

文章编号:1003-207(2015)06-0032-09

DOI:10.16381/j.cnki.issn1003-207x.2015.06.005

上证综指的股指波动:基于模糊 FEGARCH 模型及不同分布假设的预测研究

侯利强,杨善林,王晓佳,陈志强

(合肥工业大学管理学院,安徽 合肥 230009)

摘要:本文主要对 2006 年至 2011 年上证综指收益率序列的高频波动性进行预测研究。首先,针对金融数据的非线性和不确定等特性,借助模糊逻辑系统,提出一种新的金融市场波动率的预测方法—模糊 FEGARCH 模型,用来更好的针对具有非线性特性的收益率数据进行预测。其次,为了判断分布型模型和不对称型模型对预测精度的影响程度,分别采用分布型(GARCH-N,GARCH-t,GARCH-HT 和 GARCH-SGT)和不对称型(GJR-GARCH、EGARCH 和模糊 FEGARCH)的波动模型进行高级能力预测法(SPA)检测。实证结果表明,不对称模型对波动率预测的影响程度比分布假设的确定更为重要,而且模糊 FEGARCH 模型对于具有尖峰厚尾、高偏度和杠杆效应的非线性波动数据的预测能力更佳,说明了该模型的有效性与实用性。

关键词:波动性;模糊 FEGARCH 模型;预测;SPA 检验

中图分类号:N945.24 **文献标识码:**A

1 引言

自 1987 年美国的股市崩盘以来,对于金融市场波动性的建模和预测就受到了学者、从业人员和监管者的高度重视,他们在期权定价,资产分配和规避风险等金融领域发挥着重要的作用。在之后的十余年中,我们在世界金融市场目睹了许多金融机构处在破产或濒临破产的状态中,他们遭受了巨大的损失的原因是面对金融市场(如股市)的波动难以进行有效的把控。这些金融灾难进一步强调了波动性预测(在险价值的计算)对于风险管理意义的重要性。鉴于此,追求准确的预测依然是当今金融风险管理的焦点。

对于金融市场的预测关键因素是对波动率的准确预测。而收益率的波动性大多都由一些特定的因素构成,如持久性、集群波动性、时变性和尖峰数据形态等。Bollerslev^[1]意识到这一点,因此他引入了广义条件异方差模型(GARCH),这种新的方法作为一种考虑了上述所有因素的计量经济学模型,很

快就成为金融学界常用的研究波动性的工具。然而尽管 GARCH 模型有其特有的优点,学者们仍然对其未能考虑到不对称情形下的波动和不同分布假设下的波动这两类情形而提出质疑。

据笔者分析,现有的文献关于 GARCH 波动性的预测方法主要分为两类。第一类是利用 GARCH 模型的不同分布假设,预测金融市场的波动性。大量的实证证据表明股票收益率存在着非正态的分布特点,出于这个原因,学者们提出了针对误差分布的改进模型,用以减轻正态假设的不足。例如,一些学者证明采用 t 分布^[2]和 HT 分布^[3]在针对厚尾分布时是有效的;而偏度广义 t 分布(SGT)^[4]被证明是一种可以更灵活的处理尖峰厚尾和高偏度数据的分布。又如 Wilhelmsson 和 Chuang 等^[5-6]采用一系列误差分布来预测标准普尔 500 的预期收益率,用以分析 GARCH(1,1)模型的预测能力,他们引入一个含有尖峰的误差分布模型用来证明此模型具有较好的预测能力。但是,我们从中可以看出,引入时变的偏度或峰度分布并不能进一步提高预测能力。Chuang 等^[6]学者分别比较了在股票和外汇市场背景下的线性 GARCH 模型对波动性的预测能力,他们采用移动窗口估计技术,表明就波动性预测建模而言,一个相对复杂的分布并不是总是优于一个简单的分布。

收稿日期:2013-07-15; 修订日期:2014-02-21

基金项目:国家自然科学基金资助项目(71101041)

作者简介:侯利强(1984-),男(汉族),北京人,合肥工业大学管理学院,博士研究生,研究方向:金融工程、数学建模。

第二类是利用不同非对称的 GARCH 模型进行预测。通过引入包含正面和负面冲击效应的不对称性模型,用来克服 GARCH 模型未考虑到不对称情形下的波动性这个缺点,这些模型包括 Nelson^[7]提出的指数 GARCH 模型(EGARCH)和 Glosten、Jagannathan 和 Runkle^[8]提出的门限 GARCH 模型(GJR-GARCH)。同时,Engle 和 Ng^[9]的研究结果表明二次 GARCH 模型可以对股票收益率进行预测。也有学者 Taylor^[10]的研究结果已经找到了支持 GJR-GARCH 模型可以进行良好的预测的证据。同时,Yadav^[11]已经得到了 EGARCH 模型在预测股市波动问题非常有效的结论。另外,Evans 和 McMillan^[12]的研究结果表明,在考虑不对称情形下的波动性预测更为准确。总之,不对称波动性在预测中发挥着至关重要的作用。

除了不对称性和假设分布两个因素之外,在利用 GARCH 类的模型进行样本外点的预测时,替换真实波动的代理(Proxy)也是非常重要的。先前的研究中绝大多数都采用日平方收益率作为隐含波动率的代理^[13-15]。然而,平方收益率是一个有噪声的方差估计值,会降低 GARCH 模型预测波动时的准确性^[16-20]。因此,本文采用的是已实现的波动率,而不是利用平方收益率替代隐含波动率的方法进行预测。

以往主要是用 GARCH 模型族等对金融数据收益波动性进行研究,然而金融时间序列的观测数据除了具有随机性,还带有模糊性,传统的 GARCH 模型族的方法无疑忽视了收益率波动的模糊性。由于考虑到股票市场的复杂性和非线性仅通过 GARCH 族模型很难解决,而模糊建模法对于描述系统中的复杂动力学方程和不对称性是十分有效的方法,所以本文将模糊逻辑系统引入到 GARCH 预测模型中。

本文的研究思路主要为:首先,弥补先前论文仅单一考虑分布形模型(GARCH-N, GARCH-T, GARCH-HT, GARCH-SGT)或非对称型模型(GJR-GARCH、EAGRCH)的情形,对波动率分别进行预测,并采用 SPA 方法,检测不同非对称模型和不同误差分布假设的重要性;其次,还将从数据的模糊不确定性角度进行探讨,利用模糊逻辑方法,在传统 EGARCH 模型的基础上,建立非对称的 Fuzzy EGARCH(模糊 FEGARCH)模型,并用遗传(GA)算法估计模型中的参数。

本文研究的结构如下,第二节介绍基于

GARCH 模型对波动性的预测,第三节介绍模糊 FEGARCH 模型,第四节描述预测和评价的方法。结果的分析 and 主要结论分别在第五、六节给出。

2 基础 GARCH 预测模型

2.1 GARCH 类模型

对称型 GARCH(1,1)模型可以由如下方程组表示:

$$\begin{cases} y_t = \mu + \varepsilon_t \\ \varepsilon_t = \sigma_t z_t \\ z_t | \Omega_{t-1} \stackrel{iid}{\sim} F(0, 1) \\ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{cases} \quad (1)$$

这里的 μ 和 σ_t^2 分别代表条件均值和收益率的方差, ε_t 是一个增量过程,而 $F(0, 1)$ 是一个均值为 0、方差为 1 的概率密度函数。此外,为了保持条件方差的正定性和更好的平稳性, α 和 β 需要满足 $\alpha + \beta < 1$ 的约束条件。

而在 GARCH 类模型中,有两类模型非常适宜处理具有非对称特性的金融时间序列问题,它们是 GJR-GARCH 模型和 EGARCH 模型。

GJR-GARCH 模型与方程组(1)的第四式不同,其条件方差形式如下:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \delta I_{t-1} \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (2)$$

这里, I_{t-1} 表示示性函数,当 $\varepsilon_{t-1} < 0$ 时, I_{t-1} 取值为 0;当 $\varepsilon_{t-1} \geq 0$ 时,取值为 1。示性函数变量分为正向冲击(利好消息)和负向冲击(利空消息)。即在 GJR-GARCH 模型中,正向冲击可以用 α 表示,而负向冲击要受到 $\alpha + \delta$ 的作用。

EGARCH 模型的条件方差表现形式为:

$$\ln(\sigma_t^2) = \omega + \alpha \left[\gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \sqrt{\frac{2}{\pi}} \right] + \beta \ln(\sigma_{t-1}^2) \quad (3)$$

其中 γ 为非对称杠杆系数,它表明当一个负向冲击来到时,往往比一个正向冲击的影响效果要大。

2.2 分布假设

GARCH 类模型在研究中的要面对的另一个重要问题是经验分布的非正态性。一般情况下可分为有偏分布和厚尾分布两类。

首先,介绍有偏分布,这类分布主要以 SGT 分布为主。SGT 分布是一种有偏的广义 t 分布,这类分布可以有效的解决金融数据中存在的尖峰厚尾和有偏的问题。设收益率的标准化残差为 (z_t) 的 SGT 分布的概率密度函数可记为:

$$F(z_i; N, k, \lambda) = C[1 + |z_i + \delta|^k \div ((N + 1)/k)^{(1 + \text{sign}(z_i + \delta)\lambda)k\theta^k}]^{-(N+1)/k} \quad (4)$$

上式中的 $C, \theta, \rho, g, \delta$ 分别为控制参数, 它们的表达形式和意义如下所示:

$$C = \frac{1}{2}k \left(\frac{N+1}{k}\right)^{-\frac{1}{k}} B\left(\frac{N}{k}, \frac{1}{k}\right)^{-1} \theta^{-1} \quad (5)$$

$$\theta = (g - \rho^2)^{-0.5} \quad (6)$$

$$\rho = 2\lambda B\left(\frac{N}{k}, \frac{1}{k}\right)^{-1} \left(\frac{N+1}{k}\right)^{\frac{1}{k}} B\left(\frac{N-1}{k}, \frac{2}{k}\right) \quad (7)$$

$$g = (1 + 3\lambda^2) B\left(\frac{N}{k}, \frac{1}{k}\right)^{-1} \left(\frac{N+1}{k}\right)^{\frac{2}{k}} B\left(\frac{N-2}{k}, \frac{3}{k}\right) \quad (8)$$

$$\delta = \rho \quad (9)$$

其中残差 z_i 是均值为 0, 方差为 1 标准正态分布; N 是负责控制厚尾的参数, 这里 $N > 2$; k 是负责控制峰度的参数, 这里 $k > 0$; λ 用来控制偏度, 要满足约束条件 $|\lambda| < 1$; sign 是一个符号函数, 而 $B(\cdot)$ 表示的是贝塔函数。此外, 当 $k = 2$ 且 $\lambda = 0$ 时, 就是学生 t 分布; 当 $k = 2, \lambda = 0$ 且 N 趋近无穷大时, 就是正态分布 (N 分布)。

其次, 介绍厚尾分布, 我们以 HT 分布为例。这种分布是一种典型的厚尾分布。它的概率密度函数如下所示:

$$F(z_i, a_0, 1) = \frac{(1 + a_0 z_i^2)^{-1.5} \exp\left(-\frac{z_i^2}{2(1 + a_0 z_i^2)}\right)}{\sqrt{2\pi}(\Phi(a_0^{-0.5}) - \Phi(-a_0^{-0.5}))} \quad (10)$$

其中, z_i 的标准差为 1, Φ 表示标准正态分布的累积概率密度函数, a_0 是负责控制厚尾的参数, 且 a_0 满足约束条件 $0 < a_0 < 1$ 。当 $a_0 \rightarrow 0$ 时, HT 分布就会转化成一个标准正态分布; 而当 $a_0 \rightarrow 1$ 时, HT 分布的厚尾程度将达到最大。

3 模糊 FEGARCH 模型

3.1 模糊逻辑系统

在本节中, 我们将着重阐述一种新型的 GARCH 预测模型, 模糊 FEGARCH 模型。

模糊逻辑系统是由模糊化、模糊规则、模糊推理机和去模糊化四个环节共同构成的。在本节中, 我们先介绍一些模糊逻辑的相关概念及系统的构成要素。

下面介绍构成模糊逻辑系统的四个环节。

模糊化: 模糊逻辑系统的输入信号是实数域变量或语言值变量, 而模糊推理机只能处理模糊集信

号, 不能直接处理实数域信号, 因此实数域的输入信号必须模糊化后才能进行模糊推理运算。模糊化是将实数域变量 $x \in U \subset R^n$ 映射到 U 对应的模糊集 A^l 上。

模糊规则: 模糊规则是模糊逻辑系统的核心, 它由具有 IF-THEN 形式的模糊规则组成。其形式为:

IF x_1 为 A_1^l, x_2 为 A_2^l, \dots, x_n 为 A_n^l ;

THEN y 为 B^l 。

其中 $A_i^l (i = 1, 2, \dots, n)$; B^l 分别为 $U_i \subset R$ 和 $V \subset R$ 对应的模糊 $x = (x_1, x_2, \dots, x_n)^T \in U$; $y \in V$ 分别表示模糊逻辑系统的输入变量和输出结果; l 表示模糊规则的序号, 若共有 M 条规则, 则 $l = 1, 2, \dots, M$ 。

模糊推理机: 利用模糊规则库中的 IF-THEN 逻辑原理将输入信号 U 对应的模糊集 A^l 映射到输出信号 V 对应的模糊集 B^l 。输入信号 U 对应的模糊集 A^l , 经模糊推理机处理后, 得到输出信号对应的模糊 B^l :

$$\mu_{B^l}(y) = \max[\text{supmin}(\mu_{A^l}(x), \mu_{A_1^l}(x), \dots, \mu_{A_n^l}(x), \mu_{B^l}(y))]$$

去模糊化: 是将输出信号 V 对应的模糊集 B^l 映射到一个实点 $y \in V$ 。为了使该点能够很好地反映模糊集 B^l 代表的信息, 本文采用中值法清晰化模糊集合。对于同一模糊集上的模糊等级 F_1, F_2, \dots, F_m , 若其中值分别为 y_1, y_2, \dots, y_m , 某输出信号对它们的隶属度分别为 w_1, w_2, \dots, w_m , 则代表输出信号 V 指标的实数值 y 可以用下式计算得到:

$$y = (y_1 w_1 + y_2 w_2 + \dots + y_m w_m) / (w_1 + w_2 + \dots + w_m)$$

3.2 模糊 FEGARCH 模型

在本文中, 采用上述模糊逻辑系统的 IF-THEN 准则来构建模糊 FEGARCH 模型, 用来更好地解决 EGARCH 模型中的集群性问题。

下文中的前提变量是由股票市场收益率的先前值和其波动率的先前值一起构成的。

设前提变量:

$$x_1(t) = y_{t-1}, x_2(t) = \sigma_{t-1}^2$$

故而通过上式可以看出, 模糊模型可以成功的捕捉到杠杆效应和集群效应的存在, 它们与公式 (11) 的收益率 y_t 和波动率 σ_t^2 的序列密切相关。

下面, 先利用 IF-THEN 准则来描述 EGARCH 模型。

IF

前提变量 $x_1(t)$ 为 $F_{k_1}, x_2(t)$ 为 $F_{k_2}, \dots, x_n(t)$

为 F_{ki} ,

THEN

$$y_t = \sigma_t \varepsilon_t$$

$$\ln \sigma_t^2 = \omega_k + \sum_{i=1}^q \alpha_{ki} \left[\gamma \frac{y_{t-1}}{\sigma_{t-1}} + \frac{|y_{t-1}|}{\sigma_{t-1}} - \sqrt{\frac{2}{\pi}} \right] +$$

$$\sum_{j=1}^p \beta_{kj} \ln \sigma_{t-1}^2, k = 1, 2, \dots, L \quad (11)$$

其中, y_t 是模糊逻辑系统的输出变量, F_{ki} 为相应的模糊集 $i = 1, 2, \dots, n$, L 是 IF-THEN 准则的个数。

其次,对上述模糊系统采用去模糊化操作(即完成 3.1 节中的去模糊化映射过程),结果如下所示:

$$\sigma_t^2 = \frac{\sum_{k=1}^L u_k(x(t)) \left[\omega_k + \sum_{i=1}^q \alpha_{ki} y_{t-i}^2 + \sum_{j=1}^p \beta_{kj} \sigma_{t-j}^2 \right]}{\sum_{k=1}^L u_k(x(t))} \quad (12)$$

$$u_k(x(t)) = \prod_{i=1}^n F_{ki}(x_i(t)) \quad (13)$$

$$y_t^2 = \sigma_t^2 \varepsilon_t^2 = \left\{ \frac{\sum_{k=1}^L \prod_{i=1}^n \exp\left(-\frac{1}{2} \left(\frac{x_i(t) - c_{ki}}{\zeta_{ki}}\right)^2\right) \left[\omega_k + \sum_{i=1}^q \sum_{j=1}^p \alpha_{ki} \left[\gamma \frac{y_{t-i}}{\sigma_{t-j}} + \frac{|y_{t-i}|}{\sigma_{t-j}} - \sqrt{\frac{2}{\pi}} \right] + \sum_{j=1}^p \beta_{kj} \ln \sigma_{t-j}^2 \right]}{\sum_{k=1}^L \prod_{i=1}^n \exp\left(-\frac{1}{2} \left(\frac{x_i(t) - c_{ki}}{\zeta_{ki}}\right)^2\right)} \right\} \varepsilon_t^2 =$$

$$= \left\{ \sum_{k=1}^L h_k(x(t)) \left[\omega_k + \sum_{i=1}^q \sum_{j=1}^p \alpha_{ki} \left[\gamma \frac{y_{t-i}}{\sigma_{t-j}} + \frac{|y_{t-i}|}{\sigma_{t-j}} - \sqrt{\frac{2}{\pi}} \right] + \sum_{j=1}^p \beta_{kj} \ln \sigma_{t-j}^2 \right] \right\} \varepsilon_t^2 = \Psi^T z(t) \varepsilon_t^2 \quad (15)$$

其中:

$$\Psi = \begin{bmatrix} h_1 \omega_1 + \dots + h_L \omega_L \\ h_1 \alpha_{11} + \dots + h_L \alpha_{L1} \\ \vdots \\ h_1 \alpha_{1q} + \dots + h_L \alpha_{Lq} \\ h_1 \beta_{11} + \dots + h_L \beta_{L1} \\ \vdots \\ h_1 \beta_{1p} + \dots + h_L \beta_{Lp} \end{bmatrix}$$

$$z(t) = \begin{bmatrix} 1 \\ (\gamma y_{t-1}/\sigma_{t-1}) + |y_{t-1}/\sigma_{t-1}| - \sqrt{2/\pi} \\ \vdots \\ (\gamma y_{t-q}/\sigma_{t-p}) + |y_{t-1}/\sigma_{t-1}| - \sqrt{2/\pi} \\ \ln \sigma_{t-1}^2 \\ \vdots \\ \ln \sigma_{t-p}^2 \end{bmatrix}$$

最后,我们给出衡量准确性的方法。由于模糊 EGARCH 模型是一个时变非线性方程。故而,它

$x(t) = [x_1(t), x_2(t), \dots, x_n(t)]$, $F_{ki}(x_i(t))$ 是 $x_i(t)$ 在 F_{ki} 中的隶属度函数。本文采用高斯隶属度函数:

$$u_1(x(t)) = \prod_{j=1}^n F_{ki}(x_j(t)) = \prod_{j=1}^n \exp\left(-\frac{1}{2} \left(\frac{x_j(t) - c_{ki}}{\zeta_{ki}}\right)^2\right) \quad (14)$$

其中, c_{ki} 和 ζ_{ki} 分别表示在第 k 步 IF-THEN 准则中第 i 个前提变量的中心和广度。我们假设 $u_k(x(t)) \geq 0, k = 1, 2, \dots, L$ 且 $\sum_{k=1}^L u_k(x(t)) > 0$, 则可以得到 $h_k(x(t)) \geq 0, k = 1, 2, \dots, L$ 且 $\sum_{k=1}^L h_k(x(t)) = 1$ 。

其中, $h_k(x(t)) = u_k(x(t)) / \sum_{k=1}^L u_k(x(t))$ 。通过方程(12)~(14),我们可以得到模糊 FEGARCH 模型的输出函数:

可通过最小均方误差方法衡量其预测的准确性。

令:

$$E = \sum_{i=1}^N [y_i - \hat{y}_i]^2 \quad (16)$$

其中, $y_i, i = 1, 2, \dots, N$ 是上证指数的收益率, N 是观测数据的个数, \hat{y}_i 是从公式(15)中得到的估计值。很明显, E 是一个非线性函数,对于 E 中的未知参数,由于方程 E 可能出现局部最小值,所以我们采用最常用的遗传算法估计未知参数集 $\Theta = (c, \zeta, \alpha, \beta, \gamma)$, 我们只要使得 $E(\Theta)$ 全局最小,就可以利用模糊 FEGARCH 模型解决非对称性问题的同时,降低股市出现集群波动现象为预测带来的干扰。

下面将给出采用模糊模糊逻辑系统建立的模糊 FEGARCH 模型流程图,如图 1 所示。

4 预测方法及 SPA 检验方法

4.1 波动率预测方法

本节将主要介绍波动率预测的方法及流程,主

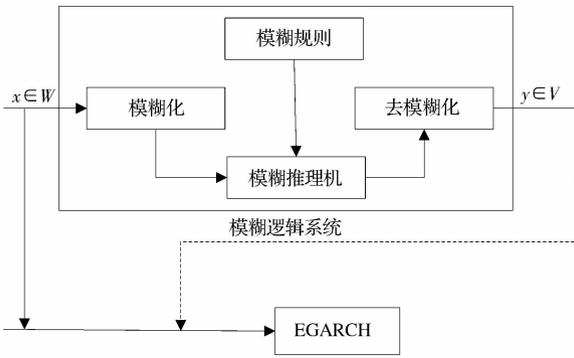


图1 模糊逻辑系统图

要流程如下所示：

步骤 1: 将数据样本总体 ($t = 1, 2, \dots, N = 1217$) 划分为参数估计样本集与预测样本集两大部分。其中, 估计集合包含 1000 个交易日的数据, 而预测集合包含剩余的 217 个交易日的数据, 即 $t = H + 1, H + 2, \dots, H + 217$ 。

步骤 2: 我们选取 $t = 1, 2, \dots, H$ 的数据作为第一个估计样本, 分别对上述七种波动率预测模型的参数进行估计, 在此估计基础之上, 获得未来 1 天的波动率的预测值, 即第 1001 天的预测值, 记为 σ_{H+1}^2 。

步骤 3: 保持总窗口长度不变, 向后递推 1 天作为第二次选取的估计样本, 再利用七种波动率预测模型预测未来 1 天的波动率, 即第 1002 天的预测值, 记为 σ_{H+2}^2 , 反复这样操作, 最终得到第 1217 天的波动率预测值, 记为 σ_{H+M}^2 。

采用已实现波动率 RV_m ($m = H + 1, H + 2, \dots, H + M$) (详见 5.1 节的样本数据说明), 替代真实的市场波动率, 与我们预测的结果进行比较, 分析预测精度之间的差异。

4.2 SPA 检验方法

本文采用四类损失函数作为各类已实现波动率模型预测精度的评价标准。这四种损失函数分别记为 L_i , $i = 1, 2, 3, 4$, 其中, L_1 为均方误差 (MSE), L_2 为平均绝对误差 (MAE), 它们是此类判断中最常用的两类损失函数形式, L_3 和 L_4 分别是均方百分比误差 (MSPE) 和平均绝对百分比误差 (MAPE)。

损失函数的定义如下所示：

令 $e_{km} = RV_m - \sigma_{km}^2$ 为相应的预测误差, 其中 σ_{km}^2 表示第 k 个模型在区间 m 对已实现波动率的预测值 ($k = 1, 2, \dots, 7$; $m = H + 1, H + 2,$

$\dots, H + M$), RV_m 为代替真实波动率的已实现波动率, 故两者之差就是预测误差。

于是我们定义的 4 类损失函数, 它们可以由以下公式表示：

$$L_1 = \frac{1}{M} \sum_{m=H+1}^{H+M} e_{km}^2 ; L_2 = \frac{1}{M} \sum_{m=H+1}^{H+M} |e_{km}|$$

$$L_3 = \frac{1}{M} \sum_{m=H+1}^{H+M} \left(\frac{e_{km}^2}{RV_m^2} \right) ; L_4 = \frac{1}{M} \sum_{m=H+1}^{H+M} \left| \frac{e_{km}}{RV_m} \right|$$

具体检验步骤如下：

步骤 1: 假设有 $J + 1$ 种类型的已实现波动率模型, 记为 M_k , $k = 0, 1, \dots, J$ 。每种波动率模型 M_k 得到下一天的波动率的预测值为 σ_{km}^2 。对每个预测值, 可以计算出相应的 L_i , $i = 1, 2, 3, 4$ 。

步骤 2: 用 M_0 表示为 SPA 检验的基础模型。因此, 对于其他的 $k = 1, 2, \dots, J$ 种类型的已实现波动率模型, 可以计算其相对于基础模型 M_0 的相对损失函数值。

记为：

$$\varphi = L_{i,0,m} - L_{i,k,m}$$

需要验证在模型 M_k , $k = 0, 1, \dots, J$ 中, 是否存在比模型 M_0 更优的模型。

步骤 3: 利用假设检验的方式, 我们定义 H_0 : 对比模型 M_k , M_0 是其中最好的模型。若 $u = E(\varphi_m) = E(\varphi_{1,m}, \varphi_{2,m}, \dots, \varphi_{J,m})$, $H_0: u \leq 0$ 。这种假设检验的检验统计量为：

$$T = \max \frac{\sqrt{M\varphi_k}}{\omega_k} \tag{17}$$

$$\text{其中 } \varphi_k = \frac{1}{M} \sum_{m=H+1}^{H+M} \varphi_{k,m}, \omega_k^2 = \text{Var}(M^{0.5} \varphi_k)$$

经上述步骤计算出的 SPA 检验的 p 值越接近于 1, 则越代表 M_0 的预测性能越好。

5 案例分析

5.1 样本数据说明

本文采用的实证研究样本为上证综指 (SSEC) 从 2006 年 6 月 30 日至 2011 年 6 月 30 日共 $N = 1217$ 个交易日中每 5min 记录一次的高频股价数据, 记为 $I_{t,d}$ ($t = 1, 2, \dots, 1217$; $d = 0, 1, \dots, 48$), 其中: $I_{t,0}$ 表示第 t 天的开盘价, $I_{t,48}$ 表示第 t 天的收盘价。上海证券交易所每天共有 4h (240min) 的连续竞价交易时间, 因此每天可以产生 48 个高频股价记录 (其中不包括 $I_{t,0}$), 样本总体的高频数据量为 58416 个。

定义:日收益率 R_t 为相邻 2 个交易日收盘价的自然对数之差,记:

$$R_t = 100(\ln I_{t,48} - \ln I_{t-1,48}) \quad t = 2, 3, \dots, N$$

由于真实市场波动率是不可观测的,所以目前公认的方法是用基于高频股价数据的实现波动率作为真实市场波动率的代理。因此,本文采用基于 5min 高频数据的实现波动率作为评价波动预测精度的基准。

根据 Andersen^[21]等的定义,第 t 天的实现波动率可表示为第 t 天内高频收益率的平方和。

即:

$$RV_t' = \sum_{d=1}^{48} R_{t,d}^2$$

需要指出的是,由于股票市场不像外汇市场那样在每个时段都连续交易,因此,能观察和记录到的高频股价数据只能反映有交易时段的市场波动状况,而无法包含无交易时段的波动信息(即股票市场从收盘到第 2 天开盘的所谓波动率)。为了解决该问题,我们采用某种尺度参数 Υ 对 RV_t' 进行尺度变换,使其能刻画股票市场全天的真实波动率,即对第 t 天的实现波动率估计为:

$$RV_t = N^{-1} \sum_{i=1}^N R_i^2 / \left[N^{-1} \sum_{i=1}^N RV_i' \right] RV_t' \quad (18)$$

5.2 试验结果

本文采用 Eviews6.0 软件中有关正态性检验的功能模块,计算出考察时段中的上证综指收益率和已实现波动率的描述性统计结果,详见表 1。

从表 1 的描述性统计结果中可以看到,所有序列都表现出明显的有偏和尖峰、厚尾的特征,同时序列之间具有显著的自相关性(全部在 1% 水平下显

著)。J-B 统计量显著,故而拒绝正态分布假设。

表 1 R_t 和 RV_t' 的描述性统计结果

	R_t	RV_t'
均值(%)	0.02445	1.945
标准差	1.438	2.476
偏度	0.594*	3.623*
峰度	5.479*	37.268*
J-B	2871*	83654*
Q(20)	32.863*	4735.98*
LM(12)	83.356	176.289

注:J-B 为 Jarque-Bera 统计量,Q(20)为之后 20 期的 Ljung-Box Q 统计量,* 表示它们在 1% 水平下均显著。

下面,我们给出 7 种不同波动率预测模型的参数估计结果,前六种使用的是极大似然参数估计法,模糊 FEGARCH 模型的参数采用遗传算法估计获得,本文设定为 IF-THEN 准则中的迭代次数 $L = 5$,初始种群是根据完全随机的方法产生的,计算迭代终止的条件是 $E(\theta)$ 最小或达到最大遗传代数。本文共循环迭代次数为 100 次。下面连续给出三张表,它们展示了模型的参数估计结果。

表 2a, 2b, 2c 给出对七种预测上证综指波动率的模型的参数,从表中可以看出,模型参数均显著,对于不同分布的 GARCH 模型而言,由于存在非对称性,它并不能完全反映出来,而 GJR-GARCH 模型就能够较好的刻画此情形,但是它不能处理非线性问题,而我们建立的模糊 FEGARCH 模型就可以更好地兼容处理这两类问题。

在上述预测模型建立之后,我们就使用 Eviews6.0 软件和 Winrats7.0 软件分别对波动率进行预测,并得到损失函数的值。通过表 3,我们可以看到各模型的损失函数值和其预测能力的排名。

表 2a 已实现波动率常用模型的参数估计结果

参数	GARCH-N	GARCH-t	GARCH-HT	GARCH-SGT	GJR-GARCH	EGARCH
μ	0.0984 [0.0082]	0.0751 [0.0377]	0.0826 [0.0394]	0.0609 [0.0394]	0.0081 [0.0296]	0.0014 [0.0386]
ω	0.0990 [0.0034]	0.1266 [0.0435]	0.0916 [0.0131]	0.0916 [0.0101]	0.1602 [0.0083]	0.0259 [0.0098]
α	0.2417 [0.0074]	0.0939 [0.0223]	0.1640 [0.0032]	0.0740 [0.0069]	-0.0384 [0.0055]	0.0905 [0.0072]
β	0.8451 [0.0049]	0.8441 [0.0347]	0.8537 [0.0026]	0.8547 [0.0177]	0.8261 [0.0061]	0.9452 [0.0141]
δ	Null	Null	Null	Null	0.2649 [0.0158]	Null
γ	Null	Null	Null	Null	Null	-1.8161 [0.1623]
N	∞	8.2370 [1.8994]	Null	8.2370 [1.8994]	Null	Null
κ	2	2	Null	2	Null	Null
λ	0	0	Null	0	Null	Null
a_0	Null	Null	0.0535 [0.0096]	Null	Null	Null

注:表中数字为各波动率模型的参数估计结果,数字后边[]内的是 t 统计量在 1% 显著性水平下的值。

表 2b 已实现波动率模糊 EGARCH 模型的参数估计结果

模型	IF-THEN 准则次数	参数	参数估计值	t 值
Fuzzy EGARCH	l = 1	ω^1	0.0416	0.8864
		α^1	0.0803	1.4689
		β^1	0.0354	4.3674
		γ^1	-0.0147	0.3584
	l = 2	ω^2	0.0025	0.3684
		α^2	0.0942	1.3681
		β^2	0.0471	9.2486
		γ^2	-0.0264	1.1382
	l = 3	ω^3	0.0359	4.3254
		α^3	0.0685	1.1368
		β^3	0.9124	1.3846
		γ^3	-0.0358	0.3858
	l = 4	ω^4	0.1395	1.3575
		α^4	0.0522	0.1658
		β^4	0.8931	3.2248
γ^4		-0.0483	1.3586	
l = 5	ω^5	0.0482	1.3687	
	α^5	0.0689	0.9824	
	β^5	0.0398	3.3548	
		γ^5	-0.0987	1.0587

注:表中数字为各波动率模型的参数估计结果,数字后边[]内的是 t 统计量在 1% 显著性水平下的值。

表 2c 模糊 FEGARCH 的中心和广度两个参数的估计结果

	l = 1	l = 2	l = 3	l = 4	l = 5
中心 (c_{11}, \dots, c_{51})	1.3584	-0.3565	-0.3874	1.3258	0.3158
广度 ($\zeta_{11}, \dots, \zeta_{51}$)	0.3848	0.3248	0.0257	0.4211	0.6126

从表 3 的实证结果可以看到,对所考察的这一时间区域的单一样本而言,当 $L = 1, 2, 3, 4, 5$ 时模糊 FEGARCH 模型在 4 种损失函数的标准下,均达到了最好的预测精度;而在不同分布的 GARCH 模型中,GARCH-SGT 分布优于其它分布,预测效果更加精确。我们同时也发现,在不同损失函数下,各模型预测精度的排名次序会有一些变化,就如同我们在 4.2 节阐述的那样。所以想要得到对预测精度更加全面,则必须对表 3 中的预测结果进行进一步的 SPA 检测。

下面,给出已实现波动率模型的 SPA 检验结果,详见表 4。

表 3 已实现波动率模型的预测精度及其排名

L_i	MSE		MAE		MSPE		MAPE	
	数值	排名	数值	排名	数值	排名	数值	排名
GARCH-N	0.0379	6	0.0839	7	0.0805	7	0.0817	7
GARCH-t	0.0415	7	0.0715	6	0.0726	6	0.0445	5
GARCH-HT	0.0331	5	0.0731	5	0.0648	5	0.0673	6
GARCH-SGT	0.0275	4	0.0624	4	0.0512	4	0.0671	5
GJR-GARCH	0.0218	2	0.0457	3	0.0354	2	0.0358	2
EGARCH	0.0224	3	0.0328	2	0.0417	3	0.0363	3
模糊 FEGARCH $L = 1, 2, 3, 4, 5$	0.0103		0.0194		0.0164		0.0279	
	0.0152		0.0235		0.0147		0.0225	
	0.0136	1	0.0174	1	0.0235	1	0.0258	1
	0.0201		0.0265		0.0222		0.0351	
	0.0125		0.0233		0.0258		0.0304	

表 4 已实现波动率模型的 SPA 检验结果

L_i	基本模型	对比模型 M_k					
	M_0	GARCH	GARCH-t	GARCH-HT	GARCH-SGT	GJR-GARCH	EGARCH
MSE		1.000	0.991	0.899	0.873	0.872	0.876
MAE	Fuzzy-	0.992	0.932	0.903	0.899	0.874	0.898
MSPE	EGARCH	0.904	0.934	0.922	0.867	0.859	0.861
MAPE		0.933	0.911	0.871	0.845	0.834	0.843
	基本模型	GARCH	GARCH-t	GARCH-HT	模糊 FEGARCH	GJR-GARCH	EGARCH
MSE		1.000	0.911	0.894	0.127	0.137	0.153
MAE	GARCH-SGT	0.985	0.962	0.877	0.101	0.153	0.168
MSPE		0.993	0.935	0.822	0.133	0.162	0.150
MAPE		0.946	0.893	0.841	0.155	0.186	0.255

注:表中数字为 SPA 检验的 p 值, p 值越大,说明与所考察的对比模型 M_k 而言,基础模型 M_0 的预测精度越高,若该值越接近于 1,则表明其基本模型的预测精度越高。

从表 4 可以看出,在七种模型的 SPA 检验中,模糊 FEGARCH 作为基本模型的 p 值都较大,甚至接近于 1,表明该模型在非对称性模型(EGARCH、GJR-GARCH)和误差分布函数模型(GARCH-N、GARCH-t、GARCH-HT 和 GARCH-SGT)中,预测精度均是最佳的。而就分布函数而言,GARCH-SGT 的预测效果相对其它三种较好,但是 GARCH-SGT 模型相对模糊 FEGARCH 等不对称性模型的 SPA 检验的 p 值较小。对于中国股票市场的特点而言,选择恰当的非对称 GARCH 模型比选择误差分布函数更为重要。

6 结语

本文主要结论有以下几点:

(1)就分布假设而言,GARCH-SGT 模型的预测精度好于其它三个分布的 GARCH 模型,其原因是该模型具控制有偏和尖峰厚尾的参数,所以较为符合中国股市的特征。然而通过本文的研究表明,具备不对称情形的 GARCH 预测模型比选择误差分布的类型更为重要。

(2)我国上证综指股票收益率序列呈现尖峰厚尾的特征,存在波动聚集性,利空消息对波动性的影响更大,即存在杠杆效应。进一步对各模型的预测能力进行了比较,结果表明 Fuzzy EGARCH 模型的预测效果最好,GJR-GARCH 模型次之。可见本文建立的模型不仅较好的解决了非对称性,并且有效的捕捉了普通 GARCH 模型无法捕捉到的模糊信息及处理非线性方程的特点,使预测效果更贴近实际。

参考文献:

- [1] Bollerslev T. Generalized autoregressive conditional heteroskedasticity [J]. *Journal of Econometrics*, 1986, 31(3):307-327.
- [2] Bollerslev T. A conditional heteroskedastic time series model for speculative prices and rates of return [J]. *Review of Economics and Statistics*, 1987, 69(3):542-547.
- [3] Hung J C, Lee M C, Liu H C. Estimation of Value-at-Risk for energy commodities via fat-tailed GARCH models [J]. *Energy Economics*, 2008, 30(3):1173-1191.
- [4] Theodossiou P. Financial data and the skewed generalized t distribution [J]. *Management Science*, 1998, (44):1650-1661.
- [5] Wilhelmsson A. GARCH forecasting performance under different distribution assumptions [J]. *Journal of Forecasting*, 2006, (25):561-578.
- [6] Chuang I Y, Lu J R, Lee P H. Forecasting volatility in the financial markets: A comparison of alternative distributional assumptions [J]. *Applied Financial Economics*, 2007, 17(3):1051-1060.
- [7] Nelson D B. Conditional heteroskedasticity in asset returns: A new approach [J]. *Econometrica*, 1991, 59(12):347-370.
- [8] Glosten L, Jagannathan R, Runkle D. On the relation between the expected value and the volatility nominal excess return on stocks [J]. *Journal of Finance*, 1993, 48(5):1779-1801.
- [9] Engle R F, Ng V K. Measuring and testing the impact of news on volatility [J]. *Journal of Finance*, 1993, 48(5): 1749-1778.
- [10] Taylor J W. Volatility forecasting with smooth transition exponential smoothing [J]. *International Journal of Forecasting*, 2004, 20(2): 273-286.
- [11] Loudon G F, Watt W H, Yadav P K. An empirical analysis of alternative parametric ARCH models [J]. *Journal of Applied Econometrics*, 2000, 15(2): 117-136.
- [12] Evans T, McMillan D G. Volatility forecasts: The role of asymmetric and long-memory dynamics and regional evidence [J]. *Applied Financial Economics*, 2007, 17(17):1421-1430.
- [13] Awartani B M A, Corradi V. Predicting the volatility of the S&P-500 stock index via GARCH models: The role of asymmetries [J]. *International Journal of Forecasting*, 2005, 21(1): 167-183.
- [14] Brooks C, Persaud G. Model choice and Value-at-Risk performance [J]. *Financial Analysts Journal*, 2002, 58(5): 87-97.
- [15] Sadorsky P. Modeling and forecasting petroleum futures volatility [J]. *Energy Economics*, 2006, 28(4): 467-488.
- [16] 于亦文. 实际波动率与 GARCH 模型的特征比较分析 [J]. *管理工程学报*, 2006, 20(2): 65-69
- [17] 杨科,陈浪南. 中国股市高频波动率跳跃的特征分析 [J]. *系统工程学报*, 2012, 27(4): 492-497.
- [18] 张小斐,田金方. 异质金融市场驱动的已实现波动率计量模型 [J]. *数量经济技术经济研究*, 2011, (9): 140-153
- [19] 王良,冯涛. 中国 ETF 基金价格“已实现”波动率、跟踪误差之间的 Granger 关系研究 [J]. *中国管理科学*, 2012, 20(1): 59-70.
- [20] 刘伟,陈敏,梁斌. 基于金融高频数据的 ETF 套利分析

[J]. 中国管理科学, 2009, 17(2): 1-7.

quency data and realized volatilities [J] . Econometrica, 2005, 73(1) : 279- 296.

[21] Andersen T G, Bollerslev T, Meddahi N. Correcting the errors: Volatility forecast evaluation using high fre-

The Shanghai Stock Index Volatility: Forecasting Research Based on Fuzzy FEGARCH Model and Different Distribution Hypothesis

HOU Li-qiang, YANG Shan-lin, WANG Xiao-jia, CHEN Zhi-qiang

(School of Management of Hefei University of Technology, Hefei 230009, China)

Abstract: In general, the transmission of volatility in the stock market is time-varying, nonlinear, and asymmetric with respect to both positive and negative results. Given this fact, the method of fuzzy logic systems is adopted to modify the threshold values for an EGARCH model. The volatility forecasting for the SSE stock index series from 2006 to 2011 is provided and the essential source of performance improvements is identified between distributional assumption and volatility specification using distribution-type (GARCH-N, GARCH-t, GARCH-HT and GARCH-SGT) and asymmetry-type (GJR-GARCH and EGARCH) volatility models through the superior predictive ability test. Such evidence strongly demonstrates that modeling asymmetric components which is the fuzzy EGARCH model is more important than specifying error distribution for improving volatility forecasts of financial returns in the presence of fat-tails, leptokurtosis, skewness, leverage effects and nonlinear effects in China stock market.

Key words: volatility; fuzzy EGARCH test model; forecasting; superior predictive ability