中国股票市场的波动率预测模型及其SPA检验

魏 宇 余怒涛
(西南交通大学经济管理学院,成都 610031;云南财经大学会计学院,昆明 650221)

摘 要:本文以上证综指的高频数据样本为例,全面探讨了各类历史波动率模型以及实现波动率模型的构建方法,同时采用滚动时间窗的样本外预测法,实证计算了不同模型假定下的指数波动率预测值,并运用基于自举法的 SPA 检验,评估了各种波动率模型对上证综指波动的预测精度。结果显示,基于 ARFIMA 的实现波动率模型以及随机波动模型(SV)具有最高的波动率预测精度,但在加入实现波动率作为附加解释变量以后,并未显著提升标准 SV 和 GARCH 模型的预测能力。

关键词:高频数据;实现波动率;随机波动模型;GARCH 模型;SPA 检验
JEL 分类:C22;C53 **文献标识码**:A **文章编号**:1002-7246(2007)07-0138-13

一、引言及文献回顾

对金融资产收益的波动率(Volatility)描述是金融学理论的核心内容之一。有关波动率大小的测度(Measurement)及其动力学特征(Dynamics)的刻画,对于金融衍生产品的定价以及金融风险的测度和管理而言,都具有极其重要的理论和实际意义。

目前对金融资产收益波动率的模型描述主要有以下三种类型:(1)第一种是一类被称之为历史波动模型(Historical volatility models,简记为 HV)的描述方法,这类波动率模型的构建基于历史收益数据,并且这些历史收益数据的时间标度(Time scales)一般较长,通常为每日、每周甚至是每月。在历史波动率模型中比较有代表性是 Engle(1982)的自回归条件异方差模型(ARCH)、Bollerslev(1986)的广义自回归条件异方差模型(GARCH)以及 Taylor(1986)的随机波动模型(Stochastic volatility model,简记为 SV)。(2)第二类对波动率的描述方法来源于期权价格数据(Option pricing data),这类波动率模型被称为隐

收稿日期:2007-02-07
作者简介:魏 宇(1975-),男,管理学博士,西南交通大学经济管理学院金融系副教授。
余怒涛(1975-),硕士,讲师,供职云南财经大学会计学院。
*作者感谢国家自然科学基金(70501025)的资助,同时感谢斯坦福大学经济系的 P. R. Hansen 博士为本文提供的 SPA 检验程序以及众多宝贵建议,感谢匿名评审人的建设性建议,当然文中错误由作者负责。

含波动率模型(Implied volatility model,简记为 IV)。(3)第三种波动率的描述方法称为实现波动率模型(Realized volatility model,简记为 RV),Andersen and Bollerslev(1998)首次指出,传统上用日收益率的平方(Squared daily return)作为日波动率(Daily volatility)的测度将会面临非常严重的测量误差和噪声(Measurement error and noise),而使用基于交易日内高频收益数据(Intradaily high - frequency return)的实现波动率(RV)作为日波动率的测度,将大大降低这些误差和噪声对真实潜在波动率过程(Underlying volatility process)的影响,并且随着高频收益频率数据的增加,这种测量的误差将会越来越小。但是,由于市场微观结构效应(Market microstructure effects)的影响,在实际运用当中,也并非高频收益率数据的频率越高越好。

从上面的讨论可以看出,对金融市场波动率的刻画可以有多种不同的模型描述方法,但是,究竟哪一种模型才是最适合中国股票市场的实际波动特征和风险状况呢? 这一问题的回答,对目前我国股市的风险监管,以及我国金融衍生产品市场的建立和发展中的很多基础性问题(比方说,如何对以中国股指为标的资产的金融期权进行精确定价等等),都具有非常重要的理论和现实意义。

近年来,国内学者对这一问题也进行了一些有益的探索。张永东,毕秋香(2003)实证比较了上海股市波动性预测模型的精度,结果表明当采用不同的预测误差统计量作为预测精度评价标准时,将会导致评价结果的显著差异,且常用的 GARCH(1,1)模型对上海股市波动性的预测效果并不理想。刘凤芹,吴喜之(2004)探讨了 SV 模型和 GARCH 模型对深圳股市波动率的预测表现。他们的研究也发现,GARCH 模型的表现明显差于 SV 模型,且 GARCH 模型的预测效果不稳定,随评价准则的不同而有显著差别。徐正国,张世英(2004)对比了实现波动率模型、SV 模型和 GARCH 模型对上海股市的波动率描述能力。他们的结果显示,实现波动率模型比 SV 和 GARCH 模型具有更高的波动率刻画精度。于亦文(2006)运用上证综指的高频数据样本,比较了实现波动率模型与 GARCH(1,1)模型对上海股市波动的解释能力,其结果同样表明,实现波动率模型比传统的 GARCH 模型提供了更好的波动率拟合。上述这些研究工作对深入探索我国股市的波动特征和风险状况奠定了坚实的实证基础。

但需要指出的是,现有研究当中仍然存在一些明显的不足和值得进一步深入的研究方向。(1)大多数研究只对比了基于低频数据的 GARCH 族模型和 SV 模型的波动率预测表现,还很少见到针对高频数据的实现波动率模型(RV)和基于低频数据的历史波动率模型(HV)的综合对比结果;(2)现有研究都是基于单一静态样本的样本内(In - the - sample)预测结果来进行的模型预测能力检验,还没有见到运用动态滚动时间窗(Rolling time windows)的样本外(Out - of - sample)预测结果来开展的相应讨论。而 White(2000)的研究指出,基于样本外预测结果的计量模型判定结论比基于样本内预测结果的模型判定更加可靠,且更具实用性;(3)现有研究采用的模型优劣判断方法大都不够严谨。Hansen and Lunde(2006)的研究表明,传统的基于单一样本的单一损失函数(Loss function)判断法(如常用的 MSE 和 MAE 等),往往会因为数据样本中的少数奇异点(Outliers)而严重影响损失函数的计算结果,进而可能导致对波动率模型优劣的错误判断。因此,就会出

现上述研究中普遍观察到的现象:当采用不同的损失函数作为模型预测精度的评价标准时,将会导致评价结果的显著差异(也就是说,在MSE标准下,也许模型甲更优,但在MAE标准下则模型乙的表现更好)。因此,已有研究结论的严谨性和稳健性(Robustness)是值得怀疑的。

基于以上认识,本文的创新点主要体现在以下三个方面:(1)我们全面对比了基于高频数据的实现波动率模型和基于低频数据的历史波动率模型对中国股市波动的预测能力,同时考虑加入RV作为附加的解释变量,对标准的SV和GARCH模型进行了扩展;(2)运用样本外的滚动时间窗预测法,计算了各类模型在1年左右时间上(250个交易日)的样本外波动率预测值;(3)进一步运用具有bootstrap特性的更加严谨和稳健的统计检验方法,即Hansen and Lunde(2005)提出的对波动率模型优劣判断的"高级预测能力检验法"(Superior predictive ability,简记为SPA),实证检验了不同波动率模型在中国股市中的适应范围和精确程度,以确保本文结论的稳健性和实用性。

当然,就中国股市情况,评判不同波动率模型的优劣,需要市场波动率的客观参考标准。显然,如上所述,传统上用日收益率的平方作为日波动率的测度标准显然已不合适。庆幸的是,Andersen等(2005)的研究指出,从根本上讲,由于潜在真实的市场波动率是不可观测的(Unobservable),因此目前公认的方法是用基于高频收益数据的实现波动率估计来作为真实市场波动率的代理变量(Proxy)和基准(Benchmark)。也就是说,在后文实证研究中,在判断模型对日波动率(Daily volatility)的预测精度高低时,主要是依据该模型计算的波动率预测值与实际估计得到的RV之间的"偏差","偏差"越小,则模型越好。显然,对"偏差"的定义不同,也可能会得到不同的模型优劣判断结果。

本文后面的结构安排是,第二部分计算上证综指的实现波动率RV,第三部分和第四部分分别介绍实现波动率模型和历史波动率模型,第五部分为预测方法及SPA检验,第六部分为实证结果,第七部分是本文的主要结论。

二、收益率描述以及实现波动率的估计

(一)数据说明以及收益率计算方法

本文研究的数据样本为上证综指(SSEC)从1999年1月19日到2003年3月31日的每5分钟高频数据(共N=1000个交易日①),记为$I_{t,d}$,$t=1,2,\cdots,N$,$d=0,1,2,\cdots,48$,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,数据来源于"中国经济研究中心(*CCER*)股票市场高频数据库"。上海证券交易所每个交易日9:30分开盘,到11:30分中午休市,然后13:00开盘,到15:00全天收盘,每天共有4个小时(即240分钟)连续竞价交易时间,因此,采用每5分钟记录一个数据的方法每天可以产生48个高频股价记录

① 本文选择的这一段样本区间包含了:1999年5月19日的5.19行情井喷;2001年6月14日国有股减持办法出台,引发股市单边大幅下挫;2001年10月22日,暂停国有股减持;2002年6月23日,停止国有股减持,引发6.24井喷等一系列重大事件。期间股市经历了一个较为完整的上升—下跌—再上升的周期,因此对我国股市的一个较为完整的波动周期具有较好的代表性。

(不包括 $I_{t,0}$),样本总体的高频数据量为 48000 个。文中的日收益率 R_t 利用相邻两个交易日的收盘价计算如下:

$$R_t = 100(\ln I_{t,48} - \ln I_{t-1,48}), t = 2,3,\mathrm{K},N \tag{1}$$

同理,本文定义第 t 天的(每 5 分钟)高频收益率 $R_{t,d}$ 为:

$$R_{t,d} = 100(\ln I_{t,d} - \ln I_{T,D-1}), d = 1,2,\mathrm{K},48 \tag{2}$$

(二)实现波动率的估计方法

根据 Andersen and Bollerslev(1998)的定义,对第 t 天的实现波动率的估计表示为第 t 天内的高频收益平方和,即:

$$RV_t' = \sum_{d=1}^{48} R_{t,d}^2 \tag{3}$$

但最近 Hansen and Lunde(2006)的研究又指出,由于股票市场并不象外汇市场那样在 24 小时连续进行交易,因此,我们能观察和记录到的高频股价数据只能反映有交易时段的(Active)市场波动状况,而无法包含无交易时段的(Inactive)市场波动信息(即股票市场从收盘到第二天开盘的所谓"Close - to - Open"波动率)。因此为了使实现波动率的估计更加准确地刻画全天的市场波动率大小,我们采用 Hansen and Lunde(2006)的建议,用某种尺度参数(Scale parameter)δ 来对 RV'进行尺度变换,即对第 t 天的实现波动率估计为:

$$RV_t = \delta RV'_t \tag{4}$$

其中:

$$\delta = \frac{N^{-1}\sum_{t=1}^{N} R_t^2}{N^{-1}\sum_{t=1}^{N} RV'_t} \tag{5}$$

表 1 是对日收益率 R_t、收益率平方 R_Ω、实现波动率估计 RV 以及对数 RV(简记为 lnRV)序列的描述性统计结果:

表 1　日收益率序列和实现波动率 RV 序列的描述性统计

	日收益率相关序列		实现波动率相关序列	
	R_t	R_t^2	RV	lnRV
均值(Mean)	0.027	2.301	2.310	0.116
标准差(Std.)	1.517	6.528	3.996	1.130
偏态系数(Skewness)	0.661***	7.730***	5.030***	0.367***
峰态系数(Kurtosis)	6.036***	77.376***	35.818***	0.445***
J - B	1590***	259421***	57672***	31***
Q (10)	9.101	108.512***	1857.223***	3298.378***
Q (20)	37.743***	152.634***	2500.586***	4901.682
Q (30)	60.214***	163.742***	2675.480***	5627.612***

* 说明:*** 代表在 1% 水平下显著,其中峰态系数 Kurtosis 为超额峰态,J - B 为 Jarque - Bera 统计量,Q(n)为滞后 n 期的 Ljung - Box Q 统计量。

从表1的描述性统计结果可以看到，所有序列都表现出明显的有偏(Skewed)和"尖峰胖尾"(Leptokurtic and fat tailed)特征，同时序列之间具有显著的自相关性(除了收益率序列 R_t 的Q(10)统计量不显著之外)，说明中国股票市场的波动较为剧烈，且市场波动具有较为显著的持续性(Persistence)或长期记忆性(Long - memory)特征。

三、实现波动率模型

Andersen 等(2001)的研究发现，在取自然对数(ln)以后，对数实现波动率(lnRV)的波动特征可以用高斯动力学过程(Gaussian dynamic process)来描述，同时 lnRV 展现出明显的长期记忆性特性。即，随着滞后期数的增加，lnRV 的相关性衰减速度要小于指数(Exponentially)衰减形式。为此，Andersen 等(2001)又建议采用自回归分整移动平均过程 ARFIMA 来描述 lnRV 的上述动力学特性。考虑到不同滞后阶数的 ARFIMA(p,d,q)模型对 lnRV 的估计结果非常接近，同时结合模型估计的 AIC(Akaike's Information Criterion)以及 BIC(Bayesian's Information Criterion)大小比较，我们这里采用 ARFIMA(1,d,1)模型来为 lnRV 建模。

我们进一步的考虑是，由于 ARFIMA(p,d,q)对不同类型时间序列的动力学特征具有很强的刻画能力，同时很多常见的时间序列模型如自回归模型 AR(p)或自回归移动平均模型 ARMA(p,q)等都可以视为 ARFIMA(p,d,q)的特例，因此，在后面的实证研究中，我们除了用 ARFIMA$(1,d,1)$来为 lnRV 建模之外，还考察了用 ARFIMA$(1,d,1)$来直接为 RV 本身建模，以增强实证结果的可对比性和可靠性。

ARFIMA(p,d,q)模型的一般形式为：

$$\Phi(L)(1-L)^d(Y-\mu)=\Theta(L)\varepsilon_t \tag{6}$$

其中 $\Phi(L)=1-\phi_1L-\phi_2L^2-\mathrm{K}-\phi_pL^p$，$\Theta(L)=1+\theta_1L+\theta_2L^2+\mathrm{K}+\theta_qL^q$ 分别为自回归滞后 p 阶算子以及移动平均滞后 q 阶算子，L 为滞后算子，$(1-L)^d$ 为分数差分算子，μ 是 Y 的均值(这里的 Y 指 RV 或 lnRV，以下对这两种波动率模型分别简记为 RV - ARFIMA 和 lnRV - ARFIMA)，同时这里假定 $\varepsilon_1 \sim NID(0,\sigma_\varepsilon^2)$。

四、历史波动率模型

(一)随机波动率模型(SV)及其扩展形式

Taylor(1986)提出了著名的随机波动模型，该模型的标准形式假定金融资产的收益率满足以下形式：

$$R_t=\mu_t+\varepsilon_t=\mu_t+\sigma_t z_t \tag{7}$$

其中 μ_t 是收益波动率的条件均值，σ_t^2 是条件方差，而假定新生量(Innovation)z_t 满足：$z_t \sim NID(0,1)$。同时由于收益率的条件均值一般很小，因此在我们的实证研究当中都假定其等于零。

与下面将要讨论的 GARCH 模型不同的是，SV 模型假定条件方差 σ_t^2 是不可观测的（Unobservable），且其服从以下的随机过程：

$$\sigma_t^2 = \sigma^{*2}\exp(h_t) \tag{8}$$

这里，真实的市场波动 σ_t^2 被假定为一个取值为正的标度因子（Scaling factor）σ^{*2} 与一种指数形式的随机过程 h_t 的乘积。其中，h_t 被认为是一种不可观测的对数波动率（log - volatility），且其满足：

$$h_t = \phi h_{t-1} + \sigma_\eta \eta_{t-1} \tag{9}$$

同时，假定 $\eta_t \sim NID(0,1)$，$h_1 \sim NID(0, \sigma_\eta^2/(1-\phi^2))$。

Koopman 等（2005）进一步指出，可以在 SV 模型中的对数波动率方程（9）中加入滞后一期的对数形式的隐含波动率（IV）或者实现波动率（RV）作为附加的解释变量，以增强模型对收益波动率的刻画能力。因此，在我们的实证研究当中，除了考虑了上面的标准 SV 模型以外，还实证分析了加入对数 RV 作为附加解释变量的扩展 SV 模型（以下简记为 SV - RV），其对数波动率 h_t 如下式所示：

$$h_t = \phi h_{t-1} + \gamma(1-\phi)\ln RV_{t-1} + \sigma_\eta \eta_{t-1} \tag{10}$$

需要说明的是，由于这里的 h_t 是对数形式的波动率，而非原始形式的市场波动，因此，公式（10）在引入 RV 作为附加解释变量时，要取其对数形式。公式（10）中的 γ 可以视为该附加解释变量的回归系数①。

（二）广义自回归条件异方差模型（GARCH）及其扩展形式

Bollerslev（1986）的广义自回归条件异方差模型（GARCH）是目前金融计量研究当中运用最广泛的波动率模型之一，该模型假定金融收益率的波动形式同（7）式。与 SV 模型不同的是，GARCH 模型认为条件方差 σ_t^2 是可观测的（Observable），其中运用最为普遍的 GARCH（1，1）模型则假定条件方差满足以下形式：

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{11}$$

与上面对 SV 模型的讨论类似，为了增强我们实证结果的可比性和可靠性，在后面的实证研究当中，我们同样考虑了在其条件方差（11）中加入滞后一期的 RV 作为附加解释变量的 GARCH 模型（以下简记为：GARCH - RV），如公式（12）所示：

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 = \gamma RV_{t-1} \tag{12}$$

显然，这里的附加解释变量 RV 是直接加入到真实的市场波动 σ_t^2 方程中去的，因此无需对其作任何形式上的变换，同理，γ 是需要估计的该附加解释变量的回归系数。

五、预测方法及 SPA 检验

（一）波动率预测方法说明

我们对上面讨论的 6 种波动率模型（即 *RV - ARFIMA*、*lnRV - ARFIMA*、*SV*、*SV - RV*、

① 同时，有关 SV 模型更深入的讨论可以参见 Shephard（1996）的研究，另外，对 SV 模型的实证估计方法中涉及的模拟极大似然估计以及 Kalman 滤波等技术可以参见 Sandmann and Koopman（1998）的研究。

GARCH 以及 *GARCH-RV*)进行滚动时间窗的"样本外预测能力检验"。具体方法如下:

(1)将数据样本总体($t=1,2,\cdots,N=1000$)划分为"估计样本"(Sample for estimation)和"预测样本"(Sample for predicting)两部分。其中,估计样本固定包含 $H=750$ 个交易日的数据,而预测样本包含最后250个交易日的数据(即 $t=H+1,H+2,\cdots,H+M$,其中 $M=250$)。

(2)第一步,我们选取 $t=1,2,\cdots,H$ 的数据作为第一个估计样本,分别对上述各种波动模型的参数进行估计,然后在此估计基础之上,获得未来1天的波动率预测,记为 $\hat{\sigma}^2_{H+1}$。也就是说,$\hat{\sigma}^2_{H+1}$ 是在前面750个样本数据的基础上对第751天的市场波动率估计。

(3)第二步,保持估计样本的时间区间长度不变($H=750$),将估计样本时间区间向后平行移动1天,即第2次选取的是 $t=2,3,\cdots,H+1$ 的数据样本作为新的估计样本,然后重新估计上述各类波动率模型的参数,并在此新的估计模型基础上获得未来1天的波动率预测,记为 $\hat{\sigma}^2_{H+2}$。

(4)同理,不断重复步骤(3),我们可以得到 $\hat{\sigma}^2_{H+3},\hat{\sigma}^2_{H+4},\hat{\sigma}^2_{H+5}\cdots$ 直到最后一次的估计样本区间为 $t=M,M+1,\cdots,H+M-1$,以获得对最后一天,即 $t=N=H+M=1000$ 的市场波动率预测 $\hat{\sigma}^2_{H+M}$。

对前述6类不同的波动率模型,分别重复进行250次模型估计①,每个模型都获得了250个未来1天的市场波动率估计,记为 $\hat{\sigma}^2_m, m=H+1, H+2,\cdots, H+M$。同理,我们记预测样本区间的实现波动率估计为 $RV_m, m=H+1,H+2,\cdots,H+M$($RV_m$ 的估计方法见第二节),并以此作为真实市场波动率的代理(Proxy),用以衡量各类波动率模型的预测精度。

(二)SPA检验方法说明

有了对市场波动的预测值 $\hat{\sigma}^2_m$ 以后,我们就可以比较模型预测值与真实市场波动率估计基准——RV_m 的偏差(或损失)究竟有多大了。然而,至于用哪一种损失函数作为衡量预测误差的标准最为合理,学术界尚未达成共识。Hansen and Lunde(2005)建议,可以尽可能多地采用不同形式的损失函数来作为预测模型精度的判断标准。基于这样的考虑,在我们的实证研究当中采用了4种不同的损失函数来分别作为各类波动率模型预测精度的评判标准。

这4种损失函数分别标记为 $L_i, i=1,2,3,4$,其中 L_1 和 L_2 分别称为平均误差平方(Mean squarederror,MSE)和平均绝对误差(Mean absolute error,MAE),它们是此类判断中最常用的两类损失函数形式,而 L_3 和 L_4 则分别是经异方差调整的MSE和MAE(Heteroskedastic adjusted MSE and MAE),其具体定义如下所示:

$$L_1: MSE = M^{-1}\sum_{m=H+1}^{H+M}(RV_m-\hat{\sigma}^2_m)^2 \tag{13}$$

① 即总共进行了6×250=1500次不同的模型估计,其中,由于对SV模型采用的是模拟极大似然估计法,因此对其估计耗时最多。在系统配置为奔腾4CPU和512M内存的计算机上,进行250次SV模型的估计总共耗时约为30分钟左右。

$$L_2: MAE = M^{-1}\sum_{m=H+1}^{H+M}|RV_m - \hat{\sigma}_m^2| \tag{14}$$

$$L_3: HMSE = M^{-1}\sum_{m=H+1}^{H+M}(1 - \hat{\sigma}_m^2/RV_m)^2 \tag{15}$$

$$L_4: HMAE = M^{-1}\sum_{m=H+1}^{H+M}|1 - \hat{\sigma}_m^2/RV_m| \tag{16}$$

需要指出的是，如果在一次实证研究中发现：采用某种 L_i 作为判断标准，得到了模型甲比模型乙的预测误差值小的话，那么我们只能判断："在这样一个特定的数据样本中，采用这一特定的损失函数 L_i 时，模型甲比模型乙的预测精确度高"。很明显，这一判断是不稳健的，且无法推广到其他类似的数据样本或者其他的损失函数判断标准。

为了解决这一问题，Hansen and Lunde(2005)提出了一种所谓的"高级预测能力检验法"(SPA)。他们的研究证明，因为采用了"自举法"(Bootstrap)，SPA 检验比类似的 White(2000)提出的 Reality Check(RC)检验法具有更加优异的模型判别能力，且 SPA 检验的结论具有更好的稳健性。也就是说，与基于一个单一样本的其它检验法相比，SPA 得到的检验结论更加可靠，且其得到的结论可以推广到其他类似的数据样本当中去。

SPA 检验过程如下：首先，假定我们有 $J+1$ 种类型的波动率模型，记为 $\mathbf{M}_k, k=0,1,\cdots,J$。每种波动率模型 $\mathbf{M}_k$ 得到的未来 1 天的波动率预测记为 $\hat{\sigma}_{k,m}^2$，其中 $m=H+1,H+2,\cdots,H+M$。对每一个预测值，我们都可计算公式(13)～(16)所定义的 4 种损失函数值，记为 $L_{i,k,m}$，其中 $i=1,2,3,4$。下面，用 $\mathbf{M}_0$ 表示作为 SPA 检验的基础模型(Base Model，即用该模型作为与其他模型的预测表现进行对比检验的基础)，因此，对于其它的 $k=1,2,\cdots,J$ 种波动率模型，我们可以计算其相对于基础模型 $\mathbf{M}_0$ 的"相对损失函数值"，记为：

$$X_{k,m} = L_{i,0,m} = L_{i,k,m} \tag{17}$$

现在需回答的问题是，在模型 $\mathbf{M}_k(k=1,2,\cdots,J)$ 中，是否存在比基础模型($\mathbf{M}_0$)表现更好的模型？为此，我们可以定义这样的零假设 $\mathbf{H}_0$："与对比模型 $\mathbf{M}_k$ 相比，基础模型 $\mathbf{M}_0$ 是表现最好的预测模型。"这一零假设可以用数学表达式表示为：

$$\max\lambda_k = E(X_{k,m}) \le 0, k=1,2,\mathrm{K},J \tag{18}$$

Hansen and Lunde(2005)证明了这一假设检验的检验统计量为：

$$T = \max\frac{\sqrt{\mathrm{MX_k}}}{\hat{\omega}_{\mathrm{kk}}}, \mathrm{k}=1,2,\mathrm{K},\mathrm{J} \tag{19}$$

其中：

$$\overline{X}_k = M^{-1}\sum_{m=H+1}^{H+M}X_{k,m}, \hat{\omega}_{kk}^2 = \mathrm{var}(\sqrt{M}\overline{X}_k) \tag{20}$$

为了获得公式(19)的 T 检验量的分布状况及其 p 值，Hansen and Lunde(2005)建议采用"自举法"(bootstrap)来获得。首先，我们需要获得一个长度为 M 的 $\mathrm{X_{k,m}}$ 新样本。要获得这样一个样本，则先要从 $\{X_{k,m}\}$ 的集合当中随机抽取一个新的子样本，而该子样本的长度来自一个服从均值为 q 的几何分布的随机数，同时控制这些子样本的组合长度为所要求的 M。

重复这样的 bootstrap 过程 B 次，可以获得 B 个长度为 M 的 $X_{k,m}$ 新样本，记为 $X^i_{k,m}$，$i=1,2,\cdots,B$。在我们后面的实证研究当中，选取 $q=0.5$ 和 $B=1000$ 次作为这一 bootstrap 过程的控制参数。对每一个 bootstrap 样本的均值表示为：

$$\bar{X}^i_k = M^{-1}\sum_{m=1}^{M} X^i_{k,m}, i=1,2,\mathrm{K},B \tag{21}$$

而所有 B 个 bootstrap 样本均值的方差估计表示为：

$$\hat{\omega}_{kk} = B^{-1}\sum_{i=1}^{B}(\bar{X}^i_k - \bar{\bar{X}}_k)^2, \bar{\bar{X}}_k = B^{-1}\sum_{i=1}^{B}\bar{X}^i_k \tag{22}$$

其次，定义 $\bar{Z}^i_k$ 为：

$$\bar{Z}^i_k = (\bar{X}^i_k - \bar{\bar{X}}_k) \times I\{\bar{\bar{X}}_k > -A_k\} \tag{23}$$

其中：

$$A_k = \frac{1}{4}M^{-4}\hat{\omega}_{kk} \tag{24}$$

而 $I\{\cdot\}$ 是一个指示函数，即当 $\{\}$ 中的条件成立时，其取值为 1，否则取值为 0。最后，可以得到如下的实证统计量：

$$T^i = \max\frac{\sqrt{M}\bar{Z}^i_k}{\hat{\omega}_{kk}}, i=1,2,\mathrm{K},B \tag{25}$$

Hansen and Lunde(2005)的研究表明，在(18)式所示的零假设条件下，公式(25)所示的实证统计量收敛于公式(19)所定义的统计检验 T。因此，该统计检验 T 的 p 值可以直接从下式得出：

$$p = B^{-1}\sum_{i=1}^{B} I\{T^i > T\} \tag{26}$$

简言之，*SPA* 检验的 p 值越大(越接近于1)，则表明越不能拒绝公式(18)所定义的零假设 H_0，说明基础模型的预测精度越高。

六、实证结果

(一)波动率模型预测结果

图 1(a)是 lnRV - ARFIMA 模型和 RV - ARFIMA 模型在预测样本区间(t = 751,752,…,1000)的波动率预测结果(分别用实线和虚线表示)，而对实际市场波动率测度的估计 RV 则用实心的小方块表示。类似地，图 1(b)和图 1(c)分别为 SV、SV - RV 以及 GARCH、GARCH - RV 模型在预测区间的波动率预测结果。

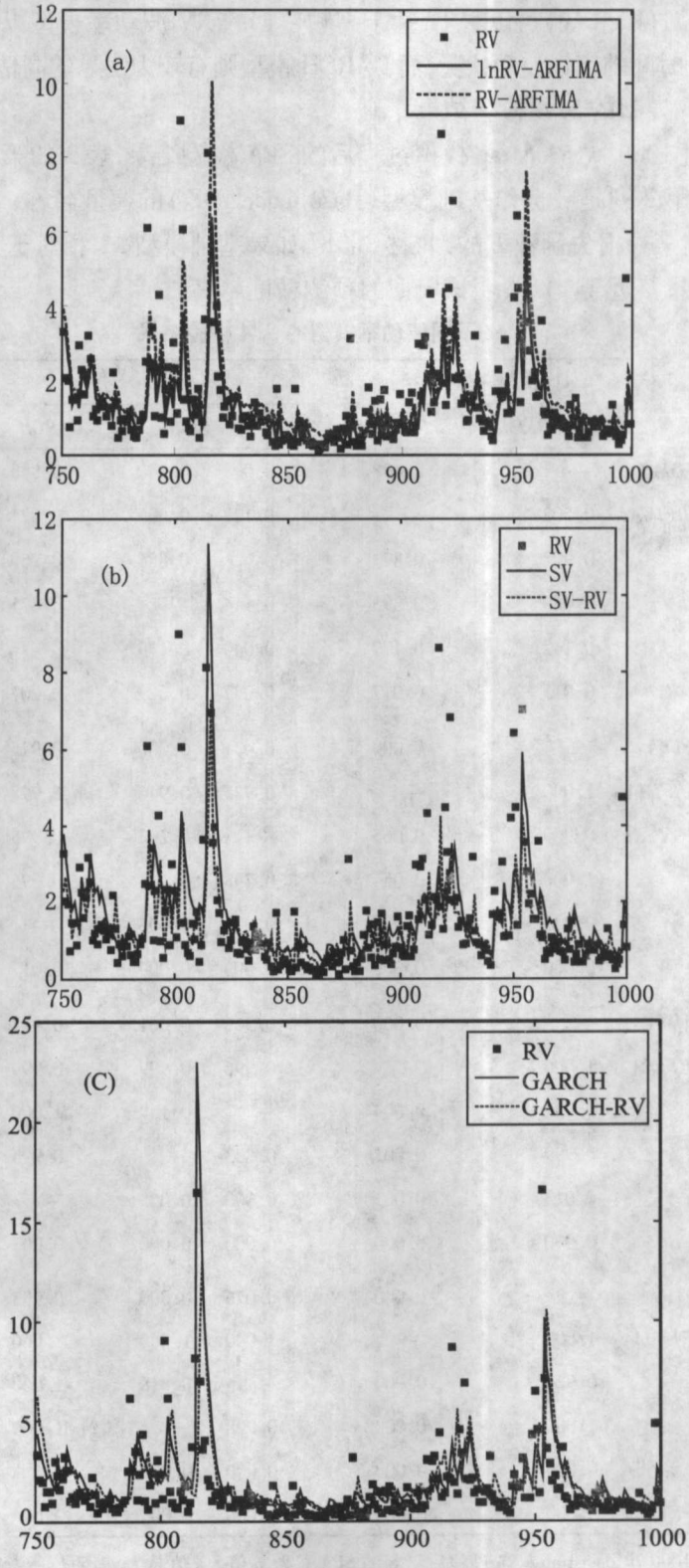

图 1　不同波动率模型在预测样本区间(t = 751,752,…,1000)的预测结果

比较图1(a)、(b)和(c)可以看出,基于高频数据的RV波动率模型以及SV模型都较好地预测了该段时间的市场波动状况,而GARCH模型则有较为明显的高估波动率的倾向。

(二)波动率模型的SPA检验结果

表2是经过1000次bootstrap模拟过程后的SPA检验结果,表2的第1列表示的是4种损失函数L_i,第2列是被选作基础模型(Base model,M_0)的模型名称,表中数字为SPA检验的p值。在某一损失函数L_i的判断标准下,如果基础模型M_0相对于其他模型的SPA检验p值越大(越接近于1),表明该模型的预测精度越高。

表2 不同波动率模型的*SPA*检验结果

		对比模型 M_k					
损失函数	基础模型 M_0	*RV-ARFIMA*	*lnRV-ARFIMA*	*SV*	*SV-RV*	*GARCH*	*GARCH-RVMSE*
MSE	*RV-ARFIMA*	-	0.498	0.093	0.540	0.858	0.985
	lnRV-ARFIMA	0.502	-	0.127	0.605	0.821	0.983
	SV	0.907	0.873	-	0.886	0.960	0.963
	SV-RV	0.460	0.395	0.114	-	0.821	0.971
	GARCH	0.142	0.179	0.040	0.179	-	0.793
	GARCH-RV	0.015	0.017	0.037	0.029	0.207	-
MAE	*RV-ARFIMA*	-	0.000	0.031	0.000	0.975	1.000
	INRV-ARFIMA	1.000	-	0.935	0.946	0.998	1.000
	SV	0.969	0.065	-	0.202	0.995	1.000
	SV-RV	1.000	0.054	0.798	-	0.995	1.000
	GARCH	0.025	0.002	0.005	0.005	-	0.983
	GARCH-RV	0.000	0.000	0.000	0.000	0.017	-
HMSE	*RV-ARFIMA*	-	0.000	0.366	0.001	0.987	0.999
	lnRV-ARFIMA	1.000	-	0.996	0.999	0.999	1.000
	SV	0.634	0.004	-	0.147	1.000	1.000
	SV-RV	0.999	0.001	0.853	-	0.995	0.999
	GARCH	0.013	0.001	0.000	0.005	-	0.970
	GARCH-RV	0.000	0.000	0.000	0.000	0.003	-
HMAE	*RV-ARFIMA*	-	0.000	0.076	0.000	0.995	1.000
	lnRV-ARFIMA	1.000	-	1.000	1.000	1.000	1.000
	SV	0.924	0.000	-	0.018	1.000	1.000
	SV-RV	1.000	0.000	0.928	-	1.000	1.000
	GARCH	0.005	0.000	0.000	0.000	-	0.996
	GARCH-RV	0.000	0.000	0.000	0.000	0.004	-

* 说明:表中数字为1000次bootstrap模拟过程后得到的SPA检验p值。p值越大,表明与所考察的对比模型M_k相比,基础模型M_0的表现越好。

从表 2 可以看出:(1)对中国股市而言,总体来说,lnRV - ARFIMA 是预测精度最高的波动率模型。这也表明,与日数据相比,高频数据当中确实蕴含着更加丰富的市场波动信息。(2)随机波动模型 SV 和加入 RV 作为附加解释变量的 SV - RV 模型的预测表现也相当不错,但加入 RV 作为解释变量以后,并没有明显提高 SV 模型对市场波动率的预测能力。(3)RV - ARFIMA 模型的预测精度表现并不突出,这也说明了 ARFIMA 模型没有很好地刻画 RV 本身的波动特征,但 RV - ARFIMA 模型的预测精度要显著高于 GARCH 模型。(4)几乎在所有的损失函数标准下,GARCH 和 GARCH - RV 模型都是表现最差的波动率模型,这也与图 1 所展示的实证结果吻合。且与 SV 模型类似,加入 RV 作为解释变量的 GARCH - RV 模型并未提升标准 GARCH 模型的预测能力。

七、主要结论

本文运用 Hansen and Lunde(2005)提出的对波动率模型预测能力的 SPA 检验法,实证考察了基于高频数据的 RV 波动率模型以及随机波动模型(SV)和 GARCH 模型对中国股市波动率的刻画和预测能力问题。

实证结果显示,就中国股市而言,历史波动率模型(特别是 GARCH 模型)对市场波动率的预测精度要明显低于基于高频数据的 RV 波动率模型。但我们同时也发现,在加入滞后期的 RV 作为附加解释变量以后,并未显著提升标准 SV 和 GARCH 模型的预测能力,这也与一些国外相关研究中的结论不符。因此,这也提醒我们,对适用于成熟资本市场的金融理论和模型,必须进行认真的比较和检验,才能判断其是否也适合我国股市这样的新兴资本市场的实际情况。

论文的检验方法和实证结果对于中国股票市场的风险管理以及即将全面推出的金融衍生产品(如股指期货和期权)的定价等问题都具有一定的理论和现实意义。当然,如何对研究中的一些与成熟资本市场结论不相吻合的实证结果进行理论解释,并进一步寻找更加适合中国股市实际风险状况和波动特征的波动率模型,仍然是我们下一步研究的主要方向。

参 考 文 献

[1]刘凤芹、吴喜之,2004:《基于 SV 模型的深圳股市波动的预测》,《山西财经大学学报》第 4 期,第 96 - 99 页。

[2]徐正国、张世英,2004:《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》,《系统工程》第 8 期,第 60 - 63 页。

[3]于亦文,2006:《实际波动率与 GARCH 模型的特征比较分析》,《管理工程学报》第 2 期,第 65 - 69 页。

[4]张永东、毕秋香,2003:《上海股市波动性预测模型的实证比较》,《管理工程学报》第 2 期,第 16 - 19 页。

[5]Andersen, T. G., and T. Bollerslev, 1998, "Answering the Skeptics: Yes, Standard Volatility Models Do Provide AccurateForecasts", *International Economic Review*39:885 - 905.

[6]Andersen, T. G., T. Bollerslevand N. Meddahi, 2005, "Correcting the Errors: Volatility Forecast Evaluation Using High Frequency Data and Realized Volatilities", *Econometrica*73: 279 - 296.

[7]Andersen, T. G., T. Bollerslev and F. X. Diebold, 2001, "The Distribution of Realized Stock Return Volatility", *Journal of Financial Economics*61:43 - 76.

[8] Bollerslev, T., 1986, "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*31:307 - 327.

[9] Engle, R. F., 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation", *Econometrica* 50: 987 - 1007.

[10] Hansen, P. R., and A. Lunde, 2005, "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1, 1)?", *Journal of Applied Econometrics* 20: 873 - 889.

[11] Hansen P. R., and A. Lunde, 2006, "Consistent Ranking of Volatility Models", *Journal of Econometrics* 131:97 - 121.

[12] Koopman, S. J., B. Jungbacker and E. Hol, 2005, "Forecasting Daily Variability of the S&P100 Stock Index Using Historical, Realized and Implied Volatility Measurements", *Journal of Empirical Finance* 12: 445 - 475.

[13] Sandmann, G., and S. J. Koopman, 1988, "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood", Journal of Econometrics87:271 - 301.

[14] Shephard, N., 1996, "Statistical Aspects of ARCH and Stochastic Volatility", In: Cox, D. R., Hinkley, D. V., Barndorff - Nielsen, O. E. (Eds.), Time Series Models in Econometrics, Finance and Other Fields, Number 65 in Monographs on Statistics and Applied Probability, Chapman and Hall, London.

[15] Taylor, S. J., 1986, "Modeling Financial Time Series", John Wiley and Sons, Chichester.

[16] White, H., 2000, "A Reality Check for Data Snooping", *Econometrica* 68:1097 - 1126.

Abstract: In this paper, one high-frequency dataset of the most important stock index in Chinese stock market is used to construct historical volatility models and realized volatility models. Based on Out-of-sample predicting results using rolling time windows method, we compare the predicting performance of different kinds of volatility models using bootstrapping SPA test. The empirical results show that, realized volatility models and stochastic volatility models are the best models for volatility forecasts in Chinese stock market. Furthermore with RV as additional explanatory variables, no improvements of predicting performance are found in SV and GARCH models.

Key Words: High-frequency Data; Realized Volatility; Stochastic Volatility Model; GARCH; SPA Test

(特约编辑:王素珍)(校对:HA)