

ARCH 模型的诊断分析^①

柯珂, 张世英

(天津大学管理学院, 天津 300072)

摘要:探讨了 ARCH 模型的诊断分析和变结构建模问题, 所提出的分整增广 GARCH-M 模型包括了几乎所有现有的 ARCH 模型, 基于分整增广 GARCH-M 模型克服了 ARCH 模型在模型设定检验, 长记忆检验和参数估计等方面的障碍, 利用分段建模方法来检测模型结构变化点和分段变化模型的选择. 最后, 以上海证券综合事业股票指数为数据, 验证了诊断分析方法的有效性.

关键词: ARCH 模型; 诊断分析; 变结构建模

中图分类号: F224.0

文献标识码: A

文章编号: 1007-9807(2001)02-0012-07

0 引言

在分析和预测时, 参数向量在所有拟合样本期和预测期是否结构相同是一个非常重要的前提. 在我国, 由于经济处于不断调整和转轨之中, 经济规律不断变化, 不同时期采用同一模型似乎是难以令人信服的. 近年来, 非线性模型变结构问题已经引起人们注意, 人们提出了一些诊断变结构点的检验方法^[10,11].

文[6]提出的 ARCH 模型及由此衍生出来的 ARCH 模型族是重要的度量金融波动性的模型, 它随经济波动呈现其模型结构的变化. 然而就目前国内外文献而言, 未见有对 ARCH 模型进行变结构研究的报道. 本文对 ARCH 模型的诊断与变结构分析问题作一探讨.

1 诊断问题与判断准则

1.1 异常点的种类

所谓异常点是指严重偏离既定模型的数据点. 按文[7]的说法, 异常点根据其性质和产生的机理, 可以分成两大类: 加性异常点 (additive out-

liers) 和革新异常点 (innovation outliers).

(1) 加性异常点 AO (additive outlier)

这是一种孤立的异常点, 往往是由某一时刻的观察或记录误差所造成的, 它们的出现并不波及邻近的观测值. 就时间序列而言, 此类异常点是非本质的, 因为它们不涉及到时间序列内在的相关结构.

(2) 革新异常点 IO (innovation outliers)

这是一种成片出现的异常点, 在某段时间内一个异常点通过随机序列的相关结构, 波及到与它邻近的一批数据点 (通常是它的后继临近点), 从而使这些临近点表现出一定的异常. 与 AO 相比, 此类异常点是更为内在或本质的一类异常点, 它产生的机理也更为复杂.

一般而言, AO 对时间序列滞后相关系数的影响是使其偏向于零; 而 IO 对滞后相关系数的影响是使其偏向于 1^[9]. 由此可见, 不同类型的异常点可能对估计量产生不同的影响, 因此区别异常点的类型很重要, 而且由于个别异常点的存在可能对模型分析的结果产生举足轻重的作用, 因此在时间序列分析中, 探察异常点并区别其类型是非常有意义的工作.

① 收稿日期: 2000-05-17; 修订日期: 2000-09-11.
基金项目: 教育部博士点专项科研基金资助项目 (2000005606).
作者简介: 柯珂 (1974-), 女, 硕士.

1.2 影响点的概念

与异常点含义接近的还有强影响点. 顾名思义, 强影响点是那些对模型的建模、应用和预测有重要影响的数据点. 显然强影响点与我们所研究的变结构点含义很接近.

虽然数据集中所有点都对模型的建立有影响, 但各个数据点对统计推断的影响大小是不相等的. 一般而言, 在分析影响大小时, 有几个基本问题要考虑.

(1)“影响的指向”, 即分析对哪一个统计量的影响. 如是对模型参数估计值的影响, 还是对拟合优度统计量的影响等, 分析的目标不同, 所考虑的影响也必然有所不同;

(2)“度量影响的尺度”, 为了定量地刻画影响的大小, 迄今已提出多种尺度. 如, 基于残差的尺度、基于拟合的尺度、基于影响函数的尺度、基于置信域的尺度、基于似然函数的尺度等. 在每一种类型中都可能有不同的统计量.

1.3 变结构点

变结构点是一类特殊的强影响点. 一般而言, 变结构点是指这样的一类点: 它们对模型的结构有如此重大的影响, 在该点之前与该点之后模型的结构有显著区别. 这里模型的结构不仅包括模型的函数形式, 还包括模型中各参数. 模型结构的变化可能是模型形式的根本改变, 以 ARCH 模型族为例, 可能是由 LGARCH 变成 EGARCH 模型; 可能是模型参数的显著改变; 还可能是变量的增减等.

变结构点与强影响点的区别在于: 如果没有强影响点, 模型的结构可能会因此发生重大变化, 但对强影响点的前后序列建立的模型可能不会有显著区别.

强影响点和异常点之间有联系也有区别. 强影响点可能是异常点, 也可能是正常点; 反之, 异常点中的革新异常点可能也是强影响点, 也可能不是, 而加性异常点则肯定不是强影响点. 以线性回归直线 $y = ax + b$ 为例, 在图 1 中, A 为强影响点, 但它与回归直线比较接近, 因此不是异常点; 而 B 点既为强影响点同时又是异常点, 如果没有 B 点, 回归直线将向 x 轴显著偏移, 说明 B 是强影响点, 但同时 B 点离回归直线较远, 因此是异常点; 而 C 点是异常点, 但不是强影响点, 因为即使

没有 C 点, 回归直线也不会有大的变化.

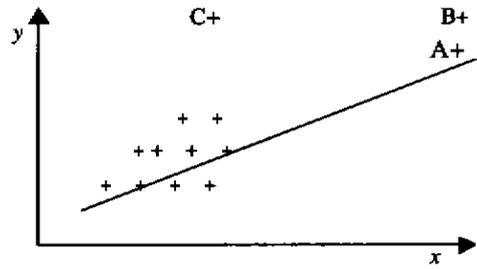


图 1 异常点与强影响点示意图

1.4 判断准则

给定 q 维空间 \mathcal{R} , 数据集 D 中的点 $z_i (i = 1, 2, \dots, N)$, 其中 N 是数据集的容量, 由 q 维向量 $(x_{i1}, x_{i2}, \dots, x_{iq})$ 定义. 不失一般性, 将 $(x_{i1}, x_{i2}, \dots, x_{iq-1})$ 设为自变量坐标 z_x , z_y 为因变量坐标. 基于数据集 D 的拟合函数设为 F , 而拟合值为 $F(\cdot)$, $\rho(z_i)$ 表示数据集 D 中距 z_i 最近的点的空间距离^①, η_ρ 为给定置信水平下该距离的临界值, η_F 是给定置信水平下拟合函数值的临界范围, 则有以下两条结论:

i. 当 $|\rho(z_i)| > \eta_\rho$ 且 $|z_{iy} - F(z_i)| \geq \eta_F$ 时, z_i 是强影响点, 同时又是异常点;

ii. 当 $|\rho(z_i)| \leq \eta_\rho$ 且 $|z_{iy} - F(z_i)| \geq \eta_F$ 时, z_i 是异常点但不是强影响点.

换种说法, z_i 是异常点, 当且仅当 $|z_{iy} - F(z_i)| \geq \eta_F$ 时; 若同时又有 $|\rho(z_i)| \geq \eta_\rho$, 则 z_i 是强影响点.

从定义出发, 这个判断准则的含义很清楚. 首先考虑到异常点的定义, 若点 z_i 的拟合函数值 $F(z_i)$ 远离拟合曲线, 则这一点是异常点, 如图 1 中的点 B 和 C; 若 z_i 是一个异常点同时又远离数据集 D , 模型建立时会因考虑到要把这一点包容进去而做一些改变. 这样, 这一点将因其特殊性而对建模产生较大的作用. 因此成为强影响点, 如图 1 中的点 B, 既是异常点又是强影响点.

当然象 A 点 (记做 z_a) 这样远离数据集的点既可能是强影响点, 也可能不是. 但却是一个值得怀疑的点, 因为理论上说, 远离数据集, 建立在数据集 D 上的模型就会设法把该点包括进来, 但如

① 这个“距离”有不同的计算方法, 如 Cook 距离^[3-5]、Welsch-Kuh 距离^[1]、Welsch 距离^[1]和修正的 Cook 距离^[4,5]等, 可根据具体问题的需要加以确定.

2.2 分段建模的思想

为了研究数据集 D 与模型符合的情况, 一个很重要的方法就是逐个考虑各数据点的作用. 具体来说, 就是逐个考虑每个数据点对于建模、模型分析和预测的影响. 传统的探察异常点的模型有两种: 第一, 数据删除模型, 即删除某个数据点如第 i 个, 研究这个点删除前后对于模型确立以及检验统计量是否有举足轻重的影响; 第二, 均值漂移模型, 即在第 i 个数据点上增加一个扰动, 这相当于因变量的均值有所漂移, 研究这个扰动对于估计量及其它统计量是否有显著影响. 文[9]证明这两种模型实际上是等价的.

实际上, 在诊断分析中, 除了希望找到异常点外, 更希望获取进一步的信息. 比如, 这个异常点的性质: 该点是加性异常点还是革新异常点; 是否是强影响点, 更甚或是变结构点. 所以借鉴数据删除的诊断思想, 我们提出用分段建模方法来探察异常点. 是因为一方面, 可以找到异常点, 并识别出该点是不是强影响点或变结构点; 另一方面, 分整增广 GARCH-M 模型概括了几乎所有现有的 ARCH 模型, 从而使分段建模方法, 成为很自然的选择.

2.3 分段建模的程序

根据异常点的定义, 首先寻找各点中偏离模型输出最大的数据点 (即满足 $\max |y_i - \hat{y}_i|$ 的数据点), 以该点为基点, 前后数据段分别建模, 考察各分段模型之间以及各自与基于数据集 D 的模型的差异是否显著.

不失一般性, 设最偏离基于数据集 D 的拟合模型的数据点为第 m 个点, 该数据点前后的数据集分别为 $D_{m_1} = \{z_i | i = 1, \dots, m-1\}$, $D_{m_2} = \{z_i | i = m+1, \dots, N\}$, 对 D_{m_1} , D_{m_2} 和 D 分别建模, 得到模型 M_{m_1} , M_{m_2} 和 M_0 . 考察模型的具体形式和模型参数等指标, 若有显著差异, 说明模型的结构发生显著改变, 即第 m 个点是变结构点; 若没有显著差异, 则要具体分析. 若能给定 η_F , 则问题变得很简单, 由判断准则, 可以判断第 m 个点是不是异常点. 若第 m 个点不是异常点, 则可以认为数据集 D 中不存在异常点, 因为根据异常点的含义, 若连偏离数据集最远的点都不是异常点, 其他点也不可能是异常点. 若第 m 个点是异常点, 就要考察这个点是不是强影响点, 毕竟这是一个值

得怀疑的点. 若该点远离数据集的其它点, 由判断准则, 问题也能迎刃而解, 这必是一个强影响点; 否则对删除数据点 z_m 的数据集 $D_1 = \{z_i | i = 1, 2, \dots, m-1, m+1, \dots, N\}$ 建立模型 M_1 , 考察模型 M_0 和 M_1 , 若 M_0 和 M_1 没有显著差异, 由强影响点的定义可以断定点 z_m 不是强影响点; 如果 M_0 和 M_1 有显著差异, 同样由定义得知, 点 z_m 是强影响点.

在实际时间序列分析之前, 有几个基本问题需要解决:

- i. η_F 和 η_p 的取值问题;
- ii. 如何判断模型结构差异是否显著.

把这两个问题结合起来解决, 设基于原数据集 D 建立的模型为 M_0 , 基于被怀疑点 z_m 被删除后的数据集 D_1 建立的模型为 M_1 , 从两个方面考察这两次建模. 一方面, 模型损失的拟合优度有多大, 这将决定第 m 个点是不是异常点; 另一方面, 这两个模型是否有显著差异, 这将用于判断第 m 个点是不是强影响点.

原假设 H_0 : z_m 不是异常点, 即 $|y_m - F(z_m)| < \eta_F$; 备择假设 H_1 : 第 m 个点是异常点, 即 $|y_m - F(z_m)| \geq \eta_F$. 模仿置信区间的设定, 取一个充分小的数 τ (比如 τ 可以等于 0.1 或 0.01). 设模型 M_0 的变异系数为 γ_0^D , M_1 的变异系数为 γ_1 , 只要 $|\gamma_0 - \gamma_1|/\gamma_0 \leq \tau$, 就可以接受原假设, 即认为点 z_m 不是异常点; 否则认为点 z_m 是异常点.

同理, 如果点 z_m 是异常点, 要验证点 z_m 是不是强影响点, 可应用判断准则. 取原假设 H_0 : z_m 不是强影响点, 即 $\rho(z_m) \leq \eta_p$; 备择假设 H_1 : z_m 是强影响点, 即 $\rho(z_m) > \eta_p$. ARCH 模型的差异可能有两种体现: 模型形式是否有变化, 即删除 z_m 后, 模型是否变成另外 31 种模型之一; 或者模型形式未变, 但参数有显著变化.

分段建模方法应用于对被怀疑点前后数据的建模, 可以由此判断该点是不是变结构点.

探测变结构点的可疑点是变结构分析的第一步, 使用的是图示法. 对数据集 D 建模并绘出图形, 观察方差的描点图, 看是否有较突兀的点. 因

① γ 是表达模型偏离实际数据程度的量度, 定义为 $\gamma = \frac{\sum_{i=1}^N (z_i - \hat{z}_i)^2}{x}$.

为二阶矩上的异常点很有可能就是一阶矩的强影响点或变结构点,这个点就是变结构点的可疑点.

3 实证研究

3.1 时间序列样本数据的选取

选取上海证券交易所综合股票指数的日收盘价格为样本数据,在此基础上实现 ARCH 模型变结构研究,并进行了分析,以检验所提方法的实用性.

样本的取值区间是从 1999 年 1 月 2 日到 1999 年 11 月 1 日,共有 214 个样本数据,即该时间序列的样本长度为 214. 给这组数据起名为 CI . 由于数据量庞大,不便在文中罗列原始数据.

3.2 实证研究

按以下步骤为数据建模.

(1) 读入数据及数据预处理

CI 数据是价格序列,设为 $\{P_t\}$, 包含各种成本,并非投资者关心的收益. 而所谓收益是今天价格与昨日价格的比价,所以先通过式(10)将价格转化为收益率

$$x_t = 100 \ln(P_t/P_{t-1}) \quad (10)$$

式中 100 只是常数,起放大数据的作用.

(2) 建模过程

对时间序列 $\{x_t\}$ 依次进行异方差检验、ARCH 检验、分整增广 GARCH-M 模型的参数估计、分整增广 GARCH-M 的近似参数检验、模型检验和诊断分析.

首先,对这 214 个数据(即数据集 D_0) 进行参数估计和检验,得到下列估计结果. 设此时的模型为 M_0 , 变异系数为 γ_0 .

$$x_t = 0.1878 + x_{t-1}(0.5969 + 0.4040 \exp(-\sigma_t^2/\mu^2)) + 0.2199\sigma_t^2 + \varepsilon_t \quad (11)$$

$$\sigma_t^2 = 0.5044 + 0.5260\varepsilon_{t-1}^2 + 0.5618\sigma_{t-1}^2 \quad (12)$$

$$\gamma_0 = 2.8881 \quad (\text{LGARCH-M 模型})$$

图 2 是基于 D_0 的方差的散点图. 从中可以找到一个非常突兀的点,其对应的横坐标是 114.

第 2 步,探索异常点;按 $|x_t - \hat{x}_t|$ 的大小排序 (\hat{x}_t 是分整增广 GARCH-M 模型在第 t 个数据点

的估计值). 发现距离拟合模型最远的点是第 114 个点,即 $m = 114$. 这是一个最可疑的异常点. 对这个点进行进一步的分析;令 $D_1 = \{x_t | t \neq 114\}$, 基于 D_1 建模 M_1 如式(13), 变异系数是 γ_1

$$x_t = 3.4952 + x_{t-1}(0.1898 + 0.8044 \exp(-\sigma_t^2/\mu^2)) + 0.2898\sigma_t^2 + \varepsilon_t \quad (13)$$

$$\ln(\sigma_t^2) = 0.7428 + 0.0720 \ln(\varepsilon_{t-1}^2) + 0.5831 \ln(\sigma_{t-1}^2) \quad (14)$$

$$\gamma_1 = 2.9157 \quad (\text{MGARCH-M 模型})$$

$|\gamma_1 - \gamma_0| = 0.0276 < 0.02881 = 1\% \gamma_0$, 这说明 114 点的加入并没有使模型刻画实际数据的能力下降. 也就是说这一点是一个正常点, 因为 114 点是最偏离模型 M_0 的, 它都不是异常点, 其它点更不可能是异常点, 所以数据集 D_0 中不存在异常点.

但是显然式(11)、(12)和式(13)、(14)是不同的两个模型. 这就表明点 114 虽然不是异常点, 却是一个强影响点, 它的存在对建模有重大作用. 进一步地, 考察一下点 114 对模型结构的影响. 应用分段建模, 设 $D_{m_1} = \{x_t | 1 \leq t \leq 113\}$, $D_{m_2} = \{x_t | 115 \leq t \leq 214\}$, 对 D_{m_1} 和 D_{m_2} 分别建模, 变异系数分别为 γ_{m_1} 和 γ_{m_2} 得式(15)~式(18).

基于 D_{m_1} 建模得到

$$x_t = 0.5920 + x_{t-1}(0.3287 + 0.4248 \exp(-\sigma_t^2/\mu^2)) + 0.3287\sigma_t^2 + \varepsilon_t \quad (15)$$

$$\sigma_t = 0.5351 + 0.1090\sigma_{t-1} + 0.5618|\varepsilon_{t-1}| + 0.7114 \max(0, -\varepsilon_t) \quad (16)$$

$$\gamma_{m_1} = 1.3142 \quad (\text{TGARCH-M})$$

基于 D_{m_2} 建模得到

$$x_t = 0.3243 + x_{t-1}(0.3653 + 0.6308 \exp(-\sigma_t^2/\mu^2)) + 0.7582\sigma_t^2 + \varepsilon_t \quad (17)$$

$$\ln(\sigma_t^2) = 0.7322 + 0.6968 \ln(\sigma_{t-1}^2) + 0.8130(|\varepsilon_{t-1}| + 0.9768\varepsilon_{t-1}) \quad (18)$$

$$\gamma_{m_2} = 1.4854 \quad (\text{EGARCH-M})$$

这两个模型在模型形式上根本不同, 说明点 114 是一个变结构点.

第 3 步, 探索其它强影响点和变结构点

为了探索其它强影响点, 需要先把点 114 屏

蔽掉,因为可能会由于它的存在,其它强影响点被“掩盖”(masking)了。反复运行强影响点和变结构点诊断计算。我们发现的第二个强影响点是点 113。事实上,在第二步中,点 113 并不突出,正是因为它的作用被近邻点 114 对时间序列的强影响力所掩盖。

对点 113 做分段建模,得到

$$\begin{aligned} \text{对于 } D_{m1}' = \{x_t | 1 \leq t \leq 112\} \text{ 建模} \\ x_t = 0.6973 + x_{t-1}(0.2007 + \\ 0.8007 \exp(-\sigma_t^2/\mu^2)) + \\ 0.7704 \sigma_t^2 + \varepsilon_t \end{aligned} \quad (19)$$

$$\begin{aligned} \sigma_t = 0.6619 - 1.6546 \sigma_{t-1} + \\ 0.9970 |\varepsilon_{t-1}| + \\ 0.6224 \max(0, -\varepsilon_t) \end{aligned} \quad (20)$$

$$\gamma_{m1} = 1.2491 \quad (\text{TGARCH-M})$$

对 $D_{m1}' = \{x_t | 115 \leq t \leq 214\}$ 建模,得

$$\begin{aligned} x_t = 0.8669 + x_{t-1}(0.5604 + \\ 0.4347 \exp(-\sigma_t^2/\mu^2)) + \\ 0.3471 \sigma_t^2 + \varepsilon_t \end{aligned} \quad (21)$$

$$\sigma_t^2 = 0.4776 + 0.5349 \varepsilon_{t-1}^2 + 0.2668 \sigma_{t-1}^2 +$$

$$0.4340 S_{t-1}^* \varepsilon_{t-1}^2 \quad (22)$$

$$\text{其中 } S_{t-1}^* = \begin{cases} 1 & \varepsilon_{t-1} \leq 0 \\ 0 & \varepsilon_{t-1} > 0 \end{cases}$$

$$\gamma_{m2} = 1.5117 (\text{GJR-GARCH-M})$$

除了成片强影响点,为了找到其它强影响点和变结构点,使用图示法。图 2 中的突兀点为进一步的探测提供了方向。经识别,除了已探测出的点 113 和 114 外,没有其它强影响点。

第四步,结果分析和经济原因探究

当把变结构点 113 和 114 对应的时间点列出时,点 113 是 1999 年 6 月 28 日,而 114 为 1999 年 6 月 29 日。这两个日子是香港回归两周年前夕,更是经历 5.19 股市狂飙后,中央采取相应措施的日子。与此形成对照的是 5.19 附近并没有变结构点;这一方面说明分整增广 GARCH-M 模型的鲁棒性很强,把系统内的各种变化因子都考虑进去了,如可能的投资者心理预期和股市的“信息潜流”等的影响;另一方面,说明变结构的根源主要来自系统外,具体到本案例,则是中央政府的干预和宏观调控。

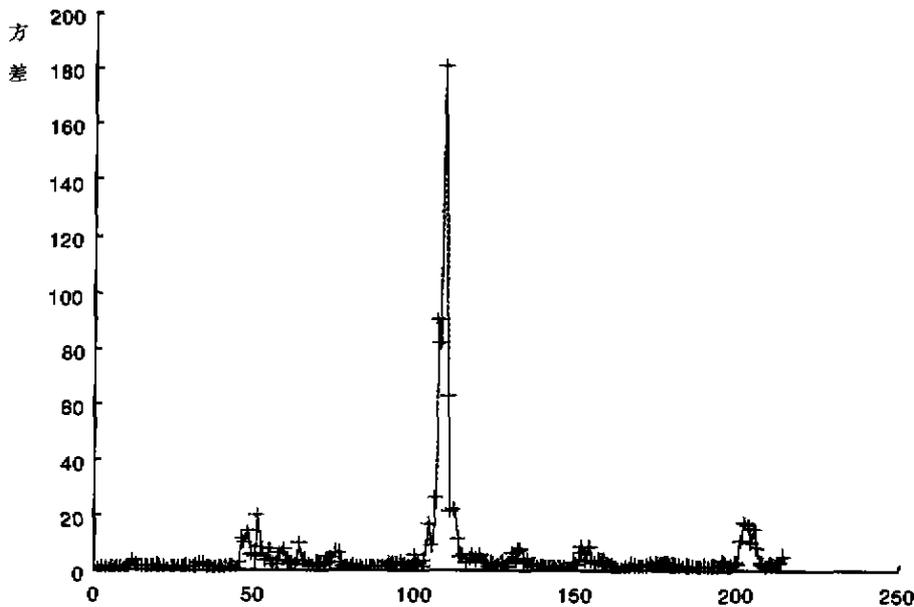


图 2 基于 D_0 的方差散点图

参 考 文 献:

- [1] Belsley C A, Kuh E, Welsch T E. Regression diagnostics[M]. New York, John Wiley, 1980
- [2] Bruce G, Martin T D. Leave-k-out diagnostics for time series[R]. Tech. Rep. 109, Dept. Of Statistics, Univ. Of

- Washington, Seattle, 1987
- [3] Cook R D. Detection of influential observations in linear regression[J]. *Econometrics*, 1977, 19: 15-18
- [4] Cook R D, Weisberg S. Residuals and influence in regression[M]. New York: Chapman and Hall, 1980
- [5] Cook R C. Influential observations in linear regression[J]. *JASA*, 1979, 74: 169-174
- [6] Engle R F. Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. Inflation[J]. *Econometrica*, 1982, 50: 987-1008
- [7] Fox A J. Outliers in time series[J]. *J. Roy. Statist. Soc. B*, 1972, 48: 39-47
- [8] Manderbrot B B. Fractional Brownian motions, fractional Brownian noises and applications[J]. *SIAM Review*, 1968, 10: 422-437
- [9] 韦博成, 鲁国斌, 史建清. 统计诊断引论[M]. 南京: 东南大学出版社, 1991, 504~507, 42-45
- [10] 张世英, 袁学民. 非线性变结构模型的贝叶斯分析[J]. *天津大学学报*, 1996, 26, (增刊): 31-38
- [11] 张世英, 李铁山. 非线性回归的诊断问题[J]. *天津大学学报*, 1998, 31(3): 257-264
- [12] 张世英, 黄建洪. 模型结构变化点检测算法[J]. *应用数学学报*, 1987, 10(3): 267-275
- [13] 张世英等. 多状态模型的贝叶斯分析[J]. *天津大学学报*, 1990, 3: 105-114

Diagnose analysis in ARCH models

KE Ke, ZHANG Shi-ying

School of Management, Tianjin University, Tianjin 300072, China

Abstract: The diagnose analysis and modeling problem with structure change in ARCH models are discussed in this paper. The paper proposes fractional integrated augmented GARCH-M model which can embrace almost all the current ARCH models. Based on the fractional integrated augmented GARCH-M model, the handicap in model specification test, long memory test and parameters estimation is overcome. The segment-modeling method is used to detect the structure changing points of the ARCH model and test the models for segment-changing models. Finally, with the composite index of closing price in Shanghai stock market, it is shown to illustrate the effectiveness of the diagnose analysis.

Key words: ARCH model; diagnose analysis; modeling with structure change