

文章编号: 1003- 207(2007)06- 0013- 07

# GARCH 类模型波动率预测评价

黄海南, 钟 伟

(北京师范大学金融研究中心, 北京 100875)

**摘 要:** GARCH 类模型已经广泛运用于波动率的预测, 但对模型的预测表现进行评价却受到了忽视, 其主要原因是缺乏合适的衡量标准。本文首先运用 GARCH 类模型对上证指数收益率进行了全面的估计及样本外预测, 然后以已实现波动率作为波动率预测的评价标准, 通过 M- Z 回归和损失函数来评价 GARCH 类模型的波动率预测表现。结果表明, 无论是样本内还是样本外, GARCH 类模型都能够较好的预测上证指数的收益波动率。其中, 偏斜 t- 分布假设下的 GJR(1, 1) 模型的预测能力最强。

**关键词:** GARCH; 已实现波动率; M- Z 回归; 损失函数

**中图分类号:** F830 **文献标识码:** A

## 1 引言

金融市场中, 无论是在金融衍生产品的定价、金融风险的测定还是在资产组合分析中, 波动率扮演了一个重要的角色。到目前为止, 测定波动率的方法有四种: 一是历史波动率; 第二种方法是隐含波动率; 第三种方法是通过 GARCH 类模型进行估计, 这种方法目前成了主流; 第四种方法是通过随机波动率(SV)模型进行估计得到。在衡量以上四种方法的优劣时, 通常有两个标准: 一方面是模型是否能刻画历史数据的特征, 即样本内预测的表现; 另一方面, 也是更重要的, 是模型能否预测未来的波动率, 即样本外预测的表现。对于第一个标准来说, 对模型的检验基本上可以做到, 经验表明, 由 Engel (1982) 发展起来的 GARCH 类模型表现较好<sup>[1]</sup>。

就样本外预测能力的评价来说, 主要有两种方法, 一是运用 Mincer 和 Zarnowitz(1969) 发展的 M- Z 回归方法(Mincer- Zarnowitz regression)<sup>[2]</sup>。二是利用损失函数。但这两种方法首先都必须对样本外事后波动率进行估计。最初, 在评估样本外的预测时, 用扰动项的平方充当事后波动率, 但模型的表现很差, 这引发了一场讨论, 并令人对 GARCH 类模型的实用性产生了怀疑。<sup>[3]</sup>事实上, Tsay

(2002) 指出, 当用 M- Z 回归方程时, 单独的  $\hat{\sigma}_{t+h}^2$  本身不能作为  $t+h$  时刻波动率的合适度量。例如, 对向前 1 步预测, 从统计观点看,  $E(\hat{\sigma}_{t+1}^2 | F_t) = \hat{\sigma}_{t+1}^2$ , 因此  $\hat{\sigma}_{t+1}^2$  是  $\hat{\sigma}_{t+1}^2$  的一致估计, 但它不是  $\hat{\sigma}_{t+1}^2$  的精确估计, 因为一个有已知均值的随机变量的单个观察不可能提供该随机变量方差的精确估计。从而, 严格的说, 这种评价波动率预测表现的方法是不合适的<sup>[4]</sup>。Andersen 和 Bollerslev (1998) 基于 Nelson (1990)、Drost 和 Werker (1996) 发展的连续时间随机波动框架, 利用高频数据构建了日内收益率的波动率(intra-day volatility) 作为事后波动率, 这就是已实现波动率(realized volatility)<sup>[5]</sup>。在把扰动项的平方替换成已实现波动率后, 作者从理论和实证两方面证明随机波动率模型的表现很好。Andersen, Bollerslev, Diebold 和 Labys (2003) 给出了已实现波动率的理论解释<sup>[6]</sup>。

国内有关波动率的研究侧重于运用 GARCH 类模型或者 SV 模型对我国股票市场的特征进行刻画, 但对波动率预测效果的研究较少。魏巍贤等 (1999) 比较了标准 GARCH, QGARCH 以及 GJR 模型对日波动率(inter-day volatility) 的预测效果, 发现 QGARCH 的预测力要胜出一筹, 但用日收益率的平方作为事后波动率, 显然不可信<sup>[7]</sup>。徐正国和张世英 (2004) 运用调整的已实现波动率, 发现 ARFIMAX 模型的预测能力最好, 但文中没有全面比较 GARCH 类模型<sup>[8]</sup>。周林 (2006) 运用已实现波动率, 并借助损失函数来衡量预测能力, 发现

收稿日期: 2007- 06- 10; 修订日期: 2007- 11- 05

作者简介: 黄海南(1979- ), 男(汉族), 湖南人, 北京师范大学金融研究中心, 博士, 研究方向: 金融工程1

GARCH 类模型的预测能力并不好,但作者没有用高频数据来估计已实现波动率<sup>[9]</sup>。因此,国内在波动率预测效果的研究方面还欠缺全面的比较和正确的衡量,还不足以得出 GARCH 类模型的波动率预测能力是否优劣的结论。基于这些认识,本文运用已实现波动率来度量事后波动率,分别运用 M- Z 回归及损失函数,全面评价 GARCH 类模型在我国股票市场收益波动率的预测表现。

## 2 模型与数据

### 2.1 GARCH 类模型介绍

#### 2.1.1 ARCH 模型

条件异方差模型可以分为两类,第一类是用确定的函数来刻画,如 ARCH 模型;第二类是用随机方程来描述,如随机波动率模型。ARCH 模型称为自回归条件异方差 (Autoregression Conditional Heteroskedasticity) 模型,是由 Engel(1982)发展起来的<sup>[10]</sup>。ARCH(q)模型可表示如下:

$$E_t = z_t R_t \quad z_t \sim i.i.d. D(0, 1)$$

$$R_t^2 = X_t + \sum_{i=1}^q A E_{t-i}^2 \quad (1)$$

其中,  $D(\cdot)$  表示有着零均值单位方差的概率密度函数。

#### 2.1.2 GARCH 模型

GARCH 模型称为广义 ARCH 模型,是 ARCH 模型的拓展,由 Bollerslev (1986)发展起来的<sup>[11]</sup>。GARCH(p, q)表示如下:

$$R_t^2 = X_t + \sum_{i=1}^q A E_{t-i}^2 + \sum_{j=1}^p B R_{t-j}^2 \quad (2)$$

但无论是 ARCH 还是 GARCH,只能反映金融时间序列的厚尾 (heavy tail) 和波动聚集 (volatility clustering) 现象,不能反映非对称性 (asymmetry)。

#### 2.1.3 EGARCH 模型

为了克服 GARCH 模型在处理金融时间序列时的一些弱点, Nelson(1991)提出了指数 GARCH (EGARCH) 模型, Bollerslev 和 Mikkelsen (1996)重新表达了 EGARCH 模型<sup>[12]</sup>:

$$\ln R_t^2 = X_t + [1 - \beta(L)]^{-1} [1 + \alpha(L)] g(z_t) \quad (3)$$

为了允许在模型中体现正负资产收益率的非对称效应,考虑加权的新息:

$$g(z_t) = H z_t + H [|z_t| - E |z_t|] \quad (4)$$

$E |z_t|$  的具体形式依赖于  $z_t$  的分布假设,详细讨论见 Tsay(2002)。

### 2.1.4 GJR- GARCH 模型

由 Glosten, Jagannathan, 和 Runkle (1993)发展起来的 GJR- GARCH 模型是另一个反映非对称性的模型<sup>[13]</sup>。模型的形式如下:

$$R_t^2 = X_t + \sum_{i=1}^q (A E_{t-i}^2 + G S_{t-i} E_{t-i}) + \sum_{j=1}^p B R_{t-j}^2 \quad (5)$$

模型表明,正的残差与负的残差对条件方差的影响是不一样的。

### 2.1.5 APARCH 模型

Ding, Granger, 和 Engle (1993)发展了更为灵活的 APARCH 模型,来捕捉金融时间序列的非对称性<sup>[14]</sup>。

$$R_t^2 = X_t + \sum_{i=1}^q A (|E_{t-i}| - G E_{t-i}) D_t + \sum_{j=1}^p B R_{t-j}^2 \quad (6)$$

其中,  $D > 0, -1 < G < 1 (i = 1, \dots, q)$ 。事实上,ARCH 模型 GARCH 模型 GJR 模型都是 APARCH 模型的特例。

## 2.2 预测效果的检验

### 2.2.1 波动率的预测

以 GARCH 模型为例,向前 h 步预测的条件方差为:

$$R_{t+h|t}^2 = X_t + \sum_{i=1}^q A E_{t+h-i|t}^2 + \sum_{j=1}^p B R_{t+h-j|t}^2 \quad (7)$$

根据上式,通过递归方法,可求出 h 步的条件方差。其他 GARCH 类模型的预测方法类似。

### 2.2.2 计算已实现波动率

由于我们感兴趣的是收益率,而不是股价,因此需要把股价转换成收益率。

$$r_t = \ln P_t - \ln P_{t-1} \quad (8)$$

利用对数收益率的性质,我们有:

$$r_t = \sum_{k=1}^K r_{k,t} \quad (9)$$

其中,  $r_{k,t}$  是第 t 天的第 k 个 5 分钟的收益率, K 是每天的 5 分钟的间隔次数。进一步假设条件方差和协方差是存在的,则

$$\text{Var}(r_t | F_{t-1}) = \sum_{k=1}^K \text{Var}(r_{k,t} | F_{t-1}) + 2 \sum_{k < j} \text{Cov}(r_{k,t}, r_{j,t} | F_{t-1}) \quad (10)$$

其中  $F_{t-1}$  是 t-1 月的已知信息,假定  $r_{k,t}$  是白噪声序列,则

$$\text{Var}(r_t | F_{t-1}) = n \text{Var}(r_{1,t}) \quad (11)$$

$\text{Var}(r_{1,t})$  可以用高频数据来估计:

$$\hat{R}^2 = \frac{n}{n-1} \sum_{k=1}^K (r_{k,t} - r_t)^2 \cup \sum_{k=1}^K r_{k,t}^2 \quad (12)$$

在以上近似中,我们假定样本均值为0。

### 2.1.2.3 预测效果检验: M- Z 回归

检验 ARCH 类模型的预测效果的最流行的方法是 M- Z 回归, 回归方程为:

$$\hat{R}^2 = a + B\hat{R}^2 + u_t \quad (13)$$

其中  $\hat{R}^2$  是事后波动率, 我们用已实现波动率来衡量,  $\hat{R}$  是预测的波动率。如果模型设定正确, 则  $E(\hat{R}) = \hat{R}$ , 那么要求  $a = 0$ , 以及  $B = 1$ 。而回归的拟合优度通常被用来测度 ARCH 类模型的预测效果。

### 2.1.2.4 预测效果检验: 损失函数

损失函数是一种更为传统的评价指标。尽管 Bollerslev et al (1994), Diebold, Lopez (1996) 和 Lopez (2001) 等指出, 与 M- Z 回归相比, 损失函数并不见得更适合来评价波动率的预测效果, 但仍不失为一种好的衡量方法。<sup>[15]</sup> 我们采用以下 MSE (Mean Squared Error), MAE (Mean Absolute Error), RMSE (Root Mean Squared Error) 以及 MAPE (Mean Absolute Percentage Error) 等 4 个损失函数来检验预测效果。各个损失函数的方差越小, 表明预测误差越小, 预测就越准确。

$$\begin{aligned} \text{MSE} &= \frac{1}{N} \sum_{t=1}^N (\hat{R}^2 - \hat{R}^2)^2 \\ \text{MAE} &= \frac{1}{N} \sum_{t=1}^N |\hat{R}^2 - \hat{R}^2| \\ \text{RMSE} &= \left[ \frac{1}{N} \sum_{t=1}^N (\hat{R}^2 - \hat{R}^2)^2 \right]^{1/2} \\ \text{MAPE} &= \frac{1}{N} \sum_{t=1}^N |1 - \hat{R}\hat{R}| \end{aligned} \quad (4)$$

## 2.3 数据

我们采用的样本是上证指数的收盘价。不同的交易制度下市场的波动会产生显著的不同, 不宜跨不同的交易制度选择数据。从 1996 年 12 月 16 日至今, 中国股市实行有 10% 涨跌停板限制的 T+1 交易制度, 故本文的样本其间从 1997 年 1 月 2 日起, 到 2006 年 9 月 29 日止。其中, 1997 年到 2004 年 12 月 31 日为日度数据, 共 1928 笔数据, 根据式 (9) 计算出日收益率, 用于模型估计。从 2005 年 1 月 2 日到 2006 年 6 月 31 日为 5 分钟的高频数据, 共 337 个交易日, 上海证券交易所每天的交易时间为上午 9:30 至 11:30 下午 13:00 至 15:00。所以共有  $337 \times 48 = 16176$  笔数据, 首先根据式 (9) 计算

出 5 分钟的收益率, 其中, 每天的第一个日内收益率为 9:35 的股价对数减去上一个交易日的 15:00 的股价对数。最后根据式 (12) 计算每日的已实现波动率。

## 3 实证分析结果

### 3.1 估计与检验结果

#### 3.1.1 ARCH 效应检验

Engel (1982) 提出了用拉格朗日乘子 (LM test) 来检验序列是否存在 ARCH 效应, 其零假设是不存在 ARCH 效应。检验结果是,  $V^2(1) = 631920 [010000]$ ,  $V^2(2) = 931135 [010000]$ ,  $V^2(3) = 1181893 [010000]$ , 小括号内为滞后阶数, 对应于自由度, 方括号中为 p 值。结果表明, p 值很小, 即显著拒绝零假设, 说明上证指数的收益率具有显著的 ARCH 效应。检验结果对滞后阶数不敏感, 事实上, 滞后阶数的选择是不重要的 (范剑青, 姚琦伟, 2005)<sup>[16]</sup>。

#### 3.1.2 拟合具有高斯误差的 GARCH 模型

我们用条件极大似然法来估计 GARCH 模型。极大似然法必须对残差的分布做出假设, 我们首先假定残差服从高斯分布, 即  $z_t \sim i.i.d. N(0, 1)$ 。分别对 GARCH, EGARCH, GJR, APARCH 等 4 个模型进行估计, 运用 AIC 确定模型的阶数, 估计结果见表 1。

结果表明, 除了 EGARCH 外, 其余模型的系数都在 1% 的显著性水平上显著。如果根据 AIC 来选择具体的模型, 应该选择 GJR(2, 1)。但还必须做出一系列的检验。

首先用 Ljung-Box Q 统计量检验标准化残差的相关性。四个模型的 Q 统计量所对应的 p 值都很大, 说明四个模型的标准化残差都没有前后相关性。

其次用 Engle 和 Ng (1993) 提出了指示偏误检验 (Sign Bias Test, SBT) 来检验模型的设定是否有误。指示偏误检验考察模型之外的其他可观测量是否能够预测波动率, 如果能够预测到, 则模型设定是有误的。此外, 还有两个类似的检验, 负面偏误检验 (Negative Size Bias Test) 以及正面偏误检验 (Positive Size Bias Test) 强调负的 (或者正的) 冲击对波动率的影响, 同时对以上三个检验共同组成联合检验 (Joint Test)。从表 1 可以看出, 四个模型的所有检验都不显著, 说明模型以外的可观测量不能预测波动率, 意味着模型设定是正确的, 因此, 指

示偏误检验不能帮助我们选择适当的模型。

表 1 高斯分布下的拟合结果

	GARCH(1, 2)	EGARCH(1, 2)	GJR(2, 1)	APARCH(2, 1)
常数	0.072385*	1.064123*	0.120436*	0.062343*
Alpha1	0.249365*	- 3.473635	0.160363*	0.200091*
Alpha2	- 0.121644*	8.849177		
Beta1	0.851302*	0.961996*	0.337453*	0.443784*
Beta2			0.415666*	0.366431*
Theta1		- 0.005281		
Theta2		0.042281		
Gamma1			0.102222*	0.152105*
Delta				0.936940*
对数似然函数	- 3353.090	- 3339.270	- 3349.133	- 3336.517
AIC	3.482458	3.470197	3.479391	3.467342
Ljung- Box Q(20)	20.7899 [0.409591]	19.9748 [0.1459504]	20.9715 [0.39881]	20.1254[0.45011]
指示偏误检验	0.56399	0.76000	0.61478	0.98442
负面偏误检验	0.43009	0.49462	0.30043	0.71017
正面偏误检验	1.22010	0.54958	0.40572	0.60996
联合检验	1.78143	0.66953	0.39593	1.10190
泊松拟合优度	159.6141*	147.1037*	145.2365*	126.9378*

\* 表示在 1% 的显著性水平上显著; 方括号中为 p 值。

最后用调整的泊松拟合优度(adjusted Pearson goodness-of-fit test)来检验残差的经验分布与理论分布(我们所假定的分布)是否一致。详细讨论见 Palm 和 Vlaar(1997)<sup>[17]</sup>。我们清楚地看到, 四个模型的泊松拟合优度检验都在 1% 的显著性水平上拒绝零假设, 即残差服从高斯分布的假设与经验分布是不一致的。因此, 我们考察残差服从其他分布的情况。

### 3.1.1.3 拟合具有 t- 分布误差的 GARCH 模型

基于以上分析, 我们假定残差服从 t- 分布, 即  $z_t \sim i. i. d. t_d$ , 其中 d 为自由度, 运用条件极大似然方法拟合 GARCH 类模型, 同样运用 AIC 确定模型

的阶数, 估计结果见表 2。

4 个模型中, AIC 最小的是 EGARCH(1, 1), 但 EGARCH(1, 1) 的系数出了常数项和 Beta1 之外都不显著, 其余三个模型中, APARCH(1, 1) 的 AIC 最小。Ljung-BoxQ 统计量检验的结果是所有模型的标准化残差都没有前后相关性。四个模型的指示偏误检验都不显著。四个模型的 t 分布为 4.6 到 5 之间, 且在 1% 的显著性水平上显著, 说明确实存在厚尾现象, 同时泊松拟合优度检验都不能拒绝零假设, 说明残差的 t 分布假设是正确的。综合来看, 在 t 分布残差假设下, 应该选择 APARCH(1, 1) 模型。

表 2 t 分布下的拟合结果

	GARCH(1, 1)	EGARCH(1, 1)	GJR(1, 1)	APARCH(1, 1)
常数	0.097212*	2.110559*	0.086323*	0.055931*
Alpha1	0.140604*	5.931894	0.076028*	0.147561*
Beta1	0.826614*	0.957737*	0.830617*	0.854208*
Theta1		- 0.001346		
Theta2		0.004374		
Gamma1			0.135663*	0.315035*
Delta				1.096309*
对数似然函数	- 3259.014	- 3242.024	- 3249.468	- 3243.350
AIC	3.384870	3.369320	3.376004	3.370695
Ljung- Box Q(20)	20.9221 [0.401732]	20.7804 [0.1410157]	20.7636 [0.411162]	20.1661 [0.447585]
指示偏误检验	0.62203	0.70514	0.53332	0.89845
负面偏误检验	1.70630	0.83706	0.66694	1.00114
正面偏误检验	0.14796	0.59649	0.62987	0.49364
联合检验	3.08932	1.70663	1.34009	2.04522
t 分布自由度	4.646*	4.906*	4.732*	4.896*
泊松拟合优度	41.9585	40.2988	45.9834	32.0000

\* 表示在 1% 的显著性水平上显著; 方括号中为 p 值。

### 3.1.1.4 拟合具有偏斜 t- 分布误差的 GARCH 模型

虽然泊松拟合优度检验都不能拒绝 t 分布残差假设, 但为了全面比较预测能力, 我们拟合具有偏斜 t- 分布 (Skewed t- distribution) 误差的 GARCH 模型。因为 t- 分布仅反映了序列的厚尾 (fat tails) 特征, 不能反映序列的非对称性 (asymmetry), 而偏斜 t- 分布是既能反映厚尾特征, 又能反映非对称性。估计的方法以及模型阶数的确定与前面一样, 拟合结果见表 3。

从表 3 看, 根据 AIC, 考虑系数的显著性, 我们应该选择 APARCH(1, 1) 模型。Ljung- Box Q 统计量检验、指示偏误检验无助于我们识别各种模型。

尽管四个模型的泊松拟合优度检验都不能拒绝零假设, 但只有 GARCH(1, 1) 和 GJR(1, 1) 的偏态系数 (Asymmetry) 在 10% 的显著性水平上显著, EGARCH(1, 1) 与 APARCH(1, 1) 的偏态系数不显著, 原因是 APARCH 模型已经捕捉了收益的非对称性。因此, 偏斜 t- 分布假设不太合适。

综合以上各种估计及检验结果, t 分布假定下的 APARCH(1, 1) 模型最合适。但 t 分布以及偏斜 t- 分布假设下的 GARCH(1, 1), GJR(1, 1) 也是可以接受的。

表 3 偏斜 t 分布下的拟合结果

	GARCH(1, 1)	EGARCH(1, 1)	GJR(1, 1)	APARCH(1, 1)
常数	0.094310*	0.170480	0.084342*	0.055089*
Alpha1	0.141375*	10.000000	0.077356*	0.148633*
Beta1	0.827685*	0.958462*	0.831299*	0.854066*
Theta1		- 0.000796		
Theta2		0.002612		
Gamma1			0.134045*	0.307934*
Delta				1.103474*
对数似然函数	- 3257.431	- 3241.699	- 3248.065	- 3242.064
AIC	3.384265	3.370020	3.375587	3.370398
Ljung- Box Q(20)	20.9083 [0.402548]	20.3464 [0.436458]	20.7543 [0.411717]	20.1408 [0.449153]
指示偏误检验	0.61524	0.72709	0.53203	0.88915
负面偏误检验	1.67589	0.81068	0.65888	0.97809
正面偏误检验	0.16716	0.55687	0.60101	0.46213
联合检验	2.99363	1.65251	1.27213	1.93705
非对称性	- 0.048766***	- 0.046942	- 0.046521***	- 0.044963
厚尾	4.615706*	4.866258*	4.706847*	4.882046*
泊松拟合优度	34.7801	35.1950	42.9544	37.6846

\*, \*\*\*, 分别表示在 1%, 10% 的显著性水平上显著; 方括号中为 p 值。

### 3.2 预测效果的评价

为了比较预测效果, 我们依据以上 3 个模型的 5 种不同的估计结果预测波动率, 进而比较各种不

同模型的预测表现。根据式 (7), 我们进行 1 步预测, 得到各种估计的预测波动率。根据式 (13) 的 M - Z 回归以及式 (14) 的损失函数, 检验结果见表 4。

表 4 GARCH 类模型预测效果的评价

	t 分布			偏斜 t- 分布	
	GARCH(1, 1)	GJR(1, 1)	APARCH(1, 1)	GARCH(1, 1)	GJR(1, 1)
M- Z 回归					
Alpha	- 0.108111	- 0.469552**	- 0.444997**	- 0.096186***	- 0.456086**
Beta	0.778550*	1.022980*	0.992290*	0.771296*	1.014193*
R2	0.307312	0.294044	0.270892	0.306987	0.295008
损失函数					
MSE	2.454	2.29	2.389	2.465	2.29
MAE	1.114	1.058	1.084	1.116	1.059
RMSE	1.566	1.513	1.546	1.57	1.513
MAPE	1.242	1.156	1.16	1.241	1.155
泊松拟合优度	55.7596**	56.4718*	42.2285	47.6884***	43.1780

\*, \*\*, \*\*\*, 分别表示在 1%, 5%, 10% 的显著性水平上显著。

在 M- Z 回归中, R<sup>2</sup> 普遍较低, 大部分实证研究中, R<sup>2</sup> 不到 10%。所以一般认为, 在 M- Z 回归中, R<sup>2</sup> 能达到 30- 40%, 即认为预测效果较好。详

细讨论见 Andersen and Bollerslev (1998)。从 M- Z 回归结果来看, 所有模型的 R<sup>2</sup> 大约为 0.3, 说明模型的预测效果较好。t- 分布假设下的 GJR(1, 1)

APARCH(1, 1) 以及偏斜 t- 分布假设下的 GJR(1, 1) 的 Beta 系数接近于 1, 且在 1% 的显著性水平上显著, 尽管 Alpha 在 5% 的显著性水平上显著不为零。从拟合优度来看, t 分布以及偏斜 t- 分布假设下的 GARCH(1, 1) 好于其他三个模型, 但其 Beta 系数显著不为 1。综合来看, t 分布以及偏斜 t- 分布假设下的 GJR(1, 1) 模型的预测能力较好。

从损失函数看, 表中所示的是 MSE, MAE, RMSE 以及 MAPE 等 4 个损失函数的方差, t 分布以及偏斜 t- 分布假设下的 GJR(1, 1) 模型的各个损失函数的方差较小, 说明其预测能力较好。

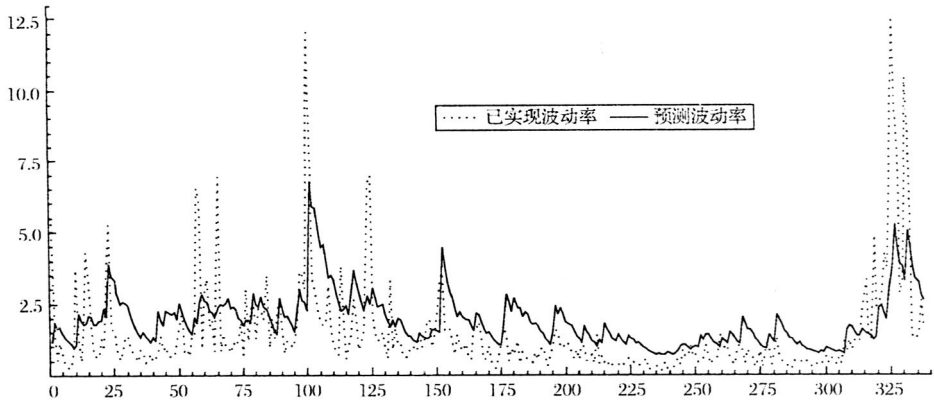


图 1 已实现波动率 vs 预测波动率

#### 4 结语

本文在高斯分布、t 分布、偏斜 t 分布等三种不同的残差分布假定下分别对 GARCH, EGARCH, GJR, APARCH 等 4 个模型进行了估计和检验, 然后对各种合适的估计结果进行了预测。本文采用 5 分钟的高频数据计算出已实现波动率, 并以此作为衡量标准, 通过 M- Z 回归和损失函数等方法对预测的效果进行了检验, 从而得到如下结论: <sup>1</sup> M- Z 回归的结果表明, 拟合优度接近 30%, 而且 Beta 系数接近于 1, 说明 GARCH 类模型的样本外预测效果较好; <sup>o</sup> 不同的 GARCH 模型的预测效果有所差别。本文的实证分析表明, 偏斜 t- 分布假设下的 GJR(1, 1) 模型的预测效果最好, 其次是 t- 分布假设下的 GJR(1, 1) APARCH(1, 1) 模型。但无论残差假定是 t 分布还是偏斜 t- 分布, GARCH(1, 1) 预测效果都较差。

需要指出得是, 以已实现波动率为衡量标准来评价模型的预测能力, 其前提条件是要运用高频数据计算出已实现波动率, 因此对数据的频度要求比

最后, 我们同样运用调整的泊松拟合优度检验标准化预测残差的经验分布与理论分布的一致性。从这个指标看, t- 分布假设下的 APARCH(1, 1) 以及偏斜 t- 分布假设下的 GJR(1, 1) 不能显著拒绝零假设, 即分布假设是合适的。

M- Z 回归结果与损失函数的检验结果是一致的, 即 t 分布以及偏斜 t- 分布假设下的 GJR(1, 1) 模型的预测能力最好。但与泊松拟合优度检验不一致。综合来看, 偏斜 t- 分布假设下的 GJR(1, 1) 模型的预测效果最好 (见图 1), 其次是 t- 分布假设下的 GJR(1, 1) APARCH(1, 1) 模型。

较高。此外, 从图 1 可看出, 即便是预测能力最强的 GJR(1, 1) 模型, 也难以预测特别大的损失, 这是所有时间序列建模者们面临的巨大挑战。

#### 参考文献:

- [1] Bollerslev, TI, RI YI Chou, K. F. Kroner. ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence [J]. Journal of Econometrics, 1992, 52: 5- 59.
- [2] Mincer, J. and V. Zarnowitz. The Evaluation of Economic Forecasts [J]. in J. Mincer, ed., Economic Forecasts and Expectations (New York: National Bureau of Economic Research), 1969.
- [3] Cumby R., S Figlewski, and J. Hasbrouck. Forecasting Volatility and Correlations with EGARCH Models [J]. Journal of Derivatives, 1993, Winter: 51- 63
- [4] Tsay, Ruey S. . Analysis of Financial Time Series [M]. John Wiley & Sons, 2002: 112- 113
- [5] Andersen T, Bollerslev T. . Towards a unified framework for high and low frequency return volatility modeling [J]. Statistical Neerlandica, 1998, 52(3): 273 - 302.

- [ 6 ] Andersen T. G. , Tim Bollerslev Francis Diebold, and P. Labys. Modeling and Forecasting Realized Volatility [ J ]. *Econometrica*, 2003, 71(2): 579- 625.
- [ 7 ] 魏巍贤, 周晓明. 中国股票市场波动的非线性 GARCH 预测模型[ J ]. *预测*, 1999, ( 5 ): 47- 49.
- [ 8 ] 徐正国, 张世英. 调整/ 已实现波动率与 GARCH 及 SV 模型对波动的预测能力的比较研究[ J ]. *系统工程*, 2004, ( 8 ): 60- 63.
- [ 9 ] 周林. 股票波动率模拟及预测效果的实证研究[ J ]. *上海经济研究*, 2006, ( 12 ): 100- 110.
- [ 10 ] Engle RF. . Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. inflation [ J ]. *Econometrica*, 1982, 45: 987- 1007.
- [ 11 ] Bollerslev, T. . Generalized Autoregressive Conditional Heteroskedasticity [ J ]. *Journal of Econometrics*, 1986, 31: 307- 327.
- [ 12 ] Bollerslev, T. , H. O. Mikkelsen. Modeling and Pricing Long- Memory in Stock Market Volatility [ J ]. *Journal of Econometrics*, 1996, 73: 151- 184.
- [ 13 ] Glosten, L. , R. Jagannathan, and D. Runkle. On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks [ J ]. *Journal of Finance*, 1993, 48: 1779- 1801.
- [ 14 ] Ding, Z. , C. W. J. Granger, and R. F. Engle. A Long Memory Property of Stock Market Returns and a New Model [ J ]. *Journal of Empirical Finance*, 1993, 1: 83- 106.
- [ 15 ] Lopez JA. . Evaluation of predictive accuracy of volatility models [ J ]. *Journal of Forecasting*, 2001, 20(2): 87- 109.
- [ 16 ] 范剑青, 姚琦伟. 非线性时间序列: 建模、预报和应用 [ M ]. 北京: 高等教育出版社, 2005: 128.
- [ 17 ] Palm, F. , and P. Vlaar. Simple Diagnostics Procedures for Modeling Financial Time Series [ J ]. *Algebra meines Statistisches Archiv*, 1997, 81, 85- 101.

## Evaluation on Volatility Forecasting Performance of GARCH2Type Models

HUANG Ha2nan, ZHONG Wei

(Financial Research Centre, Beijing Normal University, Beijing 100875, China)

Abstract: GARCH- type models have been broadly used to forecast volatility . But it's ignored to evaluate the performance of volatility forecasting. The reason is mainly lack of appropriate benchmark to evaluate. We estimate and forecast the return of SZZS using GARCH- type models. Realized volatility is computed as benchmark using 5- minuets high frequency data. Volatility forecasting performance is measured using M- Z regression and loss function. The conclusion is that GARCH type models have a very good forecasting performance both in sample and out of sample, and GJR(1, 1) under skewed t- distribution assumption is the most powerful to forecast.

Key words: GARCH; realized volatility; M- Z regression; loss function