

沪深 300 股指期货的波动率预测模型研究^①

魏 宇

(西南交通大学经济管理学院, 成都 610031)

摘要: 以沪深 300 股指期货仿真交易的 5 分钟高频数据为例, 运用滚动时间窗的样本外预测和具有 Bootstrap 特性的 SPA 检验法, 全面对比了基于日收益数据的历史波动率 (historical volatility) 模型和基于高频数据的已实现波动率 (realized volatility) 模型对波动率的刻画和预测能力. 主要实证结果显示, 已实现波动率模型以及加入附加解释变量的扩展随机波动模型是预测精度较高的波动模型, 而在学术界和实务界常用的 GARCH 及其扩展模型对沪深 300 股指期货的波动率预测能力最弱.

关键词: 沪深 300 股指期货; 已实现波动率模型; 随机波动模型; GARCH 模型; SPA 检验

中图分类号: F224 **文献标识码:** A **文章编号:** 1007-9807(2010)02-0066-11

0 引 言

经过多年的研究和论证, 我国真正意义上的第 1 只规范的金融期货——沪深 300 股指期货合约即将正式上市交易. 股指期货的推出是适应我国资本市场改革开放与发展的客观要求, 同时也是深入贯彻《国务院关于推进资本市场改革开放和稳定发展的若干意见》以及党的十七大报告中“推进金融体制改革, 发展各类金融市场”的具体体现.

作为我国资本市场发展历史上的重要里程碑, 股指期货的推出将会进一步完善我国的资本市场结构, 其提供的全面避险工具也必将大大增加我国资本市场的流动性和有效性. 在这样的背景下, 随着我国宏观经济和政治环境继续持续平稳发展, 中国股市作为“国民经济晴雨表”的功能将会真正开始发挥, 我国资本市场也必将真正走向全面理性的繁荣.

但是, 也必须清醒地看到, 由于采用的是保证

金交易 (margin trading)、逐日盯市 (marking to market) 以及强行平 (减) 仓制度等, 期货价格的波动会造成投资者获利和损失成倍放大, 因此, 期货交易的市场风险远远高于股票现货交易. 鉴于此, 为了帮助市场各方参与者更好地了解股指期货合约、熟悉交易规则与测试交易系统, 并配合做好投资者的风险教育工作, 中国金融期货交易所 (以下简记为中金所) 于 2006 年 10 月 30 日开始推出沪深 300 股指期货的仿真交易 (mock trading) 活动. 为了在最逼真的状态下进行仿真交易, 使市场的参与各方最大程度地熟悉真实交易环境, 中金所从仿真交易合约的设计到交易规则的制定, 都按照尽可能真实的状态来要求. 同时, 中金所也要求申请仿真交易的单位要向中金所交易监管部提出仿真会员资格申请, 签署有关协议后才能成为仿真交易会员. 而投资者要参与仿真交易的话, 则首先应在具有仿真会员资格的期货公司开户, 开户时要签署风险揭示书, 并登记真实姓名、身份证号码及联系方式, 且实行一户一码;

① 收稿日期: 2008-06-24; 修订日期: 2009-05-31.

基金项目: 国家自然科学基金资助项目 (70501025, 70771097, 70771095); 教育部新世纪优秀人才支持计划资助项目 (NCET-08-0826); 教育部创新团队发展计划资助项目 (PCSIRT0860).

作者简介: 魏宇 (1975—), 男, 四川攀枝花人, 博士, 副教授. E-mail: weiyu@s126.com

经过中金所批准,由期货公司分配一定的虚拟资金后,投资者才能使用正规的交易系统进行交易。因此,沪深 300 股指期货的仿真交易确实能够较为真实和客观地反应出股指期货正式推出后的期指市场波动和风险状况。

因此,无论是对即将正式进入股指期货市场的各类参与者还是我国的金融监管当局而言,对股指期货的波动特征和风险状况进行准确的刻画和科学的预测,进而探索有效的期指市场风险防范和监控手段,都具有非常重要的理论和现实意义。

目前对金融资产波动率的模型描述方法主要有以下 3 种类型: 1) 第 1 类是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^[1] 的自回归条件异方差模型 (ARCH)、Bollerslev^[2] 的广义自回归条件异方差模型 (GARCH) 以及 Taylor^[3] 的随机波动模型 (stochastic volatility model 简记为 SV)。2) 第 2 类对波动率的描述方法来源于期权定价数据 (option pricing data),这类波动率模型称为隐含波动率模型 (implied volatility model 简记为 IV)。3) 第 3 类波动率的描述方法称为已实现波动率模型 (realized volatility model 简记为 RV), Andersen 和 Bollerslev^[4] 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) 的已实现波动率 (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,也并非高频收益率的频率越高越好。

从上面的讨论可以看出,既然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,那

么究竟哪一种模型才是最适合沪深 300 股指期货的实际波动特征和风险状况呢? 近年来,国内学者对我国沪深股市股价指数的波动率模型也进行了一些有益的前期探索。举例来说,张永东,毕秋香^[5] 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^[6] 探讨了 SV 模型和 GARCH 模型对深圳股市波动率的预测表现,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^[7] 对比了已实现波动率模型、SV 模型和 GARCH 模型对上海股市的波动率描述能力,结果显示,已实现波动率模型比 SV 和 GARCH 模型具有更高的波动率刻画精度。余素红等^[8] 对比了 GARCH 模型和 SV 模型的 VaR 计算能力,发现 SV 模型具有更好的风险测度精度。于亦文^[9] 运用上证综指的高频数据样本,比较了已实现波动率模型与 GARCH (1, 1) 模型对上海股市波动的解释能力,结果同样表明,已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。上述这些研究为深入探索我国资本市场的波动特征和风险状况都奠定了坚实的实证基础。

但需要指出的是,国内现有相关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。例如: 1) 大多数现有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已实现波动率模型 (RV) 和基于低频数据的历史波动率模型 (HV) 的全面对比结果; 2) 现有研究大多采用单一静态样本的样本内 (in-the-sample) 预测结果来进行模型的预测能力检验 (包括文献 [7]), 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^[10] 的研究指出,基于样本外预测结果的计量模型判定结论比样本内预测结果更加可靠和实用; 3) 现有研究采用的模型优劣判断方法不够严谨。Hansen 和 Lunde^[11] 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“平均误差平方 - MSE”和“平均绝对误差 - 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^[5-6]: 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 (例如, 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) 有待进一步论证.

基于以上认识, 为了全面深入地探求沪深 300 股指期货的波动特征及其最优的波动率预测模型, 本文与已有研究的不同之处主要体现在以下 4 个方面: 1) 首次采用了中金所股指期货仿真交易的高频数据, 对沪深 300 股指期货的波动定量特征进行了深入分析; 2) 全面对比了基于高频数据的已实现波动率模型和基于低频数据的历史波动率模型对沪深 300 股指期货波动的预测能力, 同时考虑加入滞后期的已实现波动率估计作为附加的解释变量 (additional explanatory variables), 对标准的 SV 和 GARCH 模型进行了扩展; 3) 运用样本外 (out-of-sample) 的滚动时间窗 (rolling time windows) 预测法, 计算了各类模型在样本外的日波动率预测值; 4) 进一步运用具有 Bootstrap 特性的更加严谨和稳健的统计检验法, 即 Hansen^[12] 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), 实证评估了不同波动率模型对沪深 300 股指期货的波动率预测精度, 以确保本文结论的稳健性和实用性.

进一步, 为了在中国股市的实际环境下, 对不同波动率模型的优劣进行评判, 需获得市场波动率的客观参考标准. 如上所述, 传统上用日收益率的平方 (squared daily return) 作为日波动率 (daily volatility) 的测度标准具有明显偏差. Andersen 等^[13] 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已实现波动率估计来作为真实市场波动率的替代变量

(proxy) 和基准 (benchmark).

本文的结构安排如下: 第 1 部分, 首先介绍数据样本和已实现波动率 RV 的估计方法; 第 2 部分分别介绍已实现波动率和历史波动率模型的构建; 第 3 部分则是对滚动时间窗的样本外波动率预测方法以及 SPA 检验的说明; 第 4 部分是实证结果, 而论文的主要结论在第 5 部分.

1 数据说明、收益率描述以及已实现波动率的估计

1.1 数据说明以及收益率的计算方法

本文研究的数据样本为中金所沪深 300 股指期货仿真交易的 5 分钟高频数据, 采用近月合约价格来构造连续期货价格序列, 时间从 2006 年 11 月 1 日到 2007 年 11 月 12 日, 共 $N = 245$ 个交易日. 中金所股指期货仿真交易每天有 4 小时 30 分钟交易时间, 因此每个交易日可以记录 54 个 5 分钟高频数据, 全部样本合计 13 230 个 5 分钟高频数据, 记为 $I_{t,d}$, $t = 1, 2, \dots, N$, $d = 1, 2, \dots, 54$ 其中 $I_{t,54}$ 表示第 t 天的收盘价, 数据来源于中国金融期货交易所 (CFEEX).

文中的日收益率 (daily return) R_t 利用相邻两个交易日的收盘价计算如下

$$R_t = 100(\ln I_{t,54} - \ln I_{t-1,54}) \quad (1)$$

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

$$R_{t,d} = 100(\ln I_{t,d} - \ln I_{t,d-1}) \quad (2)$$

1.2 已实现波动率的估计方法

根据 Andersen 和 Bollerslev^[4] 的定义, 对第 t 天的已实现波动率的估计表示为第 t 天内的所有高频收益率的平方和, 即

$$RV_t^2 = \sum R_{t,d}^2 \quad (3)$$

但最近 Hansen 和 Lunde^[11] 的研究又指出, 由于股票现货与期货市场并不象外汇市场那样在 24 小时连续进行交易, 因此, 能观察和记录到的高频价格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 (即市场从第 1 天收盘到第 2 天开盘的所谓 “close-to-Open” 波动率), 因此为了使

已实现波动率的估计能更加准确地刻画真实市场波动幅度的大小, 这里采用 Hansen 和 Lunde^[11] 的建议, 用某种尺度参数 (scale parameter) δ 来对 R_t^2 进行尺度变换, 即对第 t 天的已实现波动率估计为

$$RV_t = \delta RV_t' \quad (4)$$

其中尺度参数 δ 表示为

$$\delta = \frac{N^{-1} \sum_{t=1}^N R_t^2}{N^{-1} \sum_{t=1}^N RV_t'} \quad (5)$$

表 1 是对日收益率 R_t 、收益率平方 R_t^2 、已实现波动率估计 RV_t 以及对数 RV_t (简记为 $\ln RV_t$) 序列的描述性统计结果:

表 1 日收益率序列和已实现波动率 RV 序列的描述性统计

Table 1 Descriptive statistics for daily returns and realized volatility series

	日收益率相关序列		已实现波动率相关序列	
	R_t	R_t^2	RV_t	$\ln RV_t$
Mean	0.266	10.244	10.244	1.338
Standard deviation	3.196	18.749	19.187	1.506
Skewness	0.039	2.440***	5.577***	-0.231
Kurtosis	1.394***	5.441***	41.913***	-0.153
J-B	19.908***	545.287***	19.202.720***	2.409
Q (5)	7.463	20.572**	94.861***	294.678**
Q (10)	10.564	26.133**	107.706***	313.157**
Q (20)	18.388	31.416**	118.365***	354.161**
ADF	-13.568***	-11.914**	-3.992**	-6.412**
P-P	-13.624***	-11.963***	-11.390***	-6.438**

说明: “*” 和 “***” 分别代表在 5% 和 1% 水平上显著, 其中峰态系数 Kurtosis 为超额峰态, J-B 为 Jarque-Bera 统计量, $Q(n)$ 为滞后阶数为 n 的 Ljung-Box Q 统计量, ADF 和 P-P 分别是以最小 AIC 准则确定最优检验滞后阶数后得到的 Augmented Dickey-Fuller 单位根检验以及 Phillips-Perron 单位根检验结果。

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 “尖峰胖尾” (leptokurtic and fat tailed) 特征 (超额峰态系数都非常显著), 这说明沪深 300 股指期货的波动幅度较为剧烈, 远远超出了正态分布假定的范围 (Jarque-Bera 统计量都很显著)。同时, 收益率序列之间不具有显著的自相关性; 但收益率平方序列、已实现波动率序列 RV_t 和对数已实现波动率序列 $\ln RV_t$ 则在很长的时间范围内, 都展示出非常显著的自相关性特征 (Q 统计量都很显著), 这从一方面也说明了沪深 300 股指期货的波动存在较为显著的持续性 (persistence) 或长期记忆性 (long memory) 特征。进一步, ADF 单位根检验以及 Phillips-Perron 单位根检验表明, 各序列都显著拒绝了存在单位根的原假设。因此, 可以认为各个序列是平稳的, 从而可以直接作进一步的分析和计量建模。

2 各类波动率模型描述

2.1 已实现波动率模型

Andersen 等^[14] 的研究指出, 用一般的自回归移动平均模型 (ARMA) 将无法准确地描述具有长记忆 (long memory) 特征的时间序列。因此, 他们建议采用 “自回归分整移动平均 (ARFMA) 模型” 来刻画这类具有长期记忆性时间序列的动力学特征。同时, 考虑到不同滞后阶数的 ARFMA (p, d, q) 模型对长记忆时间序列的估计结果非常接近^[15], 同时结合模型估计的 AIC (Akaike's information criterion) 大小, 本文采用 ARFMA (1, d , 1) 模型来同时为沪深 300 股指期货的 RV_t 以及 $\ln RV_t$ 序列建模, 以增强实证结果的对比性以及可靠性。

ARFMA (p, d, q) 模型的一般形式为

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

其中: Y 是待估的时间序列(文中指 RV_t 和 $\ln RV_t$ 序列); $\Phi(L) = 1 - \phi_1L - \phi_2L^2 - \dots - \phi_pL^p$, $\Theta(L) = 1 + \theta_1L + \theta_2L^2 + \dots + \theta_qL^q$ 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,同时假定 $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$, 并且, 以下对这两种波动率模型分别简记为 ARFMA-RV 和 ARFMA- $\ln RV$.

2.2 历史波动率模型

下面介绍最具代表性的两类历史波动率模型(HV)及其在本文中的实证研究方法.

1) 随机波动模型(SV)

Taylor^[3]提出了著名的随机波动模型(stochastic volatility model SV), 该模型的标准形式假定金融资产的收益率满足以下形式

$$R_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \quad (7)$$

其中 μ_t 是收益波动的条件均值(conditional mean), σ_t^2 是条件方差(conditional variance), 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, 因此在本文的实证研究当中都假定其等于零^[15].

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

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

其中, 不可观测的对数波动率(log-volatility) h_t 则满足

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

且假定

$$\eta_t \sim NID(0, 1),$$

$$h_1 \sim NID\left(0, \frac{\sigma_\eta^2}{1 - \phi^2}\right)$$

Koopman等^[15]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 1 期的对数形式的隐含波动率或者已实现波动率估计作为附加的解释变量, 以增强模型对收益波动的刻画能力. 因此, 在本文的实证研究当中, 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已实现波动率作为解释变量的扩展 SV 模型(以下简记为 SV-RV), 其对数波动率 h_t 如下式所示.

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

有关 SV 模型更深入的讨论可以参见 Shephard^[16]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 和 Koopman^[17]的研究.

2) 广义自回归条件异方差模型(GARCH)

Bollerslev^[2]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. 其对金融收益率的模型假定同样如公式(7)所示, 但与 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

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

与上面对 SV 模型的讨论类似, 为了增强实证结果的对比性和可靠性, 在后面的实证研究当中, 同样考虑了在其条件方差式(11)中加入滞后 1 期的已实现波动率估计作为附加解释变量的扩展 GARCH 模型(以下简记为: GARCH-RV), 如式(12)所示

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

3 滚动时间窗的样本外波动率预测方法及 SPA 检验说明

3.1 波动率预测方法说明

在本文的实证研究当中, 对上面讨论的 6 种主流波动率模型(即 ARFMA-RV、ARFMA- $\ln RV$ 、SV、SV-RV、GARCH 以及 GARCH-RV)进行了滚动时间窗(rolling time windows)的“样本外预测能力检验”(tests for out-of-sample predicting ability). 本文的预测方法具体如下:

1) 将数据样本总体 ($t = 1, 2, \dots, N = 245$) 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 $H = 150$ 个交易日的数据, 而预测样本则包含最后 95 个交易日的数据(即 $t = H + 1, H + 2, \dots, H + M$, 其中 $M = 95$).

2) 第 1 步, 选取 $t = 1, 2, \dots, H = 150$ 的数据作为第 1 个估计样本, 分别对上述 6 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 $\hat{\sigma}_{H+1}^2$ 。也就是说, $\hat{\sigma}_{H+1}^2$ 是在前面 150 个样本数据的基础上对第 151 天的市场波动率预测。

3) 第 2 步, 保持估计样本的时间区间长度不变 ($H = 150$), 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 $t = 2, 3, \dots, H + 1$ 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, 并在此新的模型估计基础上获得未来 1 天的波动率预测, 记为 $\hat{\sigma}_{H+2}^2$ 。

4) 同理, 不断重复步骤 3), 可以得到 $\hat{\sigma}_{H+3}^2, \hat{\sigma}_{H+4}^2, \hat{\sigma}_{H+5}^2$, 直到最后一轮的估计样本区间为 $t = M, M + 1, \dots, H + M - 1$ 以获得对最后一天, 即 $t = N = H + M = 245$ 天的市场波动率预测 $\hat{\sigma}_{H+M}^2$ 。

简言之, 对前面讨论的 6 类不同的波动率模型, 都分别重复进行了 95 次的模型估计 (共计 $6 \times 95 = 570$ 次不同的估计), 从而每个模型都获得了 95 个未来 1 天的样本外市场波动率估计, 记为, $\hat{\sigma}_m^2, m = H + 1, H + 2, \dots, H + M$ 。同理, 记预测样本区间的已实现波动率估计为 $RV_m, m = H + 1, H + 2, \dots, H + M$ 。这里对 RV_m 的估计方法来自于 1.2 节的说明, 并以此作为真实市场波动率的替代变量 (proxy), 用以衡量各类波动率模型的预测精度。

3.2 基于 Bootstrap 模拟的 SPA 检验法说明

有了以上所讨论的 6 类波动率模型及其对市场波动的预测值 $\hat{\sigma}_m^2$ 以后, 就可以比较这些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。然而需要说明的是, 到目前为止, 学术界还不清楚用哪一种损失函数 (loss function) 作为衡量预测误差的标准最为合理。因此, Hansen^[12] 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, 在本文的实证研究当中采用了 6 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 6 种损失函数分别标记为 $L_i, i = 1, 2, \dots, 6$ 其中 L_1 和 L_2 分别称为平均误差平方 (mean squared error 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。 L_3 和 L_4 分别是经异方差调整的 MSE 和 MAE (heteroskedastic adjusted MSE and MAE), 对 L_5 和 L_6 的具体含义讨论可以参考 Hansen^[12] 的研究。各损失函数的具体定义如下

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

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

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

$$L_4: HMAE = M^{-1} \sum_{m=H+1}^{H+M} \left| \frac{1 - \hat{\sigma}_m^2}{RV_m} \right| \quad (16)$$

$$L_5: QLIKE = M^{-1} \sum_{m=H+1}^{H+M} \left(\ln \hat{\sigma}_m^2 + \frac{RV_m}{\hat{\sigma}_m^2} \right) \quad (17)$$

$$L_6: R^2LOG = M^{-1} \sum_{m=H+1}^{H+M} \left(\ln \frac{RV_m}{\hat{\sigma}_m^2} \right)^2 \quad (18)$$

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。因为, 数据样本中的少数奇异点 (outliers) 往往会严重影响损失函数的计算结果, 从而引起损失函数值的异常增加, 进而可能会导致对波动率模型优劣的错误判断。深入讨论可以参见 Hansen^[12] 的研究。

为了解决这一问题, Hansen 提出了所谓的“高级预测能力检验法” (superior predictive ability test SPA)。他们的研究证明, 因为采用了所谓的“自举法” (bootstrap) 模拟, SPA 检验比类似的 White^[10] 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 (robustness)。也就是说, 与基于单一样本的其他检验法相比, 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 检验的实现过程如下: 首先, 假定有 $J + 1$

种类型的波动率模型, 记为 $m_k, k = 0, 1, \dots, J$. 每种波动率模型 m_k 得到的未来 1 天的波动率预测记为 $\hat{\sigma}_{m,k}^2$, 其中 $m = H + 1, H + 2, \dots, H + M$. 对每一个预测值, 都可以计算公式 (13) ~ (18) 所定义的 6 种损失函数值, 记为 $L_{i,m,k}$, 其中 $i = 1, 2, \dots, 6$. 下面, 用 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, 对于其他的 $k = 1, 2, \dots, J$ 种波动率模型, 可以计算其相对于基础模型 m_0 的“相对损失函数值” (relative loss function), 记为

$$X_{k,m} = L_{i,0,m} - L_{i,k,m} \quad (19)$$

现在的关键问题是: 是否在 $k = 1, 2, \dots, J$ 的模型 m_k 当中有比基础模型 (m_0) 表现更加优异的模型呢? 为了得到这一问题的答案, 首先可以定义这样的零假设 H_0 : “与对比模型 m_k 相比, 基础模型 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

$$\max \lambda_k = E(X_{k,m}) \leq 0, \quad k = 1, 2, \dots, J \quad (20)$$

Hansen^[12] 证明了这一假设检验的检验统计量可以表示为

$$T = \max \frac{\sqrt{M} \bar{X}_k}{\hat{\omega}_{kk}}, \quad k = 1, 2, \dots, J \quad (21)$$

其中:

$$\bar{X}_k = M^{-1} \sum_{m=H+1}^{H+M} X_{k,m}, \quad \hat{\omega}_{kk}^2 = \text{var}(\sqrt{M} \bar{X}_k) \quad (22)$$

为了获得公式 (21) 的 T 检验量的分布状况及其 p 值, Hansen^[12] 建议可以采用所谓的“自举过程” (bootstrap procedure) 来取得. 首先, 需要获得一个长度为 M 的 $X_{k,m}$ 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 (new subsample), 而该子样本的长度则来自一个服从均值为 q 的几何分布 (geometric distribution) 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 $X_{k,m}^i, i = 1, 2, \dots, B$. 在后面的实证研究当中, 选取 $B = 2000$ 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

$$\bar{X}_k^i = M^{-1} \sum_{m=1}^M X_{k,m}^i, \quad i = 1, 2, \dots, B \quad (23)$$

而所有 B 个 bootstrap 样本均值的方差估计表示为

$$\hat{\omega}_{kk} = B^{-1} \sum_{i=1}^B (\bar{X}_k^i - \bar{\bar{X}}_k)^2, \quad \bar{\bar{X}}_k = B^{-1} \sum_{i=1}^B \bar{X}_k^i \quad (24)$$

其次, 定义 Z_k^i 为

$$Z_k^i = (\bar{X}_k^i - \bar{\bar{X}}_k) \times I\{\bar{X}_k^i > -A_k\} \quad (25)$$

其中

$$A_k = \frac{1}{4} M^{-4} \hat{\omega}_{kk} \quad (26)$$

而 $I\{\cdot\}$ 是一个指示函数 (indicator function), 即当 $\{\cdot\}$ 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

$$T^i = \max \frac{\sqrt{M} Z_k^i}{\hat{\omega}_{kk}}, \quad i = 1, 2, \dots, B \quad (27)$$

Hansen^[12] 的研究表明, 在式 (20) 所示的零假设条件下, 式 (27) 所示的实证统计量收敛于式 (21) 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

$$p = B^{-1} \sum_{i=1}^B I\{T^i > T\} \quad (28)$$

简言之, SPA 检验的 p 值越大 (越接近于 1), 则表明越不能拒绝式 (20) 所定义的零假设 H_0 : “与对比模型 m_k 相比, 基础模型 m_0 是表现最优的预测模型.” 即说明该基础模型的预测精度越高.

4 实证结果

4.1 各类波动率模型的预测结果

各类波动率模型对未来 1 天的市场波动率估计方法如 3.1 节所示, 图 1 的 (a) 图表示的是 ARFMA-RV 模型和 ARFMA-lnRV 模型在预测样本区间 ($t = 151, 152, \dots, 245$) 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则用实心的小方块表示. 类似地, 图 1 的 (b) 图和 (c) 图分别是 SV、SV-RV 以及 GARCH、GARCH-RV 模型在预测区间的波动率预测结果.

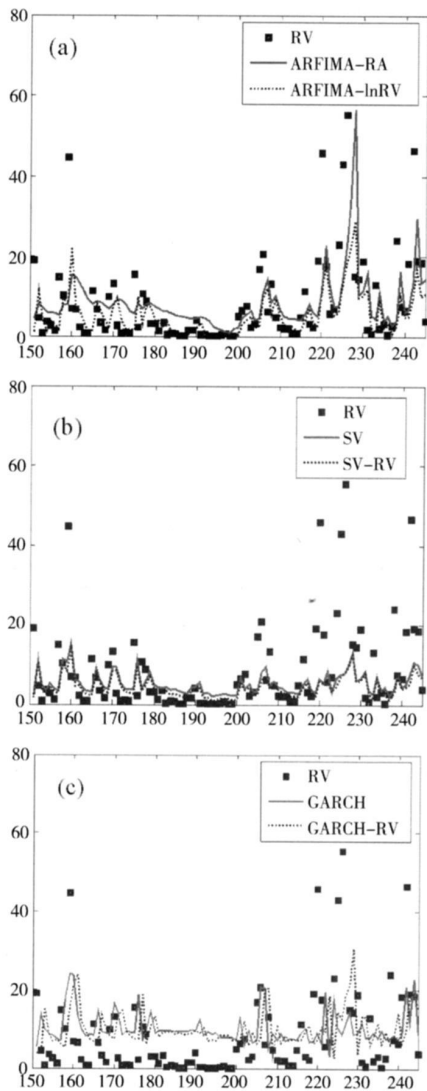


图 1 不同波动率模型在预测样本区间的预测结果
Fig 1 Forecasting results for different volatility models during forecasting sample

从图 1 中 (a)、(b) 和 (c) 图的比较可以看出, 直观来讲, 基于高频数据的 RV 波动率模型都较好地预测了该段时间内的市场波动状况, 特别是在市场出现大幅度波动的阶段, 基于高频数据的 RV 模型具有比其他 HV 模型更加优异的波动预测效果。而就 SV 和 GARCH 模型的对比来看, 总的来说, GARCH 模型似乎有明显高估波动率的倾向。

4.2 波动率模型的 SPA 检验结果

表 2 是经过 2 000 次 bootstrap 模拟得到的 SPA 检验结果, 为了节省篇幅以及清晰起见, 表中用小写字母 m 加下标来表示不同的模型名称。其中, m_1 代表 ARFMA-RV; m_2 代表 ARFMA-lnRV; m_3 代表 SV; m_4 代表 SV-RV; m_5 代表 GARCH; m_6 代表 GARCH-RV。

表 2 的第 1 列表示的是 6 种损失函数 L_i , 第 2 列是被选作基础模型 (bbase model M_0) 的模型名称, 表中数字为 SPA 检验的 p 值。 p 值越大 (越接近于 1), 则表明越不能拒绝公式 (20) 所定义的零假设 H_0 : “与对比模型 M_k 相比, 基础模型 M_0 是表现最优的预测模型。” 这也就是说, 在某一损失函数 L_i 判断标准下, 如果基础模型 M_0 相对于其它模型的 SPA 检验 p 值越大, 则表明该模型的预测精度越高。反之, 如果 p 值越小, 则有理由相信, 基础模型 M_0 的表现要劣于所考察的对比模型 M_k 。

从表中的实证结果可以看出: 1) 对沪深 300 股指期货而言, 总体来说, 基于高频数据的已实现波动率模型, 特别是 ARFMA-lnRV 模型具有最为出色的波动率预测能力。这一结果也表明, 与日数据 (daily data) 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) 加入已实现波动率估计作为附加解释变量的扩展 SV 模型 (SV-RV) 的预测表现也相当不错, 除了在 MSE 和 QLIKE 标准下, SV-RV 模型都表现出了非常精确的预测能力; 3) 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。且总体来讲, 加入已实现波动率估计作为附加解释变量的扩展 GARCH 模型并没有对普通 GARCH 模型的预测能力有显著的改进。

综上所述可以发现, GARCH、SV 和 RV 模型的波动率预测能力具有显著的差异。由于 Andersen 等^[13] 的研究发现, 潜在真实的市场波动率是不可直接观测的, 而 GARCH 的建模方法则假定其是可以观察到的市场变量, 从而对其进行直接的计量建模。这种建模方法在理论上的偏误可能是导致 GARCH 预测能力较差的重要原因之一。而 SV 的建模依据则刚好与 Andersen 等^[13] 的结论吻合, 因此 SV 的预测精度要高于 GARCH 模型。进一步, 由于 RV 模型中采用的是与 GARCH 和 SV 模型不同的高频收益数据, 而高频数据当中显然蕴涵了更多的市场波动信息, 从而可以更加真实地刻画实际市场的波动特征。另外, Andersen 等^[14] 的研究也从理论上证明了 RV 是一种对市场真实波动率更好的无偏估计, 因此基于 RV 的时间序列模型应该从理论上具有比传统方法更好的波动率预测能力。本文的实证结果与 Koopman 等^[15] 的研究结论类似。

表 2 不同波动率模型的 SPA 检验结果
Table 2 SPA results for different volatility models

损失函数	基础模型	对比模型					
		m_1	m_2	m_3	m_4	m_5	m_6
MSE	m_1	—	0.492	0.817	0.853	0.803	0.954
	m_2	0.508	—	0.879	0.913	0.864	1.000
	m_3	0.183	0.121	—	0.978	0.418	0.648
	m_4	0.147	0.087	0.022	—	0.290	0.496
	m_5	0.197	0.136	0.582	0.710	—	0.759
	m_6	0.046	0.000	0.352	0.504	0.241	—
MAE	m_1	—	0.007	0.147	0.107	0.984	1.000
	m_2	0.993	—	0.916	0.833	1.000	1.000
	m_3	0.853	0.084	—	0.080	1.000	1.000
	m_4	0.893	0.167	0.920	—	1.000	1.000
	m_5	0.016	0.000	0.000	0.000	—	0.941
	m_6	0.000	0.000	0.000	0.000	0.059	—
HMSE	m_1	—	0.053	0.094	0.055	0.936	0.946
	m_2	0.947	—	0.951	0.959	0.939	0.949
	m_3	0.906	0.049	—	0.050	0.938	0.949
	m_4	0.945	0.041	0.950	—	0.938	0.949
	m_5	0.064	0.061	0.062	0.062	—	0.119
	m_6	0.054	0.051	0.051	0.051	0.881	—
HMAE	m_1	—	0.008	0.014	0.003	0.978	0.981
	m_2	0.992	—	0.980	0.963	0.984	0.989
	m_3	0.986	0.020	—	0.016	0.987	0.991
	m_4	0.997	0.037	0.984	—	0.987	0.990
	m_5	0.022	0.016	0.013	0.013	—	0.117
	m_6	0.019	0.011	0.009	0.010	0.883	—
QLIKE	m_1	—	0.970	0.966	0.989	0.984	1.000
	m_2	0.030	—	0.137	0.303	0.207	0.299
	m_3	0.034	0.863	—	0.985	0.495	0.781
	m_4	0.011	0.697	0.015	—	0.220	0.386
	m_5	0.016	0.793	0.505	0.780	—	0.905
	m_6	0.000	0.701	0.219	0.614	0.095	—
R ² LOG	m_1	—	0.033	0.186	0.051	0.995	1.000
	m_2	0.967	—	0.952	0.933	0.992	0.996
	m_3	0.814	0.048	—	0.041	0.999	0.999
	m_4	0.949	0.067	0.959	—	0.995	0.998
	m_5	0.005	0.008	0.001	0.005	—	0.676
	m_6	0.000	0.004	0.001	0.002	0.324	—

说明: 表中数字是进行了 2 000 次 bootstrap 模拟的 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。

其中, m_1 代表 ARFMA-RV; m_2 代表 ARFMA-lnRV; m_3 代表 SV; m_4 代表 SV-RV; m_5 代表 GARCH; m_6 代表 GARCH-RV。

5 结束语

本文以中金所沪深 300 股指期货仿真交易的 5 分钟高频数据为例, 运用更加严谨和稳健的 SPA 检验法, 首次全面探讨和对比了各类主流波动率模型对其样本外日波动率 (out of sample daily volatility) 的预测能力, 主要实证结果显示: 就沪深 300 股指期货而言, 历史波动率模型 (HV) 的预测精度要明显低于基于高频数据的已实现波动率 (RV) 模型. 在所考察的各种损失函数标准下, ARFMA-hRV 模型都具有较高的波动预测精度, 而加入已实现波动率估计作为附加解释变量的扩展 SV 模型 (SV-RV) 也具有较好的预测能力.

值得注意的是, 研究同时发现, 目前在金融学术界和实务界常用的 GARCH 模型以及加入已实

现波动率估计作为附加解释变量的扩展 GARCH 模型 (GARCH-RV) 都无法对沪深 300 股指期货的波动作出令人满意的刻画和预测. 因此, 对于在成熟资本市场适用的金融理论和模型, 必须进行认真的比较和检验, 才能判定其是否也适合我国新兴资本市场的实际情况. 论文的研究方法和结论对目前我国资本市场的风险监管工作, 以及我国金融衍生产品市场的建立和发展过程中的很多基础性问题 (比方说, 如何设定与其波动特征相匹配的股指期货交易保证金水平以及如何对股指期货期权进行精确定价等等), 都具有重要的理论和现实意义. 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资本市场实际风险状况和波动特征的波动率模型, 仍然是下一步研究的主要方向.

参考文献:

- [1] Engle R F. Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation [J]. *Econometrica* 1982, 50(3): 987-1007
- [2] Bollerslev T. Generalized autoregressive conditional heteroskedasticity [J]. *Journal of Econometrics*, 1986 31(2): 307-327.
- [3] Taylor S J. *Modeling Financial Time Series* [M]. Chichester: John Wiley and Sons, 1986
- [4] Andersen T G, Bollerslev T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts [J]. *International Economic Review*, 1998 39(6): 885-905
- [5] 张永东, 毕秋香. 上海股市波动性预测模型的实证比较 [J]. *管理工程学报*, 2003, 15(2): 16-19
Zhang Yong-dong, Bi Qi-xiang. Volatility forecasting models in the Shanghai stock market: A comparison [J]. *Journal of Industrial Engineering and Engineering Management*, 2003, 15(2): 16-19 (in Chinese)
- [6] 刘凤芹, 吴喜之. 基于 SV 模型的深圳股市波动的预测 [J]. *山西财经大学学报*, 2004, 12(4): 96-99
Liu Feng-qin, Wu Xi-zhi. SV forecast of Shenzhen stock exchanges [J]. *Journal of Shanxi Finance and Economics University*, 2004, 12(4): 96-99 (in Chinese)
- [7] 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 11(8): 60-63
Xu Zheng-guo, Zhang Shi-ying. The comparative research on volatility prediction ability of adjusted realized volatility, GARCH model and SV model [J]. *Systems Engineering* 2004, 11(8): 60-63 (in Chinese)
- [8] 余素红, 张世英, 宋 军. 基于 GARCH 模型和 SV 模型的 VaR 比较 [J]. *管理科学学报*, 2004 7(5): 61-66
Yu Su-hong, Zhang Shi-ying, Song Jun. Comparison of VaR based on GARCH and SV models [J]. *Journal of Management Sciences in China*, 2004, 7(5): 61-65 (in Chinese)
- [9] 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 18(2): 65-69
Yu Yi-wen. A comparative analysis of realized volatility and GARCH model [J]. *Journal of Industrial Engineering and Engineering Management*, 2006, 18(2): 65-69 (in Chinese)
- [10] White H. A reality check for data snooping [J]. *Econometrica* 2000 68(3): 1097-1126

- [11] Hansen P R, Lunde A. Consistent ranking of volatility models[J]. *Journal of Econometrics*, 2006, 131(6): 97– 121.
- [12] Hansen P R. A test for superior predictive ability[J]. *Journal of Business and Economic Statistics*, 2005, 23(4): 365– 380.
- [13] Andersen T G, Bollerslev T, Meddahi N. Correcting the errors: Volatility forecast evaluation using high frequency data and realized volatilities[J]. *Econometrica*, 2005, 73(1): 279– 296.
- [14] Andersen T G, Bollerslev T, Diebold F X, *et al*. The distribution of realized stock return volatility[J]. *Journal of Financial Economics*, 2001, 61(5): 43– 76.
- [15] Koopman S J, Jungbacker B, Hol E. Forecasting daily variability of the S&P100 stock index using historical, realized and implied volatility measurements[J]. *Journal of Empirical Finance*, 2005, 12(4): 445– 475.
- [16] Shephard N. Statistical aspects of ARCH and stochastic volatility[M]// Cox D R, Hinkley D V, Bamdorff N, Jensen O E (Eds), *Time Series Models in Econometrics, Finance and Other Fields*, Number 65 in Monographs on Statistics and Applied Probability. London: Chapman and Hall, 1996: 1– 67.
- [17] Sandmann G, Koopman S J. Estimation of stochastic volatility models via Monte Carlo maximum likelihood[J]. *Journal of Econometrics*, 1988, 87(3): 271– 301.

Volatility forecasting models for CSI300 index futures

WEI Yu

School of Economics & Management, Southwest Jiaotong University, Chengdu 610031, China

Abstract Taking 5-minutes high-frequency mock trading data of CSI300 index futures as example, the out-of-sample daily volatility predictions of these models are calculated by using rolling predicting method, and a bootstrap SPA test is used to evaluate the predicting accuracy for different historical volatility models and realized volatility models. The empirical results show that realized volatility model based on high-frequency data and the extended SV model are superior to other models. However the GARCH and its extended model, which are popular in financial academe and practice, perform worst for volatility predicting of CSI300 index futures.

Key words CSI300 index futures; realized volatility model; SV model; GARCH model; SPA test