

# 基于 LHAR-RV-V 模型的中国股市波动性研究<sup>①</sup>

文凤华<sup>1 3</sup>, 刘晓群<sup>1</sup>, 唐海如<sup>1</sup>, 杨晓光<sup>1 2 3</sup>

(1. 长沙理工大学经济与管理学院, 长沙 410114; 2. 中国科学院数学与系统科学研究院管理、决策与信息系统重点实验室, 北京 100190; 3. 湖南省金融工程与金融管理研究中心, 长沙 410114)

摘要: 在已实现波动率异质自回归模型(HAR-RV 模型)的基础上, 基于市场微观结构的理论, 同时考虑市场波动的杠杆效应和量价关系, 构造了已实现波动率及交易量之长记忆异质自回归模型(LHAR-RV-V 模型). 利用该模型对沪深 300 指数的等时 1min 高频数据进行实证分析, 实证结果表明该模型能够较好地捕捉到我国股票市场波动的长记忆性和杠杆效应, 且杠杆效应具有一定的持续性. 此外, 过去不同周期交易量的加入不仅能够更为细微的反映量价之间的关系, 而且在一定程度上改善了模型的预测能力.

关键词: 异质市场假说; LHAR-RV-V 模型; 交易量; 杠杆效应

中图分类号: F830.91 文献标识码: A 文章编号: 1007-9807(2012)06-0059-09

## 0 引 言

股市的波动及相关特征是国内外学者研究金融衍生工具的定价、有效资产组合的选择及金融风险管理的变量, 也是金融学领域的一个研究热点. 近年来, 随着计算机及通讯技术的快速发展, 极大的降低数据记录和存储的成本, 从而使得金融高频数据日益成为研究金融资产价格波动及市场微观结构的重要手段. Andersen 和 Bollerslev<sup>[1]</sup>首次对高频时间序列的“波动率”采用“已实现”波动率(realized volatility, RV)的全新概念来度量. “已实现”波动率相比金融资产的日收益能够更为精确的度量股市的波动, 它的提出极大的推动了金融高频时间序列模型的发展, 也为从市场微观的角度来研究股市波动特征提供了可能.

以往学者对股市波动性特征的研究主要集中在低频数据的范畴和单个特征的考量. 自 Andersen 和 Bollerslev<sup>[1]</sup>提出“已实现”波动率以来, 基

于高频数据的波动率模型得到了飞速发展, 并在波动率特征刻画和波动率预测领域取得了显著的成就. Andersen<sup>[2-4]</sup>等对西方国家发达金融市场的高频金融时间序列的研究得到“已实现”波动率具有显著的长记忆性和非对称性. Andersen<sup>[5]</sup>等将已实现波动率与条件协方差矩阵结合起来进行实证分析, 研究结果表明基于日已实现波动率的长记忆高斯 VAR 模型在波动率的预测上要明显优于 GARCH 模型. Koopman<sup>[6]</sup>等建立了加入 RV 作为解释变量的扩展 SV 模型(SV-RV 模型)和扩展的 ARFIMA 模型(ARFIMA-RV 模型), 实证结果显示加入 RV 可以显著的改善传统波动模型的预测能力. Corsi<sup>[7]</sup>基于 Müller<sup>[8]</sup>的异质市场假说理论提出了 HAR-RV 模型, 即通过不同时间段(日、周、月)的已实现波动率的自回归过程来刻画异质交易行为, 该模型的偏回归系数直接度量了特定类型交易者的行为对整个波动率的边际影响. Bollerslev 和 Diebold<sup>[9]</sup>将已实现波动率分解成连续和跳跃两种不同的成分, 并建立 HAR-RV-

① 收稿日期: 2011-02-14; 修订日期: 2011-03-28.

基金项目: 国家自然科学基金资助项目(70971013; 71171024); 国家 973 计划资助项目(2010CB731405); 湖南省杰出青年基金资助项目(09JJ1010).

作者简介: 文凤华(1972—), 男, 湖南益阳人, 博士, 教授. Email: mailtowfh@126.com

CJ 模型对已实现波动率进行预测,从而极大地提高了金融资产收益波动的预测精度. McAleer 和 Medeiros<sup>[10]</sup>扩展了 HAR-RV 模型,提出了一个柔性多重平稳变换模型,用来捕捉时间序列的非线性性和长期自相关性. Louzis<sup>[11]</sup>等则更进一步考虑 HAR 模型的残差的异方差性,构建了非对称 HAR-FIGARCH 模型来预测股票指数已实现波动率. Corsi<sup>[12]</sup>在 HAR-RV 模型的基础上考虑波动率的持续和跳跃成分以及负收益率的异质结构构建了 LHAR-CJ 模型来捕捉波动率的非对称动态特征. Liu 和 Maheu<sup>[13]</sup>把 ACD 模型与 Coris<sup>[14]</sup> HAR-RV 模型结合起来,提出了 HAR-BACD 模型,该模型能够考察交易久期对波动率影响以及包含不同信息量的波动成分对即时总波动的贡献大小.

国内对已实现波动率研究起步相对较晚. 黄后川和陈浪南<sup>[15]</sup>基于高频数据估计中国股市的已实现波动率,研究了我国股市波动率的非对称性和长记忆性. 徐正国和张世英<sup>[16]</sup>对调整的“已实现波动率”的长记忆性与“杠杆”效应建立了 ARFIMAX 模型,并将其与 GARCH 及 SV 模型的波动预测能力进行了比较分析,结果发现 ARFIMAX( $p, d, q$ ) 模型对积分波动进行了最好的预测. 徐正国和张世英<sup>[17]</sup>将一维高频数据的“已实现”波动率扩展到多维高频数据情形,并针对“已实现”协方差阵建立了 FIVAR 模型来刻画波动的长记忆性及相关特征. 魏宇<sup>[18]</sup>通过利用沪深 300 股指期货的仿真交易数据对股市波动预测模型进行对比分析,得出 ARFIMA-InRV 模型和 SV-RV 模型的预测能力要优于 GARCH 族模型. 张波等<sup>[19]</sup>和唐勇,池云果<sup>[20]</sup>利用已实现波动率来研究沪指波动的相关特性,发现 HAR-RV 模型比 ARFIMA 模型更能有效地刻画沪指波动的长记忆性,且 HAR-RV 模型样本外预测效果远远优于 ARFIMA 模型. 曾增展<sup>[21]</sup>在 HAR-RV-BPV 模型里加入滞后日、周、月交易量,构建了基于已实现波动率的异质自回归模型,即 HAR-RV-BPV-cum-Vol 模型来研究量价关系,结果表明加入异质结构的交易量后模型的拟合能力和预测能力都得到显著提高.

杠杆效应与量价关系是股票市场波动的两个主要特征. 综合以上国内外对已实现波动率的研究现状,发现还没有文献在对已实现波动率的研

究中同时考虑杠杆效应与量价关系. 因此,本文旨在全面刻画股市波动特征和提高股市波动预测能力的同时,在 HAR-RV 模型的基础上,综合考虑波动的非对称性及量价关系,构建了 LHAR-RV-V 模型,并将其在中国股市进行实证分析,以进一步揭示中国股市波动的相关特征.

## 1 LHAR-RV-V 模型

### 1.1 HAR-RV 模型

Müller<sup>[8]</sup>提出了异质市场假说,该假说认为交易者是非对称的,即市场是由不同交易期限的交易者构成,他们在风险偏好、制度约束、信息获取、理性程度等方面可能存在差异. 于是,不同的交易者对市场上的同一信息会作出不同的反应,因而对价格的变化和波动的影响不同. 在异质市场假说的基础上, Corsi<sup>[7]</sup>提出了 HAR-RV 模型,通过不同时间段(日、周、月)的已实现波动率的一阶自回归过程来刻画市场上交易者的异质交易行为,具体模型为

$$RV_{t+H}^d = \beta_0 + \beta_D RV_t^d + \beta_W RV_t^w + \beta_M RV_t^m + \varepsilon_{t+H} \quad (1)$$

其中  $RV_t^d$  为  $t$  期的日已实现波动率,即  $RV_t^d = \sqrt{\sum_{j=1}^n r_j^2}$ ;  $RV_t^w$  为  $t$  期的周已实现波动率,即  $RV_t^w = \frac{1}{5}(RV_{t-5}^d + RV_{t-4}^d + \dots + RV_{t-1}^d)$ ;  $RV_t^m$  为  $t$  期的月已实现波动率,即  $RV_t^m = \frac{1}{20}(RV_{t-20}^d + RV_{t-19}^d + \dots + RV_{t-1}^d)$ . 该模型主要反映了市场的波动是一个由不同波动成分相互叠加的复杂混合波动,它是由短、中、长期投资者等交易行为共同作用的结果,其偏回归系数直接度量了特定类型交易者的行为对整个波动率的边际影响. 同时,该模型也很好地刻画了波动率的长记忆性特征.

### 1.2 LHAR-RV 模型

为了更好地刻画股市波动的总体相关特征, Corsi<sup>[12]</sup>进一步拓展了 HAR-RV 模型,构建了 LHAR-CJ 模型来探讨前期不同周期的负收益和波动中跳跃成分对当期或者未来一段时间内已实现波动率的影响. 为了进一步研究中国股市非对称性和市场交易者的异质性,通过借鉴 LHAR-CJ

模型的建模思路 结合 Crossi<sup>[7]</sup>的 HAR-RV 模型构建了 LHAR-RV 模型 具体模型为

$$RV_{t+H} = \beta_0 + \beta_D RV_t^d + \beta_W RV_t^w + \beta_M RV_t^m + \rho_D r_t^{d-} + \rho_W r_t^{w-} + \rho_M r_t^{m-} + \varepsilon_{t+H} \quad (2)$$

其中  $RV_{t+H} = \frac{1}{H}(RV_{t+1} + RV_{t+2} + \dots + RV_{t+H})$ ,  $H = 1, 5, 10, 15, 20$  (设  $t$  为当期) 分别代表未来一天、一周、两周、三周、一个月的平均已实现波动率. 另外  $r_t^{d-}$  代表  $t$  期对应负的日收益, 即  $r_t^{d-} = r_t I\{r_t < 0\}$ ;  $r_t^{w-}$  代表  $t$  期对应负的周收益, 即  $r_t^{w-} = \frac{1}{5}(r_{t-5} + r_{t-4} + \dots + r_{t-1}) I\{(r_{t-5} + r_{t-4} + \dots + r_{t-1}) < 0\}$ ;  $r_t^{m-}$  代表  $t$  期对应的负的月收益, 即  $r_t^{m-} = \frac{1}{20}(r_{t-20} + r_{t-19} + \dots + r_{t-1}) I\{(r_{t-20} + r_{t-19} + \dots + r_{t-1}) < 0\}$ .

### 1.3 LHAR-RV-V 模型

Copeland<sup>[22]</sup> 提出了连续信息到达假说. 该假说认为市场信息是分步逐渐向外扩散的, 在市场信息不断传递的过程中, 引起价格波动和交易量的变化, 并随着新信息的不断增多, 价格波动和交易量同步增大. 为了进一步细致地考察价量关系, 在连续信息到达假说的基础上, 初步认为市场的信息在当期并没有被市场的波动完全吸收, 因而对未来市场的波动有一定影响. 于是, 从对市场波动预测的角度, 提出了异质结构交易量存在的可能性, 并将其加入到 LHAR-RV 模型中构建了 LHAR-RV-V 模型.

$$RV_{t+H} = \beta_0 + \beta_D RV_t^d + \beta_W RV_t^w + \beta_M RV_t^m + \rho_D r_t^{d-} + \rho_W r_t^{w-} + \rho_M r_t^{m-} + \lambda_D \log(V_t^d) + \lambda_W \log(V_t^w) + \lambda_M \log(V_t^m) + \varepsilon_{t+H} \quad (3)$$

金融资产收益波动率的预测倾向于使用对数已实现波动率, 因为对数已实现波动率的使用可以提高模型的拟合能力和预测精度, 因此, 可以将 LHAR-RV-V 模型拓展为

$$\begin{aligned} \log(RV_{t+H}) &= \beta_0 + \beta_D \log(RV_t^d) + \beta_W \log(RV_t^w) + \beta_M \log(RV_t^m) + \\ &\rho_D r_t^{d-} + \rho_W r_t^{w-} + \rho_M r_t^{m-} + \lambda_D \log(V_t^d) + \lambda_W \log(V_t^w) + \\ &\lambda_M \log(V_t^m) + \varepsilon_{t+H} \end{aligned} \quad (4)$$

其中  $V_t^d$  为  $t$  期的日交易量, 即  $V_t^d = \sum_{j=1}^{240} v_{ij}$  ( $v_{ij}$  表示  $t$  期内第  $j$  笔交易量);  $V_t^w$  为  $t$  期对应的周交易量, 即  $V_t^w = \frac{1}{5}(V_{t-5}^d + V_{t-4}^d + \dots + V_{t-1}^d)$ ;  $V_t^m$  为  $t$  期对应的月交易量, 即  $V_t^m = \frac{1}{20}(V_{t-20}^d + V_{t-19}^d + \dots + V_{t-1}^d)$ .

模型 (4) 为所构建的 LHAR-RV-V 模型, 该模型主要采用最小二乘法并结合 Newey West 协方差修正形式进行估计, 通过对相关系数的分析, 从市场微观结构的角度来进一步剖析我国股市波动的相关特征以及量价关系变化. 此外, 对量价关系的研究还会从波动率预测能力的角度进行分析, 以此来判断异质形式交易量的加入是否显著改善了模型的预测能力.

## 2 实证分析

### 2.1 数据的获取与预处理

选取的沪深 300 指数的 1 min 高频数据来源于分析家证券投资分析系统, 样本区间为 2008 年 7 月 1 日到 2010 年 8 月 19 日之间共 518 个交易日. 由于我国证券市场正常交易日的营业时间为 9:30 ~ 11:30、13:00 ~ 15:00, 共计 4 h 的交易时间, 因而按 1 min 为最小交易时间单位进行划分, 则每天将会得到 240 个有效数据. 数据指标包括交易时间、收盘价格、交易量. 研究中剔除了交易时间外记录的数据, 还剩下 125 433 笔交易数据. 根据  $r_i = \log(p_i) - \log(p_{i-1})$  来计算高频收益率, 每个交易日的成交量为所有 240 笔交易量之和, 即  $V_t^d = \sum_{j=1}^{240} v_{ij}$ . 由于价格离散和买卖价差等市场微观结构噪音的影响, 当抽样频率很高时, 已实现波动率是隐含波动率的有偏估计, 因此, 为了提高已实现波动率的估计精度, 采用一阶自相关偏差修正方法对已实现波动率的计算方法加以修正, 即

$$RV_t^d = \sqrt{\sum_{j=1}^n r_j^2 + 2 \sum_{j=2}^n r_j r_{j-1}}$$

表 1 记录了沪深 300 指数的已实现波动率和对数已实现波动率的描述性统计量, 从表 1 中可

以看出,与已实现波动率序列相比,对数已实现波动率序列的统计特征更接近正态分布,因此,从金

融资产收益波动率建模的角度考虑,将使用对数已实现波动率对模型(1)和(4)进行估计.

表 1 沪深 300 指数 RV 和 log(RV) 的描述性统计量

Table 1 Descriptive statistics for RV and log(RV) series of Shenzhen 300 index

	$RV_t^d$	$\log(RV_t^d)$	$RV_t^m$	$\log(RV_t^m)$	$RV_t^m$	$\log(RV_t^m)$
Mean	0.008 218	-4.904 248	0.008 238	-4.873 478	0.008 319	-4.846 495
Median	0.007 367	-4.910 749	0.007 722	-4.863 747	0.007 682	-4.868 876
Maximum	0.034 204	-3.375 413	0.024 884	-3.693 53	0.017 376	-4.052 665
Minimum	0.002 696	-5.915 986	0.003 07	-5.786 078	0.003 485	-5.659 287
Std. Dev.	0.004 124	0.444 457	0.003 338	0.384 678	0.002 872	0.339 668
Skewness	2.072 865	0.277 906	1.365 62	0.072 731	0.848 741	-0.02 676
Kurtosis	10.458 01	3.073 412	6.113 52	2.946 256	3.451 975	2.840 368
Jarque-Bera	1 510.787	6.522 067	355.938 7	0.498 983	64.028 86	0.588 198

为了保证所构建的 LHAR-RV-V 模型的有效性和可靠性,进而初步验证我国股市波动的持续性以及杠杆效应的持续性.在对已实现波动率进行自相关分析的同时,从沪深 300 指数的日收益

率中提取正收益率和负收益率,并将它们与已实现波动率进行相关性检验.为了更为直观的比较三者的变化,在此将负收益率取绝对值后再进行相关性分析.于是,便可以得到相关性分析图(图 1)

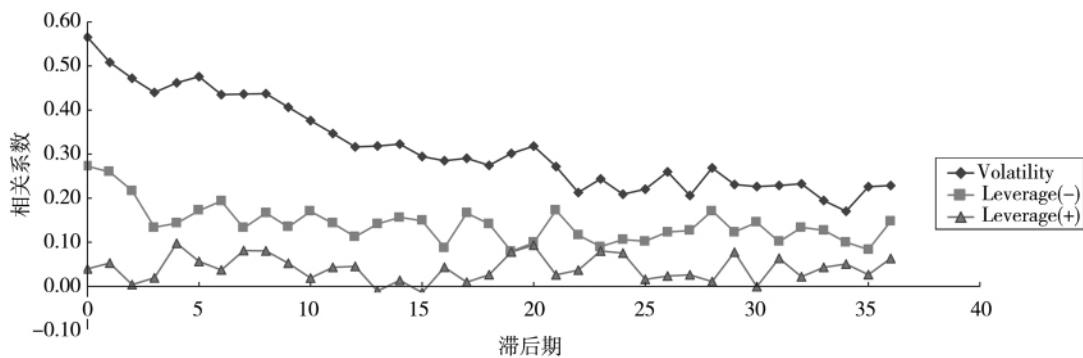


图 1 已实现波动率自相关及正、负收益与已实现波动率相关性分析

Fig. 1 Realized volatility autocorrelation, positive-negative returns and realized volatility correlation analysis

由图 1 所示,已实现波动率自相关性曲线(volatility)呈缓慢衰减过程,即已实现波动率之间的相关性强度随着时间慢慢减弱.由此可以初步推测我国股市波动率存在一定的长记忆性,已实现波动率之间相互影响具有持续性,即过去的已实现波动率对未来的已实现波动率在一定时间范围内存在着不可忽略的影响,从而可以在下文利用已实现波动率的长记忆性对模型(4)的预测能力做一个评价分析.另外,代表负收益率与已实现波动率的相关性曲线(Leverage(-))相比正收益率与已实现波动率的相关性曲线(Leverage(+))衰减的速度更慢、更平稳,即体

现利空消息变化的负收益相比利好消息的正收益对我国股市的冲击更大、持续时间更长.从而可以初步断定我国股市波动存在显著的杠杆效应,且这种杠杆效应具有一定的持续性.基于以上的初步分析认为非常有必要在高频数据的广泛应用背景下结合 LHAR-RV-V 模型对中国股票市场的杠杆效应与长记忆特征进行定量分析.

### 2.2 LHAR-RV-V 模型参数估计

由于 LHAR-RV-V 模型反映的是基于  $t$  期的异质形式的已实现波动率、收益率以及成交量对  $t + H$  期的已实现波动率的影响.因此,假如  $t$  期为当期的话,对于该模型中被解释变量  $RV_{t+H}$ ,当

$H = 1, 5, 10, 15, 20$  时, 即分别代表未来一天、一周、两周、三周、一个月的平均股市波动情况. 为了充分反映股市波动的相关特征, 对模型(4) 选取样本内不同区间( $H = 1, 5, 10, 15, 20$ ) 的已实现

波动率分别进行估计, 并在此基础上进行样本内和样本外预测分析. 主要采用最小二乘法对模型参数进行有效估计, 估计的结果如表 2 所示.

表 2 LHAR-RV-V 模型的参数估计结果

Table 2 Parameter estimation of the LHAR-RV-V model

Variable	Horizons				
	1 day	1 week	2 weeks	3 weeks	1 month
$\beta_0$	- 2. 114 406 ** [0. 830 91]	- 1. 284 567 * [0. 657 661]	- 0. 829 648 [0. 629 783]	- 0. 851 335 [0. 643 524]	- 0. 678 122 [0. 654 424]
$\beta_D$	0. 295 509 *** [0. 048 267]	0. 203 459 *** [0. 038 203]	0. 197 345 *** [0. 036 584]	0. 163 984 *** [0. 037 382]	0. 154 773 *** [0. 038 015]
$\beta_W$	0. 259 601 *** [0. 078 099]	0. 328 988 *** [0. 061 815]	0. 309 523 *** [0. 059 195]	0. 284 455 *** [0. 060 486]	0. 268 572 *** [0. 061 511]
$\beta_M$	0. 262 844 *** [0. 082 287]	0. 245 351 *** [0. 065 13]	0. 213 017 *** [0. 062 369]	0. 195 256 *** [0. 063 73]	0. 163 252 ** [0. 064 809]
$\rho_D$	- 2. 686 111 *** [0. 573 359]	- 1. 638 505 *** [0. 453 811]	- 1. 168 65 *** [0. 434 574]	- 0. 996 152 ** [0. 444 056]	- 0. 909 553 ** [0. 451 577]
$\rho_W$	- 3. 584 36 ** [1. 558 112]	- 3. 874 107 *** [1. 233 238]	- 3. 717 73 *** [1. 180 961]	- 3. 486 372 *** [1. 206 728]	- 2. 314 967 * [1. 227 168]
$\rho_M$	- 4. 032 042 [3. 209 633]	- 1. 665 596 [2. 540 408]	- 1. 824 358 [2. 432 72]	- 4. 322 752 * [2. 485 8]	- 6. 089 7 ** [2. 527 904]
$\lambda_D$	0. 183 726 *** [0. 067 87]	0. 126 868 ** [0. 053 719]	0. 153 765 *** [0. 051 442]	0. 133 372 ** [0. 052 564]	0. 134 232 ** [0. 053 454]
$\lambda_W$	0. 090 515 [0. 104 302]	0. 221 348 *** [0. 082 555]	0. 222 016 *** [0. 079 055]	0. 264 84 *** [0. 080 78]	0. 247 35 *** [0. 082 148]
$\lambda_M$	- 0. 211 733 *** [0. 077 511]	- 0. 340 093 *** [0. 061 35]	- 0. 407 786 *** [0. 058 749]	- 0. 449 862 *** [0. 060 031]	- 0. 458 098 *** [0. 061 048]
$R^2$	0. 589 509	0. 658 768	0. 651 112	0. 610 686	0. 574 340
RMSE	0. 002 963	0. 002 130	0. 001 953	0. 001 968	0. 001 932

注: 方括号内为系数的标准差, “\*\*\*”、“\*\*”、“\*” 分别指在 1%、5%、10% 的水平下显著.

由表 2 可知大多数参数的估计结果是显著的. 具体而言, 从单期来看, 系数  $\beta_D$ 、 $\beta_W$ 、 $\beta_M$  在 1% 的显著水平下都是非常显著的, 这说明了我国股市的整体波动在单期由不同周期波动相互叠加而成. 根据 Müller<sup>[8]</sup> 提出的异质市场假说理论, 由此也可以看出我国的股市波动受过去不同成分波动的共同影响, 且这些不同成分的波动由不同持有期限的投资者行为所引起的, 从而这一结果进一步证实了我国股票市场交易者异质性的存在. 此外, 如图 2 所示, 从多期来看随着样本内滚动时间窗的不断延长  $\beta_D$ 、 $\beta_W$ 、 $\beta_M$  系数值都呈现一个逐

步衰减的态势, 这说明我国股市波动存在明显的长记忆性, 即过去的股市波动在持续地影响着未来的股市波动, 这与房振明等<sup>[23]</sup> 基于高频数据所揭示的中国股市长记忆性特征不谋而合. 另外,  $\beta_W$ 、 $\beta_M$  在估计的有效区间内都明显大于  $\beta_D$  值, 这说明在长期内股市的波动主要由中长期投资者的交易行为所决定. 也即, 中长期投资者的交易行为在未来很长一段时间里不仅影响着自身的投资决策, 而且对短期投资者的行为产生不可忽略的影响. 因此, 从实证的角度也可以发现中国股市的投机成分不断减少、投资者不断趋于理性.

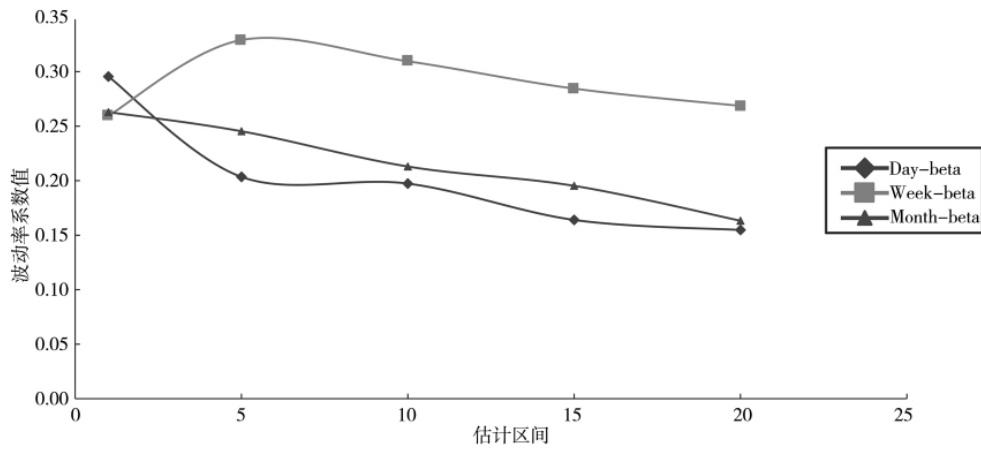


图2 不同已实现波动率估计期限波动率系数  $\beta$  的变化情况

Fig. 2 Volatility coefficients  $\beta$  change along with different realized volatility estimated periods

为了判断和分析过去不同周期的负收益对波动率的影响,可以通过对模型(4)中  $\rho_D$ 、 $\rho_W$ 、 $\rho_M$  的估计结果来加以评价. 由表2显示系数  $\rho_D$ 、 $\rho_W$ 、 $\rho_M$  的估计结果总体是显著的,即不同周期的负收益对未来波动率有着显著的影响,从单期来看,不仅过去的日负收益对未来一天的平均波动率有明显的影响(杠杆效应),而且过去周负收益和月负收益对未来的平均波动率有着不可忽略的影响,这证实了我国股票市场存在异质结构的杠杆效应,进而从非对称性的角度进一步证实了 Müller<sup>[8]</sup> 提出的异质市场假说. 此外,从多期来看  $\rho_D$  和  $\rho_W$  值在估计的有效区间内随着期限的延长而逐渐递减;而代表月负收益影响的  $\rho_M$  则在短期内对波动率的影响不明显,伴随估计区间的扩展,

其显著性逐渐增强且值的大小保持稳定增长如图3所示. 这说明我国股市中杠杆效应存在一定的持续性,市场中“利空”消息在一段时间里的叠加,对未来股市波动产生了持续性的影响,且这种影响在日和周为单位的负收益中日渐变弱,这说明“利空”消息在随着时间而被市场消化吸收了. 而在以月为单位的负收益的影响在特定期间集中不断放大,这可以从我国股市严格的卖空限制的交易制度及投资者异质信念来解释:卖空限制使得股票价格刚开始只反映了乐观投资者的预期,悲观投资者的一部分私人信息因无法及时进入市场而慢慢累积,一旦这种累积的隐藏信息在市场走弱过程中被释放出来,就有可能造成股市连续下跌甚至发生暴跌.

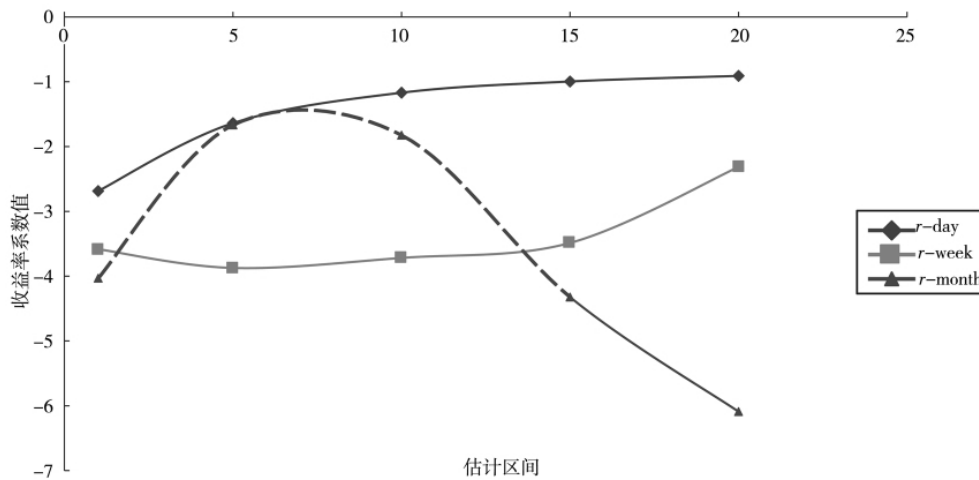


图3 不同已实现波动率估计期限收益率系数  $\rho$  的变化情况

Fig. 3 Return coefficients  $\rho$  change along with different realized volatility estimated periods

注: 图中虚线部分代表对市场的波动没有影响

对于量价关系的研究,通过引入过去异质形式的交易量来反映过去的交易量对未来波动的影响。由表2估计结果显示 $\lambda_D$ 、 $\lambda_W$ 、 $\lambda_M$ 等系数都是非常显著的,这说明了过去的交易量对未来的波动率存在一定程度的影响,交易量作为一个不可观测的信息代理变量,新的信息进入市场后是分步逐渐向外扩散的,在市场信息不断传递的过程中,引起价格和交易量的变化,并随着新信息的不断增多,价格波动和交易量同步增大。然而,新信息在传递过程中并没有被当期的市场波动完全吸收并加以充分反映,而是对未来的市场波动产生不同程度的影响,从而进一步验证了Copeland<sup>[22]</sup>提出的连续信息到达假说。具体而言,反映过去日、周交易量对未来波动的影响系数 $\lambda_D$ 和 $\lambda_W$ 估计结果都为正,说明了过去一部分未被消化的信息作为未预期信息在未来的市场中得到了反映,从而进一步加剧了未来的股市波动。尽管如此,反映过去月交易量对股市波动的影响系数 $\lambda_M$ 为负,这说明了一部分未被消化的市场信息作为预期信息在未来的市场中降低了投资者的交易预期,使一部分投资者的行为逐渐趋于理性,从而减缓了未来股市的波动。这与陈怡玲和宋逢明<sup>[24]</sup>实证研究得出中国股市存在不对称的交易量—价格变动关系结论相一致。

### 2.3 LHAR-RV-V 模型预测能力评价

以ARFIMA模型作为参照模型通过选取常见的评价指标从样本内预测和样本外预测两方面对模型HAR-RV与LHAR-RV-V的预测效果进行比较。评价的指标有:根均方误差(root mean squared error, RMSE)、平均绝对值误差(mean absolute error, MAE)、平均绝对误差百分比(mean absolute percentage error, MAPE)、Theil不等系数(Theil inequality coefficient, TIC)。若这四个指标越小,说明模型的预测值和真实值的差异就越小,预测效果就越好,反之亦然。

为了对模型的预测能力进行有效的评价,将整个样本区间分为两个部分,即前400个指标观察值作为模型的估计区间,后118个指标观测值作为模型的预测区间。从样本内预测角度来分析,通过利用对数已实现波动率分别对ARFIMA模型、HAR-RV模型及LHAR-RV-V模型在估计区间内进行有效估计,在此基础上进行样本内预测分

析,从而得出各评价指标值如表3所示。此外,在样本内预测分析的基础上,结合样本的预测区间,对三个模型的样本外预测能力进行分析,于是得出各模型样本外预测评价指标值如表4所示。

表3 各模型样本内预测能力分析

Table 3 Predictive power analysis in sample of each model

评价指标	ARFIMA	HAR-RV	LHAR-RV-V
RMSE	0.004 06	0.003 073	0.002 979
MAE	0.002 845	0.002 015	0.001 935
MAPE	38.790 23	23.295 85	22.187 86
TIC	0.233 958	0.173 408	0.167 155

表4 各模型样本外预测能力分析

Table 4 Predictive power analysis out of sample of each model

评价指标	ARFIMA	HAR-RV	LHAR-RV-V
RMSE	0.003 907	0.001 862	0.001 832
MAE	0.003 008	0.001 328	0.001 307
MAPE	42.067 490	21.267 810	21.186 160
TIC	0.405 491	0.147 072	0.143 790

通过对表3、表4各评价指标数据的整理和分析,发现与参照模型ARFIMA相比,HAR-RV和LHAR-RV-V模型无论是样本内预测能力还是样本外预测能力都明显优于ARFIMA模型。这也进一步证实了张波、钟玉洁等<sup>[19]</sup>和唐勇、池云果<sup>[20]</sup>研究得出HAR-RV模型的预测能力要比ARFIMA模型更胜一筹的结论。此外,还可以清楚地看到同时考虑异质形式的杠杆效应和交易量的LHAR-RV-V模型的RMSE、MAE、MAPE、TIC的值要略小于HAR-RV模型所对应的值,这充分说明了在HAR-RV模型中同时加入异质结构的负收益和过去不同周期的交易量可以在一定程度上改善原有模型的预测能力,即LHAR-RV-V模型有着更为精确的波动预测能力。这也进一步证实了过去交易量中未被吸收的信息有利于预测未来市场的波动情况。

## 3 结束语

在异质市场假说的理论上联系市场微观结构中的量价关系,构造了LHAR-RV-V模型,并利用高频1min数据将其在中国股市进行了实证

分析. 研究结果显示我国股票市场的波动具有明显的长记忆性和杠杆效应,且这杠杆效应具有一定的持续性和异质性,从而也进一步验证了 Müller<sup>[8]</sup> 提出的异质市场假说. 此外,对于量价关系的研究表明我国股票市场的信息在传递过程中并没有完全被当期的市场波动吸收并加以反映,

而是对未来股市的波动存在着一定的影响,这一结论可以从考虑滞后不同周期交易量后的 LHAR-RV-V 模型的预测能力相比 ARFIMA 和 HAR-RV 模型获得了改善中得到进一步证实,从而也进一步验证了 Copeland<sup>[22]</sup> 提出的连续信息到达假说.

#### 参 考 文 献:

- [1] Andersen T G, Bollerslev T. Answering the critics: Yes, ARCH models do provide good volatility forecasts [J]. *International Economic Review*, 1998, (4): 885 - 905.
- [2] Andersen T G, Bollerslev T, et al. Exchange rate returns standardized by realized volatility are (nearly) gaussian [J]. *Multinational Finance Journal*, 2000, (4): 159 - 179.
- [3] Andersen T G, Bollerslev T, Diebold F. The distribution of exchange rate volatility [J]. *Journal of American Statistical Association*, 2001, (96): 42 - 55.
- [4] Andersen T G, Bollerslev T, et al. The distribution of stock return volatility [J]. *Journal of Financial Economics*, 2001, (61): 43 - 76.
- [5] Andersen T G, Bollerslev T, Diebold F X, et al. Modeling and forecasting realized volatility [J]. *Econometrica*, 2003, 71 (2): 579 - 625.
- [6] 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 (3): 445 - 475.
- [7] Corsi F. A Simple Long Memory Model of Realized Volatility [R]. University of Southern Switzerland, 2004, <http://ssrn.com/abstract=626064>.
- [8] Müller U, Dacorogna M, Dave R, et al. Fractals and Intrinsic Time: A Challenge to Econometricians [R]. Olsen and Associates, 1993.
- [9] Andersen T G, Bollerslev T, Diebold F X. Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility [R]. NBER, 2005.
- [10] McAleer M, Medeiros M C. A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries [J]. *Journal of Econometrics*, 2008, (147): 104 - 119.
- [11] Louzis S. Forecasting stock index realized volatility with an asymmetric HAR-FIGARCH model: The case of S&P500 and DJIA stock indices [EB/OL]. 2009 <http://ssrn.com/abstract=1524861>.
- [12] Corsi R. HAR volatility modeling with heterogeneous leverage and jumps [EB/OL]. <http://ssrn.com/abstract=1316953>, 2009.
- [13] Liu Chun, Maheu J M. Intraday dynamics of volatility and duration: Evidence from the Chinese stock market [R]. Working Paper. Canada, University of Toronto Department of Economic, 2010, 401.
- [14] Corsi F. A simple long memory model of realized volatility [J]. *Journal of Financial Econometrics*, 2009, 7(2): 174 - 196.
- [15] 黄后川, 陈浪南. 股票市场波动率的高频估计与特性分析 [J]. *经济研究*, 2003, (2): 74 - 84.  
Huang Houchuan, Chen Langnan. High frequency estimation and feature analysis of stock market volatility [J]. *Economic Research Journal*, 2003, (2): 74 - 84. (in Chinese)
- [16] 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, 22(8): 60 - 63.  
Xu Zhengguo, Zhang Shiyong. The comparative research on volatility prediction ability of adjusted realized volatility, GARCH model and SV model [J]. *Systems Engineering*, 2004, 22(8): 60 - 63. (in Chinese)
- [17] 徐正国, 张世英. 多维高频数据的“已实现”波动建模研究 [J]. *系统工程学报*, 2006, 21(1): 6 - 11.  
Xu Zhengguo, Zhang Shiyong. Research of “realized” volatility modeling with multidimensional high-frequency data [J].



- Journal of Systems Engineering , 2006 , 21( 1) : 6 - 11. ( in Chinese)
- [18] 魏 宇. 沪深 300 股指期货的波动率预测模型研究 [J]. 管理科学学报 , 2010 , 13( 2) : 66 - 76.  
Wei Yu. Volatility forecasting models for CSI300 index futures [J]. Journal of Management Sciences in China , 2010 , 13( 2) : 66 - 76. ( in Chinese)
- [19] 张 波, 钟玉洁, 田金方. 基于高频数据的沪指波动的长记忆性驱动因素分析 [J]. 统计与信息论坛 , 2009 , 24( 6) : 21 - 26.  
Zhang Bo , Zhong Yujie , Tian Jinfang. Driving factors analysis for prev close volatility long memory characteristic based on high frequency data [J]. Statistic and Information Forum , 2009 , 24( 6) : 21 - 26. ( in Chinese)
- [20] 唐 勇, 池云果. 基于已实现波动率的长记忆性分析 [J]. 福州大学学报 , 2010 , ( 5) : 27 - 32.  
Tang Yong , Chi Yunguo. Long memory analysis of realized volatility [J]. Journal of Fuzhou University , 2010 , ( 5) : 27 - 32. ( in Chinese)
- [21] 曾增展. 基于高频数据股票市场已实现波动率预测研究 [D]. 台湾: 中国台湾国立云林科技大学, 2009.  
Tseng-Chan Tseng. Predicting Realized Volatility Based on High Frequency Data in Stock Markets [D]. Taiwan: National Yulin University of Science and Technology of Chinese Taiwan , 2009. ( in Chinese)
- [22] Copeland T E. A model of asset trading under the assumption of sequential information arrival [J]. Journal of Finance , 1976 , ( 31) : 1149 - 1168.
- [23] 房振明, 王春峰, 蒋祥林. 中国股市回报波动性分析—基于高频数据揭示股市的特征 [J]. 系统工程 , 2004 , 22( 2) : 13 - 19.  
Fang Zhenming , Wang Chunfeng , Jiang Xianglin. Return volatility analysis of Chinese stock market: Based on stock market characteristics revealed by high frequency data [J]. System Engineering , 2004 , 22( 2) : 13 - 19. ( in Chinese)
- [24] 陈怡玲, 宋逢明. 中国股市价格变动与交易量关系的实证研究 [J]. 管理科学学报 , 2000 , 3( 2) : 62 - 68.  
Chen Yiling , Song Fengming. The empirical research of relations between price and volume in Chinese stock market [J]. Journal of Management Sciences in China , 2000 , 3( 2) : 62 - 68. ( in Chinese)

## Research on China ' s stock market fluctuations based on LHAR-RV-V model

WEN Feng-hua<sup>1 3</sup> , LIU Xiao-qun<sup>1</sup> , TANG Hai-ru<sup>1</sup> , YANG Xiao-guang<sup>1 2 3</sup>

1. School of Economics and Management , Changsha University of Science and Technology , Changsha 410114 , China;
2. Academy of Mathematics and Systems Science , Chinese Academy of Sciences , Beijing 100190 , China;
3. Financial Engineering and Financial Management Research Center of Hunan Province , Changsha 410114 , China

**Abstract:** On the basis of heterogeneous autoregressive realized volatility model( HAR-RV model) , combining the market microstructure theory , the paper proposes a new dynamic volatility model which considers both the leverage effect and price-volume relation and which is called Leverage Heterogeneous Autoregressive Realized Volatility with Volume model( LHAR-RV-V model) . The proposed model is applied to empirical analysis with the 1 minute high frequency data of Shenzhen 300 index. The empirical results show that the model can well capture the long memory and the leverage effect characteristics , and the leverage effect has certain sustainability. In addition , the past different cycle volume joined in the model can not only reflect the relationship between the volume and price in a more sophisticated way , but also improve the predictive power of the model to some degree.

**Key words:** heterogeneous market hypothesis; LHAR-RV-V model; volume; leverage effect