

# 基于跳跃、好坏波动率的混频已实现 EGARCH 模型的波动率预测与风险度量

郭宝才<sup>1,2</sup>, 项琳<sup>1</sup>

(1. 浙江工商大学 统计与数学学院, 浙江 杭州 310018;

2. 浙江工商大学 统计数据工程技术与应用协同创新中心, 浙江 杭州 310018)

**摘要:** 为探究资产价格的跳跃行为和收益波动的非对称效应对波动率预测的影响, 以高频数据建模为视角, 基于跳跃、好坏波动率将 Realized EGARCH-MIDAS 模型进行拓展, 以提升模型的波动率预测能力与风险度量效果。运用拓展后的模型, 以沪深300指数价格高频数据为样本进行实证分析, 探究中国股票市场的波动性规律, 并采用似然函数、信息准则和基于损失函数的 DM 与 MCS 等检验方法, 综合比较了改进前后的模型对波动率及风险值的预测效果。实证结果显示: (1) 沪深300指数收益的长期波动主要来源于连续波动而非跳跃波动, 且受正连续波动影响更大, 而负跳跃对波动具有明显的负向冲击; (2) 文章提出的拓展模型均能更好地捕捉波动率的长记忆性, 在样本内估计和样本外预测上也都有更好的表现, 其中同时考虑跳跃与非对称影响的 Realized EGARCH-MIDAS-RSJ 拓展模型拥有最优的估计及预测效果。

**关键词:** Realized EGARCH-MIDAS 模型; 跳跃; 好坏波动率; 波动率预测; VaR

**中图分类号:** F830 **文献标志码:** A **文章编号:** 1000-2154(2022)05-0079-19

**DOI:** 10.14134/j.cnki.cn33-1336/f.2022.05.007

## Realized EGARCH-MIDAS Model Based on Jumps and Good-bad Volatility with Its Volatility and VaR Forecasting

GUO Baocai<sup>1,2</sup>, XIANG Lin<sup>1</sup>

(1. School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China; 2. Collaborative Innovation Center of Statistical Data Engineering Technology & Application, Zhejiang Gongshang University, Hangzhou 310018, China)

**Abstract:** To study the impact of jumps and asymmetric effects on volatility forecasting and further improve the risk measurement performance, the Realized EGARCH-MIDAS model is extended by introducing the decomposition of the realized volatility into the long-term volatility equation. Specifically, the decomposition includes the continuous-jump volatility, the good-bad volatility, and the positive-negative jump volatility with positive-negative continuous volatility. Then, high-frequency data from the CSI 300 is used for empirical analysis to explore the volatility pattern of the Chinese stock market. Different statistical and economic criteria, including the likelihood function, information criteria, and loss-functions-based tests (DM and MCS tests), are applied to assess the performance of proposed models in both volatility modeling and risk measurement. The empirical results show that: 1) The long-term volatility of the CSI 300 mainly comes from the continuous component rather than the jump component and is more affected by the positive continuous component, while the negative jump has a significant negative impact on the volatility. 2) All improved models can better capture the long-memory of volatility and thus perform better in both the in-sample fitting and out-of-sample forecasting of

**收稿日期:** 2021-10-25

**基金项目:** 浙江省自然科学基金项目(LY20G020008); 浙江省重点建设高校优势特色学科(浙江工商大学统计学)

**作者简介:** 郭宝才,男,教授,博士生导师,经济学博士,主要从事应用统计研究;项琳(通讯作者),女,博士研究生,主要从事金融时间序列研究。

volatility and risk, with the REGARCH-MIDAS-RSJ considering both jumps and asymmetries being the optimal model.

**Key words:** Realized EGARCH-MIDAS model; jump; good-bad volatility; volatility forecasting; VaR

## 一、引言

在金融资本市场中,投资组合优化、风险管理和衍生品定价等金融应用的准确实施依赖于对资产收益波动的准确建模与预测。因此,有关波动率建模的研究一直是金融计量经济学中最活跃的领域之一。其中,最常见的是 GARCH 族模型,此类模型能够很好地刻画资产收益率的异方差性和波动聚集性。然而,GARCH 族模型通常使用同一频率的数据,未能充分利用蕴含丰富波动信息的日内高频数据与周、月等长期数据,因此无法在时间尺度上更精细地刻画波动率,也并不适用于研究金融市场波动的长期驱动因素<sup>[1]</sup>。

金融市场及计算机技术的高度发达使得金融资产的高频交易普遍存在,资产价格受各方因素的影响迅速变化,这给波动率的研究带来了新的机遇与挑战。随着高频金融数据获取难度的降低,学者们构建了众多包含日内高频收益信息的已实现测度(Realized Measure),其中应用最为广泛的是由 Andersen 和 Bollerslev(1998)<sup>[2]</sup>提出的已实现波动率(Realized Volatility, RV)。已实现测度作为真实积分波动率(Integrated Volatility, IV)的估计,具有计算简便、稳健性等多重优势。为在 GARCH 模型的框架下引入包含高频信息的已实现波动率,Engle(2002)<sup>[3]</sup>将其直接作为外生变量加入波动方程中,以此构造 GARCH-X 模型,但该模型不能合理解释已实现波动率的变动。为解决这一问题,Hansen 等(2012)<sup>[4]</sup>引入测量方程提出 Realized GARCH 模型,并由 Hansen 和 Huang(2016)<sup>[5]</sup>将模型拓展为结构更加灵活的 Realized EGARCH 模型,实现了对日收益率和包含日内高频信息的已实现波动率的联合建模,在实证中取得较优的拟合与预测效果。此后,学者们<sup>[6-9]</sup>不断研究并拓展 Realized GARCH 模型,并将其广泛应用于金融市场的实证研究中。以上研究显示,日内高频价格信息对于 GARCH 族模型波动率预测精度的提升具有重要作用,然而上述波动率模型在不大量增加模型参数的前提下,并无法充分捕捉波动率的长记忆性,故存在一定的局限性。

受市场结构、交易机制和投资者构成等多方因素的影响,金融资本市场通常存在异质性特征(Muller 等,1997)<sup>[10]</sup>,导致资产波动呈现一定程度的相异性,即存在波动持续性(长记忆性),具体表现为波动序列存在正的和缓慢衰减的自相关函数。考虑到忽略长记忆性将导致模型无法准确地刻画长期波动率,Huang 等(2016)<sup>[11]</sup>将包含日、周、月已实现波动率的 HAR 结构<sup>[12]</sup>引入波动方程,构建了能够捕捉波动长记忆性的 Realized HAR GARCH 模型,但 HAR 结构较为固定且回溯时间范围较短,无法刻画超过一个月的自相关性。为此,王天一等(2018)<sup>[13]</sup>利用混频数据抽样(Mixed Data Sampling, MIDAS)结构<sup>[14]</sup>改进 Realized GARCH 模型,能够有效地提取数据中的长期波动信息。Borup 和 Jakobsen(2019)<sup>[15]</sup>进一步将条件方差乘法分解为短期和长期成分,并引入 MIDAS 结构的长期波动方程,提出 Realized EGARCH-MIDAS 模型,该模型整合了 Realized EGARCH 与 MIDAS 模型各自的优势,在充分利用高频信息的基础上能够灵活有效地刻画资产收益波动率的聚集性与长记忆性,在实证中显示出较强的拟合效果和预测能力。然而,资产价格除了存在以上讨论的波动聚集性及长记忆性之外,还具有一些无法忽视的典型特征,如跳跃行为与其收益率波动的非对称效应。

在现实生活中,经济、政治和社会等各方面出现的突发事件不断影响着各国金融市场,这使得资产价格呈现出非连续的跳跃式变化,这种由异常消息冲击而导致的资产价格极端波动的现象被称为跳跃行为<sup>[16]</sup>。资产价格跳跃行为的研究首先需要解决的问题是如何识别与区分价格变动中的连续波动成分与跳跃成分。随着已实现测度渐近理论性质研究的不断深入,学者们基于高频数据在二次变差理论的框架下提出了众多非参数跳跃检验方法,包括 Barndorff-Nielsen 和 Shephard(2004,2006)<sup>[17-18]</sup>提出的 BNS 跳跃检测法,Andersen 等(2007)<sup>[19]</sup>提出的 ABD 跳跃检测法,At-Sahalia 和 Jacod(2009)<sup>[20]</sup>提出的 ASJ 检测法,Bollerslev 等(2013)<sup>[21]</sup>提出的 TOD 检测法以及 Andersen 等(2012)<sup>[22]</sup>提出的 ADS 跳跃检测法等,而其中最为广泛应用的是 ADS 跳跃检测法,这一方法对跳跃及微观噪声同时稳健。在利用跳跃检测方法识别出显著性跳跃之后,已实现波动率可以分解为离散跳跃部分与连续波动部分。在此基础上,学者们开始研究

这两个波动成分对于模型波动率预测的影响。Santos 和 Ziegelmann (2014)<sup>[23]</sup>、Ma 等 (2019)<sup>[24]</sup> 在 MIDAS 模型中引入基于不同跳跃检测法得到的跳跃波动与连续波动,以此提升了原有模型的波动率预测精度。Pan 等 (2020)<sup>[25]</sup> 则在 Engle 等 (2013)<sup>[26]</sup> 的 GARCH-MIDAS 模型中引入收益率的跳跃性,提出了 GARCH-Jump-MIDAS 模型。上述研究证实了在波动率建模中考虑跳跃行为的重要性,也由此启发本文进一步在 Realized (E)GARCH 模型框架下考虑跳跃行为对条件方差的影响,而 Realized EGARCH-MIDAS 模型长期方程中的 MIDAS 结构为在模型中引入跳跃波动提供便利。

此外,大量的实证研究表明金融资产收益率序列的波动常常存在非对称效应,即波动率对坏消息和好消息的反应不同。考虑到非对称效应是影响资产价格波动的重要因素,Nelson (1991)<sup>[27]</sup> 将非对称效应引入 GARCH 模型,随后学者们<sup>[28-29]</sup> 不断丰富非对称 GARCH 族模型,并在实证中检验了此类非对称模型在波动率拟合和预测上的良好性能。而在高频数据领域,波动非对称性的研究同样引起了学者们的关注。Barn-dorff-Nielsen 等 (2010)<sup>[30]</sup> 区分正负收益提出已实现半方差 (Realized Semi-variance, RS), 将已实现波动率分解成上涨半方差 (Positive Realized Semi-variance,  $RS^+$ ) 与下跌半方差 (Negative Realized Semi-variance,  $RS^-$ ), 亦被称为“好”波动率与“坏”波动率<sup>[31]</sup>, 以刻画波动的非对称性。在此基础上,学者们将已实现半方差<sup>[32-33]</sup> 引入包含已实现波动率的 GARCH-MIDAS 模型框架中,研究好坏波动率对未来条件方差的不对称冲击,并验证了考虑非对称效应能够提升模型的波动率预测能力。值得注意的是,Realized EGARCH-MIDAS 模型的测量方程及波动方程中的杠杆函数也能够刻画收益率对条件方差的非对称性影响,但是忽略了长期方程中已实现波动率的非对称效应。因此,本文考虑将 Realized EGARCH-MIDAS 模型的长期波动方程中的已实现波动率分解为好坏波动率,以完善模型的非对称结构。

当前,在经济全球化、金融自由化及金融创新加速的背景下,金融市场波动率的影响因素越来越交错复杂,极端事件频繁发生,导致资产价格的跳跃性增强,收益的波动率也普遍呈现出长记忆性与非对称性。在这样的背景下,学者们<sup>[34-35]</sup> 开始探讨在单一模型框架下同时研究以上这些因素对波动率的影响,而研究结果也证实同时考虑跳跃与非对称性能够提升模型的波动率刻画能力。为此,考虑到 Realized EGARCH-MIDAS 模型能够很好地刻画波动率的长记忆特征,而已实现波动率既能分解为连续、跳跃波动,又能分解为好、坏波动率,本文提出 Realized EGARCH-MIDAS 模型的三组拓展模型,分别考虑跳跃、好坏波动率以及综合这两因素提出的正负跳跃与正负连续波动的影响,并在偏 t 分布的假定下将模型应用于国内股票市场,以沪深300指数为实证研究对象,综合刻画其收益率波动的聚集性、偏峰厚尾性、长记忆性与非对称效应以及价格的跳跃特征,并进一步比较各模型的波动率及风险预测能力。

本研究的主要贡献为:第一,运用稳健的 ADS 跳跃检测法识别沪深300指数价格序列每日的跳跃波动,并将之引入 Realized EGARCH-MIDAS 模型,以探究跳跃行为对条件方差的影响;第二,考虑波动的非对称效应,将模型长期方程中的已实现波动率分解为好波动率与坏波动率,以此完善 Realized EGARCH-MIDAS 模型的非对称结构;第三,根据已实现波动率的渐近理论性质,基于跳跃检验提取正负跳跃与正负连续波动,并引入 Realized EGARCH-MIDAS 模型中,以考虑跳跃行为与非对称效应的综合影响;第四,通过样本内拟合与样本外滚动预测,结合似然函数、信息准则与基于损失函数的 DM、MCS 检验法综合比较了以上模型在波动率估计和预测上的效果,并以此为基础进行风险度量。

本文其余部分结构安排如下:第二部分为理论介绍,包括已实现波动率的理论基础,以及波动率基础模型和本文提出的改进模型,并给出估计方法;第三部分基于以上理论进行实证研究,包括模型样本内的估计以及波动率、风险值的样本外预测,并对实证结果进行评估;第四部分概括主要研究结论并作出展望。

## 二、理论与模型

### (一) 已实现波动率的极限理论

自 Black-Scholes 期权定价模型提出以来,一般都假定资产价格过程服从连续路径,然而金融市场中不时出现的极端事件证实这一假设存在一定的局限性。为此,Merton (1976)<sup>[36]</sup> 在连续路径的扩散过程中引

入了不连续的跳跃过程对资产价格序列进行建模。在跳跃存在的无风险套利市场中,当资产对数价格  $X_s = \log(p_s)$  ( $p_s$  表示  $s$  时刻的资产价格) 服从连续时间跳跃扩散过程时,对数收益率在第  $t$  日的二次变差过程可以表示为:

$$[X, X]_t = \int_{t-1}^t \sigma_s^2 ds + \sum_{t-1 < s \leq t} \kappa_s^2 \quad (1)$$

其中,  $\sigma_s$  是满足通常假设的瞬时波动过程,  $\kappa_s$  表示对数价格过程的跳跃幅度。由式(1)可知金融资产的波动率分为两个部分:连续样本路径方差(积分波动率  $IV_t = \int_{t-1}^t \sigma_s^2 ds$ ) 和离散的跳跃方差。

将每个交易日分为  $n$  个时间段,令  $\Delta$  为价格数据的抽样间隔( $n = 1/\Delta$ ),第  $t$  日  $s$  时刻的对数收益率  $r_s = X_s - X_{s-\Delta}$ 。已实现波动率由日内高频收益率的平方和定义:

$$RV_t = \sum_{s=1}^n r_{t-1+s\Delta}^2 \quad (2)$$

Barndorff-Nielsen 和 Shephard(2002)<sup>[37]</sup> 证实,若资产价格过程存在跳跃,已实现波动率依概率收敛于二次变差过程。因此,只要得到积分波动率  $IV$  的一致估计量,就可以剥离已实现波动率中的跳跃部分。

基于这一思路,Barndorff-Nielsen 和 Shephard(2004)<sup>[17]</sup> 提出已实现双幂次变差(RBV)作为积分波动率的一致估计,并基于已实现波动率与已实现双幂差的差值进一步构建了BNS跳跃检测法<sup>[18]</sup>。然而,RBV对市场微观结构噪声及零收益率较为敏感,其估计的有偏性导致BNS检测缺乏稳健性。为此,Andersen等(2012)<sup>[22]</sup> 基于最近邻截断方法构建对跳跃及微观噪声均稳健的中位数已实现波动率(MedRV)来估计积分波动率(IV):

$$MedRV_t = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left( \frac{n}{n-2} \right) \times \sum_{s=2}^{n-1} med(|r_{t-1+(s-1)\Delta}|, |r_{t-1+s\Delta}|, |r_{t-1+(s+1)\Delta}|)^2 \quad (3)$$

由此形成了ADS跳跃检测法,其检验统计量  $Z_t^{med}$  定义如下:

$$Z_t^{med} = n^{1/2} \frac{(RV_t - MedRV_t) RV_t^{-1}}{[0.96 \max(1, MedTQ_t \cdot MedRV_t^{-2})]^{1/2}} \quad (4)$$

其中,  $MedTQ_t$  表示中位数已实现四次幂差,是四次积分波动  $\int_{t-1}^t \sigma_s^4 ds$  的一致估计:

$$MedTQ_t = \frac{3\pi n}{9\pi + 72 - 52\sqrt{3}} \left( \frac{n}{n-2} \right) \times \sum_{s=2}^{n-1} med(|r_{t-1+(s-1)\Delta}|, |r_{t-1+s\Delta}|, |r_{t-1+(s+1)\Delta}|)^4 \quad (5)$$

若资产价格过程不存在日内跳动,则  $Z_t^{med}$  渐近服从标准正态分布,因此通过判断  $Z_t^{med}$  是否超过临界值  $\Phi_\alpha$  ( $\Phi_\alpha$  为标准正态分布的上  $\alpha$  分位数) 可以检测跳跃是否显著存在。

因此,可将已实现波动率分解为如下离散跳跃部分与连续波动部分:

$$CJ_{t,\alpha} = I(Z_t^{med} > \Phi_\alpha) \cdot (RV_t - MedRV_t)^+ \quad (6)$$

$$CRV_{t,\alpha} = I(Z_t^{med} \leq \Phi_\alpha) \cdot RV_t + I(Z_t^{med} > \Phi_\alpha) \cdot MedRV_t \quad (7)$$

其中,“+”表示取跳跃方差的非负值(负值则取为0), $I(\cdot)$  为示性函数,当括号中条件成立时取值为1,否则取值为0。

进一步,考虑到资产价格波动普遍存在非对称效应,Barndorff-Nielsen等<sup>[30]119-120</sup> 为区分正负收益率提出上涨已实现半方差和下跌已实现半方差:

$$RS_t^+ = \sum_{s=1}^n r_{t-1+s\Delta}^2 I(r_{t-1+s\Delta} > 0) \quad (8)$$

$$RS_t^- = \sum_{s=1}^n r_{t-1+s\Delta}^2 I(r_{t-1+s\Delta} < 0) \quad (9)$$

Patton 和 Sheppard(2013)<sup>[31]</sup> 将其称为“好”“坏”波动率,用以反映正、负收益的不同特征。从定义可以看出如下关系式:  $RV_t = RS_t^+ + RS_t^-$ 。此外,Barndorff-Nielsen等(2010)<sup>[30]</sup> 还证明了已实现半方差的渐进性。

$$RS_t^+ \xrightarrow[\Delta \rightarrow 0]{P} \frac{1}{2} \int_{t-1}^t \sigma_s^2 ds + \sum_{t-1 < s \leq t} \kappa_s^2 I(\kappa_s > 0)$$



$$RS_t^- \xrightarrow{\Delta \rightarrow 0} \frac{P}{2} \int_{t-1}^t \sigma_s^2 ds + \sum_{t-1 < s \leq t} \kappa_s^2 I(\kappa_s < 0)$$

此外,在风险管理实践中,投资者对于正跳跃与负跳跃有着不同的态度和风险控制措施,分析正、负跳跃对波动率的影响有助于投资者更深入地了解股票市场的内在运行机制。因此,根据中位数已实现波动率 (MedRV) 与已实现半方差的渐近理论性质,将已实现跳跃波动分解为由正向收益驱动的正跳跃部分与由负向收益形成的负跳跃部分,剥离出的正跳跃与负跳跃可以分别记为:

$$CJ_{t,\alpha}^+ = I(Z_t^{med} > \Phi_\alpha) \cdot \left( RS_t^+ - \frac{1}{2} MedRV_t \right)^+ \quad (10)$$

$$CJ_{t,\alpha}^- = I(Z_t^{med} > \Phi_\alpha) \cdot \left( RS_t^- - \frac{1}{2} MedRV_t \right)^+ \quad (11)$$

由此也可以得到正连续波动部分与负连续波动部分如下:

$$CRV_{t,\alpha}^+ = I(Z_t^{med} \leq \Phi_\alpha) \cdot RS_t^+ + I(Z_t^{med} > \Phi_\alpha) \cdot \frac{1}{2} MedRV_t \quad (12)$$

$$CRV_{t,\alpha}^- = I(Z_t^{med} \leq \Phi_\alpha) \cdot RS_t^- + I(Z_t^{med} > \Phi_\alpha) \cdot \frac{1}{2} MedRV_t \quad (13)$$

从定义可以看出如下关系式:  $CRV_{t,\alpha} = CRV_{t,\alpha}^+ + CRV_{t,\alpha}^-$ 。由此,在充分考虑跳跃与非对称性的基础上,已实现波动率 RV 既可以分解为离散跳跃 CJ 与连续波动 CRV,又可以分解为好、坏波动率,并可进一步分解为正负跳跃与正负连续波动。

## (二) Realized EGARCH-MIDAS 模型

为充分利用日内高频信息并同时捕捉波动率的长记忆性,Realized EGARCH-MIDAS 模型利用乘法分量结构与 MIDAS 结构合理整合包含日内高频信息的已实现波动率与较低频率的变量,在实证中显示出较强的波动性刻画能力。滞后阶数  $p = q = 1$  的对数 Realized EGARCH-MIDAS 模型为:

$$r_t = \sigma_t z_t, \quad \sigma_t^2 = h_t g_t \quad (14)$$

$$\log h_t = \beta \log h_{t-1} + \tau(z_{t-1}) + \alpha u_{t-1} \quad (15)$$

$$\log x_t = \xi + \phi \log \sigma_t^2 + \delta(z_t) + u_t \quad (16)$$

$$\log g_t = \omega + \lambda \sum_{k=1}^K \Gamma_k(\gamma) y_{t-1,k}^{(N)} \quad (17)$$

以上四个方程分别被称为“收益方程”“GARCH 方程”“测量方程”和“长期方程”。其中,  $r_t$  表示均值为 0 的第  $t$  日资产价格收益率; 收益率的条件方差  $\sigma_t^2 = Var[r_t | F_{t-1}]$ ,  $F_t = \sigma(r_t, x_t, r_{t-1}, x_{t-1}, \dots)$ ; 条件方差  $\sigma_t^2$  被乘法分解为两个分量,其中  $h_t$  为短期成分,在条件方差中捕捉日度(高频)波动,  $g_t$  捕捉条件方差的长期(低频)波动,称为长期成分或基线波动;  $\{z_t, t \geq 1\}$  为零均值方差为 1 的独立同分布残差序列,  $\{u_t, t \geq 1\}$  独立同分布,  $E u_t = 0, Var(u_t) = \sigma_u^2$ , 且  $z_t$  与  $u_t$  相互独立;  $x_t$  表示已实现测度,本文采用已实现波动率  $RV_t$  [式(2)] 作为  $x_t$ 。另外,  $\tau(\cdot)$  与  $\delta(\cdot)$  为杠杆函数,反映不同方向收益率对短期波动率冲击的非对称性,参照 Hansen 等(2012)<sup>[4]</sup>,采用如下形式:

$$\tau(z_t) = \tau_1 z_t + \tau_2 (z_t^2 - 1) \quad (18)$$

$$\delta(z_t) = \delta_1 z_t + \delta_2 (z_t^2 - 1) \quad (19)$$

长期方程则采用 MIDAS 结构来考察已实现波动率的多期历史信息对条件方差的影响,其中  $y_{t-1,k}^{(N)} = (1/N) \sum_{i=1}^N \log x_{t-N(k-1)-i}$  为滞后的对数已实现波动测度的  $N$  日均值,  $N$  值决定了数据馈入低频分量的频率,取  $N = 5$  对应于周线均值,  $N = 22$  对应于月线均值。另外,  $\Gamma_k(\gamma)$  为参数向量  $\gamma$  的非负权函数,满足  $\sum_{k=1}^K \Gamma_k(\gamma) = 1$ ,采用单参数 Beta 权函数形式:

$$\Gamma_k(\gamma) = \frac{(1 - k/K)^{\gamma-1}}{\sum_{j=1}^K (1 - j/K)^{\gamma-1}} \quad (20)$$

另外, MIDAS 分量的滞后长度  $K$  的选择参照 Borup 和 Jakobsen(2019)<sup>[15]</sup>, 采用  $K = 52$  与  $K = 12$  分别对应周线  $N = 5$  与月线  $N = 22$ , 即考察总长度大约为一整年的滞后已实现波动率的影响。此外, 值得注意的是, 当长期方程中  $\lambda = 0$  时, Realized EGARCH-MIDAS 模型就退化为 Realized EGARCH 模型。在后文的讨论中, 将 Realized EGARCH 模型与 Realized EGARCH-MIDAS 模型分别简记为 REGARCH 模型与 REGARCH-MIDAS 模型。

### (三) Realized EGARCH-MIDAS 模型的拓展

REGARCH-MIDAS 模型借鉴 GARCH-MIDAS 模型的构建思路, 将条件方差乘法分解为短期和长期分量以捕捉波动性, 其中短期分量的动态特征与 REGARCH 模型保持一致, 长期成分则引入 MIDAS 结构进行建模。值得注意的是, 针对 GARCH-MIDAS 模型的长期方程, 学者们提出了不同的改进形式, 如引入影响波动的宏观经济变量<sup>[38-39]</sup>, 或是考虑波动率的跳跃行为<sup>[25]</sup>与非对称性效应<sup>[33]</sup>的影响等, 这些改进均在一定程度上提升了模型的波动率预测效果。基于这一思路, 本文将 REGARCH-MIDAS 模型的长期方程进行拓展, 以提升模型整体的波动性刻画能力。具体来说, 从跳跃性与非对称性两个角度进行改进。

首先, 考虑跳跃行为对波动率的影响。在波动率模型的研究中, 越来越多的学者关注于极端收益引起的跳跃行为对波动的影响, 如孙洁(2014)<sup>[40]</sup>、Laurent 和 Shi(2020)<sup>[41]</sup>等。这些研究均表明, 建立一个考虑跳跃影响下收益异常波动的波动率模型至关重要。为此, 参考 Ma(2019)<sup>[24]</sup>的建模思路, 将 REGARCH-MIDAS 模型的长期方程式(17)的已实现波动率按照式(6)与式(7)进行跳跃分解, 得到拓展后的长期方程为:

$$\log g_t = \omega + \lambda^J \sum_{k=1}^K \Gamma_k(\gamma) y_{t-1,k}^{(N,J)} + \lambda^C \sum_{k=1}^K \Gamma_k(\gamma) y_{t-1,k}^{(N,C)} \quad (21)$$

其中, 跳跃部分  $y_{t-1,k}^{(N,J)} = (1/N) \sum_{i=1}^N \log(CJ_{t-N(k-1)-i} + 1)$  (参考唐勇和崔金鑫(2018)<sup>[42]</sup>, 取对数处理时将跳跃波动取值加 1), 用于捕捉由极端收益引起的变化,  $y_{t-1,k}^{(N,C)} = (1/N) \sum_{i=1}^N \log(CRV_{t-N(k-1)-i})$  表示连续的已实现波动成分。由此, 利用已实现波动率的跳跃分解来考察跳跃行为对条件方差的影响, 并将拓展后的模型简记为 REGARCH-MIDAS-CJ 模型。

其次, 探究波动率的非对称特征。参考 Patton 和 Sheppard(2013)<sup>[31]</sup>的处理, 将 REGARCH-MIDAS 模型长期方程式(17)的已实现波动率按照式(8)与式(9)分解为好坏波动率, 得到拓展后的长期方程为:

$$\log g_t = \omega + \lambda^+ \sum_{k=1}^K \Gamma_k(\gamma) y_{t-1,k}^{(N,+)} + \lambda^- \sum_{k=1}^K \Gamma_k(\gamma) y_{t-1,k}^{(N,-)} \quad (22)$$

其中,  $y_{t-1,k}^{(N,+)} = (1/N) \sum_{i=1}^N \log(RS_{t-N(k-1)-i}^+)$  与  $y_{t-1,k}^{(N,-)} = (1/N) \sum_{i=1}^N \log(RS_{t-N(k-1)-i}^-)$  表示  $N$  日的平均好、坏波动率。由此, 利用已实现波动率的非对称性分解来显示正、负收益对长期波动的非对称影响, 并将拓展后的模型记为 REGARCH-MIDAS-RS 模型。

进一步, 在同一模型框架内同时考虑跳跃行为与非对称性的影响, 将 REGARCH-MIDAS 模型长期方程式(17)的已实现波动率按照式(10)至式(13)分解为正跳跃、负跳跃与正连续波动、负连续波动四个部分, 得到拓展后的长期方程为:

$$\log g_t = \omega + \lambda^{J+} \sum_{k=1}^K \Gamma_k(\gamma) y_{t-1,k}^{(N,J+)} + \lambda^{J-} \sum_{k=1}^K \Gamma_k(\gamma) y_{t-1,k}^{(N,J-)} + \lambda^{C+} \sum_{k=1}^K \Gamma_k(\gamma) y_{t-1,k}^{(N,C+)} + \lambda^{C-} \sum_{k=1}^K \Gamma_k(\gamma) y_{t-1,k}^{(N,C-)} \quad (23)$$

其中, 正跳跃部分  $y_{t-1,k}^{(N,J+)} = (1/N) \sum_{i=1}^N \log(CJ_{t-N(k-1)-i}^+ + 1)$  与负跳跃部分  $y_{t-1,k}^{(N,J-)} = (1/N) \sum_{i=1}^N \log(CJ_{t-N(k-1)-i}^- + 1)$  分别用于捕捉由好、坏波动率中由极端收益引起的变化; 正连续部分  $y_{t-1,k}^{(N,C+)} = (1/N) \sum_{i=1}^N \log(CRV_{t-N(k-1)-i}^+)$  与负连续部分  $y_{t-1,k}^{(N,C-)} = (1/N) \sum_{i=1}^N \log(CRV_{t-N(k-1)-i}^-)$  分别表示  $N$  日的平均好、坏波动率中的连续已实现波动成分。由此, 利用已实现波动率的非对称性跳跃分解来显示正跳跃、负跳跃对长期波动的非对称影响, 并将拓展后的模型记为 REGARCH-MIDAS-RSJ。

#### (四) 模型参数估计方法

本文采用极大似然法(Maximum Likelihood Estimation, MLE)对模型的具体参数进行估计,具体的估计步骤参考 Hansen 和 Huang(2016)<sup>[5]</sup>。REGARCH-MIDAS 族模型的对数似然函数可以表示为:

$$\ell(r, x; \Theta) = \sum_{i=1}^T \ell_i(r_i, x_i; \Theta) = \sum_{i=1}^T [\ell_i(r_i; \Theta) + \ell_i(x_i | r_i; \Theta)] \quad (24)$$

式(24)中  $T$  为样本期总长度,  $\Theta = (\beta, \tau_1, \tau_2, \alpha, \xi, \phi, \delta_1, \delta_2, \omega, \lambda^*, \gamma, \eta, \nu, \sigma_u^2)^T$  是模型的参数向量,其中 REGARCH-MIDAS 模型中的  $\lambda^* = \lambda$ , REGARCH-MIDAS-CJ 模型的  $\lambda^* = (\lambda^c, \lambda^j)$ , REGARCH-MIDAS-RS 模型的  $\lambda^* = (\lambda^+, \lambda^-)$ , REGARCH-MIDAS-RSJ 模型的  $\lambda^* = (\lambda^{j+}, \lambda^{j-}, \lambda^{c+}, \lambda^{c-})$ 。给定测量方程残差分布  $u_i \sim N(0, 1)$ , 可进一步由测量方程得到  $\ell_i(x_i | r_i; \Theta)$  的表达式为:

$$\ell_i(x_i | r_i; \Theta) = -\frac{1}{2} \left( \log 2\pi + \log \sigma_u^2 + \frac{u_i^2}{\sigma_u^2} \right) \quad (25)$$

另外,参照王天一和黄卓(2012)<sup>[6]</sup>将收益残差  $z_i$  的分布设定为零均值、单位方差的偏  $t$  分布,即  $z_i \sim skt(\eta, \nu)$ , 这一设定是基于金融市场收益率分布普遍存在的有偏及厚尾性质。在已知  $z_i$  分布的基础上,进一步可由均值方程计算得到半对数似然函数  $\ell_i(r_i; \Theta)$  的表达式为:

$$\ell_i(r_i; \Theta) = -\frac{1}{2} \left\{ \log(bc) + \log \sigma_i^2 + (\eta + 1) \left\{ \begin{array}{l} \log \left[ 1 + \frac{1}{\eta - 2} \left( \frac{bz_i/\sigma_i + a}{-\nu} \right)^2 \right] \mathbb{I}(z_i < l) + \\ \log \left[ 1 + \frac{1}{\eta - 2} \left( \frac{bz_i/\sigma_i + a}{\nu} \right)^2 \right] \mathbb{I}(z_i \geq l) \end{array} \right\} \right\} \quad (26)$$

其中,  $a = 4\lambda c(\eta - 2)/(\eta - 1)$ ,  $b^2 = 1 + 3\lambda^2 - a^2$ , 且  $c = \frac{\Gamma((\eta + 1)/2)}{\sqrt{\pi(\eta - 2)}\Gamma(\eta/2)}$ 。另外,参数  $\eta$  控制尾部厚度,满足  $2 < \eta < \infty$ , 而  $0 < \nu < 2$  为偏度参数,当  $1 < \nu < 2$  时分布呈现右偏,当  $0 < \nu < 1$  时呈现左偏。

在得到对数似然函数  $\ell(r, x; \Theta)$  的具体表达式之后,模型参数的最大似然估计值可以通过以下方式获得:

$$\hat{\Theta} = \arg \max_{\Theta} \ell(r, x; \Theta)$$

此外,将短期波动的初始值设置为其无条件均值,即  $\log h_0 = 0$ 。为了初始化长期波动  $\log g_t$ , 将 MIDAS 结构中滞后的对数已实现测度均取值为  $\log x_1$ 。

### 三、实证分析

在这一部分,我们研究 REGARCH-MIDAS 模型与基于跳跃、好坏波动率拓展的 REGARCH-MIDAS-CJ 模型, REGARCH-MIDAS-RS 模型和 REGARCH-MIDAS-RSJ 模型的实证拟合效果及预测表现。首先选取实证数据,并进行初步的统计检验与分析,确认模型对于实证数据的适配性,在此基础上给出上述模型周线与月线设定下共 8 个模型的参数估计结果。进一步,对比这些模型在不同准则下的样本外预测性能,以探索引入跳跃与好坏波动率是否能提高 REGARCH-MIDAS 模型的预测能力。

#### (一) 数据选择与基本统计分析

沪深 300 指数作为衡量中国股市的最佳单项指标之一,被广泛应用于波动率模型的实证研究<sup>[8-9]</sup>中。本文选取沪深 300 指数的价格数据作为样本,时间跨度为 2015 年 1 月 5 日到 2021 年 3 月 3 日(共计 1500 个交易日),数据来源于 JoinQuant 数据库。此外,为减少微观结构噪声的影响,本文采用 5 分钟的取样频率,由此得到的高频数据总量为 72000 个(每天 48 个高频价格记录)。采用“收盘价—收盘价”计算对数收益率(数值放大 100 倍以便观测处理),并按照式(2)计算得到第  $t$  日的已实现波动率  $RV_t$ 。图 1 给出了指数

序列的日收盘价、收益率和已实现波动率的时序图以及对数已实现波动率的自相关函数图。从图中可以看到,样本期内沪深300指数的价格波动比较明显,收益率序列也存在明显的波动集聚现象。选取的样本区间涵盖了指数的牛市、熊市与平稳期,比较有代表性。此外,对数已实现波动率的ACF呈现缓慢衰减的趋势,这表明波动具有持续性,存在一定的长记忆性。

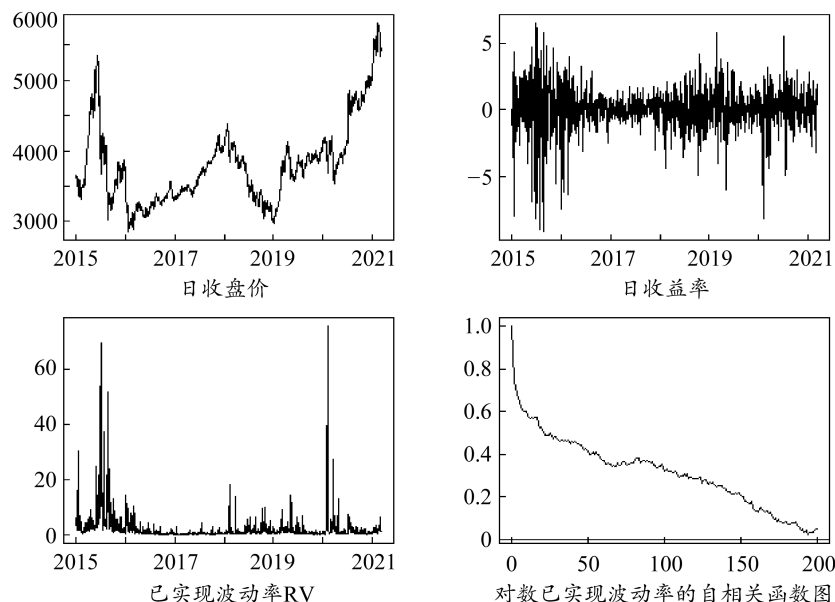


图1 沪深300指数日收盘价、日收益率和已实现波动率的时序图以及对数已实现波动率的自相关函数图

另外,表1列出了沪深300指数的日收益率和日对数已实现波动率的描述性统计信息。从J-B检验的结果,结合偏度与峰度值,可以看出收益率与对数已实现波动率的分布均偏离正态分布,这说明设定模型残差服从非正态分布的合理性。ADF检验的结果显示两个序列均不存在单位根过程,可以判定序列都是平稳的。Ljung-Box检验证实,两个样本序列均存在明显的自回馈效应,可以进一步证实沪深300指数的波动存在长记忆特征。此外,对数收益率的单样本 $t$ 检验的 $p$ 值为0.604,可以判定收益率的均值是显著等于零的,那么在模型的收益方程剔除常数项是合理的。

表1 沪深300指数日对数收益率序列和对数已实现波动率的描述性统计

	均值	最小值	最大值	标准差	偏度	峰度	J-B值	ADF	LBQ(20)
$r_t$	0.027	-9.154	6.499	1.536	-1.011	6.204	2669.744	-28.284	74.633
$\log RV_t$	-0.035	-2.716	4.323	1.104	0.633	0.545	119.275	-9.446	11082.930

注:表中峰态系数为超额峰态,J-B值为Jarque-Bera统计量,ADF为Augmented Dickey-Fuller test的统计量,LBQ(n)为滞后阶数为 $n$ 的Ljung-BoxQ统计量

接下来,对沪深300指数的已实现波动率序列进行分解:首先,利用ADS检测法,按照式(4)定义的跳跃检验统计量对指数在样本期内每一个交易日是否存在跳跃行为进行检验,检验的显著性水平设定为 $\alpha = 0.05$ ;其次,根据检验结果,按照公式(6)和(7)将已实现波动率进行跳跃分解,得到离散的显著性跳跃 $CJ_t$ 和连续波动率 $CRV_t$ ;进一步,区分日内正负收益,依据公式(8)和(9)计算每日的好波动率 $RS_t^+$ 与坏波动率 $RS_t^-$ ,以此将已实现波动率进行非对称分解;最后,按照公式(10)和(11)将跳跃波动分解为正跳跃 $CJ_t^+$ 与负跳跃 $CJ_t^-$ ,并得到正连续波动 $CRV_t^+$ 与负连续波动 $CRV_t^-$ 。图2展示了沪深300指数这八种日高频波动率序列的时序图。

从图2可以发现,各序列均存在明显的波动集聚现象,在2015年至2016年这段时间普遍出现大幅波动。具体来看,样本期内沪深300指数的离散跳跃分布密集,计算得出跳跃显著的交易日共计480日,占比超过30%,而连续波动率的聚集性相比跳跃会更强一些。此外,沪深300指数在样本期内由负收益构成的



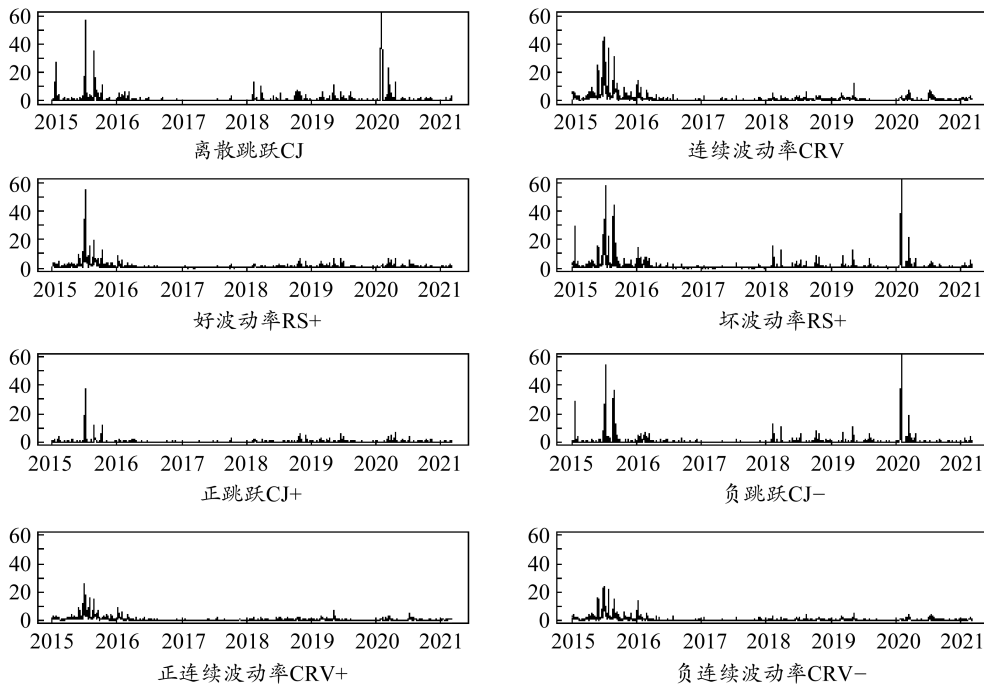


图 2 沪深 300 指数各日波动率序列的时序图

坏波动率相较受正收益影响的好波动率振幅更大,但收益波动出现负跳跃的天数要小于出现正跳跃的天数。从图中还可以看到,2020 年初由新冠肺炎疫情引发的指数波动的主要成分为离散跳跃,而且这一波动主要是由负收益导致的,即负跳跃。

### (二) 样本内估计与结果分析

在本节,基于沪深 300 指数的全样本数据,对周线与月线(周线: $N = 5, K = 52$ ;月线: $N = 22, K = 12$ )设定下的 REGARCH-MIDAS 模型与拓展的 REGARCH-MIDAS-CJ 模型、REGARCH-MIDAS-RS 模型与 REGARCH-MIDAS-RSJ 模型共计 8 个模型进行参数估计,并利用似然函数与信息准则来评估模型的拟合效果,具体的估计结果见表 2。

表 2 模型参数估计结果

	REG-MIDAS (weekly)	REG-MIDAS (monthly)	REG-MIDAS- CJ(weekly)	REG-MIDAS- CJ(monthly)	REG-MIDAS- RS(weekly)	REG-MIDAS- RS(monthly)	REG-MIDAS- RSJ(weekly)	REG-MIDAS- RSJ(monthly)	
$\beta$	0.543(0.000)	0.856(0.000)	0.508(0.000)	0.874(0.000)	0.505(0.000)	0.831(0.000)	0.522(0.000)	0.856(0.000)	$\beta$
$\tau_1$	-0.090(0.000)	-0.072(0.000)	-0.094(0.000)	-0.072(0.000)	-0.113(0.000)	-0.095(0.000)	-0.102(0.000)	-0.079(0.000)	$\tau_1$
$\tau_2$	0.058(0.000)	0.066(0.000)	0.060(0.000)	0.069(0.000)	0.060(0.000)	0.067(0.000)	0.061(0.000)	0.069(0.000)	$\tau_2$
$\alpha$	0.176(0.000)	0.259(0.000)	0.165(0.000)	0.255(0.000)	0.140(0.000)	0.242(0.000)	0.152(0.000)	0.243(0.000)	$\alpha$
$\xi$	-0.475(0.000)	-0.475(0.000)	-0.474(0.000)	-0.461(0.000)	-0.478(0.000)	-0.471(0.000)	-0.476(0.000)	-0.459(0.000)	$\xi$
$\varphi$	0.967(0.000)	0.960(0.000)	0.976(0.000)	0.961(0.000)	0.978(0.000)	0.970(0.000)	0.984(0.000)	0.969(0.000)	$\varphi$
$\delta_1$	-0.068(0.000)	-0.070(0.000)	-0.066(0.000)	-0.069(0.000)	-0.066(0.000)	-0.072(0.000)	-0.065(0.000)	-0.069(0.000)	$\delta_1$
$\delta_2$	0.146(0.000)	0.148(0.000)	0.148(0.000)	0.150(0.000)	0.146(0.000)	0.148(0.000)	0.147(0.000)	0.148(0.000)	$\delta_2$
$\omega$	0.507(0.000)	0.544(0.000)	0.756(0.000)	1.285(0.000)	1.226(0.000)	1.196(0.000)	1.438(0.000)	1.845(0.000)	$\omega$
$\lambda$	0.933(0.000)	0.881(0.000)	—	—	—	—	—	—	—
$\lambda^J$	—	—	-0.194(0.016)	-2.461(0.000)	—	—	0.159(0.051)	-0.512(0.057)	$\lambda^{J+}$
$\lambda^C$	—	—	0.962(0.000)	1.201(0.000)	—	—	-0.405(0.004)	-2.142(0.000)	$\lambda^{C-}$
$\lambda^+$	—	—	—	—	0.698(0.000)	1.392(0.002)	0.537(0.000)	0.882(0.021)	$\lambda^{C+}$
$\lambda^-$	—	—	—	—	0.247(0.005)	-0.443(0.033)	0.420(0.002)	0.226(0.074)	$\lambda^{C-}$
$\gamma$	41.817(0.000)	11.442(0.001)	55.434(0.000)	9.196(0.000)	51.414(0.000)	12.539(0.000)	56.709(0.000)	11.066(0.000)	$\gamma$
$\eta$	4.819(0.000)	4.797(0.000)	4.898(0.000)	4.989(0.000)	4.847(0.000)	4.902(0.000)	4.929(0.000)	5.052(0.000)	$\eta$
$\nu$	0.927(0.000)	0.926(0.000)	0.930(0.000)	0.928(0.000)	0.928(0.000)	0.927(0.000)	0.930(0.000)	0.928(0.000)	$\nu$
$\sigma_u^2$	0.310(0.000)	0.313(0.000)	0.295(0.000)	0.302(0.000)	0.303(0.000)	0.309(0.000)	0.293(0.000)	0.302(0.000)	$\sigma_u^2$

(续表 2)

	REG-MIDAS (weekly)	REG-MIDAS (monthly)	REG-MIDAS- CJ(weekly)	REG-MIDAS- CJ(monthly)	REG-MIDAS- RS(weekly)	REG-MIDAS- RS(monthly)	REG-MIDAS- RSJ(weekly)	REG-MIDAS- RSJ(monthly)	
$\ell(r, x)$	-3636.533	-3642.629	-3597.906	-3613.795	-3619.528	-3632.239	-3594.159	-3613.755	$\ell(r, x)$
AIC	7301.107	7313.258	7225.813	7261.509	7269.056	7294.479	7222.318	7257.590	AIC
BIC	7375.482	7387.634	7312.632	7351.823	7348.745	7374.167	7305.501	7337.278	BIC
VR	0.844	0.620	0.871	0.708	0.859	0.670	0.875	0.712	VR

注:括号里为对应参数  $t$  检验的  $p$  值;最右侧一列为 REGARCH-MIDAS-RSJ 模型的参数

接下来,根据表 2 的估计结果,从以下几个角度展开分析:

**1. 总体拟合效果分析。**从表 2 可知,已实现波动率无论按照周线还是月线聚合,相比 REGARCH-MIDAS 模型,基于跳跃、好坏波动率改进的三个模型在全似然值  $\ell(r)$  与信息准则 AIC 与 BIC 上均有更好的表现,其中 REGARCH-MIDAS-RSJ 模型的样本内表现最优,这说明将长期方程中的已实现波动率进行分解能够提升 REGARCH-MIDAS 模型的整体拟合效果,而同时考虑波动率的跳跃性与非对称性对模型拟合效果的提升最大。另外,周线下的 REGARCH-MIDAS 族模型普遍比月线下的同类模型拥有更优的总体拟合表现,这一结果与 Borup 和 Jakobsen(2019)<sup>[15]</sup> 的研究一致。

**2. 参数合理性分析。**从估计结果来看,所有模型参数的检验  $p$  值都小于 0.1,且大部分接近于 0,参数显著性较强,相应的符号和大小也与理论假设相符,且不同模型下参数估计值比较接近,说明此类模型对于拟合沪深 300 指数收益率的波动具有一定的稳健性。具体来看,所有模型参数  $\beta$  的估计值均小于并偏离 1,周线下各模型的  $\beta$  值更是在 0.5 左右,这是因为参数  $\beta$  仅反映短期波动的影响,而按照周线聚合的长期方程能更好地刻画波动持续性。模型中的杠杆参数都显著且符号相反( $\tau_1$ 、 $\delta_1$  为负, $\tau_2$ 、 $\delta_2$  为正),表明沪深 300 指数收益波动在短期内具有明显的杠杆效应,且正向(负向)收益的冲击会抑制(加剧)波动。测量方程中的参数  $\xi$  均为负,且参数  $\varphi$  均显著且接近于 1,表明条件方差与已实现波动率具有高度一致性。长期方程中参数  $\lambda$  (包括各模型的不同  $\lambda$  参数)均显著,这说明以 MIDAS 结构引入的各已实现波动序列的历史信息对于模型波动率的刻画是有意义的。所有模型的偏度估计值  $\nu$  均小于 1,峰度估计值  $\eta > 3$ ,说明条件分布呈左偏厚尾形态,进一步显示将残差设定为服从偏  $t$  分布的合理性。最后,相比 REGARCH-MIDAS 模型,三个拓展模型的参数  $\sigma_u^2$  估计值均更小,显示了更优的整体拟合效果。

**3. 跳跃波动与连续波动的贡献分析。**由表 2 可知,REGARCH-MIDAS-CJ 模型中,参数  $\lambda^J$  与  $\lambda^C$  均显著且符号相反( $\lambda^J$  为负, $\lambda^C$  为正),说明历史的跳跃波动信息对于沪深 300 指数收益的条件方差在长期来看具有抑制作用,而持续较高的连续波动则会导致较高的长期波动性。另外,按月线聚合的已实现波动序列相比周线对于条件方差的持续性影响相对更强一些。以上结果显示,当沪深 300 指数收益率出现显著的跳跃时,可能会导致收益波动的短期震荡,但长期来看这种影响持续较短,股票市场的长期波动主要来源于收益率的日常连续波动。

**4. 好坏波动率的贡献分析。**表 2 中 REGARCH-MIDAS-RS 模型的参数  $\lambda^+$  和  $\lambda^-$  都是显著的,说明沪深 300 指数的好坏波动率具有不对称的波动冲击,而  $\lambda^+$  显著为正,且显著性和绝对值都高于  $\lambda^-$ ,说明由日内正向收益形成好波动率相比由日内负向收益形成的坏波动率对收益率条件方差的长期影响更强。同样,按月线聚合的已实现波动序列相比周线的持续性影响相对更强。以上结果显示,沪深 300 指数收益的长期波动性与过去的好波动率的关系更密切,这是因为正收益导致的持续好波动率使得投资者在长期时间范围内期望股价继续上涨,而对市场的波动造成更持久的影响。

**5. 正、负跳跃的贡献分析。**表 2 中 REGARCH-MIDAS-RSJ 模型的参数  $\lambda^{J+}$  和  $\lambda^{J-}$  显著,说明沪深 300 指数的跳跃波动存在非对称性,而  $\lambda^{J-}$  显著为负,且显著性和绝对值都高于  $\lambda^{J+}$ ,说明由负收益导致的负跳跃对于沪深 300 指数收益的长期波动具有明显的负向冲击,而正跳跃对波动的持续性影响相对较弱。另外,参数  $\lambda^{C+}$  和  $\lambda^{C-}$  显著为正,但  $\lambda^{C+}$  的显著性和绝对值要高于  $\lambda^{C-}$ ,说明正、负连续波动对于沪深 300 指数的长期波动都具有正向影响,且正连续波动的影响更大。对于这些结果一个可能的解释是我国股票市场的投资者对于收益率波动的正跳跃和负跳跃存在不同的态度,当股票价格大幅下跌时,虽然短时间内会造成较

大的波动,但长期来看会降低投资者的市场参与积极性,导致负跳跃的影响程度逐渐降低,而出现收益的正跳跃时,投资者往往会情绪高涨,市场参与度随之上升,从而对波动性造成持续性影响。

**6. 波动持续性分析。**图 3 展示了所有模型的拟合条件方差  $\sigma_t^2$  (灰色部分) 以及长期波动  $g_t$  (黑色线条)。可以看到,周线下的 REGARCH-MIDAS 族模型拟合的长期波动与条件方差比较接近,而月线下各个模型的长期波动部分则比较平滑。为了更准确地说明模型拟合的条件方差中有多少可归因于长期因素,本文计算方差比率  $VR = \frac{Var(\log g_t)}{Var(\log \sigma_t^2)}$ ,表 2 最后一行的计算结果显示,长期因素对于条件方差波动的贡献较大,周线下模型的长期波动方差占比均接近 90%,而月线下模型的已实现波动率是月度聚合的,这导致波动被平滑,因此具有较小的方差比率。另外,基于跳跃、好坏波动率改进的三个模型的方差比率相比同时间线下的 REGARCH-MIDAS 更高,其中 REGARCH-MIDAS-RSJ 模型的方差比率最高,这一结果进一步说明对长期方程中的已实现波动率进行分解能够提升模型对于波动持续性的捕捉能力,且同时考虑两个因素对模型捕捉长记忆性的改善效果最优。

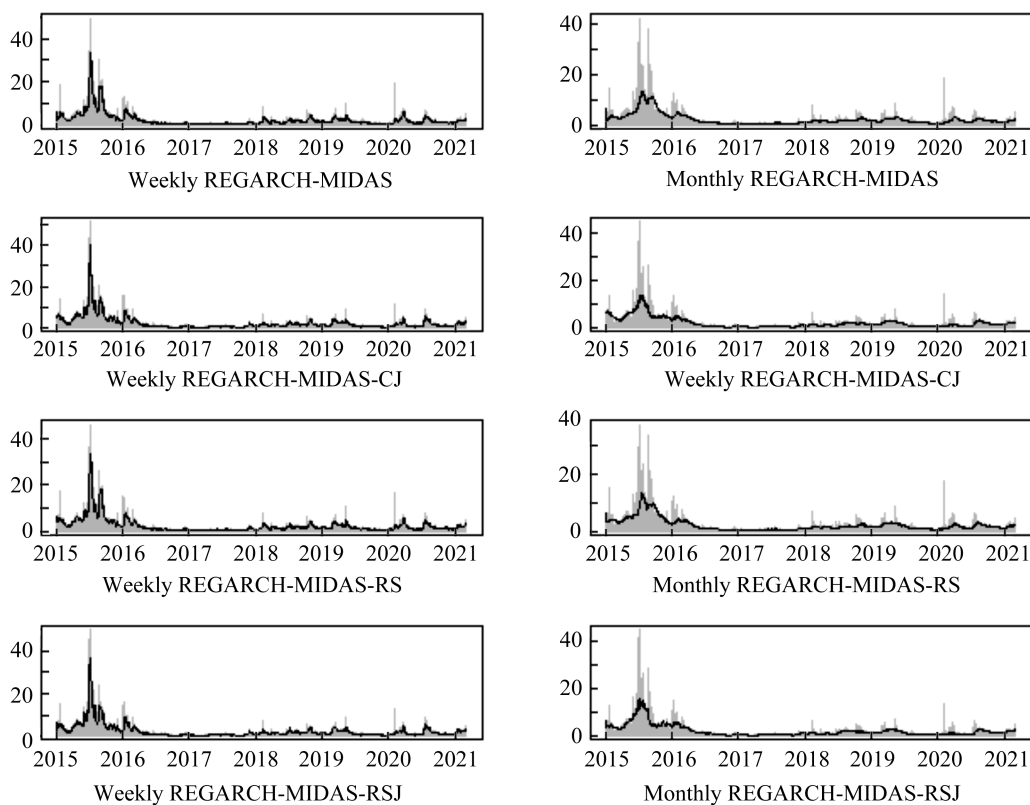


图 3 各模型的拟合条件方差  $\sigma_t^2$  以及长期波动  $g_t$  时序图

### (三) 条件方差的自相关函数

为了更清晰地展示以上模型对于波动持续性的刻画能力,本节通过自相关函数(ACF)图来研究比较各个模型的自相关结构。图 4 展示了周线下 REGARCH-MIDAS 模型及其拓展模型的对数条件方差的模拟 ACF 和样本 ACF。其中实线为模型通过样本实际拟合得到的条件波动率的样本 ACF(SACF),虚线则为利用表 2 中的估计参数获得的模拟 ACF(月线下每个模型的 ACF 图形及相关结论与周线下类似)。其中,模拟选择的样本长度为 2000(约 8 年),在上文定义的残差分布下分别进行 10000 次 Monte Carlo 模拟。一般来说,模拟的 ACF 与样本 ACF 靠得越近,则理论模型与实际模型之间的内部一致性就越高,能够更好地捕捉条件方差的相依结构。

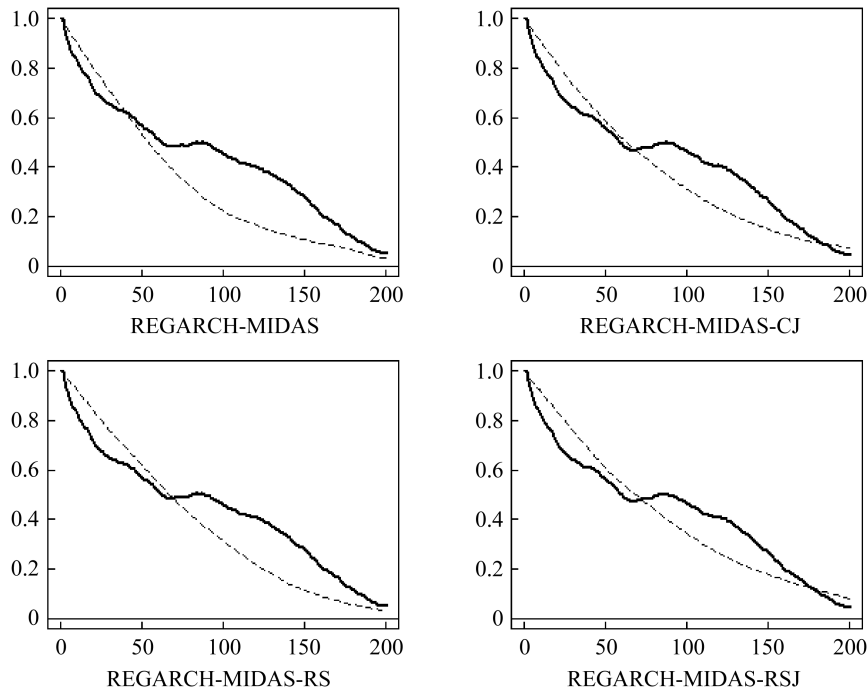


图4 周线下各模型对数条件方差的模拟(虚线)和样本(实线)自相关函数

从图4可以看到,引入乘法分量结构的 REGARCH-MIDAS 族模型能够在较长时间范围上捕捉波动率的相依性,总体而言这种刻画能力在滞后 70 期范围内更佳。此外,相比 REGARCH-MIDAS 模型,基于跳跃、好坏波动率改进的三个模型的模拟 ACF 与样本 ACF 曲线整体上都更接近,这说明本文提出的改进模型在一定程度上具有更强的波动率长记忆性捕捉能力。另外,REGARCH-MIDAS-CJ 和 REGARCH-MIDAS-RSJ 模型的两条 ACF 曲线在滞后 70 期至 200 期的范围内比较靠近,其中 REGARCH-MIDAS-RSJ 模型的 ACF 曲线最为贴近,说明在模型的长期波动方程中考虑跳跃分解能够比较明显地提升模型在长期范围内捕捉波动持续性的能力。

#### (四) 样本外波动率预测及检验结果

由于市场参与者更关心的是模型对未来的预测能力,而不是对过去的分析能力,所以模型预测能力的样本外表现相比样本内表现更重要<sup>[43]</sup>。因此,本文通过样本外预测对所提出的 REGARCH-MIDAS-CJ、REGARCH-MIDAS-RS 与 REGARCH-MIDAS-RSJ 模型的性能进行研究比较,以此探讨在模型中考虑跳跃行为和非对称效应是否可以提高收益波动的预测能力。具体地,本文采用固定窗口大小为 1200(约占数据集总长度的 80%) 的滚动窗口对模型进行重新估计,并在每一次估计后计算未来  $k$  日(最大预测步长为 22 日)的样本外条件方差预测值  $\hat{\sigma}_{t+k|t}^2$ 。其中,样本外的一步预测( $k = 1$ )可以直接通过模型中的 GARCH 方程[式(15)]和长期方程[式(17)、式(21-23)]计算得到。而对于多步预测( $k > 1$ ),首先由 GARCH 方程[式(15)]可以得到:

$$\log h_{t+k} = \beta^k \log h_t + \sum_{j=1}^k \beta^{j-1} \left( \tau(z_{t+k-j}) + \alpha u_{t+k-j} \right) \tag{27}$$

所以,对数条件方差  $\log \sigma_{t+k|t}^2$  的表达式为:

$$\begin{aligned} \log \sigma_{t+k|t}^2 &\equiv E[\log \sigma_{t+k}^2 | F_t] = E[\log h_{t+k} | E_t] + E[\log g_{t+k} | F_t] \\ &= \beta^k \log h_t + \beta^{k-1} \left( \tau(z_t) + \alpha u_t \right) + \log g_{t+k|t} \end{aligned} \tag{28}$$

因此,已知  $t$  时刻的  $h_t, u_t$  和  $z_t$  就可以直接计算短期成分对预测的贡献。为了获取长期成分  $g_{t+k|t}$ ,本文通过模拟方法来获得预测值,首先采取 bootstrap 方法从  $z_t$  和  $u_t$  的经验分布(在同一历史观察窗口中获得)中重新采样,并利用测量方程(16)计算得到已实现波动率的预测值,然后通过长期方程估计的参数递归



得到长期波动的预测值。需要说明的是,对于 REGARCH-MIDAS-CJ 模型,在得到已实现波动率的预测值后,需重新进行一次跳跃检验(仍采用 ADS 检验,且其中的 MedRV 与 MedTQ 由前日的原样本数据计算得到)以将其分解为跳跃波动与连续波动;同样,对于 REGARCH-MIDAS-RS 模型,也需要将得到的已实现波动率预测值进一步分解为好、坏波动率,两者具体的比例与前日原样本数据计算的两个指标的比例一致;REGARCH-MIDAS-RSJ 模型则需进行跳跃分解与非对称分解两个步骤。最后,在进行  $M$  次 ( $M = 10000$ ) 模拟后得到预测的条件方差为:

$$\sigma_{t+k|t}^2 = \frac{1}{M} \sum_{m=1}^M \exp(\log \sigma_{t+k|t,m}^2) \tag{29}$$

接下来,与学者们之前的研究(王天一,2018<sup>[13]</sup>;Borup 和 Jakobsen,2019<sup>[15]</sup>)保持一致,在稳健损失函数族中选取对称损失函数均方误差(MSE)和非对称的准似然函数(QLIKE)来评估预测性能。损失函数的表达式如下表示:

$$L_{t+k}^{(MSE)}(\hat{\sigma}_{t+k}^2, \sigma_{t+k|t}^2) = (\hat{\sigma}_{t+k}^2 - \sigma_{t+k|t}^2)^2 \tag{30}$$

$$L_{t+k}^{(QLIKE)}(\hat{\sigma}_{t+k}^2, \sigma_{t+k|t}^2) = \frac{\hat{\sigma}_{t+k}^2}{\sigma_{t+k|t}^2} - \log\left(\frac{\hat{\sigma}_{t+k}^2}{\sigma_{t+k|t}^2}\right) - 1 \tag{31}$$

其中,  $\hat{\sigma}_{t+k}^2$  为条件方差的真实值,取调整后的已实现核估计作为代理变量(Huang 等,2016<sup>[11]</sup>;Borup 和 Jakobsen,2019<sup>[15]</sup>),即  $\hat{\sigma}_t^2 = cRK_t$ ,其中  $c = \frac{\sum_{t=1}^T r_t^2}{\sum_{t=1}^T RK_t}$ 。作出这一调整的原因在于:已实现波动率是“开盘价—

收盘价”收益率方差的度量,而通过模型得到的预测条件方差度量的是“收盘价—收盘价”的收益率方差。

为了评估本文提出的三个改进模型是否显著地提升了 REGARCH-MIDAS 模型的预测性能,进一步进行 Diebold-Mariano 检验。DM 检验对两组模型的预测值损失函数进行比较,检验其中一组模型预测的损失函数是否比另一组有显著减小,原假设为两个模型具有相同的预测精度,备择假设为第二组模型的预测精度更高,具体的统计量定义如下:

$$S = \frac{\bar{d}}{\sqrt{\frac{\hat{LRV}}{T}}} \tag{32}$$

其中,  $\bar{d} = \sum_{t=1}^T d_t/T$ ,  $d_t$  为两组模型间的损失函数差值,  $\hat{LRV} = \gamma_0 + 2 \sum_{j=1}^H K\left(\frac{j}{H}\right) \text{cov}(d_t, d_{t-j})$ ,  $K(\cdot)$  为用于 LRV 的 Newey-West 估计的 Bartlett 核函数,  $H$  是由 Andrews(1991)<sup>[44]</sup> 估计的最佳带宽参数。

表 3 给出了以上 8 个不同波动率模型的样本外滚动预测的损失函数结果,以及 6 个改进模型相对基准 REGARCH-MIDAS 模型的损失函数 DM 检验结果。

表 3 REGARCH-MIDAS 模型与对应拓展模型样本外波动率预测的损失函数及 DM 检验结果

预测步长		$k = 1$		$k = 5$		$k = 22$	
损失函数		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
REGARCH-MIDAS	weekly	1.000	1.000	1.000	1.000	1.000	1.000
	monthly	1.050	1.047	1.014	1.025	0.974	0.937
REGARCH-MIDAS-CJ	weekly	0.987*	0.930*	0.972**	0.906**	0.897***	0.899***
	monthly	0.991*	0.948*	0.977**	0.888*	0.641***	0.788***
REGARCH-MIDAS-RS	weekly	0.989*	0.941*	0.975**	0.903*	0.911***	0.862**
	monthly	0.998*	0.984*	0.992**	0.950*	0.612***	0.785**
REGARCH-MIDAS-RSJ	weekly	0.979*	0.923**	0.966***	0.820**	0.725***	0.804***
	monthly	0.989*	0.944*	0.972**	0.852**	0.612***	0.735***

注:表中的数值为损失函数标准化的结果,具体方法参考陈声利等(2018)<sup>[34]</sup>,以 REGARCH-MIDAS(weekly)模型为基准模型,将其余模型的损失函数均值除以基准模型的损失函数均值,值小于1表示对应模型的预测能力有所提高,加粗表示对应模型的损失函数均值最小。\*表示相应的损失函数 DM 检验对应  $p$  值小于 0.1,即在 10% 的水平下认为损失函数存在显著的差异,相应地 \*\*与 \*\*\* 分别表示对应  $p$  值小于 0.05 与  $p$  值小于 0.01

从表3可得,在所有预测范围上,同一时间线下,基于跳跃、非对称拓展的 REGARCH-MIDAS-CJ 模型、REGARCH-MIDAS-RS 模型与 REGARCH-MIDAS-RSJ 模型的两个损失函数均值皆比对应的 REGARCH-MIDAS 模型的更小,结合 DM 检验结果(均至少在0.1的水平下具有显著差异),可以说明本文提出的改进模型均显著改善了 REGARCH-MIDAS 模型的预测性能,且在预测期较长时,改进效果更为显著(相应  $p$  值更小),而其中 REGARCH-MIDAS-RSJ 模型相比单独考虑跳跃行为或非对称效应的另外两个改进模型对原有模型预测精度的提升更加明显。另外,表3的结果显示,随着预测步数的增加,损失函数的比值逐渐减小,这表明在模型中引入时间推移的 MIDAS 结构是有效的,能够明显地提升模型在长时间范围上的预测精度。值得注意的是,在短期预测步长上( $k=1, k=5$ ),周线下的模型比月线下的对应模型普遍拥有更小的损失函数均值,而在长期预测步长上( $k=22$ ),月线下的模型预测表现相对更优。

值得注意的是,在采用某种损失函数作为模型的评判标准时,可能会存在波动率模型优劣的错误判断问题(Hansen 和 Lunde, 2005)<sup>[45]</sup>。为使本文的研究结论有更强的稳健性,进一步进行“模型可信集”(Model Confidence Set, MCS)检验。MCS 检验是 Hansen 等(2011)<sup>[46]</sup>提出的一种基于损失函数的模型检验方法,其检验的过程如下:

第一步,设  $M_0$  为最开始用于比较的  $m_0$  个模型构成的集合,按照损失函数的定义计算第  $i$  个模型损失函数值  $L_{i,h}$ ,  $i=1, \dots, m_0$ ,  $h$  为预测序列长度,并计算任意两个模型  $u$  和  $v$  预测值的相对损失函数值  $d_{uv,h}$ ;

第二步,在显著性水平  $\alpha$  上,以两个模型具有相同的预测能力作为原假设,构造相应的统计量进行等价检验,若原假设不能被拒绝,则定义  $\hat{M}_{1-\alpha}^* = M$ , 否则,按照 Hansen 等(2011)<sup>[46]</sup> 定义的剔除准则将集合  $M_0$  中预测能力较差的模型予以删除;

第三步,对于新的模型集合  $M$ ,重复第二步,在每一次等价检验  $\delta_M$  中原假设可以统一表示为:

$$H_{0,M}: E(d_{uv,h}) = 0, \text{ for all } u, v \in M \subset M_0,$$

并按剔除准则  $e_m$  在模型集中进行持续的检验,直到没有模型被剔除该集合为止,得到“最优模型集”。

对于第二步中构造的统计量,采用较为常用的范围统计量  $T_R = \max_{u,v \in M} \frac{|\bar{d}_{uv}|}{\sqrt{\text{var}(\bar{d}_{uv})}}$  和半二次方统计量  $T_{SQ} = \max_{u,v \in M} \frac{(\bar{d}_{uv})^2}{\text{var}(\bar{d}_{uv})}$ , 其中  $\bar{d}_{uv}$  表示模型  $u$  和  $v$  的损失函数值差的均值。

此外,参照 Hansen 等(2011)<sup>[46]</sup>,本文采用 Bootstrap 法来获得统计量  $T_R$  和  $T_{SQ}$  的估计值及相应的  $p$  值,且检验的控制参数按如下设置:块长度  $k$  由对所有损失差异进行自回归处理后得到的最大有效参数数确定;模拟次数  $B=10000$  次;检验的显著性水平  $\alpha$  取值为 0.1。表4展示了所有模型波动率预测的 MCS 检验结果。从表4可知,在10%的显著性水平上,大部分模型在两种损失函数下的不同步数波动率预测通过了 MCS 检验,这说明 REGARCH-MIDAS 族模型能够很好地预测沪深300指数收益率的波动情况。具体来看,本文基于跳跃、好坏波动率拓展的三个模型在日( $k=1$ )、周( $k=5$ )、月( $k=22$ )三个预测范围内均改进了原始 REGARCH-MIDAS 模型的预测性能,具体体现在 MCS 的检验  $p$  值都更大。这说明本文对于 REGARCH-MIDAS 模型做出的改进能使得其在预测精度方面得到提升。其中,REGARCH-MIDAS-RSJ 模型的预测效果是最好的(成为预测最优模型的次数最多),这说明在 REGARCH-MIDAS 模型的长期方程中同时考虑跳跃与非对称性的影响,能够使模型的波动率预测准确性得到实质性的改进。另外,对于较短时间范围的波动率预测(日、周),周线下的模型普遍拥有更好的预测表现;而对于月度这种较长时间范围的预测,月线下的模型能够提供更准确的预测结果。

综合以上所有检验结果(表3、表4)上可知,相比 REGARCH-MIDAS 模型,本文基于跳跃、好坏波动率提出的三个拓展模型能够为沪深300指数提供更准确的收益率波动预测结果,且预测表现最优的是综合考虑这两个因素的 REGARCH-MIDAS-RSJ 模型。这一预测性能的提升可能在于:拓展的模型包含反映跳跃引起的大幅波动信息(REGARCH-MIDAS-CJ),或者包含关于收益波动的非对称信息(REGARCH-MIDAS-RS),又或者考虑正、负跳跃的波动信息(REGARCH-MIDAS-RSJ),而这些丰富的信息能够提升模型对长期波动的刻画,进而对条件方差的预测产生影响。

表 4 不同模型波动率预测的 MCS 检验结果

MSE 损失函数							
预测步长		k = 1		k = 5		k = 22	
检验统计量		$T_R$	$T_{SQ}$	$T_R$	$T_{SQ}$	$T_R$	$T_{SQ}$
REGARCH-MIDAS	weekly	0.874	0.806	0.490	0.429	0.044	0.031
	monthly	0.775	0.695	0.063	0.027	0.065	0.024
REGARCH-MIDAS-CJ	weekly	0.912	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.152	0.361
	monthly	0.874	<b>1.000</b>	0.702	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
REGARCH-MIDAS-RS	weekly	0.898	<b>1.000</b>	0.847	<b>1.000</b>	0.077	0.310
	monthly	0.841	0.997	0.490	0.600	<b>1.000</b>	<b>1.000</b>
REGARCH-MIDAS-RSJ	weekly	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.854	<b>1.000</b>
	monthly	0.877	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>

QLIKE 损失函数							
预测步长		k = 1		k = 5		k = 22	
检验统计量		$T_R$	$T_{SQ}$	$T_R$	$T_R$	$T_{SQ}$	$T_R$
REGARCH-MIDAS	weekly	0.782	0.796	0.596	0.374	0.017	0.054
	monthly	0.077	0.155	0.203	0.335	0.062	0.077
REGARCH-MIDAS-CJ	weekly	0.993	<b>1.000</b>	0.999	<b>1.000</b>	0.348	0.732
	monthly	0.442	<b>1.000</b>	0.249	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
REGARCH-MIDAS-RS	weekly	0.978	<b>1.000</b>	0.775	<b>1.000</b>	0.196	0.663
	monthly	0.422	0.248	0.249	0.593	0.757	0.819
REGARCH-MIDAS-RSJ	weekly	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.948	<b>1.000</b>
	monthly	0.830	<b>1.000</b>	0.596	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>

注:本表报告了各模型在两种损失函数下的 MCS 检验的  $p$  值

### (五) 风险值的预测及检验

考虑到对于投资者与监管机构而言,资产的最大损失是较为直观地反映资产风险的指标,所以本节在波动率预测的基础上,通过计算 VaR 与 ES 这两个风险度量指标来进一步判断以上模型的预测效果。

金融市场上测度风险最主流的方法是在险价值<sup>[47]</sup> (Value at Risk, VaR),具体的计算这里采用的是模型法,即利用 REGARCH-MIDAS 族模型中的条件方差  $\sigma_t^2$  来估计 VaR,具体的计算公式为:

$$VaR_{t+1} = t_{\alpha}^{skt} \cdot \hat{\sigma}_{t+1} \quad (33)$$

其中,  $t_{\alpha}^{skt}$  为偏  $t$  分布的上  $\alpha$  分位数,  $\hat{\sigma}_{t+1}$  为向前一步预测的收益率波动条件标准差。

在实际的风险管理中, VaR 经常低估真实损失,为此引入期望损失<sup>[48]</sup> (Expected Shortfall, ES),其定义为损失超过 VaR 时的预期损失:

$$ES_{t+1} = E(r_{t+1} | r_{t+1} \leq VaR_{t+1}) \quad (34)$$

由此,利用上文得到的向前一步条件方差预测值,采用文献研究中常用的5%的覆盖水平,通过公式(33)与(34)可以计算得到沪深300指数样本外的一步预测风险值 VaR 与 ES。

图5展示了预测区间沪深300指数的样本对数收益率以及由 REGARCH-MIDAS-RSJ (Weekly) 模型计算得到的对应预测风险值,其中实线为 VaR 预测值,虚线为 ES 预测值。可以看出,在预测区间内,由 REGARCH-MIDAS-RSJ (Weekly) 模型预测的风险值与样本收益率波动形态一致,且能够覆盖大部分的负向收益。经过进一步计算可得,预测区间内 VaR 预测失败的天数占总预测天数的比例(失败率)约为0.047,接近给定的覆盖率  $\alpha$ ,且通过了后验测试,条件覆盖(Conditional Coverage, CC) 检验<sup>[49]</sup> 的  $p$  值约为0.858。以上结果说明 REGARCH-MIDAS-RSJ (Weekly) 模型产生了有效的

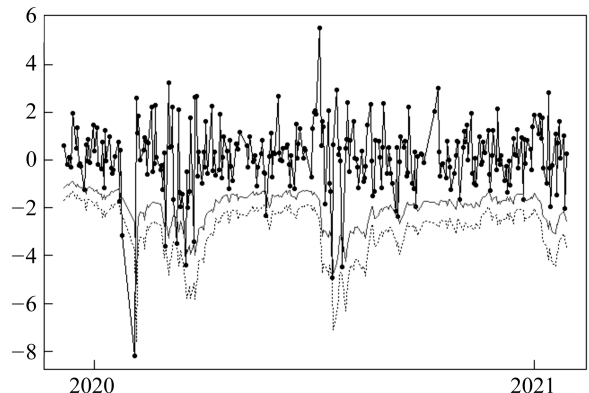


图5 预测区间沪深300指数对数收益率及其预测风险值时序图

VaR 风险预测(其余模型也均通过了 VaR 的后验测试,详细结果限于篇幅未列出)。

接下来,为评价和比较各模型的 VaR 预测效果,进一步进行损失值检验。首先,采用分位数损失函数<sup>[50]</sup>(Quantilie Loss, QL)来评价各模型预测 VaR 的能力,具体定义如下:

$$QL_{\alpha} = (\alpha - L_{t+1})(r_{t+1} - VaR_{t+1}) \quad (35)$$

其中,  $L_{t+1} = I(r_{t+1} < VaR_{t+1})$ 。

其次,为了评估以上模型预测 VaR 和 ES 的综合能力,采用 Fissler 和 Ziegel(2016)<sup>[51]</sup>提出的 FZ 损失函数,并进一步参考 Patton 等(2019)<sup>[52]</sup>,假定 VaR 与 ES 严格为负,且满足  $ES_{t+1} \leq VaR_{t+1} < 0$ ,由此得到以下损失函数:

$$FZ_{\alpha}^{(0)} = \frac{1}{\alpha ES_{t+1}} L_{t+1}(r_{t+1} - VaR_{t+1}) + \frac{VaR_{t+1}}{ES_{t+1}} + \log(-ES_{t+1}) - 1 \quad (36)$$

同样,为了更稳健地评估不同模型的风险预测能力差异,在如上定义的两类损失函数下,进行 MCS 检验,检验结果如表5所示。

表5 不同模型风险值预测的 MCS 检验结果

检验统计量		QL 损失函数		FZ 损失函数	
		$T_R$	$T_{SQ}$	$T_R$	$T_{SQ}$
REGARCH-MIDAS	weekly	0.957	0.964	0.780	<b>1.000</b>
	monthly	0.387	0.574	0.284	0.605
REGARCH-MIDAS-CJ	weekly	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
	monthly	0.763	<b>1.000</b>	0.768	0.999
REGARCH-MIDAS-RS	weekly	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
	monthly	0.470	0.716	0.462	0.709
REGARCH-MIDAS-RSJ	weekly	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
	monthly	0.952	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>

注:本表报告了各模型在两种损失函数下的 MCS 检验的  $p$  值

从表5的检验结果可知,在10%的显著性水平下,所有模型均通过了 MCS 检验,且有较多的模型 MCS 检验  $p$  值为1,这说明利用 REGARCH-MIDAS 族模型来预测沪深300指数收益率的风险值是稳健的,此类模型是一个很有效的股票市场风险测度工具。此外,周线下的模型相比月线下的对应模型普遍提供了更准确的样本外一步的风险预测值,其中本文提出的三个拓展模型(Weekly)在两种损失函数下的 MCS 检验  $p$  值均为1,显示了十分卓越的风险预测能力。综合来看,三个拓展模型(REGARCH-MIDAS-CJ、REGARCH-MIDAS-RS 与 REGARCH-MIDAS-RSJ)出现最优预测(MCS 检验  $p$  值为1)的次数都比 REGARCH-MIDAS 模型(1次)更多,分别为5次、4次与7次,这说明本文将 REGARCH-MIDA 模型进行跳跃、非对称拓展能够提升原有模型的风险预测能力,预测精度最高的模型是 REGARCH-MIDAS-RSJ 模型,这也说明在同一模型框架下同时考虑跳跃与非对称性的对于模型风险预测能力的提升最大。

## 四、结论及展望

在高频金融数据建模的背景下,考虑到金融资产价格普遍存在的跳跃行为,以及收益率波动的长记忆性与非对称性,本文基于跳跃、好坏波动率对 REGARCH-MIDAS 模型进行拓展,提出 REGARCH-MIDAS-CJ 模型、REGARCH-MIDAS-RS 模型与 REGARCH-MIDAS-RSJ 模型,提升了模型对收益率条件方差的刻画能力。以沪深300指数价格高频数据为样本,应用以上模型实证分析揭示了中国股票市场的波动性规律,进一步对收益率的波动性与风险值进行预测,以损失函数的 DM 及 MCS 检验来考察所提出的模型是否能达到



更高的预测精度。其中,模型的改进基于三种形式的已实现波动分解:连续波动与跳跃波动、好与坏波动率以及正负跳跃与正负连续波动。此外,考虑到收益的偏峰厚尾性,设定残差服从偏  $t$  分布,并利用稳健的 ADS 检测法来甄别跳跃。本文的实证结果及相关结论有如下几点:

首先,样本内结果表明,本文提出的 REGARCH-MIDAS 拓展模型均能够更好地捕捉波动率的长记忆性,由此提升了原模型的整体估计表现,且 REGARCH-MIDAS-RSJ 模型的样本内表现最佳,即同时考虑跳跃与非对称波动的影响对模型的拟合效果提升最大。另外,周线下的 REGARCH-MIDAS 族模型普遍比月线下的同类模型拥有更优的样本内估计效果。参数估计结果显示,沪深300指数的长期波动主要来源于收益率的日常连续波动而非跳跃波动,其中由正向收益形成的正连续波动的影响更大;另外,沪深300指数的波动具有明显的杠杆效应,负向(正向)收益的冲击会加剧(抑制)短期波动,而负跳跃对于指数的长期波动具有明显的负向冲击。

其次,样本外波动率滚动预测的损失函数及 DM 与 MCS 检验结果显示,REGARCH-MIDAS 族模型能够提供比较准确的沪深300指数收益率波动的预测值。相比 REGARCH-MIDAS 模型,本文提出的三个拓展模型在预测性能上均有所提升,而 REGARCH-MIDAS-RSJ 模型的预测效果是最优的,这说明将 REGARCH-MIDAS 模型长期方程中的已实现波动率同时进行跳跃与非对称分解能够使模型的预测准确性得到实质性的改进。另外,对于 REGARCH-MIDAS 族模型,周线下的模型能够提供更准确的日、周(短时间)波动率预测,而对于月度这种较长时间范围的预测,月线下的模型预测精度更高。

最后,风险值预测及相应的损失函数 MCS 检验结果显示,REGARCH-MIDAS 族模型是很有效的风险预测工具,而本文基于跳跃、好坏波动率提出的三个拓展模型均显示了十分卓越的风险预测能力,其中预测精度最高的是 REGARCH-MIDAS-RSJ 模型,这也说明在同一模型框架下同时考虑跳跃与非对称性对于模型风险预测能力的提升最大。

综合以上研究结论,本文最主要的贡献在于进一步扩展和丰富了 REGARCH-MIDAS 模型,引入了跳跃和好坏波动率以完善原有模型的波动性结构,并证实改进后的模型能够为沪深300指数提供更准确的波动率预测结果,为市场参与者进行风险管理提供参考。本文提出的模型是基于混频数据建立的,不仅包含日内高频价格信息,还融合了日、周、月的已实现波动信息,并在模型的改进中进一步细分了长期波动率的波动形式,或是跳跃引起的大幅波动信息,或是关于收益波动的非对称信息,或是综合考虑这两种信息,而这些丰富的信息能够提升模型的波动率预测准确性。在之后的研究中,可以在模型中引入更丰富的已实现测度,如广义已实现测度等,或者在长期方程中进一步融合其他对波动有影响的外生变量,如宏观经济变量、不确定性变量等。另外,也可将创新的模型应用于其他资产(如债券、汇率、商品期货等)的波动性研究,并可进一步利用模型来进行投资策略选取、衍生品定价等金融应用。

#### 参考文献:

- [1] GHYSELS E, SANTA-CLARA P, VALKANOV R. Predicting volatility: getting the most out of return data sampled at different frequencies[J]. *Journal of Econometrics*, 2006, 131(1): 59-95.
- [2] ANDERSEN T G, BOLLERSLEV T. Answering the skeptics: yes, standard volatility models do provide accurate forecasts[J]. *International Economic Review*, 1998, 39(4): 885-905.
- [3] ENGLE R F. New frontiers for Arch models[J]. *Journal of Applied Econometrics*, 2002, 17(5): 425-446.
- [4] HANSEN P R, HUANG Z, SHEK H H. Realized GARCH: a joint model for returns and realized measures of volatility[J]. *Journal of Applied Econometrics*, 2012, 27(6): 877-906.
- [5] HANSEN P R, HUANG Z. Exponential GARCH modeling with realized measures of volatility[J]. *Journal of Business and Economic Statistics*, 2016, 34(2): 269-287.
- [6] 王天一, 黄卓. 高频数据波动率建模——基于厚尾分布的 Realized GARCH 模型[J]. *数量经济技术经济研究*, 2012(5): 149-161.
- [7] 黄友珀, 唐振鹏, 周熙雯. 基于偏  $t$  分布 Realized GARCH 模型的尾部风险估计[J]. *系统工程理论与实践*, 2015(9): 2200

- 2208.
- [8]于孝建,王秀花.基于混频已实现 GARCH 模型的波动预测与 VaR 度量[J].统计研究,2018(1):104-116.
- [9]蒋伟,顾研.基于广义已实现测度的 Realized GARCH 模型改进及应用[J].数量经济技术经济研究,2019(7):156-173.
- [10]MULLER U A,DACOROGNA M M,DAVE R D,et al. Volatilities of different time resolutions——analyzing the dynamics of market components[J]. *Journal of Empirical Finance*,1997,4(2):213-239.
- [11]HUANG Z,LIU H,WANG T. Modeling long memory volatility using realized measures of volatility: a realized HAR GARCH model[J]. *Economic Modelling*,2016,52:812-821.
- [12]CORSI F. A simple approximate long-memory model of realized volatility[J]. *Journal of Financial Econometrics*,2009,7(2):174-196.
- [13]王天一,刘浩,黄卓.基于混频数据抽样的已实现波动率长记忆模型[J].系统工程学报,2018(6):812-822.
- [14]GHYSELS E,SINKO A,VALKANOV R. MIDAS regressions: further results and new directions[J]. *Econometric Reviews*,2007,26(1):53-90.
- [15]BORUP D,JAKOBSEN J S. Capturing volatility persistence: a dynamically complete realized EGARCH-MIDAS model[J]. *Quantitative Finance*,2019,19(11):1839-1855.
- [16]MAHEU J M,MCCURDY T H,ZHAO X. Do jumps contribute to the dynamics of the equity premium? [J]. *Journal of Financial Economics*,2013,110(2):457-477.
- [17]BARNDORFF-NIELSEN O E,SHEPHARD N. Power and bipower variation with stochastic volatility and jumps[J]. *Journal of Financial Economics*,2004,2(1):1-37.
- [18]BARNDORFF-NIELSEN O E,SHEPHARD N. Econometrics of testing for jumps in financial economics using bipower variation[J]. *Journal of Financial Econometrics*,2006,4(1):1-30.
- [19]ANDERSEN T G,BOLLERSLEV T,DIEBOLD F X. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility[J]. *The Review of Economics and Statistics*,2007,89(4):701-720.
- [20]AT-SAHALIA Y,JACOD J. Testing for jumps in a discretely observed process[J]. *The Annals of Statistics*,2009,37(1):184-222.
- [21]BOLLERSLEV T,TODOROV V,LI S Z. Jump tails, extreme dependencies, and the distribution of stock returns[J]. *Journal of Econometrics*,2013,172(2):307-324.
- [22]ANDERSEN T G,DOBREV D,SCHAUMBURG E. Jump-robust volatility estimation using nearest neighbor truncation[J]. *Journal of Econometrics*,2012,169(1):75-93.
- [23]SANTOS D G,ZIEGELMANN F A. Volatility forecasting via MIDAS, HAR and their combination: an empirical comparative study for IBOVESPA[J]. *Journal of Forecasting*,2014,33(4):284-299.
- [24]MA F,LIAO Y,ZHANG Y,et al. Harnessing jump component for crude oil volatility forecasting in the presence of extreme shocks[J]. *Journal of Empirical Finance*,2019,52:40-55.
- [25]PAN Z,BU R,LIU L,et al. Macroeconomic fundamentals, jump dynamics and expected volatility[J]. *Quantitative Finance*,2020,20(8):1345-1371.
- [26]ENGLE R F,GHYSELS E,SOHN B. Stock market volatility and macroeconomic fundamentals[J]. *The Review of Economics and Statistics*,2013,95(3):776-797.
- [27]NELSON D B. Conditional heteroskedasticity in asset returns: a new approach[J]. *Econometrica*,1991,59(2):347-370.
- [28]AWARTANI B M,CORRADI V. Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries[J]. *International Journal of Forecasting*,2005,21(1):167-183.
- [29]陈浪南,黄杰鲲.中国股票市场波动非对称性的实证研究[J].金融研究,2002(5):67-73.
- [30]BARNDORFF-NIELSEN O E,KINNEBROCK S,SHEPHARD N. Measuring downside risk-realized semivariance [M]//BOLLERSLEV T,RUSSEL J,WATSON M. Volatility and time series econometrics. Oxford: Oxford University Press,2010:117-136.
- [31]PATTON A J,SHEPPARD K. God volatility, bad volatility: signed jumps and the persistence of volatility[J]. *Review of Economics and Statistics*,2013,97(3):683-697.

- [32] PAN Z, LIU L. Forecasting stock return volatility: a comparison between the roles of short-term and long-term leverage effects[J]. *Physica A: Statistical Mechanics and its Applications*, 2018, 492: 168–180.
- [33] WANG L, MA F, LIU J, et al. Forecasting stock price volatility: new evidence from the GARCH-MIDAS model[J]. *International Journal of Forecasting*, 2020, 36(2): 684–694.
- [34] 陈声利, 关涛, 李一军. 基于跳跃、好坏波动率与百度指数的股指期货波动率预测[J]. *系统工程理论与实践*, 2018(2): 299–316.
- [35] 蔡光辉, 应雪海. 基于跳跃、好坏波动率和马尔科夫状态转换的高频波动率模型预测[J]. *系统科学与数学*, 2020(3): 521–546.
- [36] MERTON R C. Option pricing when underlying stock returns are discontinuous[J]. *Journal of Financial Economics*, 1976, 3(1): 125–144.
- [37] BARNDORFF-NIELSEN O E, SHEPHARD N. Estimating quadratic variation using realized variance[J]. *Journal of Applied Econometrics*, 2002, 17(5): 457–477.
- [38] 夏婷, 闻岳春. 经济不确定性是股市波动的因子吗? ——基于 GARCH-MIDAS 模型的分析[J]. *中国管理科学*, 2018(12): 1–11.
- [39] CONRAD C, KLEEN O. Two are better than one: volatility forecasting using multiplicative component GARCH-MIDAS models[J]. *Journal of Applied Econometrics*, 2020, 35(1): 19–45.
- [40] 孙洁. 考虑跳跃和隔夜波动的中国股票市场波动率建模与预测[J]. *中国管理科学*, 2014(6): 114–124.
- [41] LAURENT S, SHI S. Volatility estimation and jump detection for drift-diffusion processes[J]. *Journal of Econometrics*, 2020, 217(2): 259–290.
- [42] 唐勇, 崔金鑫. 中国股票市场最优套期保值比率研究——基于高阶矩 HAR 模型[J]. *系统科学与数学*, 2018(9): 1036–1054.
- [43] 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[J]. *系统工程理论与实践*, 2020(5): 1113–1133.
- [44] ANDREWS D W K. Heteroskedasticity and autocorrelation consistent covariance matrix estimation[J]. *Econometrica*, 1991, 59(3): 817–858.
- [45] HANSEN P R, LUNDE A. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? [J]. *Journal of Applied Econometrics*, 2005, 20(7): 873–889.
- [46] HANSEN P R, LUNDE A, NASON J M. The model confidence set[J]. *Econometrica*, 2011, 79(2): 453–497.
- [47] JORION P. Measuring the risk in value at risk[J]. *Financial Analysts Journal*, 1996, 52(6): 47–56.
- [48] ACERBI C, TASCHE D. On the coherence of expected shortfall[J]. *Journal of Banking and Finance*, 2002, 26(7): 1487–1503.
- [49] CHRISTOFFERSEN P F. Evaluating interval forecasts[J]. *International Economic Journal*, 1998, 39(4): 841–862.
- [50] KOENKER R. Quantile regression[M]. Cambridge: Cambridge University Press, 2005: 712–718.
- [51] FISSLER T, ZIEGEL J F. Higher order elicibility and Osband's principle[J]. *The Annals of Statistics*, 2016, 44(4): 1680–1707.
- [52] PATTON A J, ZIEGEL J F, CHEN R. Dynamic semiparametric models for expected shortfall (and value-at-risk) [J]. *Journal of Econometrics*, 2019, 211(2): 388–413.



(责任编辑 王 权)