

基于非对称 ARCH 模型的海西州地区 支气管炎发病研究

马亮亮, 田富鹏

西北民族大学计算机科学与信息工程学院, 兰州 730030

摘要 基于非对称 ARCH 模型的理论, 通过建立 TARCH、TARCH-M、EGARCH 和 EGARCH-M 模型研究青海省海西州地区支气管炎月发病率和平均气温之间的相互关系。研究通过 EViews 软件对海西州地区支气管炎发病病例和平均气温监测登记资料的统计分析, 并利用原数据建立 ARCH、TARCH-M、EGARCH 和 EGARCH-M 模型, 通过对模型中的残差序列进行独立性检验和预测, 确定所建立的 ARCH、TARCH-M、EGARCH 和 EGARCH-M 模型的合理性, 并从中选出了最优模型 TARCH(1, 1)-M。结果表明, 此模型可以很好地描述海西州地区支气管炎月发病率与平均气温的关系, 还可以较好地模拟支气管炎月发病率变化趋势, 预测未来的支气管炎发病趋势, 是一种预测精度较高的预测模型。该模型可为支气管炎研究提供有参考价值的数据。

关键词 非对称 ARCH 模型; ARCH 效应检验; 独立性检验; EViews 软件

中图分类号 O212.4

文献标识码 A

文章编号 1000-7857(2009)24-0051-05

Incidence of Bronchitis in Haixizhou Region Based on the Asymmetric ARCH Model

MA Liangliang, TIAN Fupeng

School of Computer and Information, Northwest University for
Nationalities, Lanzhou 730030, China

Abstract Based on the theory of the asymmetric ARCH model, the ARCH, TARCH-M, EGARCH and EGARCH-M models are established to study the relationship between the monthly incidence rate of bronchitis and the average temperature in Haixizhou Region. The EViews software is used for statistical analysis of the data of registered cases of bronchitis and the average temperature in Haixizhou Region, Qinghai Province, at the same time, the independence tests and forecasts for the residue series are carried out to validate the established models. The optimized model, TARCH(1, 1)-M model, is obtained. The results show that this model can be best used to describe the relationship between the monthly incidence rate of bronchitis and the average temperature, to simulate the variation trend of the monthly incidence rate and to predict the future incidence trend, with a high prediction precision.

Keywords asymmetric ARCH model; ARCH LM test;

independence test; EViews software

支气管炎是一种危害人们健康的发病率较高的疾病。由于人们生活水平的提高、工作压力的增大和不良的饮食习惯及其他原因, 近年来支气管炎发病率有增加的趋势。通过建立数学模型, 研究支气管炎与平均气温的相互关系, 有助于更好地认识和预防支气管炎, 有效地降低支气管炎对人们的危害, 从而保障人们的生活质量^[1-2]。

1 基础数据和统计学方法

1.1 病历资料来源

全部发病资料取自青海省海西州第一人民医院。经过核对、补漏, 从而保证资料的准确和完整。

1.2 统计学方法

运用 Excel 2003 及 EViews 3.1 对 2001 年 1 月到 2007 年 12 月青海省海西州地区支气管炎发病资料进行整理分析, 统计出了海西州地区支气管炎月发病率和月平均气温, 通过建立 TARCH、TARCH-M、EGARCH 和 EGARCH-M 模型, 对支气管炎的发病情况进行研究。

2 理论与模型介绍

2.1 非对称 ARCH 模型理论

2.1.1 TARCH 模型理论

对于通常的回归模型^[3]

收稿日期: 2009-09-24

基金项目: 国家自然科学基金项目 (60673192)

作者简介: 马亮亮, 研究方向为数学模型, 电子信箱: mll198684@126.com; 田富鹏 (通信作者), 教授, 研究方向为数学模型, 电子信箱: tfp@xbmu.edu.cn

$$y_i = x_i \beta + \varepsilon_i \quad (1)$$

若在式(1)右边增加一项 h_i (或用标准差 $[\sqrt{h_i}]$ 代替), 则该模型变为 ARCH-M (ARCH-in-mean) 模型, 其表达式为

$$y_i = x_i \beta + \gamma h_i + \varepsilon_i \quad (2)$$

如果 h_i 的结构为 $h_i = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{i-i}^2 = \alpha_0 + \alpha(B) \varepsilon_{i-i}^2$, 则该模型称为 ARCH-M(q) 模型^[4]。其中, q 为条件方差 h_i 的最大滞后阶数。

TARCH (Threshold ARCH) 模型具有如下形式的条件方差^[5]:

$$h_i = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{i-i}^2 + \varphi \varepsilon_{i-1}^2 d_{i-1} + \sum_{j=1}^p \theta_j h_{i-j} \quad (3)$$

其中, 参数 p 为随机扰动项平方 ε_i^2 的最大滞后阶数, q 为条件方差 h_i 的最大滞后阶数, d_i 是一个名义变量,

$$d_i = \begin{cases} 1 & \varepsilon_i < 0 \\ 0 & \text{其他} \end{cases} \quad (4)$$

由于引入 d_i , 增长信息 ($\varepsilon_i > 0$) 和下降信息 ($\varepsilon_i < 0$) 对条件方差的作用效果不同。增长时 $\varphi \varepsilon_{i-1}^2 d_{i-1} = 0$, 其影响可用系数 $\sum_{i=1}^q \alpha_i$ 代表, 下降时为 $\sum_{i=1}^q \alpha_i + \varphi$ 。若 $\varphi \neq 0$, 则说明信息作用是非对称的, 而当 $\varphi > 0$ 时, 认为存在杠杆 (leverage) 效应^[6]。

2.1.2 TARCH-M 模型的理论

基于 TARCH 模型理论, 如果 h_i 的结构为 $h_i = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{i-i}^2 + \varphi \varepsilon_{i-1}^2 d_{i-1} + \sum_{j=1}^p \theta_j h_{i-j}$, 则模型(2)称为 TARCH-M(p, q) 模型^[7]。其中, $p \geq 0, q > 0; \alpha_0 > 0, \alpha_i \geq 0, i = 1, 2, \dots, q; \theta_j \geq 0, j = 1, 2, \dots, p$ 。为保证 TARCH-M(p, q) 是宽平稳的, 存在参数约束条件 $\alpha(B) + \varphi(B) + \theta(B) < 1$ ^[8]。

2.1.3 EGARCH 模型的理论

EGARCH 模型即指数 (Exponential) GARCH 模型, 由 Nelson 在 1991 年提出^[9]。模型的条件方差表达式为

$$\ln(h_i) = \alpha_0 + \sum_{j=1}^p \theta_j \ln(h_{i-j}) + \sum_{i=1}^q \left(\alpha_i \left| \frac{\varepsilon_{i-i}}{\sqrt{h_{i-i}}} \right| + \varphi_i \frac{\varepsilon_{i-i}}{\sqrt{h_{i-i}}} \right) \quad (5)$$

模型中条件方差采用了自然对数形式, 意味着 h_i 非负且杠杆效应是指数型的^[10]。若 $\varphi \neq 0$, 说明信息作用非对称。当 $\varphi < 0$ 时, 杠杆效应显著^[11]。其中, $p \geq 0, q > 0; \alpha_0 > 0, \alpha_i \geq 0, i = 1, 2, \dots, q; \theta_j \geq 0, j = 1, 2, \dots, p$ 。为保证 EGARCH(p, q) 是宽平稳的, 存在参数约束条件 $\alpha(B) + \varphi(B) + \theta(B) < 1$ ^[12]。

2.1.4 EGARCH-M 模型的理论

基于 EGARCH 模型理论, 如果 h_i 的结构满足

$$\ln(h_i) = \alpha_0 + \sum_{j=1}^p \theta_j \ln(h_{i-j}) + \sum_{i=1}^q \left(\alpha_i \left| \frac{\varepsilon_{i-i}}{\sqrt{h_{i-i}}} \right| + \varphi_i \frac{\varepsilon_{i-i}}{\sqrt{h_{i-i}}} \right)$$

则模型(2)称为 EGARCH-M(p, q) 模型^[13]。其中, $p \geq 0, q > 0$;

$\alpha_0 > 0, \alpha_i \geq 0, i = 1, 2, \dots, q; \theta_j \geq 0, j = 1, 2, \dots, p$ 。为保证 EGARCH-M(p, q) 是宽平稳的, 存在参数约束条件 $\alpha(B) + \varphi(B) + \theta(B) < 1$ ^[14]。

3 非对称 ARCH 模型的应用

基于非对称 ARCH 模型思想, 利用 EViews 软件对支气管炎的月发病率和月平均气温进行分析, 建立了 TARCH、TARCH-M、EGARCH 和 EGARCH-M 模型, 并对所建立模型的残差序列进行独立性检验和预测, 从中选出最优模型。

3.1 ARCH 效应检验

设序列 x_t 和 y_t 分别为海西州地区 2001 年 1 月到 2007 年 12 月支气管炎月平均气温和月发病率。以月平均气温 x 为因变量, 为了考查变量间的动态影响, 故采用分布滞后模型 (通过反复试验, 选取了一个相对较好的模型形式), 其形式为

$$x_t = b_1 y_t + b_2 y_{t-1} + b_3 y_{t-2} + b_4 x_{t-1} + b_5 x_{t-3} + \varepsilon_t \quad (6)$$

其参数估计和相关检验结果如表 1 所示。

表 1 分布滞后模型估计和检验结果

Table 1 Results of distribution lag model forecast and test

变量	系数	均值标准 误差	t 统计量	概率
y	0.121 848	0.134 062	0.908 894	0.366 3
y(-1)	0.315 590	0.138 907	2.271 942	0.025 9
y(-2)	0.143 183	0.133 152	1.075 331	0.285 6
x(-1)	1.175 525	0.032 965	35.660 17	0.000 0
x(-3)	-0.513 28	0.0342 13	-15.021 48	0.000 0
决定系数	0.932 283	因变量平均值	5.522 469	
调整后的判定系数	0.928 719	因变量标准差	9.200 551	
回归后的标准误差	2.456 405	赤池信息量 (AIC)	4.695 015	
残差平方和	458.578 2	施瓦茨准则 (SC)	4.842 821	
似然估计	-185.148 1	DW 统计量	1.848 783	

从表 1 可知, 所有系数的显著性检验对于 90% 置信度均通过, 且该模型的拟合优度达 0.932 4, 整体效果较好。建立的分布滞后模型的表达式为

$$x_t = 0.1218 y_t + 0.3156 y_{t-1} + 0.1432 y_{t-2} + 1.1755 x_{t-1} - 0.5139 x_{t-3} + \varepsilon_t \quad (7)$$

以下对模型(7)中的残差序列做 ARCH 效应检验。当 χ^2 检验阶数 q 为 9 时, ARCH 效应的检验结果如表 2 所示。

表 2 ARCH 效应检验结果

Table 2 Results of ARCH LM test

统计量	统计量值	概率	概率值
F 统计量	1.201 473	相伴概率	0.310 452
LM 统计量	10.692 48	相伴概率	0.297 378

由表 2 可知,在 $q=9$ 时的相伴概率 $p=0.1597$,小于显著性水平 $\alpha=0.05$,且在该模型的残差序列做 ARCH 效应的 LM 检验时,发现 $q=9$ 时的相伴概率 $p=0.0730$,序列存在高阶 ARCH 效应,即存在 TARCH、TARCH-M、EGARCH 和 EGARCH-M 效应。

3.2 模型建立

分别对分布滞后模型(7)的残差序列建立 TARCH 模型、TARCH-M 模型、EGARCH 模型和 EGARCH-M 模型。根据 AIC 和 SC 准则,并配合残差独立性检验,比较发现,建立模型 TARCH(1,1)、TARCH(1,1)-M、EGARCH(1,1)和 EGARCH(1,1)-M 较为适宜。其条件方差的数学表达式分别为

$$h_t=10.1323-0.1256\varepsilon_{t-1}^2+0.0308\varepsilon_{t-1}^2d_{t-1}-0.6488h_{t-1} \tag{8}$$

$$h_t=0.9832-0.2480\varepsilon_{t-1}^2+0.4205\varepsilon_{t-1}^2d_{t-1}-0.7970h_{t-1} \tag{9}$$

其中,

$$d_t=\begin{cases} 1 & \varepsilon_t<0 \\ 0 & \text{其他} \end{cases} \tag{10}$$

$$\ln(h_t)=1.0937+0.1248\ln(h_{t-1})+0.3524\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right|+0.4898\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \tag{11}$$

$$\ln(h_t)=2.3369-0.5887\ln(h_{t-1})-0.0705\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right|+0.1823\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \tag{12}$$

其中,模型(9)和(12)均是在回归模型(1)中加入标准差 $\sqrt{h_t}$ 得到的结果。式(8)中,0.030 8 为杠杆效应系数 φ 估计值。由于 $\varphi=0.0308\neq 0$,说明信息作用是非对称的,且 $\varphi=0.0308>0$,因此存在杠杆效应。式(9)中,0.420 5 为杠杆效应系数 φ 的估计值。由于 $\varphi=0.4205\neq 0$,说明信息作用是非对称的,且 $\varphi=0.4205>0$,因此存在杠杆效应。式(11)中,0.489 8 为杠杆效应系数 φ_1 的估计值,0.352 4 为 α_1 估计值,0.142 8 为 θ_1 估计值。由于 $\varphi=0.4898\neq 0$,说明信息作用是非对称的,且 $\varphi=0.4898>0$,则杠杆效应不显著。式(12)中,0.182 3 为杠杆效应系数 φ_1 估计值,-0.070 5 为 α_1 估计值,-0.588 7 为 θ_1 估计值。由于 $\varphi=0.1823\neq 0$,说明信息作用是非对称的,且 $\varphi=0.1823>0$,则杠杆效应不显著。

3.3 模型检验

在模型 TARCH(1,1)、TARCH(1,1)-M、EGARCH(1,1)和 EGARCH(1,1)-M 中,参数 $\alpha_1+\varphi+\theta_1=-0.1256+0.0308-0.6488=-0.7436<1$, $\alpha_1+\varphi+\theta_1=-0.2480+0.4205-0.7970=-0.6245<1$, $\alpha_1+\varphi_1+\theta_1=0.3524+0.4898+0.1248=0.967<1$, $\alpha_1+\varphi_1+\theta_1=-0.0705+0.1823-0.5887=-0.4769<1$ 均满足参数约束条件,即满足平稳条件。下面分别对模型(8)、(9)、(11)和(12)中的残差序列进行独立性检验,以确定所建模型的合理性。模型 TARCH(1,1)、TARCH(1,1)-M、EGARCH(1,1)和 EGARCH(1,1)-M 的残差独立性检验结果分别如表 3~表 6 所示。

由表 3~表 6 可知,模型(8)、(9)、(11)和(12)的 SIC 和 SC

表 3 TARCH(1,1)模型残差独立性检验结果
 Table 3 Results of TARCH(1, 1) model residual independence test

检验名称	检验值	检验名称	检验值
决定系数	0.932 169	因变量平均值	5.522 469
调整后的判定系数	0.924 633	因变量标准差	9.200 551
回归后的标准误差	2.525 838	AIC	4.769 491
残差平方和	459.349 9	SC	5.035 541
似然估计	-184.164 4	DW 统计量	1.846 564

表 4 TARCH(1,1)-M 模型残差独立性检验结果
 Table 4 Results of TARCH(1, 1)-M model residual independence test

检验名称	检验值	检验名称	检验值
决定系数	0.953 511	因变量平均值	5.522 469
调整后的判定系数	0.947 618	因变量标准差	9.200 551
回归后的标准误差	2.105 74	AIC	4.388 457
残差平方和	314.824	SC	4.684 068
似然估计	-167.732 5	DW 统计量	1.858 889

表 5 EGARCH(1,1)模型残差独立性检验结果
 Table 5 Results of EGARCH(1, 1) model residual independence test

检验名称	检验值	检验名称	检验值
决定系数	0.930 485	因变量平均值	5.522 469
调整后的判定系数	0.922 762	因变量标准差	9.200 551
回归后的标准误差	2.557	AIC	4.747 451
残差平方和	470.753 9	SC	5.013 5
似然估计	-183.271 7	DW 统计量	1.868 977

表 6 EGARCH(1,1)-M 模型残差独立性检验结果
 Table 6 Results of EGARCH(1, 1)-M model residual independence test

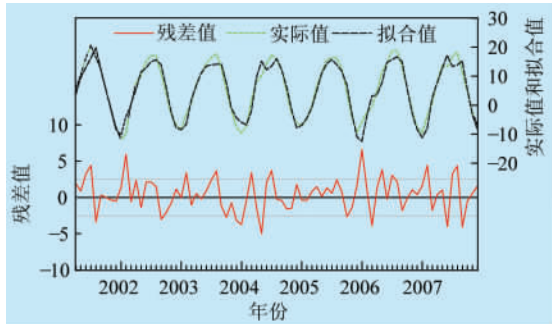
检验名称	检验值	检验名称	检验值
决定系数	0.948 997	因变量平均值	5.522 469
调整后的判定系数	0.942 532	因变量标准差	9.200551
回归后的标准误差	2.205 605	AIC	4.527 391
残差平方和	345.393 2	SC	4.823 002
似然估计	-173.359 3	DW 统计量	2.203 212

值都较小,可以认为模型较好地拟合了数据。其中,模型 TARCH(1,1)-M(式(9))的 AIC 和 SC 值(4.388 5,4.684 1)最小,因此与模型 TARCH(1,1)、EGARCH(1,1)和 EGARCH(1,1)-M 相比较,模型 TARCH(1,1)-M 的拟合度更高,即模型 TARCH(1,1)-M 更优^[15]。

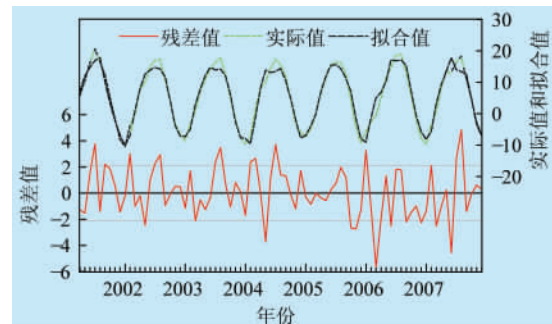
4 预测

对模型(8)、(9)、(11)和(12)中的残差序列进行拟合的结果如图1所示。可以看出,4个模型的残差序列拟合结果

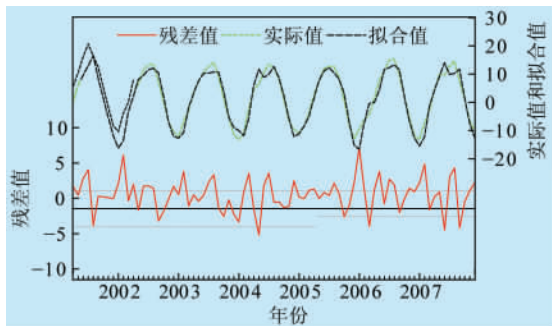
整体上均不错,但 TARCH(1,1)-M 模型残差序列的拟合效果最好。因此,基于原始数据建立的 TARCH(1,1)-M 模型是最优的^[16]。



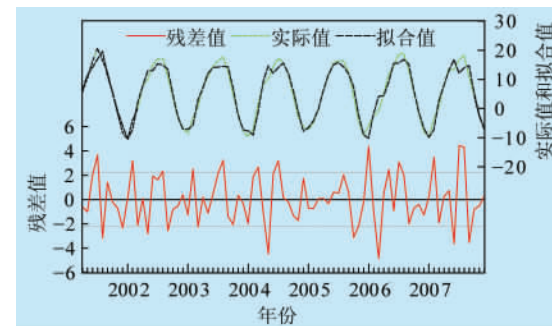
(a) TARCH(1,1)



(b) TARCH(1,1)-M



(c) EGARCH(1,1)



(d) EGARCH(1,1)-M

图1 4个模型的实际拟合残差序列结果

Fig. 1 Results of four models for actual fitted residuals

5 结论

由以上分析可知,在序列 x_t 与 y_t (支气管炎月平均气温和月发病率)之间可以建立分布滞后模型,且通过对建立的分布滞后模型中的残差序列进行 ARCH 效应检验,发现残差序列存在 TARCH、TARCH-M、EGARCH 和 EGARCH-M 效应,因此可对分布滞后模型的残差项建立 TARCH、TARCH-M、EGARCH 和 EGARCH-M 模型。最后,通过对建立的模型中的残差序列进行独立性检验和预测,确定了 TARCH(1,1)、TARCH(1,1)-M、EGARCH(1,1)和 EGARCH(1,1)-M 模型的合理性,并且从其中选出了最优模型 TARCH(1,1)-M^[17]。

参考文献 (References)

- [1] 朱新玲, 黎鹏. 汇率波动非对称效应的计量检验——基于非对称 ARCH 类模型[J]. 统计教育, 2008, 7: 31-33.
Zhu Xinling, Li Peng. *Statistical Thinktank*, 2008, 7: 31-33.
- [2] 刘毅, 陈佳, 吴润衡. 基于 TARCH 模型的 VaR 方法对上海股市的分析[J]. 北方工业大学学报, 2007, 19(1): 66-69.
Liu Yi, Chen Jia, Wu Runheng. *Journal of North China University of Technology*, 2007, 19(1): 66-69.
- [3] 王雅云. 上证 A 股与 B 股指数的波动性差异研究——基于 TARCH 模型的实证分析[J]. 太原城市职业技术学院学报, 2006, 5: 67-68.
Wang Yayun. *Journal of Taiyuan Urban Vocational College*, 2006, 5: 67-68.

- [4] 张广玉, 丁俊君. TARCH-M 模型在测度上海证券市场风险中的应用[J]. 中南财经政法大学学报, 2004, 5: 114-116.
Zhang Guangyu, Ding Junjun. *Journal of Zhongnan University of Economics and Law*, 2004, 5: 114-116.
- [5] 张雪蓉, 徐全智, 杨晋浩. TARCH-M 模型在上证指数波动率的实证分析[J]. 成都大学学报: 自然科学版, 2006, 3(25): 171-174.
Zhang Xuerong, Xu Quanzhi, Yang Jinhao. *Journal of Chengdu University: Natural Science Edition*, 2006, 3(25): 171-174.
- [6] 李俊强, 杜安. 利率上升对股票市场的影响——基于 TARCH 模型的实证研究[J]. 东方企业文化, 2007, 2: 92-95.
Li Junqiang, Du An. *Oriental Enterprise Culture*, 2007, 2: 92-95.
- [7] 王新宇, 宋学锋. 间接 TARCH-CAViaR 模型及其 MCMC 参数估计与应用[J]. 系统工程理论与实践, 2008, 46(6): 46-51.
Wang Xinyu, Song Xuefeng. *Systems Engineering-Theory & Practice*, 2008, 46(6): 46-51.
- [8] 张璇, 蔡梅. 中国股市季节效应的 EGARCH-M 研究[J]. 武汉理工大学学报, 2004, 26(7): 94-96.
Zhang Xuan, Cai Mei. *Journal of Wuhan University of Technology*, 2004, 26(7): 94-96.
- [9] 王淑良. 具有 EGARCH-SGED 误差项的时序的单位根检验[J]. 山东理工大学学报: 自然科学版, 2007, 21(2): 34-36.
Wang Shuliang. *Journal of Shandong University of Technology: Natural Science Edition*, 2007, 21(2): 34-36.
- [10] 靳庭良. 误差项的 EGARCH 效应对 DF 单位根检验可靠性的影响[J]. 河北师范大学学报: 自然科学版, 2006, 30(4): 391-395.

- Jin Tingliang. *Journal of Hebei Normal University: Natural Science Edition*, 2006, 30(4): 391-395.
- [11] 金秀, 许宏宇. 基于 EGARCH-VaR 的半参数法及实证研究 [J]. 东北大学学报: 自然科学版, 2007, 28(1): 141-144.
- Jin Xiu, Xu Hongyu. *Journal of Northeastern University: Natural Science Edition*, 2007, 28(1): 141-144.
- [12] 何宜庆, 曹慧红, 侯建荣. 我国沪深两市股指收益率的 EGARCH 效应分析[J]. 统计与决策, 2005, 8: 90-91.
- He Yiqing, Cao Huihong, Hou Jianrong. *Statistics and Decision*, 2005, 8: 90-91.
- [13] 秦伟良, 王颖, 姚如一. 基于 EGARCH-Copula 的金融市场条件相依分析[J]. 统计与决策, 2008, 15: 112-114.
- Qin Weiliang, Wang Ying, Yao Ruyi. *Statistics and Decision*, 2008, 15: 112-114.
- [14] 倪青山. 基于 EGARCH 模型上证指数收益率波动研究[J]. 金融经济,

2008, 15: 78-79.

Ni Qingshan. *Finance & Economy*, 2008, 15: 78-79.

- [15] 马亮亮, 田富鹏. 基于肺水肿相关因素的因子分析[J]. 河北北方学院学报: 自然科学版, 2009, 25(4): 53-55.

Ma Liangliang, Tian Fupeng. *Journal of Hebei North University: Natural Science Edition*, 2009, 25(4): 53-55.

- [16] 马亮亮, 田富鹏. 基于脑梗塞相关因素的逐步回归分析[J]. 南阳理工学院学报: 自然科学版, 2009, 1(3): 114-116.

Ma Liangliang, Tian Fupeng. *Journal of Nanyang Engineering College: Natural Science Edition*, 2009, 1(3): 114-116.

- [17] 易丹辉. 数据分析与 EViews 应用 [M]. 北京: 中国统计出版社, 2005, 193-199.

Yi Danhui. *Data analysis and application of EViews* [M]. Beijing: China Statistics Press, 2005, 193-199.

(责任编辑 杨书卷)

第31届国际影像科学大会(ICIS2010)

The 31st International Congress on Imaging Science

时间: 2010年5月12-16日

地点: 北京市

主办单位: 中国感光学会(CSIST)
国际影像科学委员会(ICIS)

联系电话: 010-82543686, 82543687
传真: 010-82543687
电子信箱: icis2010@csist.org.cn
通信地址: 北京市海淀区中关村北一条2号 (100190)
会议网站: <http://www.csist.org.cn/chinese/tz-icis2010.htm>