

引入测量误差和马尔科夫转换机制的 高频波动率建模与预测^{*}

李红权 张馨心

(湖南师范大学商学院, 长沙 410081)

摘要 基于现实金融市场同时存在已实现波动率测量误差问题以及市场结构突变现象, 将马尔科夫转换机制引入经测量误差修正的模型, 构建了具有时变特征的 MRS-HARQ 族模型, 以期进一步提高 HAR 框架下波动率模型的预测精度. 以沪深 300 指数 2005 年 4 月 – 2021 年 7 月 5 分钟高频数据为样本的实证研究证实: 相比于 HAR 和 HARQ(F) 族模型, 结合测量误差和马尔科夫转换机制构建的 MRS-HARQ 族模型具有更优的拟合效果和预测能力, 能够更准确地刻画中国股市真实波动的复杂特征, 并发现高波动状态持续的时间都不长; 更进一步在文章所探讨的 16 种高频波动率模型中, 结合已实现正负半方差的 MRS-HARQ-RS 模型和结合符号跳跃的 MRS-HARQ-SJ 模型拥有最佳的样本外预测能力. 研究结论在金融风险建模、资产定价和金融预测领域有较好的应用价值.

关键词 HAR-RV, 测量误差, 马尔科夫转换机制, 波动率预测, MCS 检验.

MR(2000) 主题分类号 91G20, 91G70

DOI 10.12341/jssms21557

The Forecasting Performance of the High-Frequency Volatility Models Based on Errors and Markov Regime-Switching

LI Hongquan ZHANG Xinxin

(School of Business, Hunan Normal University, Changsha 410081)

Abstract Based on the existence of both realized volatility measurement errors and sudden changes in market structure in real financial markets, the Markov regime-switching is introduced into the model corrected by errors, and the MRS-HARQ-type models with time-varying characteristics are constructed to further improve the forecasting accuracy of volatility models under the HAR framework. The empirical study with the sample of 5-minute high-frequency data of the CSI 300 index from April

^{*} 国家自然科学基金面上项目 (71871092), 宏观经济大数据挖掘与应用湖南省重点实验室资助课题.

收稿日期: 2021-10-03, 收到修改稿日期: 2021-11-27.

编委: 文凤华.

2005 to July 2021 confirms that, compared with the HAR-type and HARQ(F)-type models, the MRS-HARQ-type models constructed by combining errors and Markov regime-switching have better goodness-of-fit and forecasting performance, and can more accurately portray the complex characteristics of the real volatility of China's stock market, and finds that the high volatility states do not last long; furthermore among the 16 high-frequency volatility models explored in this paper, the MRS-HARQ-RS model combining the realized semivariance and the MRS-HARQ-SJ model combining signed jump possess the best out-of-sample forecasting performance. The findings have good applications in the fields of financial risk modeling, asset pricing and financial forecasting.

Keywords HAR-RV, errors, Markov regime-switching, volatility forecasting, MCS test.

1 引言

股市波动率的建模和预测对金融资产定价、风险管理及投资组合有着极为重要的意义. 有关股市波动率的早期研究主要集中在低频数据领域, 代表性成果有 ARCH 模型^[1]、GARCH 模型^[2]和 SV 模型^[3]. 随着高频数据越来越容易获取, 基于高频数据的波动率建模成为新的研究方向. 在高频波动率建模领域, 已实现波动率^[4] (realized volatility, RV) 是度量真实波动率最有代表性的方法. 为了更好地对波动率进行预测, Corsi^[5] 考虑波动率的长记忆性和异质性, 提出了已实现波动率异质自回归模型 (heterogeneous auto regressive model of realized volatility, HAR-RV). 与传统的低频模型相比, HAR-RV 模型对未来波动率具有更好的预测能力, 得到了广泛认可和应用.

之后, 学者们主要围绕如何提高模型预测精度对 HAR-RV 模型进行了扩展研究. 代表性研究有, Andersen 等^[6] 考虑跳跃因素, 构建了 HAR-J 模型; Barndorff-Nielsen 等^[7] 将 RV 分解成已实现正半方差和负半方差; Patton 和 Sheppard^[8] 在此基础上构建了 HAR-RS 模型, 发现其预测能力优于 HAR-J 模型; 进一步, 提出符号跳跃概念并将其引入模型, 构建了 HAR-SJ 模型. 国内研究也基于 HAR 及其扩展模型对金融市场的波动率进行预测, 均证明了 HAR 族模型预测精度的优越性^[9-11].

虽然上述 HAR 族模型对未来波动率具有较好的预测能力, 但这些模型没有考虑 RV 估计中测量误差的影响. 依据二次变差理论, 当抽样频率无限大时, RV 是 IV 的一致估计量. 但由于现实中抽样间隔无法趋近于零, RV 与 IV 之间总存在测量误差, 且误差越大, RV 估计越不准确. 因此, 在建模中考虑测量误差因素尤为必要. Bollerslev 等^[12] 基于渐进分布理论, 将已实现四次变差 (realized quarticity, RQ) 对 HAR 模型进行测量误差修正, 构建了 HARQ 模型. 已有文献^[13-16] 对 HARQ 模型进行了扩展, 更全面地分析测量误差对模型预测的作用. 此外, 由于宏观经济政策改变、突发事件以及经济周期等因素, 市场波动率可能出现结构突变或机制转化, 从而使得上述固定参数模型可能出现估计和预测偏误. 为此, 很多学者将马尔科夫转换机制引入到波动率分析与预测, 为构建具有机制转换的波动率模型提供了理论和实证依据^[17-21]. 基于现实市场同时存在已实现波动率测量误差的问题以及金融市场结构突变现象, 本文尝试将基于测量误差修正的波动率模型与马尔科夫转换机制相结合, 以克

服原有固定参数模型未考虑测量误差时变特征和波动率结构变化的不足,以期进一步提高 HAR 框架下波动率模型的预测精度,并捕捉市场波动率的动态变化特征.

具体而言,本文利用已实现四次变差 RQ 对 HAR 族 (HAR-RV、HAR-J、HAR-RS 和 HAR-SJ) 模型进行修正,考察了测量误差对已实现波动率预测的影响,在此基础上,将马尔科夫转换机制引入经测量误差修正的模型,构建了 MRS-HARQ 族模型,以提高模型的预测精度.以我国沪深 300 指数 5 分钟高频数据为样本的实证研究表明, MRS-HARQ 族模型具有更好的拟合效果和预测能力,能够更准确地刻画我国股市真实波动情况.

2 预测模型

2.1 HAR-RV 族模型

Andersen 等^[6]定义已实现波动率为日内高频收益率的平方和

$$RV_t = \sum_{j=1}^{1/\Delta} r_{(t-1)+j\cdot\Delta, \Delta}^2. \quad (2.1)$$

基于异质市场假说, Corsi^[5] 构建了 HAR-RV 模型, 其表达式为

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \varepsilon_t. \quad (2.2)$$

式中 RV_{t-1} 、 $RV_{t-1|t-5}$ 和 $RV_{t-1|t-22}$ 分别代表日、周均和月均已实现波动率, 即 $RV_{t-j|t-h} = \frac{1}{h} \sum_{i=j}^h RV_{t-i}$.

考虑跳跃波动, Andersen 等^[6] 构建了 HAR-J 模型

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_4 J_{t-1} + \varepsilon_t. \quad (2.3)$$

式中 $J_t = \max(RV_t - BV_t, 0)$ 表示跳跃, $BV_t = u_1^{-2} \sum_{j=2}^{1/\Delta} (|r_{(t-1)+j\cdot\Delta}| |r_{(t-1)+(j-1)\cdot\Delta}|) \rightarrow \int_0^t \sigma^2(s) ds$ 为已实现二次幂变差, $u_1 = (2/\pi)^{0.5} \simeq 0.7979$.

Barndorff-Nielsen 等^[7] 提出了已实现正负半方差

$$RS^+ = \sum_{i=1}^n r_i^2 I\{r_i > 0\}, \quad (2.4)$$

$$RS^- = \sum_{i=1}^n r_i^2 I\{r_i < 0\}, \quad (2.5)$$

其中, RS^+ 为已实现正半方差, RS^- 为已实现负半方差, 分别测量了与正负收益相对应的波动率.

Patton 和 Sheppard^[8] 基于已实现半方差构建了 HAR-RS 模型, 并证明 HAR-RS 模型比 HAR-J 模型预测效果更好, 其表达式为

$$RV_t = \beta_0 + \beta_1^+ RS_{t-1}^+ + \beta_1^- RS_{t-1}^- + \beta_2^+ RS_{t-1|t-5}^+ + \beta_2^- RS_{t-1|t-5}^- + \beta_3^+ RS_{t-1|t-22}^+ + \beta_3^- RS_{t-1|t-22}^- + \varepsilon_t. \quad (2.6)$$

进一步, Patton 和 Sheppard^[8] 提出符号跳跃 $SJ \equiv RS^+ - RS^-$, 并构建 HAR-SJ 模型

$$RV_t = \beta_0 + \beta_{1,SJ} SJ_{t-1} + \beta_{1,BV} BV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \varepsilon_t. \quad (2.7)$$

2.2 HARQ(F) 族模型

由于微观结构噪声的存在, RV 与 IV 间存在测量误差 η_t , 即 $RV_t = IV_t + \eta_t$. 测量误差可以根据 Barndorff-Nielsen 和 Shephard^[22] 提出的渐进分布理论来描述, $\eta_t \sim N(0, 2\Delta IQ_t)$, 其中 IQ 为积分四次变差. IQ 可以用已实现四次变差 RQ 进行一致地估计, 其定义为

$$RQ_t = \frac{n}{3} \sum_{i=1}^n r_{t,i}^4. \quad (2.8)$$

Bollerslev 等^[12] 构造了随时间变化的回归系数 $\beta_{i,t} = \beta_i + \beta_{i,Q} RQ_t^{\frac{1}{2}}$, 可以直接根据相应测量误差的大小成比例调整系数, 对 HAR-RV 模型中的日、周均和月均波动率进行测量误差修正, 得到 HARQF 模型

$$\begin{aligned} RV_t = & \beta_0 + (\beta_1 + \beta_{1,Q} RQ_{t-1}^{\frac{1}{2}}) RV_{t-1} + (\beta_2 + \beta_{2,Q} RQ_{t-1|t-5}^{\frac{1}{2}}) RV_{t-1|t-5} \\ & + (\beta_3 + \beta_{3,Q} RQ_{t-1|t-22}^{\frac{1}{2}}) RV_{t-1|t-22} + \varepsilon_t, \end{aligned} \quad (2.9)$$

$$RQ_{t-1|t-k} = \frac{1}{k} \sum_{j=1}^k RQ_{t-j}. \quad (2.10)$$

日、周和月均波动率的回归系数会随测量误差的增大而减小, 从而降低测量误差对预测的影响. 时间跨度 k 越长, $RQ_{t-1|t-k}$ 的解释能力会减弱, 仅修正解释能力较强的日波动率得到 HARQ 模型

$$RV_t = \beta_0 + (\beta_1 + \beta_{1,Q} RQ_{t-1}^{\frac{1}{2}}) RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \varepsilon_t. \quad (2.11)$$

HARQ 模型背后的动机同样适用于其他基于已实现波动率的预测模型. 特别是对于公式 (2.3) 中定义的 HAR-J 模型, 很容易以类似于 HARQ 模型的方式进行修正, 从而得到 HARQ(F)-J 模型

$$RV_t = \beta_0 + (\beta_1 + \beta_{1,Q} RQ_{t-1}^{\frac{1}{2}}) RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_4 J_{t-1} + \varepsilon_t, \quad (2.12)$$

$$\begin{aligned} RV_t = & \beta_0 + (\beta_1 + \beta_{1,Q} RQ_{t-1}^{\frac{1}{2}}) RV_{t-1} + (\beta_2 + \beta_{2,Q} RQ_{t-1|t-5}^{\frac{1}{2}}) RV_{t-1|t-5} \\ & + (\beta_3 + \beta_{3,Q} RQ_{t-1|t-22}^{\frac{1}{2}}) RV_{t-1|t-22} + \beta_4 J_{t-1} + \varepsilon_t. \end{aligned} \quad (2.13)$$

在 HAR-RS 和 HAR-SJ 模型中, 变量 RV^+ 、 RV^- 和 SJ 的渐近分布是未知的, 但一个与其渐近方差相关性很强的变量在对其系数进行测量误差修正方面应该很有效, 因为回归系数可以自动弥补任何比例差异, 所以可以用 RQ 来代替 RV^+ 、 RV^- 和 SJ 的渐近方差, 从而对其进行测量误差修正. 由此定义 HARQ(F)-RS 和 HARQ(F)-SJ 模型

$$\begin{aligned} RV_t = & \beta_0 + (\beta_1^+ + \beta_{1,Q}^+ RQ_{t-1}^{\frac{1}{2}}) RS_{t-1}^+ + (\beta_1^- + \beta_{1,Q}^- RQ_{t-1}^{\frac{1}{2}}) RS_{t-1}^- + \beta_2^+ RS_{t-1|t-5}^+ \\ & + \beta_2^- RS_{t-1|t-5}^- + \beta_3^+ RS_{t-1|t-22}^+ + \beta_3^- RS_{t-1|t-22}^- + \varepsilon_t, \end{aligned} \quad (2.14)$$

$$\begin{aligned} RV_t = & \beta_0 + (\beta_1^+ + \beta_{1,Q}^+ RQ_{t-1}^{\frac{1}{2}}) RS_{t-1}^+ + (\beta_1^- + \beta_{1,Q}^- RQ_{t-1}^{\frac{1}{2}}) RS_{t-1}^- + (\beta_2^+ + \beta_{2,Q}^+ RQ_{t-1|t-5}^{\frac{1}{2}}) RS_{t-1|t-5}^+ \\ & + (\beta_2^- + \beta_{2,Q}^- RQ_{t-1|t-5}^{\frac{1}{2}}) RS_{t-1|t-5}^- + (\beta_3^+ + \beta_{3,Q}^+ RQ_{t-1|t-22}^{\frac{1}{2}}) RS_{t-1|t-22}^+ \\ & + (\beta_3^- + \beta_{3,Q}^- RQ_{t-1|t-22}^{\frac{1}{2}}) RS_{t-1|t-22}^- + \varepsilon_t, \end{aligned} \quad (2.15)$$

$$RV_t = \beta_0 + (\beta_{1,SJ} + \beta_{1,Q} RQ_{t-1}^{\frac{1}{2}}) SJ_{t-1} + \beta_{1,BV} BV_{t-1} + \beta_2 RV_{t-1|t-5}$$

$$+ \beta_3 RV_{t-1|t-22} + \varepsilon_t, \quad (2.16)$$

$$\begin{aligned} RV_t = & \beta_0 + (\beta_{1,SJ} + \beta_{1,Q} RQ_{t-1}^{\frac{1}{2}}) SJ_{t-1} + \beta_{2,BV} BV_{t-1} + (\beta_{2,SJ} + \beta_{2,Q} RQ_{t-1|t-5}^{\frac{1}{2}}) SJ_{t-1|t-5} \\ & + \beta_{2,BV} BV_{t-1|t-5} + (\beta_{3,SJ} + \beta_{3,Q} RQ_{t-1|t-22}^{\frac{1}{2}}) SJ_{t-1|t-22} \\ & + \beta_{3,BV} BV_{t-1|t-22} + \varepsilon_t. \end{aligned} \quad (2.17)$$

2.3 MRS-HARQ 族模型

为进一步提高模型的拟合精度和预测能力, 考虑中国股市波动中存在的机制转换现象, 本文将马尔科夫转换机制引入到形式更简单、待估参数更少的 HARQ 族模型, 构建 MRS-HARQ 族模型, 设定两种机制—高波动和低波动.

首先, 定义 MRS-HARQ 模型

$$\begin{aligned} RV_t = & \beta_0^{s_i} + (\beta_1^{s_i} + \beta_{1,Q}^{s_i} RQ_{t-1}^{\frac{1}{2}}) RV_{t-1} + \beta_2^{s_i} RV_{t-1|t-5} + \beta_3^{s_i} RV_{t-1|t-22} \\ & + \varepsilon_t^{s_i}, \varepsilon_t^{s_i} \sim (0, \sigma_{\varepsilon, s_i}^2), \end{aligned} \quad (2.18)$$

其中, s_i 为二值变量, i 取值为 1 表示与动荡期相关的高波动率状态, 取值为 2 时表示平稳时期相关的低波动率状态. 两状态下的转移概率和转移概率矩阵为

$$p_{11} = p(s_i = 1 | s_{i-1} = 1), \quad p_{22} = p(s_i = 2 | s_{i-1} = 2), \quad P = \begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix}. \quad (2.19)$$

类似地, 定义 MRS-HARQ-J、MRS-HARQ-RS 和 MAR-HARQ-SJ 模型

$$\begin{aligned} RV_t = & \beta_0^{s_i} + (\beta_1^{s_i} + \beta_{1,Q}^{s_i} RQ_{t-1}^{\frac{1}{2}}) RV_{t-1} + \beta_2^{s_i} RV_{t-1|t-5} + \beta_3^{s_i} RV_{t-1|t-22} \\ & + \beta_4^{s_i} J_{t-1} + \varepsilon_t^{s_i}, \end{aligned} \quad (2.20)$$

$$\begin{aligned} RV_t = & \beta_0^{s_i} + (\beta_1^{+,s_i} + \beta_{1,Q}^{+,s_i} RQ_{t-1}^{\frac{1}{2}}) RS_{t-1}^+ + (\beta_1^{-,s_i} + \beta_{1,Q}^{-,s_i} RQ_{t-1}^{\frac{1}{2}}) RS_{t-1}^- + \beta_2^{+,s_i} RS_{t-1|t-5}^+ \\ & + \beta_2^{-,s_i} RS_{t-1|t-5}^- + \beta_3^{+,s_i} RS_{t-1|t-22}^+ + \beta_3^{-,s_i} RS_{t-1|t-22}^- + \varepsilon_t^{s_i}, \end{aligned} \quad (2.21)$$

$$\begin{aligned} RV_t = & \beta_0^{s_i} + (\beta_{1,SJ}^{s_i} + \beta_{1,Q}^{s_i} RQ_{t-1}^{\frac{1}{2}}) SJ_{t-1} + \beta_{1,BV}^{s_i} BV_{t-1} + \beta_2^{s_i} RV_{t-1|t-5} \\ & + \beta_3^{s_i} RV_{t-1|t-22} + \varepsilon_t^{s_i}. \end{aligned} \quad (2.22)$$

参考 Hamilton 和 Susmel^[23] 提出的滤波过程和 Kim^[24] 提出的平滑算法, 本文采用极大似然估计方法对 MRS-HARQ 族模型参数估计时, 极大似然估计函数表达式如下

$$\begin{aligned} \ln L = & \sum_{t=1}^T \ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(\frac{RV_t - \beta_0^{s_i} - (\beta_1^{s_i} + \beta_{1,Q}^{s_i} RQ_{t-1}^{\frac{1}{2}}) RV_{t-1} - \beta_2^{s_i} RV_{t-1|t-5} - \beta_3^{s_i} RV_{t-1|t-22}}{-2\sigma^2} \right) \right). \end{aligned} \quad (2.23)$$

3 实证分析

3.1 数据说明

本文以沪深 300 指数为研究对象, 选取 2005 年 4 月 8 日 (首次发布) 至 2021 年 7 月 31 日的日内 5 分钟高频时间序列数据进行建模和预测, 共 3967 个交易日, 每个交易日有 48 个

5 分钟数据, 数据共计 190, 416 条. 数据来源 Wind 数据库, 使用 R 处理数据. 为便于观测, 将对数收益率放大 100 倍, 即 $r_t = 100(\ln p_t - \ln p_{t-1})$, 图 1 和图 2 为沪深 300 指数的收盘价序列和对数收益率序列.

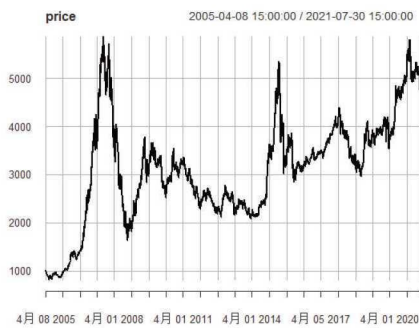


图 1 收盘价序列

(Figure 1 Closing price series)

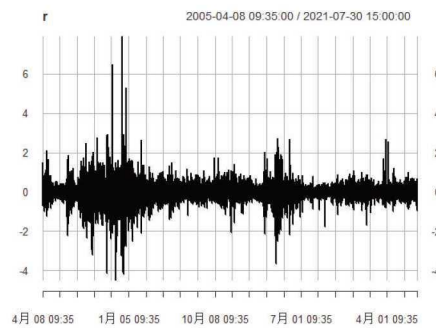


图 2 对数收益率序列

(Figure 2 Logarithmic return series)

图 1 显示沪深 300 指数从 2006 年起快速上涨, 于 2017 年 10 月达到历史最高点, 之后又持续下跌, 至 2008 年 11 月跌破 2000 点后开始回升, 直至 2009 年 7 月这一轮牛市结束; 此外, 2014 年至 2015 年也发生了一轮极端牛熊市; 可以看到右图的对数收益率在这两段时间波动明显变大.

3.2 高频波动率特征分析

图 3 和图 4 分别描述了已实现波动率 (RV) 的时序图和自相关情况. 从图 2 看, 在 2008 年和 2015 年两个时期, 沪深 300 指数波动率很大, 分别对应了次贷危机和 2015 年股灾. 从图 3 可以看出, RV 有很强的自相关性和持续性, 呈现缓慢衰减趋势, 持续到 200 阶以上均显著不为零.

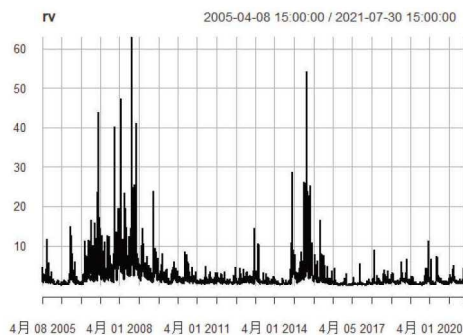


图 3 时序图

(Figure 3 Sequence diagram)

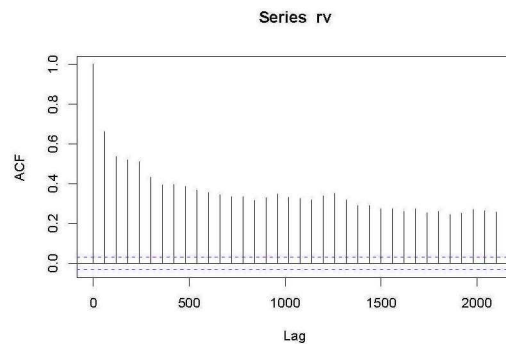


图 4 自相关系数图

(Figure 4 Autocorrelation coefficient diagram)

表 1 为各项高频波动率变量描述性统计结果.

表 1 沪深 300 指数高频波动率描述性统计结果

(Table 1 Descriptive statistical results of high-frequency volatility of CSI 300 index)

	RV	RQ	BV	J	RS ⁺	RS ⁻	SJ
Mean	2.1971	73.1921	1.8847	0.3516	1.1089	1.0882	-0.0716
Maximum	62.9055	64144.0800	51.5559	58.5638	62.9055	31.9439	16.2489
Minimum	0.0781	0.0070	0.0627	0.0000	0.0000	0.0000	-13.2621
Std.deviation	3.8076	1323.3270	3.1013	1.6637	2.3092	2.0281	1.3586
Skewness	6.6129	37.7195	6.3435	20.9673	12.4705	6.1390	0.7313
Kurtosis	66.7539	1614.8620	63.3995	584.7093	246.5980	56.7263	34.3701
LB(20)	12930.74***	649.41***	13255.95***	2388.64***	5621.42***	10551.99***	129.4932***
ADF	-3.6966	-6.7010	-3.7687	-6.9736	-3.5876	-4.0587	-12.2160

注: LB(20) 为滞后 20 期的 Ljung-Bos 统计量; ADF 为 AugmentedDickey-Fuller 单位根检验; *(**) 分别表示在 5% (1%) 置信水平下显著.

由表 1 可以得出: 1) 所有高频波动率变量的峰度和偏度均大于零, 具有尖峰、右偏和非正态特征; 2) 所有高频波动率变量 LB(20) 统计量均显著, 说明这些波动率变量存在显著的长记忆特征, 其中且 BV 的自反馈效应最强; 3) ADF 单位根检验表明各序列都是平稳时间序列.

3.3 样本内估计与预测

本部分使用沪深 300 指数全样本数据, 对所有模型进行样本内估计与预测, 以探究测量误差和结构变化因素是否对沪深 300 指数未来波动率具有预测作用. 采用两种常用的损失函数 MSE 和 MAE 作为预测精度的定量指标, 定义如下

$$MSE = \frac{1}{M} \sum_{m=H+1}^{H+M} (RV_m - \hat{\sigma}_m)^2, \quad (3.1)$$

$$MAE = \frac{1}{M} \sum_{m=H+1}^{H+M} |RV_m^2 - \hat{\sigma}_m^2|, \quad (3.2)$$

式中 $\hat{\sigma}_m$ 是预测值, M 为预测样本数量.

表 2 给出了 HARQ(F) 族模型的样本内估计结果.

表 2 HARQ(F) 族模型的参数估计结果

(Table 2 Parameter estimation results of HARQ(F)-type models)

系数	(1)			(2)			(3)			(4)		
	HAR-RV	HAR-Q	HAR-QF	HAR-J	HAR-Q-J	HAR-QF-J	HAR-RS	HAR-Q-RS	HAR-QF-RS	HAR-SJ	HAR-Q-SJ	HAR-QF-SJ
β_0	0.2386 ***	0.2187 ***	0.1097 *	0.2000 ***	0.2298 ***	0.1516 **	0.2431 ***	0.1775 ***	0.1388 **	0.1791 ***	0.2491 ***	0.3283 ***
$\beta_1^{(+)}$	0.3946 ***	0.4281 ***	0.327 ***	0.4724 ***	0.4384 ***	0.3464 ***	0.2642	-0.133 5***	-0.204 3***	0.3973 ***	0.3879 ***	0.3386 ***

续表 2 HARQ(F) 族模型的参数估计结果

(Table 2 Parameter estimation results of HARQ(F)-type models (Continued))

系数	(1)			(2)			(3)			(4)		
	HAR-RV	HAR-Q	HAR-QF	HAR-J	HAR-Q-J	HAR-QF-J	HAR-RS	HAR-Q-RS	HAR-QF-RS	HAR-SJ	HAR-Q-SJ	HAR-QF-SJ
$\beta_2^{(+)}$	0.2586 ***	0.2519 ***	0.514 ***	0.2381 ***	0.2388 ***	0.4853 ***	-0.010 1	-0.026 6	0.4937 ***	0.3568 ***	0.2787 ***	0.3892 ***
$\beta_3^{(+)}$	0.2383 ***	0.2261 ***	0.1492 ***	0.2326 ***	0.2734 ***	0.1704 ***	-0.098 3***	-0.101 8	-0.701 6***	0.2214 ***	0.2669 ***	0.2512 ***
β_J			-0.213 4***	-0.482 0***	-0.398 5***							
β_1^-						0.4976 ***	1.0512 ***	1.0790 ***				
β_2^-						0.6269 ***	0.6857 ***	0.2656 **				
β_3^-						0.5129 5***	0.462 4***	1.080 4***				
$\beta_{1,SJ}$										-0.048 2**	-0.578 5***	-0.568 8***
$\beta_{5,SJ}$												0.0572
$\beta_{22,SJ}$												-1.164 6***
$\beta_{1,Q}^{(+)}$		-0.001 3**	0.0006 ***		0.0015 ***	0.002 ***		0.0031 ***	0.0035 ***		0.0043 ***	0.0049 ***
$\beta_{2,Q}^{(+)}$			-0.003 7***			-0.003 4**			-0.006 7***			-0.003 2***
$\beta_{3,Q}^{(+)}$			-0.0002			0.0006			0.0252 ***			0.0377 ***
$\beta_{1,Q}^-$								-0.011 2***	-0.011 1***			
$\beta_{2,Q}^{(-)}$									0.0080 **			
$\beta_{3,Q}^{(-)}$									-0.030 7***			
Adj-R ²	0.4904	0.4910	0.5020	0.4950	0.4980	0.5066	0.5097	0.5433	0.5468	0.4808	0.5305	0.5313
MSE	1.0000	0.9987	0.9764	0.9908	0.9846	0.9673	0.9615	0.8950	0.8873	1.0186	0.9209	0.9184
MAE	1.0000	1.0002	0.9916	0.9978	0.9963	0.9896	0.9880	0.9441	0.9376	1.0015	0.9611	0.9491

注: 角标 *, **, *** 分别代表 10%, 5%, 1% 的显著性水平. 损失函数值为标准化结果, 标准化过程是以 HAR-RV 作为基准模型, 将其余模型的损失函数值除以 HAR-RV 损失函数值.

根据表 2, 可以得出的结论有

1) 离散跳跃的回归系数 β_J 显著为负, 说明沪深 300 指数波动率具有均值回复现象. β_1^+ , β_2^+ 和 β_3^+ 的系数总体上为负, 而 β_1^- , β_2^- 和 β_3^- 均为正, 说明已实现正半方差对沪深 300 指数未来波动率有抑制作用, 而已实现负半方差对沪深 300 指数未来波动率具有正向冲击, β_1^+ , β_2^+ 和 β_3^+ 的绝对值和显著性均低于 β_1^- , β_2^- 和 β_3^- , 说明正负半方差对沪深 300 指数未来波动率的影响是不对称的, 二者中起主要作用的是与沪深 300 指数日内负收益率相关的已实现负半方差. 回归系数 $\beta_{1,SJ}$ 显著为负, 说明符号跳跃对沪深 300 指数未来波动率具有负向冲击. 按照 Adj- R^2 由大到小和损失函数值由小到大对四组模型排序, 得到 3、4、2、1 的结果, 说明相对于基础模型, 已实现半方差分解的模型拟合效果和样本内预测能力最佳, 引入符号跳跃的模型次之, 然后是含跳跃波动的模型, 基于基准模型 HAR-RV 的拓展都是合理的.

2) 相比干的 HAR 族模型, HARQ(F) 族模型的 Adj- R^2 值均有提高, 总体上具有更小的 MSE 与 MAE 值, 且回归系数 $\beta_{i,Q}$ ($i = 1, 2, 3$) 的显著性很强, 说明 RQ 对 HAR 族有很好的修正效果; 对比 HARQ 与 HARQF 族模型可知, 后者的 Adj- R^2 值更大, 损失函数值更小, 说明 RQ 多级修正比 RQ 单级修正效果更好, 考虑测量误差因素能提高模型对沪深 300 指数波动率的拟合效果和样本内预测能力.

表 3 为 MRS-HARQ 族模型参数估计结果.

表 3 MRS-HARQ 族模型参数估计

(Table 3 Parameter estimation results of MRS-HARQ-type models)

模型	(1) MRS-HARQ		(2) MRS-HARQ-J		(3) MRS-HARQ-RS		(4) MRS-HARQ-SJ	
状态	1	2	1	2	1	2	1	2
β_0	1.8836***	0.2737***	1.9698***	0.2702***	1.7223***	0.2553***	2.0603***	0.2669***
$\beta_1^{(+)}$	0.4777***	0.1754***	0.4731***	0.2434***	-0.2002**	0.0788***	0.3283***	0.2976***
$\beta_2^{(+)}$	0.2383***	0.2228***	0.2241***	0.1973***	-0.0388	0.2576***	0.2795***	0.1687***
$\beta_3^{(+)}$	0.0835	0.1832***	0.1268*	0.1868***	-0.1492	-0.0907	0.1586**	0.1742***
β_J			-0.3636***	-0.3336***				
β_1^-					0.9813***	0.4065***		
β_2^-					0.6800***	0.1708***		
β_3^-					0.3274	0.4466***		
$\beta_{1,SJ}$							-0.5121***	-0.1280***
$\beta_{1,Q}^{(+)}$	-0.0025***	0.0025***	-0.0009	0.0035***	0.0036***	-0.0003***	0.0025***	0.0037***
$\beta_{1,Q}^-$					-0.0092***	-0.0054***		
R^2	0.2848	0.9087	0.2899	0.9112	0.3802	0.7722	0.3223	0.9133
AIC	10384.80		10330.43		10247.22		10235.38	
BIC	10530.40		10505.14		10509.29		10410.10	
loglik	-5182.402		-5153.213		-5105.608		-5105.691	
MSE	0.8585		0.8343		0.7861		0.7936	
MAE	0.8974		0.8631		0.8332		0.8401	

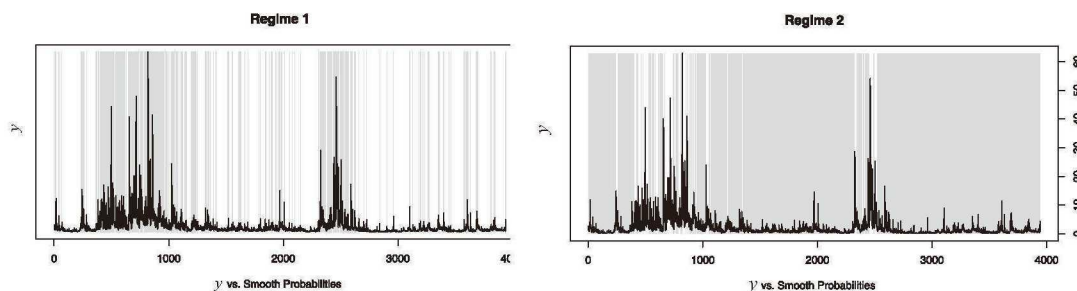
注: 角标 *, **, *** 分别代表 10%, 5%, 1% 的显著性水平. 损失函数值为标准化结果, 标准化过程是以 HAR-RV 作为基准模型, 将其余模型的损失函数值除以 HAR-RV 损失函数值.

结合表 3 分析可知

1) 回归系数 β_1 、 β_2 、 β_3 总体上显著性较强, 说明 MRS-HARQ 族模型能够刻画金融序列的长记忆性, 历史波动率包含了对沪深 300 指数未来波动率的预测信息; 回归系数 β_1 在高波动状态下的绝对值均大于在低波动状态下的绝对值, β_3 则情况相反, 说明高波动状态下短期波动率对沪深 300 指数未来波动率的影响更大, 而低波动状态下长期波动率的影响更大. 在高波动状态下, 回归系数 β_1^+ 、 β_2^+ 和 β_3^+ 之和为负, 回归系数 β_1^- 、 β_2^- 和 β_3^- 均为正, 且整体上 β_i^- 比 β_i^+ ($i = 1, 2, 3$) 显著性更强, 绝对值更大, 说明当模型处于高波动状态时, 已实现半方差具有更强的不对称性. 在高波动状态下 $\beta_{1,SJ}$ 的绝对值更大, 说明符号跳跃在高波动状态时对沪深 300 指数波动率的影响更强. 回归系数 $\beta_{1,Q}$ 总体上显著性较强, 正负号符合预期, 对沪深 300 指数波动率具有较好的修正效果.

2) 似然值、AIC 值、BIC 值都是对马尔科夫转换机制模型拟合优度的刻画. 根据表 3, MRS-HARQ-RS 和 MRS-HARQ-SJ 模型拟合优度相对较高. 由 4 个模型的损失函数值可知, MRS-HARQ-RS 的样本内预测精度最高, MRS-HARQ-SJ 模型次之, MRS-HARQ-J 模型再次, 都优于 MRS-HARQ 模型. MRS-HARQ 族模型的损失函数值在 4 组中都最小, 一致性优于 HAR 和 HARQ 族模型, 说明结合测量误差与马尔科夫转换机制构建的模型对沪深 300 指数未来波动率具有更好的样本内预测能力.

图 5 展示了日预测下 MRS-HARQ 模型的平滑概率状态图.



注: Regime1 表示高波动状态, Regime2 表示低波动状态

图 5 对数收益率序列

(Figure 5 Logarithmic return series)

阴影部分表示沪深 300 指数处于该状态的时刻, 与实际波动率的状态相吻合. 两状态的转移概率矩阵为

$$P = \begin{pmatrix} 0.8092 & 0.0608 \\ 0.1908 & 0.9392 \end{pmatrix},$$

可见, 低波动状态持续的概率 ($P^{22}=0.9392$) 大于高波动状态持续的概率 ($P^{11}=0.8092$), 即高波动状态较低波动状态更容易转换. 每种状态的平均持续天数可根据其转移概率计算得到, 即 i 状态平均持续天数 $d = 1 / (1 - p^{ii})$ ($i, j = 1, 2$ 且 $i \neq j$). 据此计算, MRS-HARQ 模型下沪深 300 指数高波动状态的平均持续天数为 5.2410 天, 低波动状态的平均持续天数为 16.4474 天. 类似地, MRS-HARQ-J、MRS-HARQ-RS 和 MRS-HARQ-SJ 模型下沪深 300 指数高波动低波动状态的平均持续天数分别为: (5.0302, 16.0514)、(5.2165, 16) 和 (5.0454, 16.1031). 可见, 沪深 300 指数高波动状态持续的时间都较短.

3.4 样本外预测和 MSC 检验

为了评估 MRS-HARQ 族模型的样本外预测能力, 本文采用滚动时间窗技术^[25], 滚动时间窗的样本长度固定为 1000 个交易日, 从第 1001 个交易日 (2009 年 6 月 23 日) 开始预测, 到第 3967 个交易日 (2021 年 7 月 31 日) 结束预测. 表 4 给出了 16 个高频波动率模型样本外预测的损失函数值.

表 4 样本外预测的损失函数值
(Table 4 Out of sample predicted loss function value)

组别	预测模型	$h = 1$		$h = 5$		$h = 22$	
		MSE	MAE	MSE	MAE	MSE	MAE
1	HAR-RV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	HARQ	0.9696	0.9905	0.9752	0.9979	<u>0.9863</u>	0.9866
	HARQF	1.1426	1.0904	1.0660	1.0980	0.9886	0.9988
	<u>MRS-HARQ</u>	<u>0.9385</u>	<u>0.9167</u>	0.7226	0.8056	0.9877	<u>0.9664</u>
	HAR-J	1.0040	0.9931	0.9944	0.9997	0.9889	0.9867
2	HARQ-J	0.9676	0.9789	0.9678	0.9995	0.9842	0.9783
	HARQF-J	1.1408	1.0855	1.0698	1.1023	0.9739	0.9842
	MRS-HARQ-J	0.9205	<u>0.9159</u>	<u>0.7750</u>	<u>0.8652</u>	<u>0.5174</u>	<u>0.6786</u>
	HAR-RS	0.9604	0.9674	0.9024	0.9484	0.8135	0.8545
3	HARQ-RS	0.9758	0.9883	0.9374	1.0052	0.8086	0.8359
	HARQF-RS	1.0952	1.0228	0.8285	<u>0.8840</u>	0.8294	0.7132
	MRS-HARQ-RS	<u>0.9417</u>	0.8985	<u>0.7895</u>	0.8849	<u>0.5302</u>	<u>0.5282</u>
	HAR-SJ	0.9549	0.9673	1.0136	1.0087	0.9780	0.9799
4	HARQ-SJ	0.9659	0.9785	1.0082	1.0373	0.9892	0.9904
	HARQF-SJ	1.0419	0.9980	0.9235	0.9989	0.8965	0.8897
	<u>MRS-HARQ-SJ</u>	<u>0.9468</u>	<u>0.9563</u>	<u>0.8252</u>	<u>0.9087</u>	0.5073	<u>0.6585</u>

注: 下划线表示在各组预测模型中损失函数值最小者, 加粗数据表示所有预测模型中损失函数值最小者.

表 4 的实证结果发现

- 1) 在日、周和月三个预测期下, 除个别情况下 (HAR-J 日预测和 HAR-SJ 周预测), HAR-J、HAR-RS 和 HAR-SJ 模型的两个损失函数值均更小, 说明跳跃、正负半方差和符号跳跃模型能降低样本外预测偏差.
- 2) 对比 HAR 族模型和 HARQ(F) 族模型, 发现在绝大多数情况下, HARQ 族模型的损失函数值更小, 说明测量误差单级修正能提高模型的样本外预测精度; 只有在周预测或月预测下, HARQF 族模型的损失函数值更低, 说明在长期预测时测量误差多级修正效果明显.
- 3) 观察四组模型可知, 在日、周和月三个预测期下, 绝大多数 MRS-HARQ 族模型具有最小的损失函数值, 总体上优于 HAR 族与 HARQ(F) 族模型, 说明将测量误差修正和马尔科夫转换机制结合能进一步提高样本外预测精度.

通过损失函数值的大小可以判断预测模型的优劣, 但仍需要对预测模型进行严格的统计检验, 才能得到具有稳健性的结论. 因此, 本文采用 Hansen 等^[26] 提出的模型可信集 (Model

Confidence Set, MCS) 检验上述预测模型的优劣性, 检验统计量 T_R 和 T_{\max} 定义如下

$$T_R = \max_{u,v \in M} \frac{|\bar{d}_{uv}|}{\sqrt{\text{var}(d_{uv})}}, \quad \bar{d}_{uv} = \frac{1}{m} \sum_{t=1}^m d_{uv,t}, \quad (3.3)$$

$$T_{\max} = \max_{u \in M} \frac{\bar{d}_{u,\cdot}}{\sqrt{\text{var}(d_{uv})}}, \quad \bar{d}_{u,\cdot} = \frac{1}{m-1} \sum_{v \in M} \bar{d}_{uv}. \quad (3.4)$$

设定 MCS 检验的显著性水平 α 为 0.1, Bootstrap 算法模拟次数为 5000. MCS 检验结果见表 5.

表 5 MCS 检验结果
(Table 5 MCS test results)

组别	预测模型	$h = 1$				$h = 5$				$h = 22$			
		MAE		MSE		MAE		MSE		MAE		MSE	
		T_{\max}	T_R	T_{\max}	T_R	T_{\max}	T_R	T_{\max}	T_R	T_{\max}	T_R	T_{\max}	T_R
1	HAR-RV	0.43	0.44	0.80	0.79	0.33	0.32	0.16	0.12	0.16	0.15	0.24	0.20
		76	70	56	58	50	38	74	18	08	58	14	40
	HARQ	0.89	0.87	0.98	0.98	0.32	0.31	0.16	0.19	0.13	0.13	0.23	0.20
		24	54	84	76	62	60	48	52	50	02	70	12
	HARQF	0.38	0.39	0.61	0.64	0.82	0.79	0.69	0.59	0.19	0.17	0.27	0.23
		90	24	94	58	42	62	06	52	18	72	90	00
	MRS-HARQ	<u>1.00</u>	<u>1.00</u>	<u>0.99</u>	<u>0.99</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>0.33</u>	<u>0.31</u>	<u>0.47</u>	<u>0.40</u>
		<u>00</u>	<u>00</u>	<u>04</u>	<u>06</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>06</u>	<u>60</u>	<u>20</u>	<u>18</u>
2	HAR-J	0.78	0.76	0.77	0.76	0.52	0.47	0.35	0.24	0.16	0.16	0.26	0.22
		54	44	46	38	56	80	78	36	50	02	56	32
	HARQ-J	0.89	0.87	0.98	0.98	0.32	0.31	0.28	0.19	0.12	0.12	0.24	0.20
		24	54	72	76	42	00	94	52	84	38	86	86
	HARQF-J	0.44	0.44	0.59	0.61	0.82	0.79	0.71	0.62	0.16	0.16	0.26	0.22
		68	34	24	10	66	80	76	04	88	28	66	04
	MRS-HARQ-J	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>0.92</u>	<u>0.90</u>	<u>0.93</u>	<u>0.89</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>
		<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>66</u>	<u>62</u>	<u>64</u>	<u>54</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>
3	HAR-RS	0.99	0.98	0.99	0.99	0.33	0.33	—	—	—	—	—	—
		26	92	38	32	74	68	—	—	—	—	—	—
	HARQ-RS	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	0.27	0.26	—	—	—	—	—	—
		<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	28	36	—	—	—	—	—	—
	HARQF-RS	0.99	0.98	0.69	0.67	0.92	0.90	0.69	0.59	—	—	0.27	0.40
		26	92	00	88	66	62	06	52	—	—	90	18
	MRS-HARQ-RS	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>
		<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>

续表 5 MCS 检验结果
(Table 5 MCS test results (Continued))

组别	预测模型	$h = 1$				$h = 5$				$h = 22$			
		MAE		MSE		MAE		MSE		MAE		MSE	
		T_{\max}	T_R	T_{\max}	T_R	T_{\max}	T_R	T_{\max}	T_R	T_{\max}	T_R	T_{\max}	T_R
4	HAR-SJ	0.99	0.98	<u>1.00</u>	<u>1.00</u>	0.82	0.79	0.69	0.59	0.15	0.14	0.24	0.22
		26	92	<u>00</u>	<u>00</u>	42	62	06	52	40	98	86	04
	HARQ-SJ	0.99	0.98	0.98	0.98	0.82	0.79	0.69	0.59	0.19	0.17	0.26	0.22
		26	92	72	76	42	62	06	52	18	72	56	04
	HARQF-SJ	<u>1.00</u>	<u>1.00</u>	0.98	0.98	0.32	0.29	0.10	—	0.13	0.11	—	—
		<u>00</u>	<u>00</u>	72	76	68	38	84	—	58	32	—	—
	MRS-HARQ-SJ	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>	<u>1.00</u>
		<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>	<u>00</u>

注: 下划线表示各组预测模型中 p 值最大者, 加粗数据表示所有预测模型中的最佳模型, “—” 表示模型被剔除出最优集。

根据表 5 可以得到以下结论

1) 根据组 1, HAR-RV、HARQ 和 HARQF 模型中, 日预测下的 HARQ 模型 p 值最大, 周和月预测下的 HARQF 模型 p 值最大. HAR-J 和 HAR-RS 模型组与组 1 的结果相同. 根据组 4, 相对于 HAR-SJ 模型, HARQ-SJ 和 HARQF-SJ 模型只有在个别预测期和在日预测下损失函数下具有更大的 p 值. 说明除了 HAR-SJ 模型外, 对 HAR 族模型进行测量误差修正有助于提高模型对中国股市波动率的预测能力, 当预测期较短 (1 天) 时, 单级修正具有更好的效果, 当预测期较长 (周或月) 时, 多级修正效果更好. 理论上, 当进行多级修正时, 除了用 RQ 对中国股市日波动率进行修正, 还分别用周和月均 RQ 对中国股市周波动率和月波动率进行了修正. 在多步预测时, 对中国股市未来一周 (月) 波动率的预测贡献和解释作用最大的是周 (月) 均波动率, 因而修正周 (月) 波动率对预测中国股市未来一周 (月) 波动率的影响也是最大的. 所以, 在中 (长) 期预测时, 经测量误差修正的周 (月) 系数对中国股市未来周 (月) 波动率的预测贡献要大于修正的日系数, 此时多级修正效果要优于单级修正. 损失函数和 MCS 检验的结果均与理论预期一致.

2) 在四组模型中, 结合测量误差和马尔科夫转换机制的 MRS-HARQ 族模型在三个预测期和两种损失函数下均具有最大的 p 值, 表现出最优的样本外预测能力. 在 16 个模型中, MRS-HARQ-RS 和 MRS-HARQ-SJ 模型的样本外预测能力最佳. 结合测量误差和马尔科夫转换机制的 MRS-HARQ 族模型表现出最优的样本外预测能力, 考虑测量误差和结构变化的波动率预测模型对中国股市波动率预测具有显著优势. 该结果也证实了中国股市波动率中确实存在结构变化特征, 而马尔科夫转换机制可以有效地刻画中国股市波动率的结构变化.

3) 在日预测期下, 按照 p 值由大到小对 HAR 族模型排序, 得到 HAR-SJ、HAR-RS、HAR-J、HAR-RV 的结果, 说明引入符号跳跃的模型样本外预测能力最佳, 已实现半方差分解的模型次之, 再之后是含跳跃波动的模型, 都优于基础模型.

4 结 语

本文利用已实现四次变差对 HAR 族模型进行修正, 考虑了测量误差对已实现波动率预测的影响, 在此基础上, 将马尔科夫转换机制引入经测量误差修正的模型, 构建具有时变特征的 MRS-HARQ 族模型. 采用 MCS 检验方法比较各组模型的预测能力发现, 考虑测量误差和结构变化因素的波动率模型对中国股市波动率的样本内拟合效果和样本内外预测能力都明显优于不考虑这两个因素的波动率模型, 能够更准确地刻画股市真实波动情况. 主要的贡献点和研究结论如下

1) 相比于 HAR 和 HARQ(F) 族模型, 结合测量误差修正和马尔科夫转换机制构建的 MRS-HARQ 族模型对中国股市波动率具有更高的拟合优度和预测精度, 也能准确地测度出在不同波动率状态下对中国股市未来波动率的影响, 并发现中国股市波动率高波动状态持续的时间都不长. 在本文所探讨的 16 种高频波动率模型中, 结合已实现正负半方差 MRS-HARQ-RS 模型和结合符号跳跃的 MRS-HARQ-SJ 模型拥有最佳的样本外预测能力.

2) 我们的研究工作还证实已实现四次变差能够显著提升 HAR 族模型对中国股市未来波动率的样本内外预测能力. 预测期较短 (1 天) 时, 单级修正后的模型具有更好的样本外预测能力, 当预测期较长 (周或月) 时, 多级修正效果更好.

3) 此外我们的研究工作还表明: 跳跃波动、已实现半方差、符号跳跃均能显著提高 HAR-RV 模型对中国股市波动率的拟合优度和预测能力; 已实现半方差对中国股市未来波动率的影响是不对称的, 已实现负半方差具有显著的正向冲击, 进一步引入马尔科夫转换机制, 在高波动状态下已实现半方差的不对称性增强, 已实现负半方差起到主要影响作用; 符号跳跃对中国股市未来波动率有显著的负面冲击, 进一步引入马尔科夫转换机制, 在高波动状态下符号跳跃的负面冲击增强, 并且对中国股市未来短期波动率的解释能力更强.

本文的研究工作拓展了金融波动率建模分析方法, 进一步提升了波动率测度的精度和动态性能, 结论可以应用在金融市场波动测量、资产定价和风险管理等领域, 对投资者和监管层认识与把握金融市场波动行为也具有实践指导价值.

参 考 文 献

- [1] Engle R F. Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, 1982, **50**: 987–1007.
- [2] Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 1986, **31**(3): 307–327.
- [3] Taylor S J. Modeling Financial Time Series. Chichester: John Wiley & Sons, 1986.
- [4] Andersen T G, Bollerslev T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 1998, **39**(4): 885–905.

- [5] Corsi F. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 2009, **7**(2): 174–196.
- [6] Andersen T G, Bollerslev T, Diebold F X. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 2007, **89**(4): 701–720.
- [7] Barndorff-Nielsen O E, Kinnebrock S, Shephard N. Measuring Downside Risk-Realised Semivariance. New York: Oxford University Press, 2010.
- [8] Patton A J, Sheppard K. Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 2015, **97**(3): 683–697.
- [9] 刘威仪, 江含宇, 张天玮, 等. 基于高频极值数据的波动率建模与预测. 系统工程理论与实践, 2020, **40**(12): 3095–3111.
(Liu W Y, Jiang H Y, Zhang T W, et al. Volatility modeling and forecasting based on high frequency extreme value data. *Systems Engineering — Theory & Practice*, 2020, **40**(12): 3095–3111.)
- [10] 龚旭, 曹杰, 文风华, 等. 基于杠杆效应和结构突变的 HAR 族模型及其对股市波动率的预测研究. 系统工程理论与实践, 2020, **40**(5): 1113–1133.
(Gong X, Cao J, Wen F H, et al. The HAR-type models with leverage and structural breaks and their applications to the volatility forecasting of stock market. *Systems Engineering — Theory & Practice*, 2020, **40**(5): 1113–1133.)
- [11] 马锋, 魏宇, 黄登仕, 等. 基于马尔科夫状态转换和跳跃的高频波动率模型预测. 系统工程, 2016, **34**(1): 10–16.
(Ma F, Wei Y, Huang D S, et al. The forecasting performance of the high-frequency volatility-models with the Markov-switching regime and jump. *Systems Engineering*, 2016, **34**(1): 10–16.)
- [12] Bollerslev T, Patton A J, Quaedvlieg R. Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 2016, **192**(1): 1–18.
- [13] 刘光强. 我国股票市场高频波动预测研究 —— 基于 ARQ 及 HARQ 模型的实证分析. 西南交通大学学报(社会科学版), 2017, **18**(4): 82–87.
(Liu G Q. The study of high frequency fluctuations forecasting for Chinese stock market: The empirical research with ARQ and HARQ Modes. *Journal of Southwest Jiaotong University (Social Sciences)*, 2017, **18**(4): 82–87.)
- [14] 宋亚琼, 王新军. 基于动态估计误差的中国股市波动率建模与预测. 中国管理科学, 2017, **25**(9): 19–27.
(Song Y Q, Wang X J. Modeling and forecasting volatility of Chinese stock market based on dynamic estimation errors. *Chinese Journal of Management Science*, 2017, **25**(9): 19–27.)
- [15] 陈声利, 李一军, 关涛. 基于四次幂差修正 HAR 模型的股指期货波动率预测. 中国管理科学, 2018, **26**(1): 57–71.
(Chen S L, Li Y J, Guan T. Forecasting realized volatility of Chinese Stock Index Futures based on approved HAR models with median realized quarticity. *Chinese Journal of Management Science*, 2018, **26**(1): 57–71.)
- [16] 赵华, 肖佳文. 考虑微观结构噪声与测量误差的波动率预测. 中国管理科学, 2020, **28**(4): 48–60.
(Zhao H, Xiao J W. Volatility forecasting in the presence of microstructure noise and measurement error. *Chinese Journal of Management Science*, 2020, **28**(4): 48–60.)
- [17] Ma F, Wahab M I M, Huang D S, et al. Forecasting the realized volatility of the oil futures market: A regime switching approach. *Energy Economics*, 2017, **67**: 136–145.
- [18] Duan Y Y, Chen W, Zeng Q, et al. Leverage effect, economic policy uncertainty and realized volatility with regime switching. *Physica A-Statistical Mechanics and Its Applications*, 2018, **493**: 148–154.
- [19] Yu M, Song J. Volatility forecasting: Global economic policy uncertainty and regime switching. *Physica A: Statistical Mechanics and Its Applications*, 2018, **511**: 316–323.

- [20] Xu W, Wang J, Ma F, et al. Forecast the realized range-based volatility: The role of investor sentiment and regime switching. *Physica A: Statistical Mechanics and Its Applications*, 2019, **527**: 121422.
- [21] Aha A, Cyh B, Iwm A. Modelling the volatility of TOCOM energy futures: A regime switching realised volatility approach. *Energy Economics*, 2019, 104434, DOI: 10.1016/j.eneco.2019.06.019.
- [22] Barndorff-Nielsen O E, Shephard N. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 2002, **17**(5): 457–477.
- [23] Hamilton J D, Susmel R. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 1994, **64**(1/2): 307–333.
- [24] Kim C J. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 1994, **60**(1/2): 1–22.
- [25] 魏宇. 沪深 300 股指期货的波动率预测模型研究. 管理科学学报, 2010, **13**(2): 66–76.
(Wei Y. Volatility forecasting models for CSI300 index futures. *Journal of Management Sciences in China*, 2010, **13**(2): 66–76.)
- [26] Hansen P R, Lunde A, James M N. The model confidence set. *Econometrica*, 2011, **79**(2): 453–497.