

文章编号:1671-4229(2022)04-0068-12

# 基于混频数据的投资者情绪与股市波动效应的研究

吴文诗<sup>a</sup>, 宋泽芳<sup>a,b\*</sup>, 张兴发<sup>a,b</sup>, 李元<sup>a,b</sup>  
(广州大学 a. 经济与统计学院; b. 岭南统计科学研究中心, 广东 广州 510006)

**摘要:** 根据投资者情绪对股市波动具有重要影响这一观点,引入投资者情绪的传统 GARCH 类波动率模型出现因不同频率数据建模而产生的效率损失问题。文章基于混频数据结构,分别从不同行业、不同情绪状态和不同经济阶段 3 个角度切入,引入自适应权重形式的广义自回归条件异方差混频数据抽样模型(GARCH-MIDAS-adapt),对中国股市日度波动率进行估计与预测比较。实证结果表明,自适应权重形式融合的混频数据结构可以更好地解释投资者情绪对股市产生的长期波动作用,不同行业表现出有显著的解释力和预测力。此外,在不同行业下,情绪低落时对股市的冲击更大。

**关键词:** 投资者情绪; 混频数据; 股市波动; GARCH-MIDAS-adapt; 自适应权重  
**中图分类号:** F 830; O 212 **文献标志码:** A

## Study on effects of investor sentiment and stock market volatility based on mixed data

WU Wen-shi<sup>a</sup>, SONG Ze-fang<sup>a,b\*</sup>, ZHANG Xing-fa<sup>a,b</sup>, LI Yuan<sup>a,b</sup>  
(a. School of Economics and Statistics; b. Center of Statistical Science of Lingnan;  
Guangzhou University, Guangzhou 510006, China)

**Abstract:** According to the view that investor sentiment has an important impact on stock market volatility, the traditional GARCH-type volatility model with investor sentiment has the problem of efficiency loss caused by different frequency data modeling. Based on the mixed frequency data structure, a generalized autoregressive conditional heteroscedasticity mixed frequency data sampling model with an adaptive weight function (GARCH-MIDAS-adapt) is used to estimate and forecast the comparison of daily volatility of the Chinese stock market from three perspectives of different industries, different sentiment states and different economic stages. The results show that mixed data structure with adaptive weight can better explain the long-term volatility of investor sentiment on the stock market, and different industries show significant explanatory and predictive power. In addition, in different industries, the impact of depression on the stock market is larger.

**Key words:** investor sentiment; mixed frequency data; stock market volatility; GARCH-MIDAS-adapt; adaptive weight

在金融时间序列研究中,波动率指标一直是学者们感兴趣的变量之一,因为它的正确度量 and 预测有利于金

融资产定价、套期保值和风险管理。影响波动率的因素很多,最早认为影响金融市场波动的主要因素来自宏观

收稿日期: 2021-09-09; 修回日期: 2021-12-11

基金项目: 国家自然科学基金重点资助项目(11731015); 国家自然科学基金青年资助项目(11701116); 广州市基础与应用基础研究资助项目(202201010276)

作者简介: 吴文诗(1998—),女,硕士研究生. E-mail: wu\_wenshi\_55@163.com

\* 通信作者. E-mail: song\_zefang@163.com

引文格式: 吴文诗, 宋泽芳, 张兴发, 等. 基于混频数据的投资者情绪与股市波动效应的研究[J]. 广州大学学报(自然科学版), 2022, 21(4): 68-79.

经济带来的系统性冲击。随着市场的日益变化,以及行为金融的诞生,投资者情绪对市场波动的主导作用也日益受到学者们的认可,因而围绕投资者情绪与市场波动关系的理论及应用研究一直在不断丰富中。

Bollerslev<sup>[1]</sup>提出 GARCH 模型以及基于它的各种扩展模型(GJR-GARCH、TGARCH 等)在捕捉波动率的集聚、非对称等特点方面给出了很好的表现,很长一段时间,许多学者将它们作为波动率的基准模型。值得关注的是,也有学者将投资者情绪作为经济变量引入到 GARCH 类模型中,展开了投资者情绪对市场波动的影响研究。Beaumont 等<sup>[2]</sup>将个体投资者情绪引入 GARCH-M 模型进行研究,发现波动性在看涨和看跌两种不同情绪中有显著的非对称影响,情绪看跌时,波动性会增加更多。文风华等<sup>[3]</sup>将投资者情绪分为积极与消极情绪,并分别作为股票收益率的解释变量构造 AR-GARCH 模型,结果发现不同投资者情绪波动对股票收益率的波动有显著冲击。

在大部分的研究中,同一个模型内使用的数据是相同频率的,例如在考虑投资者情绪与市场收益之间的关系时,往往都是同为日度或月度数据。但是实际中频率不同的数据更是常见。如对投资者情绪而言,Backer 等<sup>[4]</sup>利用主成分分析方法所构造的情绪度量指标基本是月度频率的数据,沿用它的构造方法,易志高等<sup>[5]</sup>利用封闭式基金折价、交易量、IPO 数量、上市首日收益、消费者信心指数和新增投资者开户数 6 个单项情绪指标构造综合情绪指数,也是以月度频率统计的。

但是研究中所分析的股价数据往往是周度、日度甚至是更高频的数据,为了避免不同频率数据纳入一个模型中,学者们往往就只能迁就一方变量,选择同频的数据进行分析,这就使得研究不够完整,一定程度上会损失一些重要的高频或低频信息,使得结果不尽人意。

为了能够直接对原始混频数据进行分析,一些针对混频数据的模型应运而生。其中, Ghysels 等<sup>[6]</sup>提出混频抽样回归(MIDAS)模型,构造将数据与不同的采样频率结合起来的回归模型,实现了将高频数据与低频数据的融合。如何将混频数据引入到分析市场波动的 GARCH 模型中呢? Engle 等<sup>[7]</sup>从波动分解的思想给混频波动研究提供了思路,提出了市场总波动具有长期波动和短期波动两部分的观点,并将低频信息作为长期波动引入到 GARCH 模型。基于该理论思想, Engle 等<sup>[8-9]</sup>将自己 2008 年提出的 Spline-GARCH 模型与 Ghysels 的混频回归相结合,提出了广义自回归条件异方差混频数据抽样模型(GARCH-MIDAS),模型中代表波动率的条件方差以乘法的形式分解成长期波动和短

期波动两部分,其中代表低频信息的变量作为了长期波动的解释变量。

GARCH-MIDAS 模型设定经济变量的影响根据时间顺序依次递减,即固定采用 Beta 权重形式融合低频变量的长期波动影响。但固定的权重形式或许会因为欠缺对信息本身特征考虑而存在信息损失问题。Ghysels 等<sup>[10]</sup>指出在宏观经济领域进行预测时应该采用阿尔蒙指数滞后多项式的权重形式。于扬等<sup>[11]</sup>依据不同的混频结构,给出了不同情形下混频回归模型的权重形式。这些研究表明,选择合适的权重形式可以使得研究结果更可靠,但多数研究都是将权重固定于某种函数形式。王江涛等<sup>[12]</sup>在研究高频时间序列建模时,考虑一种不依赖于函数形式的自适应方法选择权重,利用交易特征自动调整在不同交易时期的权重分配,结果表明自适应方法下的权重能更好地结合函数自变量描述股价波动特征。这为混频数据的融合提供了一个新的思考方向。自适应权重在混频数据结构中能否发挥更好的作用?本文考虑将自适应的方法融合到 GARCH-MIDAS 模型中。

另外,围绕 GARCH-MIDAS 模型的实证研究中,多以引入低频的宏观经济变量为主,如 Conrad 等<sup>[13]</sup>使用 GARCH-MIDAS 模型提取加密货币的长期与短期波动。石强等<sup>[14]</sup>利用 GARCH-MIDAS 研究日度股市波动与月度宏观经济因子之间的联系。然而针对低频非宏观经济变量的投资者情绪对高频收益的波动性分析研究较少。目前姚尧之等<sup>[15]</sup>利用 GARCH-MIDAS 模型研究投资者情绪与股票价格行为之间的关系发现,混频情绪能够显著影响收益的长期波动,但研究依旧沿用依靠时间顺序这一经济特征,没有从信息本身特点出发进行考虑。考虑到个别股票可能存在研究偏差,本文选择 10 个行业股票作为主体,针对投资者情绪在不同行业、不同情绪状态以及不同经济阶段下的情况,在 GARCH-MIDAS 模型基础上引入交易特定时期自带特征的自适应权重形式构造模型 GARCH-MIDAS-adapt,探讨投资者情绪与股市波动之间效应的多类因子混频模型。

## 1 方 法

### 1.1 投资者情绪度量

目前国内构造投资者情绪指标(IS)的方法有多种,本文采用易志高等利用主成分分析法将封闭式基金折价(CEFD)、交易量(TURN)、IPO 数量(IPON)、上市首日收益(IPOR)、新增投资者开户数(NIA)以及消费者信心指数(CCI)合成的月度综合指标。由于不同指标对投资者情绪的反映可能存在时间上的“提前”或“滞后”,最终运

用主成分方法构造了投资者情绪指数作为情绪度量<sup>[5]</sup>。

### 1.2 GARCH-MIDAS-adapt 模型

为了保证长期波动的值为正, Engle 等提出对数形式的 GARCH-MIDAS 模型<sup>[10]</sup>, 结构如下:

$$r_{i,t} = \mu + \sqrt{\tau_t} \cdot g_{i,t} \varepsilon_{i,t} \quad \forall i = 1, \dots, N_{i,t} \quad (1)$$

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta g_{i-1,t} \quad (2)$$

$$\log \tau_t = m + \theta \sum_{k=1}^K \varphi_k(\omega_1, \omega_2) X_{t-k} \quad (3)$$

$$\varphi_k(\omega) = \frac{(k/K)^{\omega_1-1} (1-k/K)^{\omega_2-1}}{\sum_{j=1}^K (j/K)^{\omega_1-1} (1-j/K)^{\omega_2-1}} \quad (4)$$

其中,  $r_{i,t}$  表示某金融资产第  $t$  个月的第  $i$  天对数收益率,  $N_{i,t}$  表示第  $t$  个月的天数。度量波动率的条件方差  $\tau_t \cdot g_{i,t}$  分解成了两部分, 分别为短期波动  $g_{i,t}$  和长期波动  $\tau_{i,t}$ 。低频经济变量  $X_t$  及它的滞后项  $X_{t-k}$  反映长期波动。上述模型待估参数为:  $\mu, \alpha, \beta, m, \theta, \omega_1, \omega_2$ , 平稳性条件:  $0 < \alpha + \beta < 1, 0 < \alpha, \beta < 1$ 。

利用 Beta 权重函数  $\varphi_k(\omega)$  (式 4) 对经济变量长期波动的影响进行综合,  $K$  是模型最优下的滞后阶数。考虑到 Beta 权重函数权重分配随时间滞后递减的形式可能会忽略特定时期信息, 因此, 本文仿照 GARCH-MIDAS 方法, 利用以下自适应权重函数进行提炼:

$$B(K, \lambda) = \frac{(x_t)^\lambda}{\sum_{k=1}^K (x_k)^\lambda} \quad (5)$$

其中,  $x_t$  是某种与收益率相关的低频变量。结合式 (1) ~ (3) 及 (5), 构建了 GARCH-MIDAS-adapt 模型。借鉴王江涛等<sup>[12]</sup> 权重函数自变量方法, 本文选用月度振幅数据作为自适应权重函数的自变量。该函数中  $K$  是模型最优下的滞后阶数,  $\lambda$  决定了权重函数的大致形态。为刻画低频变量  $x_t$  变化与长期波动具有同步变化的效果, 将  $\lambda$  设定为  $\lambda > 0$ 。

### 1.3 统计推断

#### 1.3.1 参数估计

本文使用拟极大似然估计(QMLE)的方法得到各参数估计值, 模型的似然函数为

$$LLF = -\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^{N_{i,t}} \left[ (2\pi) + \ln(g_{i,t} \tau_t) + \frac{(r_{i,t} - \mu)^2}{g_{i,t} \tau_t} \right] \quad (6)$$

#### 1.3.2 模型预测

为了突显引入投资者情绪与选择自适应权重函数的显著作用, 就必须考虑模型的预测效果。针对混频数

据下的预测, 本文借鉴郑挺国等<sup>[16]</sup> 使用的方法, 计算高频下(日度)的长期波动和短期波动后, 利用计算已实现波动率的方法(下文式 12)得到模型预测低频下(月度)的已实现波动率  $\hat{\sigma}_t^2$ , 直接利用高频收益率数据下计算得到的已实现波动率  $\sigma_t^2$  作为预测标准。本文主要考虑 2 个预测评价指标:

预测绝对误差:

$$MAE = \frac{1}{T} \sum_{t=1}^T | \hat{\sigma}_t^2 - \sigma_t^2 | \quad (7)$$

预测方根误差:

$$RMSE = \left( \frac{1}{T} \sum_{t=1}^T (\hat{\sigma}_t^2 - \sigma_t^2)^2 \right)^{1/2} \quad (8)$$

进一步将 GARCH-MIDAS 模型预测误差与 GARCH (1,1) 模型预测误差相比得到相对预测绝对误差(rMAE)和相对预测方根误差(rRMSE)以评价模型的预测效果。

为了说明 GARCH-MIDAS-adapt 模型样本外预测结果的稳健性, 本文借鉴 Hansen<sup>[17]</sup> 提出的检验预测能力的 SPA 检验方法, 通过设定基准预测模型(GARCH(1,1)) 优于其他预测模型(GARCH-MIDAS-adapt) 的原假设, 利用均方和绝对损失函数  $L(\xi_t, \delta_{k,t})$  定义相对损失变量  $d_{k,t} = L(\xi_t, \delta_{0,t}) - L(\xi_t, \delta_{k,t})$ , 进而构造相对损失变量的检验统计量, 并利用 bootstrap 算法抽样得到相应  $P$  值。其中  $\xi_t$  是随机变量,  $\delta_{k,t}$  表示第  $k$  个预测模型,  $k=0$  表示基准预测模型。

#### 1.3.3 经济来源贡献率

为进一步探讨长期波动存在的价值意义, 衡量投资者情绪对股市波动的贡献, 本文借鉴 Engle 等对长期波动的衡量方法, 基于长期波动与总波动的关系, 分别采用比率 1:  $Var(\log(\tau)) / Var(\log(\tau g))$  与比率 2:  $Var(\log(\tau)) / Var(\log(\tau^{rw} g^{rw}))$  来说明经济变量的解释波动能力。其中比率 2 分母固定为滚动时间的已实现波动率的条件方差, 将模型归一化<sup>[9]</sup>, 进而得到每个模型对股市波动的解释能力差异。

## 2 实证分析

### 2.1 数据来源与描述性统计

为了更全面了解投资者情绪与整个股市波动的影响, 本文选取上证 10 个行业的每日收盘价为研究对象, 分别为能源、材料、工业、可选(可选消费)、消费(主要消费)、医药、金融、信息、电信和公用行业。时间区间为 2009 年 1 月 9 日至 2021 年 2 月 23 日, 每个行业的样本量为 2 949。以上数据均来源于同花顺网站, 并对收盘

价进行以下处理得到各行业对数收益率:

$$r_{j,d} = 100 \times (\log(p_{j,d}) - \log(p_{j,d-1})) \quad (9)$$

$j=1,2,\dots,10$ , 分别对应 10 个行业,  $p_d$  和  $p_{d-1}$  分别代表在  $d$  期和  $d-1$  期的日度收盘价。

投资者情绪指标数据来源于国泰安数据库,运用主成分分析,使用下式得到情绪指数 IS,其中  $t$  表示第  $t$  个月。

$$IS_t = \beta_1 CEFD_t + \beta_2 TURN_{t-1} + \beta_3 IPON_t + \beta_4 IPOR_t + \beta_5 NIA_{t-1} + \beta_6 CCI_t \quad (10)$$

图 1 为 10 个行业收盘价与标准化投资者情绪指标的基本走势。表 1 为收益率与投资者情绪指标描述性统计结果。在单位根(ADF)检验与拉格朗日乘子(LM)检验中所有统计量显著,说明序列平稳且具有 ARCH 效应。

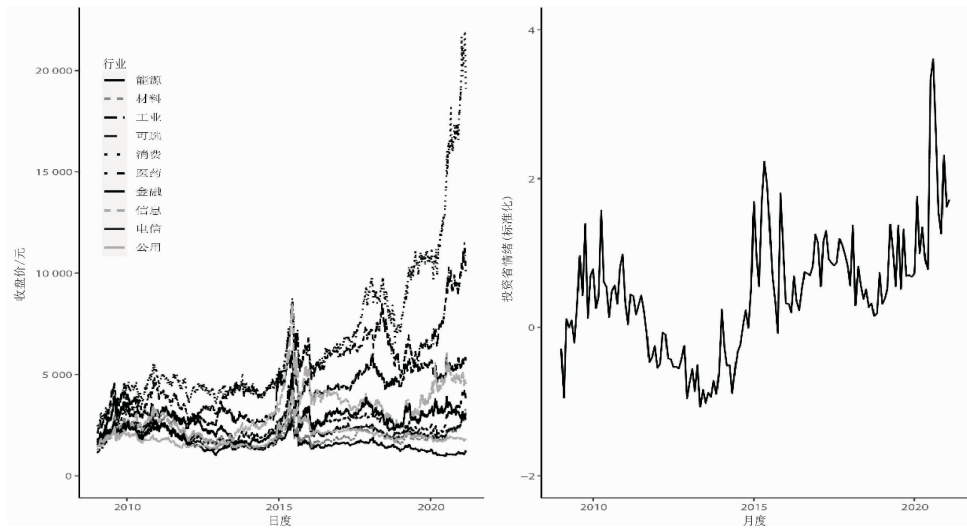


图 1 10 个行业收盘价与投资者情绪指标时序图

Fig. 1 Tendency of closing prices of 10 industries and investor sentiment index

表 1 描述性统计

Table 1 Descriptive statistics

行业	平均值	标准差	中位数	最大值	最小值	ADF 检验	LM 检验
能源	-0.01	1.82	0.00	8.69	-10.53	-14.36 ***	300.30 ***
材料	0.03	1.87	0.07	7.98	-9.65	-13.84 ***	511.52 ***
工业	0.02	1.67	0.03	8.22	-9.46	-13.56 ***	567.05 ***
可选	0.04	1.69	0.09	6.82	-9.68	-13.79 ***	453.59 ***
消费	0.07	1.63	0.12	6.70	-9.00	-13.97 ***	397.31 ***
医药	0.06	1.69	0.08	7.36	-9.00	-13.43 ***	486.79 ***
金融	0.03	1.65	-0.02	8.68	-10.28	-14.58 ***	258.17 ***
信息	0.04	2.09	0.10	7.40	-10.23	-13.94 ***	353.01 ***
电信	0.02	2.00	0.13	7.66	-10.53	-14.27 ***	501.15 ***
公用	0.01	1.46	0.02	7.04	-8.78	-14.19 ***	630.05 ***
IS	41.08	6.54	40.98	65.06	30.49	-	-

注:\*\*\*、\*\*、\* 分别表示在 1%、5%、10% 水平上显著,下同

## 2.2 GARCH-MIDAS-adapt 模型实证结果

因为本文实证数据研究所得各行业的 BIC 值具有随滞后阶数  $K$  增加而逐渐减小的特点,所以所有模型固定滞后阶数为 12 个月。根据月度振幅数据与月度投资者情绪关系,振幅数据选择固定滞后阶数分配权重。

### 2.2.1 引入投资者情绪的 GARCH-MIDAS-adapt 模型估计结果

将投资者情绪  $IS$  作为经济变量  $X$  引入到 GARCH-

MIDAS-adapt 模型中,构建了引入投资者情绪的 GARCH-MIDAS-adapt 模型,记为 M1。结果显示,各个行业参数  $\theta$  显著且表现出的正反向作用表明作为长期波动的投资者情绪会影响股市波动的上升或下降。参数  $\lambda$  显著也说明了自适应权重函数有效性,振幅更大时期的信息所带来的情绪变化对股市波动影响更大,符合经济规律,见表 2。

表 2 引入投资者情绪的 GARCH-MIDAS-adapt 模型估计结果  
Table 2 Estimation results of GARCH-MIDAS-adapt model with investor sentiment

项目	能源	材料	工业	可选	消费	医药	金融	信息	电信	公用
$\mu$	-0.024 7*** (0.000 0)	0.015 7*** (0.000 0)	-0.009 2*** (0.000 0)	0.020 1*** (0.000 0)	0.064 3*** (0.001 2)	0.051 1*** (0.000 4)	0.025 6*** (0.000 3)	-0.013*** (0.000 0)	0.006 1*** (0.000 2)	-0.002 1*** (0.000 0)
$\alpha$	0.063 6*** (0.000 0)	0.066 7*** (0.000 0)	0.072 0*** (0.000 0)	0.062 7*** (0.000 0)	0.069 2*** (0.000 0)	0.051 0*** (0.000 9)	0.072 6*** (0.000 3)	0.042 8*** (0.000 0)	0.067 4*** (0.000 2)	0.074 0*** (0.000 0)
$\beta$	0.923 0*** (0.000 0)	0.918 7*** (0.000 0)	0.910 9*** (0.000 1)	0.931 7*** (0.000 0)	0.913 7*** (0.000 0)	0.946 2*** (0.000 7)	0.916 4*** (0.000 1)	0.950 5*** (0.000 0)	0.909 9*** (0.000 3)	0.918 6*** (0.000 0)
$m$	0.286 2*** (0.000 0)	-0.506 1*** (0.000 0)	1.275 5*** (0.000 0)	-0.570 3*** (0.000 0)	-1.004 9*** (0.003 1)	-4.513 4*** (0.009 2)	0.812 2*** (0.000 7)	-1.353 3*** (0.000 0)	-0.730 1*** (0.000 5)	0.593 8*** (0.000 0)
$\theta$	0.019 3*** (0.000 0)	0.039 0*** (0.000 0)	-0.010 2*** (0.000 0)	0.037 2*** (0.000 0)	0.046 7*** (0.004 8)	0.126 0*** (0.000 2)	0.003 6*** (0.000 4)	0.061 2*** (0.000 0)	0.049 4*** (0.000 0)	0.000 6*** (0.000 0)
$\lambda$	9.998 8*** (0.000 4)	2.021 7*** (0.000 0)	4.187 9*** (0.000 3)	9.562 8*** (0.000 0)	5.361 1*** (0.000 1)	1.000 2*** (0.001 8)	13.824 9*** (0.005 9)	2.697 3*** (0.000 0)	29.995 4*** (0.012 7)	5.545 6*** (0.000 1)
LLF	-4 999.29	-5 070.04	-4 760.01	-4 787.48	-4 814.64	-4 856.02	-4 739.67	-5 519.85	-5 378.36	-4 163.01
BIC	10 052.43	10 193.99	9 572.57	9 627.89	9 682.02	9 764.20	9 532.48	11 093.99	10 811.61	8 377.12
LLF_beta	-5 039.64	-5 112.00	-4 795.36	-4 827.71	-4 856.92	-4 896.74	-4 779.60	-5 565.89	-5 424.59	-4 196.36
BIC_beta	10 133.28	10 277.99	9 642.68	9 708.57	9 766.80	9 845.20	9 612.83	11 186.60	10 904.04	8 443.85
LLF_g	-5 616.00	-5 675.50	-5 299.3	-5 348.30	-5 361.35	-5 389.86	-5 317.45	-6 107.24	-5 936.75	-4 704.38
BIC_g	112 63.95	11 382.95	10 630.56	10 728.55	10 754.65	10 811.67	10 666.85	12 246.44	11 905.46	9 440.71

注:括号中的值为稳健性标准差。LLF、BIC 表示 GARCH-MIDAS-adapt 模型下的似然函数值和 BIC 值,LLF\_beta、BIC\_beta 表示 GARCH-MIDAS 模型下的似然函数值和 BIC 值,LLF\_g、BIC\_g 表示 GARCH(1,1)模型下的似然函数值和 BIC 值

通过与 GARCH(1,1)、GARCH-MIDAS 模型的极大似然值 LLF、BIC 值比较,发现无论什么行业下,GARCH-MIDAS-adapt 模型下的似然函数值都大于 GARCH(1,1)和 GARCH-MIDAS 模型,其 BIC 值也小于另两个模型。

说明引入自适应权重和非宏观经济变量的投资者情绪指标对股市波动拟合效果有明显提升。以上结论表明,投资者情绪对股市波动的影响具有一定的研究意义。

表 3 反映了投资者情绪对波动的贡献程度。

表 3 GARCH-MIDAS-adapt 模型中投资者情绪的贡献率  
Table 3 The contribution of investor sentiment in GARCH-MIDAS-adapt model

项目	能源	材料	工业	可选	消费	医药	金融	信息	电信	公用
比率 1	4.54	9.67	0.69	8.33	22.18	<b>67.96</b>	0.13	22.29	19.95	<b>0.00</b>
比率 2	4.14	8.86	0.69	8.05	22.28	<b>73.06</b>	0.13	24.04	21.57	<b>0.00</b>

注:黑体部分给出占比最高与最低的贡献率。

从表 2 可以看到,各行业股市波动对投资者情绪的反映程度不同,从参数  $\theta$  水平值来看,除工业行业外,其余行业的参数  $\theta$  为正说明投资者情绪波动对大部分行业股市波动有正向促进作用。从绝对值来看,医药行业受投资者情绪的影响最大,其对应的表 3 贡献率也是最大的。这让人们对医药行业的传统认知有了全新的了解,日常生活中医药行业与经济景气关联度较小,人们潜意识认为金融、材料和工业等这些行业受投资者情绪影响会更大,但从长期波动角度来看,医药行业作为与人们生活紧密相关的行业,才是受投资者情绪影响最大的,这也从侧面说明了长期波动与短期波动的区别,以往研究股市波动大多是基于传统 GARCH 类模型,关注短期波动而忽略了长期波动在这过程中的作用。公用

行业自带的政策约束性使得该行业股市发展受到一定牵制,结果也表明该行业所受影响最小,贡献率占比也是最小的。整体来看,作为长期波动的投资者情绪在大部分行业股市波动中的作用不可忽略。

### 2.2.2 扩展 GARCH-MIDAS-adapt 模型估计结果

为着重比较投资者情绪对股市波动的影响,在 GARCH-MIDAS-adapt 模型(M1)下扩展了以下 M2-M7 模型:

$$\log \tau_t = m + \theta \sum_{k=1}^K B(K, \lambda) \Delta IS_{t-k} \quad (11)$$

$$\log \tau_t = m + \theta \sum_{k=1}^K B(K, \lambda) RV_{t-k} \quad (12)$$

$$\log \tau_{i^{(rw)}} = m^{(rw)} + \theta^{(rw)} \sum_{k=1}^K B(K, \lambda) RV_{i-k}^{(rw)} \quad (13)$$



图 2 中的各模型贡献率整体表现出更多的长期信息加入贡献率更高的趋势,尤其医药行业贡献率显著突出。且不同模型对不同行业贡献率趋势接近,如比率 2 中的能源、材料、电信、可选、消费和电信行业。

动受多方面的影响,分别将模型 M1、M2 和 M6 组合,模型 M1、M3 和 M5 组合,大部分行业贡献率有显著提升。因此,对行业股市波动中考虑多个因子可以更全面了解潜在的长期波动。

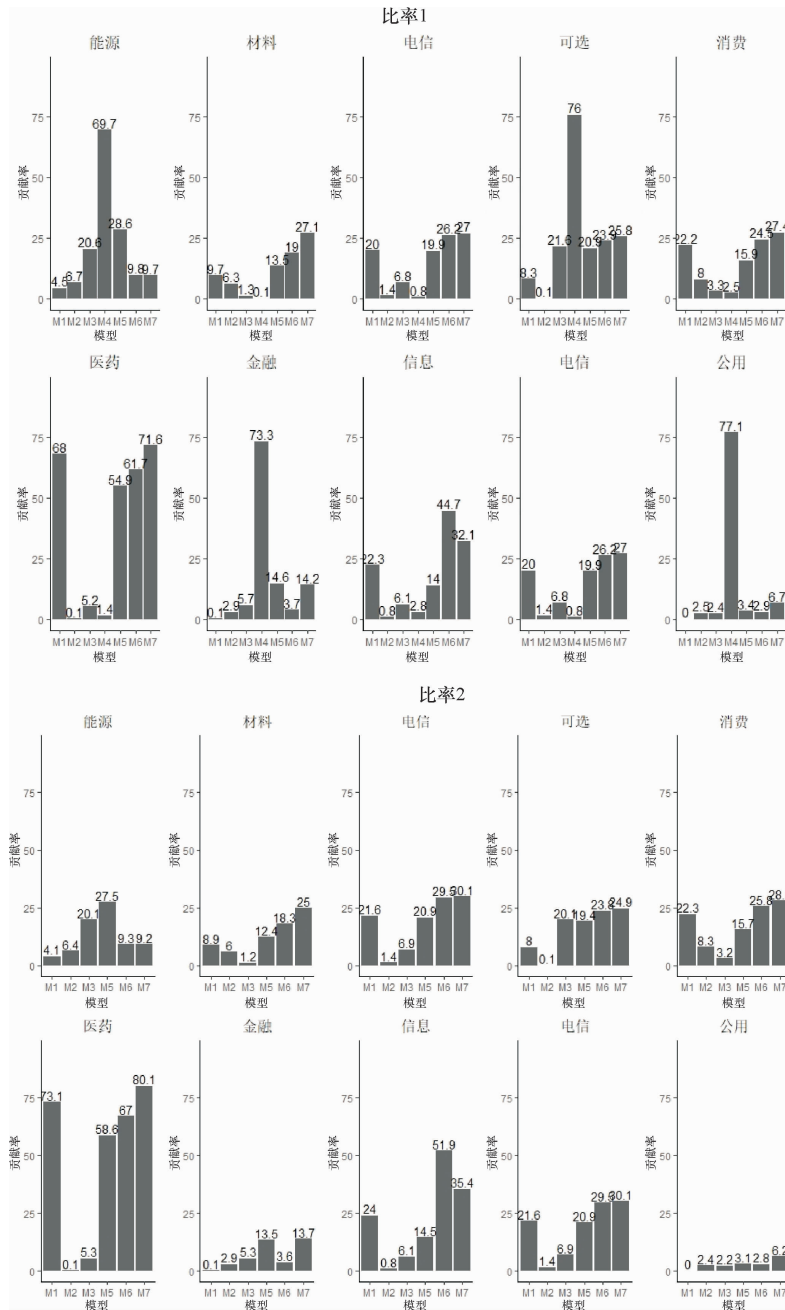


图 2 衡量经济来源贡献率

Fig. 2 Measuring the contribution of economic sources

2.2.3 扩展 GARCH-MIDAS-adapt 模型预测结果

分别以 GARCH(1,1)、GARCH-MIDAS 为基准模型进行预测对比,样本外预测固定为 24 个月。图 3 中给出了 10 个行业 rMAE 和 rRMSE 值。当结果小于 1 时,说明在该行业下该模型的预测效果可能优于基准模型。

从模型比较来看,图 3 中模型依次有 6、5、8、0、7、7、8 个行业优于 GARCH(1,1) 模型。可以看出,虽然模型 M4 预测效果不如 GARCH(1,1),但与投资者情绪因子有关的模型 M1、M2、M5、M6 和 M7 中有 50% 以上行业因为加入了该因子而改善预测精度,同时也说明了混频

模型的构建可以提高预测精度。投资者情绪作为长期波动,在预测过程中起到了一定作用。但与图 2 中的多因子贡献率占比更高的结论相比,多因子的预测效果并没有优于单因子,而是基本保持与单因子持平的状态。因此,预测效果图给出了更多参考。

图 4 中模型分别有超过 60% 行业优于 GARCH-MIDAS,除能源行业,其余行业在 7 种模型下预测结果小于 0 或接近 1。因此,自适应权重函数相较于 Beta 函数具有更佳预测效果。

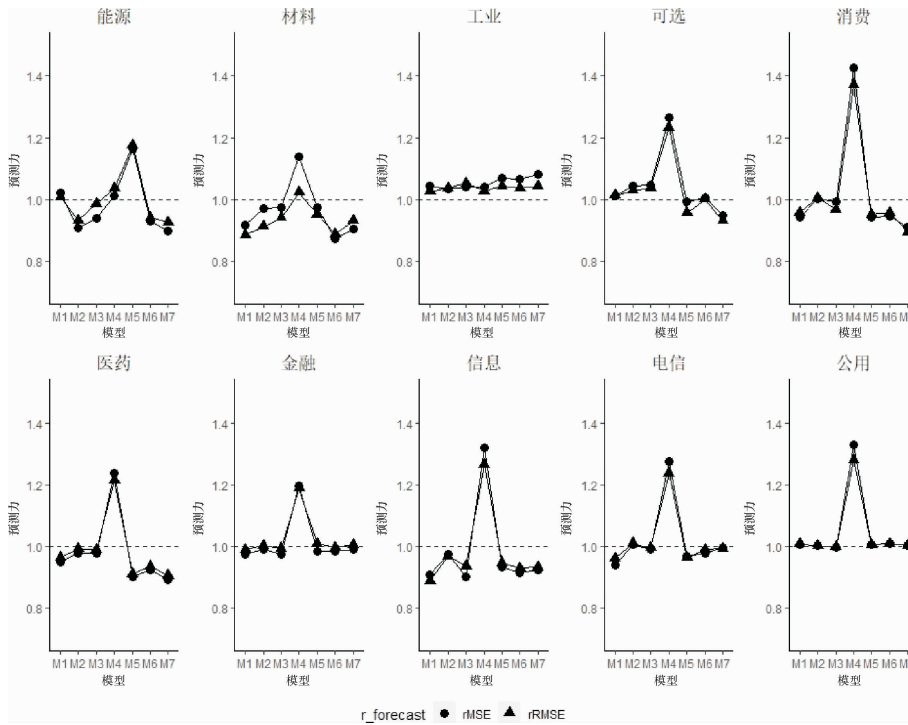


图 3 扩展 GARCH-MIDAS-adapt 模型与 GARCH(1,1) 模型相比预测结果

Fig. 3 The prediction results of extended GARCH-Midas-adapt model are compared with GARCH (1,1) model

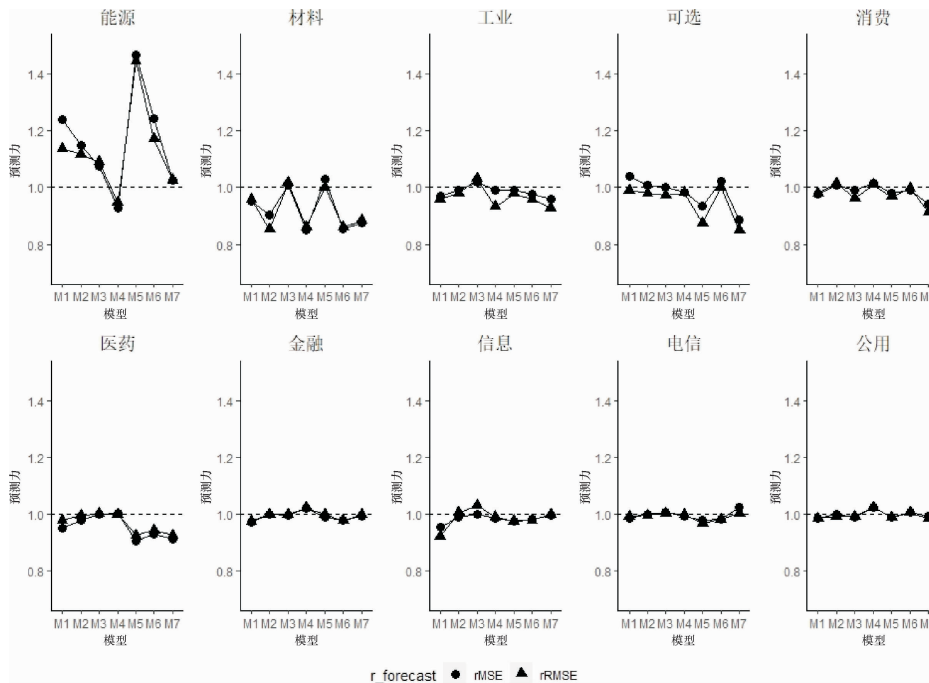


图 4 扩展 GARCH-MIDAS-adapt 模型与 GARCH-MIDAS 模型相比预测结果

Fig. 4 The prediction results of extended GARCH-MIDAS-adapt model are compared with GARCH-MIDAS model

从混频数据角度出发,为了说明构造的 GARCH-MIDAS-adapt 模型预测效果且具有稳健性,仅从预测效果好的行业个数判断并不稳妥。本文借鉴 Hansen 的模型预测能力检验方法(SPA)<sup>[17]</sup>,构造两个原假设:①单因子 GARCH-MIDAS-adapt 模型预测效果优于 GARCH(1,1)模型;②多因子 GARCH-MIDAS-adapt 模型预测效果优于 GARCH(1,1)模型;采用 10 000 次运算的 Bootstrap 方法得到 SPA 检验相应的  $P$  值在表 5 中列出。大部分行业结果具有显著性,表明无论单因子还是多因子 GARCH-MIDAS-adapt 模型,该预测能力都优于 GARCH(1,1)模型,且具有一定的稳健性。

综合模型的估计、经济来源贡献率以及预测效果这 3 个方面来看,投资者情绪的加入能对股市波动进行更好的描述,自适应权重函数也发挥了显著的作用。

### 2.3 不同情绪状态下的 GARCH-MIDAS 模型实证结果

以投资者标准化情绪指标(NIS)作为基准划分为情绪高涨期和情绪低落期,在不同情绪期分别构造 GARCH-MIDAS-adapt 模型,探讨不同情绪状态对股市波

动的影响。表 6 给出主要参数估计结果,整体来看,不同的情绪状态下参数  $\theta$  水平值、绝对值截然不同。

表 5 单因子和多因子 GARCH-MIDAS-adapt 模型预测能力 SPA 检验结果( $P$  值)

Table 5 SPA test results of predictive ability of single factor and multi factor GARCH-MIDAS-adapt models ( $P$  value)

行业	SPA(MSE)-	SPA(RMSE)-	SPA(MSE)-	SPA(RMSE)-
	单因子	单因子	多因子	多因子
能源	0.272 8	0.121 1	0.067 1*	0.019 2**
材料	0.000 0***	0.000 0***	0.031 2**	0.000 0***
工业	0.046 9**	0.273 0***	0.112 4	0.031 8**
可选	0.092 4***	0.042 1**	0.000 0***	0.000 0***
消费	0.130 9	0.115 7	0.172 4	0.000 0***
医药	0.000 0***	0.000 0***	0.000 0***	0.018 0**
金融	0.000 0***	0.000 0***	0.000 0***	0.020 6**
信息	0.152 4	0.040 0**	0.000 0***	0.000 0***
电信	0.000 0***	0.000 0***	0.000 0***	0.000 0***
公用	0.000 0***	0.000 0***	0.712 1	0.592 7

表 6 不同投资者情绪状态下的主要参数估计

Table 6 Estimation of main parameters in different investor sentiment states

行业	回归参数	NIS > 0 (高涨期 C1)			NIS ≤ 0 (低落期 C2)		
		M1	M2	M6	M1	M2	M6
能源	$\theta_1$	0.024 3***	-9.839 1***	0.033 6***	0.108 2***	17.360 4***	0.150 3
	$\theta_2$			-8.154 1***			-9.215 6***
材料	$\theta_1$	0.060 8***	-2.206 1***	0.068 5***	<b>0.110 4***</b>	9.999 9***	0.213 2
	$\theta_2$			-4.633 5***			-33.838 9***
工业	$\theta_1$	-0.032 8***	-4.186 8***	0.012 4***	0.106 2**	9.999 9***	0.059 5***
	$\theta_2$			-4.263 6***			9.639 6**
可选	$\theta_1$	-0.003 2***	-3.180 6**	0.065 9***	0.050 3***	9.999 3***	0.087 3***
	$\theta_2$			-5.410 5***			-10.030 5***
消费	$\theta_1$	0.062 1***	-2.909 3***	0.064 8***	0.003 9***	2.711 5***	0.030 5**
	$\theta_2$			-4.390 4***			0.227 2***
医药	$\theta_1$	<b>0.073 9***</b>	0.836 9***	0.096 2***	0.108 5***	-2.999 9***	0.225 2*
	$\theta_2$			-5.388 9***			-11.152 9***
金融	$\theta_1$	0.019 8***	-3.978 5***	0.029 1***	0.088 5	39.999 7***	0.085 4*
	$\theta_2$			-2.960 2***			40.284 2
信息	$\theta_1$	0.053 4***	-7.574 2***	0.105 1***	-0.062 6***	-21.229 5***	0.071 8
	$\theta_2$			-3.730 1***			-34.307 5***
电信	$\theta_1$	0.070 9***	-9.164 6***	0.074 2***	0.044 5***	-1.136 3***	0.095 4***
	$\theta_2$			-12.976 2***			-8.100 0
公用	$\theta_1$	0.003 8***	-3.177 2***	0.011 9***	0.099 7***	-9.999 7***	0.067 4***
	$\theta_2$			-3.104 4***			-8.311 4***

不同情绪状态下,模型 M1、M2 分别有 3、7 个行业表现出反向作用。且在低落期,模型 M1、M2 的参数绝对值都有 8 个行业大于高涨期状态,模型 M6 中也存在相同特征。表 6 中大部分行业在投资者情绪低落时的影响要大于高涨期的影响,表现更为强烈。结合股市波动中存在的杠杠效应,结果合理。对比全样本下医药行业受投资者情绪影响最大的结论来看,不同情绪下的估计结果有很大区别。在高涨期时,医药行业受投资者情绪影响最大,在低落期时,材料行业受影响最大。

### 2.4 不同经济阶段的 GARCH-MIDAS-adapt 模型实证结果

在 2015 年,多渠道推动股权融资以及投资者大幅度增加资金杠杠使得在 2015 年 6 月之间融资呈爆发式增长。之后多方平台的加入以及实施投资者减少资金杠杠的政策,使得股价出现崩溃式下跌。从图 1 中

也可以看出 2015 年各行业收盘价暴涨后暴跌的趋势,且情绪在 2015 年 6 月前后有明显正负效应。因此,本文将全样本分别分成 2009.01 - 2015.06 和 2015.7 - 2021.02 两个时间子样本,探讨不同经济阶段投资者情绪所产生的影响在长期波动上是否与整体有所不同。

综合表 2、表 7 中结果进行比较,全样本与子样本之间在参数绝对值、参数正负取向上都显著不同。因此,不同经济阶段的研究具有一定研究意义。在 T1、T2 时模型 M1、M2 参数绝对值最大的行业都与全样本结果不同,投资者情绪的长期波动影响出现显著变化。

表 7 的不同阶段对比显示,大部分行业间无明显增减,但能源行业参数在两阶段差异明显,且 M1、M2 模型还表现出相反的增长趋势,即不同指标在两个阶段表现为相反作用。参数绝对值差距也可看出股民们对 2015 年国际油价大跌引发的暴涨暴跌有强烈的心理斗争与博弈。

表 7 2 个子样本主要参数估计

Table 7 Estimation of main parameters of two sub samples

行业	回归参数	2009.01 - 2015.06 (T1)			2015.7 - 2021.02 (T2)		
		M1	M2	M6	M1	M2	M6
能源	$\theta_1$	0.101 1***	-0.078 3***	0.087 5***	0.028 9***	-27.117 9***	0.059 5
	$\theta_2$			8.906 8***			0.614 6***
材料	$\theta_1$	0.098 7***	-1.897 6***	0.091 3***	0.083 3***	-0.728 6***	0.087 5***
	$\theta_2$			-1.367 9***			-13.836 5***
工业	$\theta_1$	0.085 3***	-3.203 0***	0.080 1***	-0.016 8***	-0.999 9***	0.005 9
	$\theta_2$			-4.055 8***			-8.890 4***
可选	$\theta_1$	0.121 9***	3.022 1***	0.078 8***	0.095 9***	4.838 3***	0.107 4
	$\theta_2$			1.507 8***			3.205 4***
消费	$\theta_1$	0.059 7***	2.326 4***	0.058 6	0.100 9***	-0.999 9***	0.090 7***
	$\theta_2$			5.223 3***			-0.787 9***
医药	$\theta_1$	0.082 0***	-0.999 9***	0.100 6	0.096 5***	6.892 8***	0.117 2***
	$\theta_2$			-0.999 8**			3.203 5***
金融	$\theta_1$	0.028 1***	-1.267 2***	0.032 4***	0.073 0***	-0.999 9***	0.075 5***
	$\theta_2$			-1.284 6***			0.615 3***
信息	$\theta_1$	0.036 0***	-4.216 5**	0.034 4***	<b>0.116 5***</b>	-9.778 1***	0.113 0***
	$\theta_2$			-1.529 0***			3.263 8
电信	$\theta_1$	<b>0.131 0***</b>	-2.075 4***	0.083 7***	0.104 9***	4.811 6***	0.095 7***
	$\theta_2$			-3.609 5***			-15.660 7***
公用	$\theta_1$	0.090 3***	<b>-4.635 0***</b>	0.071 6***	-0.002 7***	4.560 2***	0.059 2***
	$\theta_2$			-4.845 9***			5.971 1***

综合表 7、图 5 中的结果,T1 阶段的整体贡献率更突出,尤其是可选和电信行业。但金融行业在 C2 阶段

的贡献率不降反升,经济事件的影响在金融行业得到充分体现。相比全样本,投资者情绪也在 C2 阶段大部分

行业参数绝对值达到最大值,从侧面反映出投资者情绪在经历经济事件后产生的显著影响。综合全样本与子

样本可以发现更多隐含的长期波动信息。

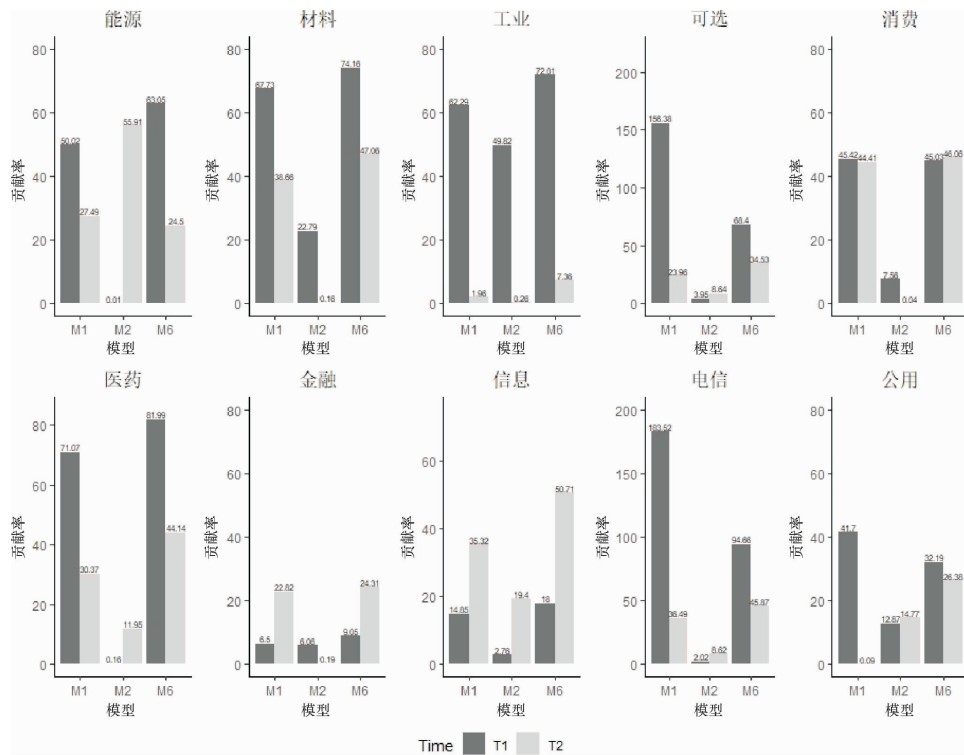


图 5 不同经济阶段贡献率(比率 1)

Fig. 5 Contribution of different economic stages (ratio 1)

从波动分解角度来看,图 1 中的各行业收盘价走势图明显在 2015 年 6 月前的趋势更加平缓,而在之后走势具有十分明显的下降后上涨趋势,变化幅度极大。综上子样本的结果,给出一个合理的结论:长期波动在趋势平缓或非平缓阶段都能发挥其作用。全样本更突出投资者情绪作为长期波动的特性,子样本更突出不同经济阶段下各行业的表现情况。

### 3 结 论

本文引入投资者情绪构建 GARCH-MIDAS-adapt 模型,探讨投资者情绪对股市波动的影响。实证结果表明:①投资者情绪作为股市波动里的长期波动成分对大部分行业具有研究意义,其中医药行业受投资者情绪影响最大;②相比于 GARCH(1,1) 和 GARCH-MIDAS 模

型,加入混频信息和自适应权重对各行业股市波动的拟合和预测都有不同程度的提升;③在不同情绪状态下,高涨期对医药行业影响更大,低落期对材料行业影响更大。从拟合效果来看,低落期相较于高涨期对股市波动作用更大;④长期波动不区分股价市场平缓与非平缓阶段,不同经济阶段同样存在影响;⑤投资者情绪与投资者情绪增长率之间在个别行业展现出不同向作用,对投资者情绪的研究给出了多一个因素的参考。

实证结论表明,GARCH-MIDAS-adapt 模型解决了股票收益与投资者情绪之间的不同数据频率问题,自适应权重的方法也提供了更准确的信息变化。在不同行业、不同情绪状态以及不同经济阶段,股市波动反映不同。因此,综合多方面情况考虑,可以更有效全面了解到各行业股市波动。作为影响股市波动因素中的一员,投资者情绪充当着不可或缺的角色。

### 参考文献:

- [1] Bollerslev T. Generalized autoregressive conditional heteroskedasticity[J]. Journal of E-conometrics, 1986, 31(3):307-327.

- [2] Beaumont B, Daele M V, Frijns B, et al. Investor sentiment, mutual fund flows and its impact on returns and volatility[J]. *Managerial Finance*, 2008, 34(11):772-785.
- [3] 文凤华,肖金利,黄创霞,等. 投资者情绪特征对股票价格行为的影响研究[J]. *管理科学学报*, 2014, 17(3):60-69.
- [4] Backer M, Wurgler J. Investor sentiment and the cross-section of stock returns[J]. *The Journal of Finance*, 2006, 61(4):1645-1680.
- [5] 易志高,茅宁. 中国股市投资者情绪测量研究: CICSI 的构建[J]. *金融研究*, 2009(11):174-184.
- [6] Ghysels E, Santaclara P, Valkanov R. The MIDAS touch: Mixed data sampling regression models[J]. *Cirano Working Papers*, 2004, 5(1):512-517.
- [7] Engle R F, Lee G. A permanent and transitory component model of stock return volatility[M]//Engle R F, White H. *Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive W. J. Granger*. London: Oxford University Press, 1999: 475-497.
- [8] Engle R F, Jose G R. The Spline-GARCH model for low-frequency volatility and its global macroeconomic causes[J]. *Review of Financial Studies*, 2008, 21:1187-1222.
- [9] Engle R F, Ghysels E, Sohn B. Stock market volatility and macroeconomic fundamentals[J]. *Review of Economics and Statistics*, 2013, 95:776-797.
- [10] Ghysels E, Santaclara P, Valkanov R. There is a risk-return trade-off after all[J]. *Journal of Financial Economics*, 2004, 76(3):509-548.
- [11] 于扬,王维国. 混频数据回归模型的分析技术及其应用[J]. *统计与信息论坛*, 2015, 30(12):22-30.
- [12] 王江涛,黄立玮,崔翔宇. 基于自适应权重的混频 GARCH 模型及其应用[J]. *统计研究*, 2021, 38(5):97-108.
- [13] Conrad C, Custovic A, Ghysels E. Long-and short-term cryptocurrency volatility C-components: A GARCH-MIDAS analysis[J]. *Journal of Risk and Financial Management*, 2018, 11(2):23. doi:10.3390/jrfm11020023.
- [14] 石强,杨一文,刘雅凯. 基于 GARCH-MIDAS 模型的宏观经济与股市波动关系[J]. *计算机工程与应用*, 2019, 55(15):257-262,270.
- [15] 姚尧之,王坚强,刘志峰. 混频投资者情绪与股票价格行为[J]. *管理科学报*, 2018, 21(2):104-113.
- [16] 郑挺国,尚玉皇. 基于宏观基本面的股市波动度量与预测[J]. *世界经济*, 2014, 37(12):118-139.
- [17] Hansen P R. A test for superior predictive ability[J]. *Journal of Business and Economic Statistics*, 2005, 23(4):365-380.

【责任编辑:周 全】