q_n = q_n)] p(Q \mid R,\Theta^g)}_{(c)} \quad (8)$$

4.1.2 M-步骤

对 $H(\Theta,\Theta^g)$ 的最大化可以通过分别最大化(8) 式中的 (a)、(b) 和 (c) 3 个部分来实现。通过拉格朗日乘子法求解可得初始概率 π_i、转移概率 a_{ij}、正态分布权重系数 c_{ik}、均值参数 μ_{ik}、标准差参数 σ_{ik} 的最优值分别为

$$\pi_i^{update} = P(Q_1 = i \mid R,\Theta^g) \quad (9)$$

$$a_{ij}^{update} = \frac{\sum_{n=1}^{N-1} P(Q_n = i, Q_{n+1} = j \mid R,\Theta^g)}{\sum_{n=1}^{N-1} P(Q_n = i \mid R,\Theta^g)} \quad (10)$$

$$c_{ik}^{update} = \frac{\sum_{n=1}^{N} P(Q_n = i, X_{in} = k \mid R,\Theta^g)}{\sum_{n=1}^{N} P(Q_n = i \mid R,\Theta^g)} \quad (11)$$

$$\mu_{ik}^{update} = \frac{\sum_{n=1}^{N} r_n P(Q_n = i, X_{in} = k \mid R,\Theta^g)}{\sum_{n=1}^{N} P(Q_n = i, X_{in} = k \mid R,\Theta^g)} \quad (12)$$

$$\sigma_{ik}^{update} = \sqrt{\frac{\sum_{n=1}^{N} (r_n - \mu_{ik}^{update})^2 P(Q_n = i, X_{in} = k \mid R,\Theta^g)}{\sum_{n=1}^{N} P(Q_n = i, X_{in} = k \mid R,\Theta^g)}} \quad (13)$$

其中,$X_{in} = k$ 表示 R_n 对应于状态 i 下混合正态分布的第 k 个组成成分。由(9)式 ~(13)式可知,参数集的更新涉及到如下 3 个概率的计算,即

$P(Q_n = i \mid R,\Theta^g)$

$P(Q_n = i, Q_{n+1} = j \mid R,\Theta^g)$

$P(Q_n = i, X_{in} = k \mid R,\Theta^g)$

本研究将使用 Baum-Welch 算法进行求解。

4.2 Baum-Welch 算法

定义前向过程 $\alpha_i(n) = p(r_1,\cdots,r_n, Q_n = i \mid \Theta)$,则 $\alpha_i(n)$ 满足如下递推关系,即

$$\alpha_i(1) = \pi_i p(r_1 \mid Q_1 = i) = \pi_i \sum_{k=1}^{K} c_{ik}\varphi_{ik}(r_1 \mid \mu_{ik},\sigma_{ik})$$

$$\alpha_j(n+1) = [\sum_{i=1}^{M} \alpha_i(n) a_{ij}] p(r_{n+1} \mid Q_{n+1} = j) \quad (14)$$
$$= [\sum_{i=1}^{M} \alpha_i(n) a_{ij}] \sum_{k=1}^{K} c_{jk}\varphi_{jk}(r_{n+1} \mid \mu_{jk},\sigma_{jk})$$

定义后向过程 $\beta_i(n) = p(r_{n+1},\cdots,r_N \mid Q_n = i,\Theta)$,则 $\beta_i(n)$ 满足如下递推关系,即

$$\beta_i(N) = 1$$

$$\beta_i(n) = \sum_{j=1}^{M} a_{ij} p(r_{n+1} \mid Q_{n+1} = j)\beta_j(n+1) \quad (15)$$
$$= \sum_{j=1}^{M} a_{ij}[\sum_{k=1}^{K} c_{jk}\varphi(r_{n+1} \mid \mu_{jk},\sigma_{jk})]\beta_j(n+1)$$

根据(14) 式和(15) 式,在当前参数集 Θ^g 下,可以递推计算出 $\{\alpha_i(n)\}_{M\times M}$ 和 $\{\beta_i(n)\}_{M\times N}$。且由 $\alpha_i(n)$ 和 $\beta_i(n)$ 的定义和模型的条件独立性假设可得 $\alpha_i(n)\beta_i(n) = p(R, Q_n = i \mid \Theta)$。

在此基础上,容易计算出 HMM-MND 模型参数更新的关键概率值为

$$P(Q_n = i \mid R,\Theta^g) = \frac{\alpha_i(n)\beta_i(n)}{\sum_{j=1}^{M} \alpha_j(n)\beta_j(n)} \quad (16)$$

$$P(Q_n = i, Q_{n+1} = j \mid R,\Theta^g)$$
$$= \frac{\alpha_i(n) a_{ij} \sum_{k=1}^{K} c_{jk}\varphi_{jk}(r_{n+1} \mid \mu_{jk},\sigma_{jk})\beta_j(n+1)}{\sum_{j=1}^{M} \alpha_j(n)\beta_j(n)} \quad (17)$$

$$P(Q_n = i, X_{in} = k \mid R,\Theta^g)$$
$$= \frac{\alpha_i(n)\beta_i(n)}{\sum_{j=1}^{M} \alpha_j(n)\beta_j(n)} \frac{c_{ik}\varphi_{ik}(r_n \mid \mu_{ik},\sigma_{ik})}{\sum_{l=1}^{K} c_{il}\varphi_{il}(r_n \mid \mu_{il},\sigma_{il})} \quad (18)$$

其中,c_{il} 为状态为 i 时的混合正态分布第 l 个组成成

分的权重系数；$\varphi_{il}(r_n|\mu_{il},\sigma_{il})$ 为状态为 i 时的混合正态分布第 l 个组成成分(均值为μ_{il}，标准差为σ_{il}) 的概率密度函数。

把(16)式~(18)式的计算结果代入(9)式~(13)式，即可实现 HMS-MND 模型的一次参数更新。E-M 算法不断重复 E-步骤和 M-步骤，直至参数更新带来的似然函数值增长比例低于预设的门限值，或者已经达到预设的参数更新次数上限，此时的一组最新参数就是 E-M 算法估计得到的 HMM-MND 模型最优参数。

4.3 似然比检验

使用似然比检验(likelihood ratio test，LRT)评估将马尔科夫状态转移结构引入中国股票指数对数收益率分布模型的意义。似然比检验的零假设(H_0)为，股指的对数收益率服从简单的混合正态分布模型，因此具有参数集$\Theta_0=\{\mu_k,\sigma_k,c_k\}_{1\times K}$，即 $H_0:\Theta=\Theta_0$。似然比检验的备择假设(H_1)为，股指的对数收益率服从隐马尔科夫状态转移-混合正态分布模型，因此具有参数集 $\Theta_1=\{\{a_{ij}\}_{M\times M},\{\pi_i\}_{1\times M},\{\mu_{ik},\sigma_{ik},c_{ik}\}_{M\times K}\}$，即 $H_1:\Theta=\Theta_1$。

用L_0和 L_1 分别表示零假设和备择假设下对数似然函数的最大值，则似然比检验的统计量$LR=-2(L_0-L_1)\to\chi^2_m$，这里卡方分布的自由度 m 等于备择假设和零假设的模型待估计参数个数差($m=|\Theta_1|-|\Theta_0|$)。由 HMM-MND 模型满足 $\sum_{j=1}^{M}a_{ij}=1$、$\sum_{i=1}^{M}\pi_i=1$、$\sum_{k=1}^{K}c_{ik}=1$ 可知，$\{a_{ij}\}_{M\times M}$ 对应待估计的参数个数为$M\times(M-1)$，$\{\pi_i\}_{1\times M}$ 对应待估计的参数个数为$(M-1)$，$\{c_{ik}\}_{M\times K}$ 对应待估计的参数个数为$M\times(K-1)$，$\{c_k\}_{1\times K}$ 对应待估计参数个数为$(K-1)$。因此有$m=M^2+3MK-M-3K$。

根据卡方分布临界值表可以判断模型差异是否显著，从而对在中国股票指数对数收益率分布模型中引入马尔科夫状态转移结构的意义做出评价。

5 仿真和实证结果及分析

5.1 数值仿真结果

本研究通过 MATLAB 实现 HMS-MND 模型参数估计的 E-M 算法。首先通过数值仿真，验证算法逼近真实分布模型参数的能力和参数估计的收敛性能。考察多组不同特征的 HMS-MND 模型参数设定，在每组参数设定下均进行20 次数值仿真，每次都首先通过随机数发生器产生符合参数设定的 HMS-MND 序列，序列长度为$N=1\ 500$；然后，用 E-M 算法对序列的 HMS-MND 模型参数进行估计。表 1 给出其中一组参数设定下20 次数值仿真的统计结果，其余各组仿真的参数估计性能与之类似。可见，实现的 E-M 算法能够有效地识别序列分布中隐含的马尔科夫状态转移结构，对 HMS-MND 模型的参数进行估计。

表1 HMS-MND 模型的参数估计性能数值仿真结果

Table 1 Simulation Results of Parameter Estimation Performance for HMS-MND Model

	参数列表	真实参数值	估计值	估计标准差	估计相对误差
马尔科夫状态转移结构参数	$[\pi_1,\pi_2]$	[1,0]	[1,0]	[0,0]	[0,0]
	a_{11}	0.950	0.949	0.010	0.148%
	a_{12}	0.050	0.051	0.010	2.810%
	a_{21}	0.050	0.050	0.009	0.630%
	a_{22}	0.950	0.950	0.009	0.033%
混合正态分布权重参数	c_{11}	0.200	0.199	0.018	0.475%
	c_{12}	0.800	0.801	0.018	0.119%
	c_{21}	0.340	0.309	0.090	9.054%
	c_{22}	0.660	0.691	0.090	4.664%
混合正态分布均值参数	μ_{11}	0.009	0.009	0.002	0.444%
	μ_{12}	0.004	0.004	0.000	0.429%
	μ_{21}	0.039	0.038	0.004	2.833%
	μ_{22}	0.014	0.015	0.004	7.857%
混合正态分布标准差参数	σ_{11}	0.020	0.021	0.001	2.300%
	σ_{12}	0.002	0.002	0.000	0.667%
	σ_{21}	0.030	0.028	0.004	8.133%
	σ_{22}	0.050	0.051	0.002	1.530%

5.2 实证结果及分析

本研究探讨在中国股票指数对数收益率分布模型中是否存在马尔科夫状态转移结构,因此以中国沪深两市代表性股票指数为研究对象,表2给出选择的11种股票指数及其相应的收益率数据时间段和数据频率。上证50指数、深证100指数、沪深300指数和上证红利指数的数据开始日期均选择为其正式推出后的第一个完整交易周的周一;其余7种沪深两市股票指数均有较长的历史数据记录,考虑到在此之前中国证券市场经历了1996年12月16日涨跌停板制度的推出和2001年至2002年6月24日国有股减持的启动到停止等影响股价和收益率的重大政策性事件,因此将收益率数据开始日期选择为此后第一个完整交易周的周一,即2002年7月1日。

为了研究在中国股票指数对数收益率分布模型中是否存在马尔科夫状态转移结构,对表2中的股票指数使用E-M算法分别拟合下列两种分布模型。零假设:简单混合正态分布模型,两个正态分布的混合,具有参数集$\Theta_0=\{\mu_k,\sigma_k,c_k\}_{1\times2}$;备择假设:HMS-MND模型,两种隐状态,每个状态均为两个正态分布的混合,具有参数集$\Theta_1=\{\{a_{ij}\}_{2\times2},\{\pi_i\}_{1\times2},\{\mu_{ik},\sigma_{ik},c_{ik}\}_{2\times2}\}$。相应的,零假设的待估计参数个数为$|\Theta_0|=5$,备择假设的待估计参数个数为$|\Theta_1|=13$。因此,似然比检验统计量对应卡方分布的自由度为8。表3给出对表2中11种指数的对数日收益率进行HMS-MND模型参数估计和似然比检验的结果,其中最后一行的LR即为似然比检验的统计量。

表2　收益率数据列表

Table 2　Research Data List

名称	数据开始日期	数据终止日期	数据频率
上证综合指数	2002-07-01	2010-10-29	日、周
上证A股指数	2002-07-01	2010-10-29	日、周
上证B股指数	2002-07-01	2010-10-29	日、周
上证50指数	2004-01-05	2010-10-29	日、周
上证180指数	2002-07-01	2010-10-29	日、周
深证成份指数	2002-07-01	2010-10-29	日、周
深证A股指数	2002-07-01	2010-10-29	日、周
深证B股指数	2002-07-01	2010-10-29	日、周
深证100指数	2003-01-06	2010-10-29	日、周
沪深300指数	2002-01-07	2010-10-29	日、周
上证红利指数	2005-01-10	2010-10-29	日、周

数据来源:Wind资讯,下同。

表3　HMS-MND模型参数估计和似然比检验结果(对数日收益率)

Table 3　HMS-MND Model Parameter Estimates and Likelihood Ratio Test Results for Stock Indexes' Log Daily Returns

	上证综合指数	上证A股指数	上证B股指数	上证50指数	上证180指数	深证成份指数	深证A股指数	深证B股指数	深证100指数	沪深300指数	上证红利指数
a_{11}	0.994	0.994	0.981	0.997	0.996	0.995	0.993	0.970	0.994	0.995	0.997
a_{12}	0.006	0.006	0.019	0.003	0.004	0.005	0.007	0.030	0.006	0.005	0.004
a_{21}	0.011	0.012	0.042	0.006	0.006	0.009	0.010	0.039	0.009	0.008	0.004
a_{22}	0.989	0.989	0.958	0.994	0.994	0.991	0.990	0.961	0.991	0.992	0.997
c_{11}	0.701	0.741	0.716	0.683	0.723	0.591	0.821	0.762	0.956	0.953	0.894
c_{12}	0.299	0.259	0.284	0.317	0.277	0.409	0.179	0.238	0.044	0.047	0.106
c_{21}	0.383	0.384	0.527	0.399	0.570	0.554	0.531	0.549	0.531	0.556	0.637
c_{22}	0.617	0.616	0.474	0.601	0.430	0.447	0.469	0.451	0.469	0.444	0.363
μ_{11}	0.0003	0.0002	0.001	-0.001	-0.001	0.001	-0.000	0.002	0.001	-0.000	0.001
μ_{12}	0.0002	0.0004	0.000	0.004	0.003	-0.000	0.002	0.000	0.002	0.007	-0.001
μ_{21}	0.008	0.008	-0.003	0.008	-0.005	-0.004	0.009	0.002	-0.006	-0.005	-0.005
μ_{22}	-0.004	-0.004	0.003	-0.004	0.009	0.007	-0.008	-0.003	0.009	0.008	0.011
σ_{11}	0.014	0.014	0.009	0.010	0.009	0.016	0.013	0.009	0.012	0.011	0.011
σ_{12}	0.005	0.004	0.020	0.019	0.018	0.008	0.004	0.018	0.029	0.029	0.026
σ_{21}	0.012	0.012	0.042	0.015	0.031	0.032	0.015	0.016	0.032	0.031	0.033
σ_{22}	0.029	0.029	0.017	0.031	0.013	0.016	0.031	0.033	0.015	0.014	0.012
LR	31.503***	137.514***	-533.977	89.176***	94.976***	-140.568	215.558***	-516.054	703.692***	935.181***	550.916***

注:***为在显著性水平0.010时统计显著,下同。

从表 3 的似然比统计量 LR 取值情况可知，除上证 B 股指数、深证成份指数、深证 B 股指数外，其余 8 种沪深两市股票指数的对数日收益率分布模型在引入马尔科夫状态转移结构后拟合能力都取得显著提高。本研究认为，深证成份指数的似然比统计量 LR 小于零，主要是由于其中 B 股成分的影响。类似的，上证综合指数由于受到其中 B 股成分的影响，其似然比统计量要小于上证 A 股指数的似然比统计量。从其余 8 种指数拟合的 HMS-MND 模型参数可以看到，指数日收益率的分布倾向于停留在其所处的状态($a_{11} > 0.99, a_{22} > 0.98$)，但仍以一个非零的概率在不同状态间切换；两种隐状态的混合正态分布组成成分在均值和方差上有较大差异，具有不同的高阶矩特征。以上证50 指数为例，HMS-MND 模型分布状态 1 对应混合正态分布组成成分为 $N(-0.001, 0.010)$ 和 $N(0.004, 0.019)$，相应权重分别为0.683 和 0.317，使用该参数生成100 组与研究数据等长的随机序列，记录各序列的偏度和峰度并计算其平均值可知，该混合正态分布刻画的偏度大致为0.357，峰度大致为4.512。HMS-MND 模型分布状态 2 对应混合正态分布组成成分为 $N(0.008, 0.015)$ 和 $N(-0.004, 0.031)$，相应权重分别为0.399 和 0.601，使用该参数生成100 组与研究数据等长的随机序列，记录各序列的偏度和峰度并计算其平均值可知，该混合正态分布刻画的偏度大致为 -0.331，峰度大致为3.819，与状态 1 的混合正态分布具有显著的高阶矩差别。其余指数对数日收益率的分布也有相似的结论。

类似的，对沪深两市主要指数的对数周收益率也进行两种假设模型的拟合，并计算似然比统计量 LR，以考察去除周日历效应的收益率分布模型中是否存在马尔科夫状态转移结构。表 4 给出对数周收益率的实证结果，可以看到除深证 B 股指数，其余10 种指数的对数周收益率分布模型中都存在显著的马尔科夫状态转移结构。

表4 HMS-MND 模型参数估计和似然比检验结果(对数周收益率)

Table 4 HMS-MND Model Parameter Estimates and Likelihood Ratio Test Results for Stock Indexes′ Log Weekly Returns

	上证综合指数	上证 A 股指数	上证 B 股指数	上证 50 指数	上证 180 指数	深证成份指数	深证 A 股指数	深证 B 股指数	深证 100 指数	沪深 300 指数	上证红利指数
a_{11}	0.965	0.965	0.935	0.994	0.995	0.983	0.958	0.549	0.965	0.996	0.982
a_{12}	0.035	0.035	0.065	0.007	0.005	0.017	0.042	0.451	0.036	0.004	0.019
a_{21}	0.019	0.019	0.133	0.007	0.008	0.023	0.024	0.363	0.019	0.007	0.005
a_{22}	0.981	0.981	0.867	0.993	0.993	0.977	0.976	0.637	0.981	0.993	0.995
c_{11}	0.756	0.759	0.987	0.876	0.819	0.808	0.597	0.664	0.642	0.776	0.813
c_{12}	0.244	0.241	0.013	0.124	0.181	0.192	0.403	0.337	0.358	0.224	0.187
c_{21}	0.639	0.637	0.957	0.899	0.860	0.763	0.776	0.931	0.774	0.848	0.669
c_{22}	0.361	0.363	0.043	0.101	0.140	0.237	0.224	0.069	0.226	0.152	0.331
μ_{11}	-0.013	-0.013	0.002	-0.003	-0.001	-0.003	-0.011	-0.013	-0.007	-0.001	-0.010
μ_{12}	-0.007	-0.007	-0.079	0.023	0.004	-0.017	-0.015	-0.055	-0.016	0.004	-0.036
μ_{21}	0.004	0.004	-0.001	0.000	-0.002	0.017	0.010	0.022	0.010	-0.002	0.010
μ_{22}	0.020	0.020	0.069	0.047	0.046	0.021	0.020	0.102	0.018	0.046	0.005
σ_{11}	0.020	0.020	0.027	0.024	0.023	0.027	0.018	0.021	0.062	0.023	0.076
σ_{12}	0.068	0.068	0.006	0.049	0.046	0.072	0.066	0.040	0.009	0.047	0.016
σ_{21}	0.042	0.042	0.077	0.056	0.056	0.050	0.044	0.025	0.043	0.057	0.045
σ_{22}	0.017	0.017	0.002	0.010	0.017	0.012	0.008	0.017	0.012	0.019	0.022
LR	22.102***	18.520***	309.570***	223.872***	201.713***	113.654***	104.274***	-2.938	125.365***	185.934***	39.818***

表5　HMS-MND 模型的偏度和峰度刻画能力评估(对数日收益率)

Table 5　Evaluation of HMS-MND Model Using Skewness and Kurtosis Statistics(log daily returns)

	上证综合指数	上证A股指数	上证50指数	上证180指数	深证A股指数	深证100指数	沪深300指数	上证红利指数
真实偏度	-0.250	-0.250	-0.192	-0.284	-0.531	-0.374	-0.241	-0.503
HMS-MND 模型偏度	-0.390	-0.381	-0.217	-0.368	-0.599	-0.438	-0.349	-0.434
真实峰度	6.214	6.212	5.528	6.029	5.681	5.490	6.089	5.393
HMS-MND 模型峰度	5.408	5.336	5.161	5.507	5.594	5.289	5.639	5.433

表6　HMS-MND 模型的偏度和峰度刻画能力评估(对数周收益率)

Table 6　Evaluation of HMS-MND Model Using Skewness and Kurtosis Statistics(log weekly returns)

	上证综合指数	上证A股指数	上证B股指数	上证50指数	上证180指数	深证成份指数	深证A股指数	深证100指数	沪深300指数	上证红利指数
真实偏度	-0.018	-0.015	0.284	0.003	-0.082	-0.246	-0.295	-0.252	-0.097	-0.189
HMS-MND 模型偏度	-0.027	-0.027	-0.069	0.140	-0.052	-0.148	-0.214	-0.112	-0.071	-0.088
真实峰度	4.519	4.524	5.892	4.556	4.718	4.829	4.586	4.553	4.692	4.350
HMS-MND 模型峰度	5.219	5.208	4.633	4.232	4.208	4.972	4.835	4.278	4.264	4.101

为了考察HMS-MND模型刻画收益率序列尖峰、厚尾特征及非对称性的能力,对于实证表明收益率分布具有显著马尔科夫状态转移结构的每一种沪深两市股票指数(表3和表4中LR统计量标注***)均使用表3和表4中E-M算法估计所得的HMS-MND模型参数,生成400组与相应指数收益率数据等长的随机序列,并计算各组序列的偏度和峰度。表5和表6分别记录了使用表3和表4模型参数的400次随机试验的平均值,并给出使用真实对数收益率数据计算出的偏度和峰度,作为性能评估的参考。

比较表5和表6中的真实数据偏度、峰度与模型实现的偏度、峰度可以看出,HMS-MND模型对于收益率分布的非对称性及尖峰、厚尾特征均具有较强的刻画能力。此外,表5中8种股票指数的对数日收益率真实数据均表现出一定程度的左偏,除上证红利指数外,模型对其余7种指数收益率左偏程度的刻画均略强于真实数据。另一方面,模型对上证红利指数尖峰、厚尾特征的刻画稍强于真实数据,对其余7种指数峰度值的刻画略低于真实数据。本研究推断,HMS-MND模型在实现偏度和峰度上面的这种误差在风险管理等实际应用中具有一定的抵消性(峰度刻画不足导致的峰度风险可以由超额刻画的负偏度来弥补),本研究也将在后续的研究中结合实际应用对此进一步展开探讨。表6中的结果并未呈现表5中的上述特征,本研究认为,这与本研究的对数周收益率历史数据较短(<450)、HMS-MND模型参数估计以及真实数据统计结果很容易被若干大的奇异值影响有一定关系,下一步的研究中将考虑使用香港、美国等成熟市场的指数进行低频收益率的研究。

6　结论

本研究将马尔科夫状态转移结构引入中国股票指数对数收益率分布的刻画,建立隐马尔科夫状态转移-混合正态分布(HMS-MND)模型,并使用E-M算法和Baum-Welch算法给出模型的参数估计方法。使用2002年7月1日至2010年10月29日沪深两市11种主要股票指数对数日、周收益率数据的实证结果表明,大部分主要股票指数的对数收益率分布中都存在统计显著的马尔科夫状态转移结构,各隐状态的混合正态分布组成成分在均值和方差上有较大差异,具有不同的高阶矩特征,且HMS-MND模型可以较好地刻画对数收益率分布的尖峰、厚尾和非对称性。

本研究结果表明,中国股票指数收益率的分布存在较显著的时间异质性,并且该时间异质性可以通过引入马尔科夫状态转移结构较好刻画,该结论对于中国资本市场的风险管理以及沪深300股指期

货等金融衍生产品的定价等问题具有一定的指导意义。后续研究将把马尔科夫状态转移结构引入中国股票指数的波动率模型,以刻画股指波动的阶段性特征,实现对波动率的更准确预测,并在此基础上探索风险管理和金融资产估价方法的合理调整;此外,还将以 HMS-MND 模型为基础对中国个股的收益率分布模型进行研究,并进一步应用于投资组合的设计和优化。

参考文献:

[1] 封建强,王福新. 中国股市收益率分布函数研究[J]. 中国管理科学,2003,11(1):14-21.
Feng Jianqiang, Wang Fuxin. A research on return distribution function of Chinese stock-market [J]. Chinese Journal of Management Science, 2003, 11 (1):14-21. (in Chinese)

[2] 楼小飞,周林,施红俊. 中国股市投资组合收益分布的实证研究[J]. 同济大学学报:自然科学版,2005,33(3):414-417.
Lou Xiaofei, Zhou Lin, Shi Hongjun. Empirical study on return distribution of portfolio in China security market [J]. Journal of Tongji University: Natural Science, 2005,33(3):414-417. (in Chinese)

[3] 曹志广,王安兴,杨军敏. 股票收益率非正态性的蒙特卡罗模拟检验[J]. 财经研究,2005,31 (10):34-41,52.
Cao Zhiguang, Wang Anxing, Yang Junmin. The test of non-normal distribution of stock returns with Monte Carlo simulation and the explanation [J]. Journal of Finance and Economics, 2005,31(10):34 -41,52. (in Chinese)

[4] 徐绪松,侯成琪. 非正态稳定分布条件下的投资组合模型:均值-尺度参数模型[J]. 系统工程理论与实践,2006,26(9):1-9.
Xu Xusong, Hou Chengqi. A portfolio selection model conditional on non-normal stable distributions: Mean-scale parameter model [J]. Systems Engineering-theory & Practice, 2006,26(9):1-9. (in Chinese)

[5] 赵秀娟,张洪水,黎建强,汪寿阳. 一个基于非对称 Laplace 分布和 DEA 的证券投资基金评价方法[J]. 系统工程理论与实践,2007,27(10):1 -10.
Zhao Xiujuan, Zhang Hongshui, Li Jianqiang, Wang Shouyang. A method for evaluating mutual funds' performance based on asymmetric Laplace distribution and DEA approach [J]. Systems Engineering-theory & Practice, 2007,27(10):1-10. (in Chinese)

[6] 宋丽娟,杨虎. 基于 APARCH-Laplace 模型的 VaR 和 CVaR 方法[J]. 统计与决策,2008(14): 16-19.
Song Lijuan, Yang Hu. VaR and CVaR methods based on APARCH-Laplace model [J]. Statistics and Decision, 2008(14):16-19. (in Chinese)

[7] 郑玉华,崔晓东. 基于混合正态分布的风险价值度量[J]. 财经问题研究,2009(5):68-73.
Zheng Yuhua, Cui Xiaodong. VaR measurement based on mixture of normal distributions [J]. Research on Financial and Economic Issues, 2009(5): 68-73. (in Chinese)

[8] Cheung Y W, Erlandsson U G. Exchange rates and Markov switching dynamics [J]. Journal of Business and Economic Statistics, 2005,23(3):314-320.

[9] Dueker M, Neely C J. Can Markov switching models predict excess foreign exchange returns? [J]. Journal of Banking & Finance, 2007,31(2):279-296.

[10] Kim C J. Markov-switching models with endogenous explanatory variables [J]. Journal of Econometrics, 2004,122(1):127-136.

[11] Kim C J. Markov-switching models with endogenous explanatory variables Ⅱ: A two-step MLE procedure [J]. Journal of Econometrics, 2009,148(1):46-55.

[12] Lin S K, Wang S Y, Tsai P L. Application of hidden Markov switching moving average model in the stock markets: Theory and empirical evidence [J]. International Review of Economics and Finance, 2009, 18 (2):306-317.

[13] Erlwein C, Benth F E, Mamon R. HMM filtering and parameter estimation of an electricity spot price model [J]. Energy Economics, 2010, 32 (5): 1034 - 1043.

[14] Dias J G, Vermunt J K, Ramos S. Mixture hidden Markov models in finance research [C] // Fink A, Lausen B, Seidel W, Ultsch A. Advances in Data Analysis, Data Handling and Business Intelligence. Berlin: Springer, 2010:451-459.

[15] Geweke J, Amisano G. Hierarchical Markov normal mixture models with applications to financial asset returns [J]. Journal of Applied Econometrics, 2011,26 (1):1-29.

[16] 刘金全,郑挺国. 利率期限结构的马尔科夫区制转移模型与实证分析[J]. 经济研究,2006 (11):82-91.
Liu Jinquan, Zheng Tingguo. Markov regime switching model and empirical analysis of the term structure of interest rates [J]. Economic Research Journal, 2006 (11):82-91. (in Chinese)

[17] 吴吉林,陶旺升. 基于机制转换与随机波动的我国短期利率研究[J]. 中国管理科学,2009, 17(3):40-46.
Wu Jilin, Tao Wangsheng. Markov-regime switching and stochastic volatility model of short-term interest rate in China [J]. Chinese Journal of Management Science, 2009,17(3):40-46. (in Chinese)

[18] 高金余,陈翔. 马尔可夫切换模型及其在中国

股市中的应用[J]. 中国管理科学, 2007, 15(6):20-25.
Gao Jinyu, Chen Xiang. Markov switching model and its application to the stock market of China [J]. Chinese Journal of Management Science, 2007, 15(6):20-25. (in Chinese)

[19] 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.
Yan Taihua, Chen Mingyu. Analyze the SSE (Shanghai Stock Exchange) composite index week returns ratio with time series analysis method based on Markov switching model[J]. Chinese Journal of Management Science, 2009, 17(6):33-38. (in Chinese)

[20] Kon S J. Models of stock returns: A comparison[J]. Journal of Finance, 1984, 39(1):147-165.

[21] McLachlan G, Peel D. Finite mixture models [M]. New York: John Wiley & Sons, Inc., 2000:7-24.

[22] Lin B H, Yeh S K. On the distribution and conditional heteroscedasticity in Taiwan stock prices [J]. Journal of Multinational Financial Management, 2000, 10(3/4):367-395.

[23] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm [J]. Journal of the Royal Statistical Society. Series B, 1977, 39(1):1-38.

[24] Bilmes J A. A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models [R]. Berkeley CA: International Computer Science Institute, 1998:1-13.

[25] Welch L R. Hidden Markov models and the Baum-Welch algorithm[J]. IEEE Information Theory Society Newsletter, 2003, 53(4):1-10.

A Study of the Distribution of Chinese Stock Index Returns Based on Markov Switching Model

Qu Hui, Xiao Binqing
School of Management and Engineering, Nanjing University, Nanjing 210093, China

Abstract: The distribution of financial asset returns is critical for investment and risk management. Since the state changes in the underlying economic and financial environments could lead to structural changes in financial asset returns' distribution, we propose the hidden Markov switching – mixed normal distribution (HMS-MND) model for Chinese stock index log returns' distribution, where the underlying Markov switching structure captures the structural changes in distribution, and the returns' distribution at each state is modeled as mixed normal. Parameter estimation is realized using the E-M algorithm and the Baum-Welch algorithm. The log daily and weekly returns on 11 major stock indexes of Shanghai and Shenzhen stock exchanges from 2002. 7. 1 to 2010. 10. 29 are used for empirical tests. Parameter estimates as well as likelihood ratio test results show that, most stock indexes have significant Markov switching structures in their log returns' distributions, and that HMS-MND model could well describe the statistics of stock index log returns. Therefore, we argue that introducing Markov switching structure is a promising direction for stock index return related researches.

Keywords: return; Markov switching; mixed normal distribution; E-M algorithm; likelihood ratio test

Received Date: November 8^{th}, 2010　　**Accepted Date**: July 5^{th}, 2011

Funded Project: Supported by the National Natural Science Foundation of China(70932003, 70901037), the Fundamental Research Funds for the Central Universities(1107011810, 1118011804), the Natural Science Foundation of Jiangsu Province(BK2011561), the Research Funds of Jiangsu Eastern Coast and Yangtse River Area Development Institute(Y201002), the Funds of Humanity and Sociology Project, Ministry of Education of China(09YJCZH061) and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry

Biography: Dr. Qu Hui, a Jiangsu Nantong native(1981 –), graduated from Cornell University and is an associate professor in the School of Management and Engineering at Nanjing University. Her research interests include computational finance, etc. E-mail: linda59qu@nju.edu.cn □