

基于 GARCH-MIDAS 模型的中国股票市场崩盘风险预测

陈倩¹, 王锦烽², 高翔³

¹深圳技术大学商学院 广东深圳

²易方达基金管理有限公司 广东广州

³上海商学院财务金融学院 上海

【摘要】将广义自回归条件异方差模型 (generalized autoregressive conditional heteroskedasticity model, 简称 GARCH) 与混频数据抽样模型 (mixed data sampling, 简称 MIDAS) 相结合构建单因子、三因子 GARCH-MIDAS 模型来预测波动率并计算波动率信息比率, 通过 Logit 模型和 1996 至 2019 年中国股票市场数据, 验证了波动率信息比率指标与中国股票市场崩盘风险之间的关系。结果表明: 中国股票市场日度波动率服从 GARCH(1,1) 过程且与宏观经济变量水平值及波动率显著相关, 波动率信息因子与国内股票市场崩盘风险显著负相关且较为稳健, 三因子 GARCH-MIDAS 模型拟合能力优于单因子模型, 包含波动率信息因子的 Logit 模型能有效预测国内股市崩盘风险。

【关键词】股市崩盘风险; 过度自信; 波动率信息比率; GARCH-MIDAS 模型

【基金项目】本研究受国家自然科学基金“操作风险动态量化方法研究: 从微观机构到宏观系统 (71501117)”资助

Predicting Crash Risk in the Chinese Stock Market Based on GARCH-MIDAS Modeling

Qian Chen¹, Jinfeng Wang², Xiang Gao³

¹School of Business, Shenzhen Technology University, Guangdong

²E Fund Management Co., Ltd., Guangdong

³Faculty of Professional Finance and Accountancy, Shanghai Business School, Shanghai

【Abstract】This paper utilizes the single-factor and three-factor GARCH-MIDAS model to predict Chinese market volatility and calculate the volatility information ratio, which purports to measure investor overconfidence and irrational behaviors. The ratio's predictivity of market crash is investigated with a Logit specification with Chinese stock market index data from 1996-2019. The empirical results indicate that the daily volatility of Chinese stock market follows a GARCH(1,1) process driven by the level and fluctuation of macroeconomic indicators. There exists a significantly negative and robust relationship between our information ratio and market crash risk in China. Moreover, the three-factor GARCH-MIDAS model outperforms the single-factor one in forecasting the crash risk in China's stock market when our proposed volatility information ratio is incorporated.

【Keywords】Market Crash Risk; Overconfidence; Volatility Information Ratio; GARCH-MIDAS

1 引言

当今世界各个经济体间的经济金融往来愈加密切, 每个国家在享受着全球化网络带来红利的同时, 金融风险传导速度也在不断加快。作为全球第二大经济体, 中国的股票市场也与全球资本市场息息相关。2008 年爆发的全球金融危机席卷了包括中国在

内的各国股票市场, 许多市场中的股票价格不仅发生了剧烈波动, 而且波动持续时间延续至今。2015 年, 中国股票市场就经历了一次严重的股价急速下跌, 在 2015 年 6 月至 7 月的短短一个月, 中国股市价格下跌超过了 1000 点, 累计跌幅达 24%, 并且出现了罕见的“千股跌停”现象。面对这样的崩盘风

险, 尽管我国政府迅速开展一系列救市措施, 但是长达一个月的极速下跌, 已经严重影响了国内金融体系的稳定和很多企业的正常运营¹ (叶康涛等, 2019)^[1]。由此可见, 股票市场价格急速下跌引发的崩盘风险有着巨大破坏力, 如果我们能找到合理的指标, 准确预测股票市场的崩盘风险, 那么这无疑不仅有益于市场投资者的风险管理, 也对我国股票市场的良性发展及金融监管体制的完善有积极意义。

基于上述思考, 在已有研究的基础上, 本文提出波动率信息比率作为预测股票市场崩盘风险的指标。波动率信息比率具体为 GARCH-MIDAS 模型拟合的股市波动率估计值与经典 GARCH(1,1)模型拟合的股市波动率估计值二者之比, 本文通过 GARCH-MIDAS 模型将股市波动率分解为长期和短期两部分分别建模, 通过单因子 GARCH-MIDAS 模型探究验证了国内宏观经济变量水平值及波动率对股票市场波动率的显著影响关系, 同时运用主成分分析从一组宏观经济变量中提取第一主成分, 构建三因子 GARCH-MIDAS 模型, 最终运用 Logit 模型验证了波动率信息比率指标与中国股票市场崩盘风险有显著的负向关系, 同时基于 ROC 曲线和 AUC 值对模型的预测能力进行了检验。

本文的边际贡献有以下三点: ①探究验证了国内宏观经济变量水平值与波动率对股票市场波动率之间的显著关系; ②与已有文献直接加入一组宏观经济变量不同, 本文运用主成分分析提取出一组宏观经济变量的主成分作为宏观经济环境的代理变量; ③创新性地提出国内股市崩盘现象预测指标——波动率信息比率, 并证明了波动率信息比率与国内股票市场崩盘概率的负相关性。

2 文献综述

2.1 股票市场崩盘现象

本文中的股票市场崩盘风险主要是由于股票价格急速下跌而引发的, 它具体是指无任何信息前兆的情况下, 市场指数或者个股价格突然大幅度下降。崩盘现象存在着三大特点: 无信息支撑、收益不对称性和传染性。这三个特点都与股票市场中完

全理性预期均衡、不完全信息理性预期均衡以及行为金融有关。因此, 下文从这三方面展开综述:

首先, 在完全理性预期均衡框架下, 市场往往是有效的, 而股市波动率与股市收益率之间传导路径则各有不同。基于“有效市场假说”下, 股市波动率向股市收益率传导的方向, Pindyck (1983)^[2]提出波动率反馈假说, 即当消息到达股票市场时, 股市波动率上升, 而投资者的风险厌恶要求为上升的波动率提供风险溢价, 从而当消息正面时, 正面消息效应会被风险溢价效应中和, 而当消息负面时, 风险溢价效应将加剧负面消息影响, 使股票市场收益率进一步下降, 从而市场崩盘风险出现频率往往比市场暴涨次数多。

完全信息理性预期的假设在现实中并不理想, 部分学者对股市崩盘研究基于不完全信息理性预期均衡框架提出了不同意见, 认为股市崩盘是由于种种因素而未能参加场内交易的知情者所持有的负面私人信息的集中释放。Romer (1992)^[3]首次提出股市崩盘并不是投资者的非理性行为结果, 也非由负面的外部消息释放所诱发, 而是一个因股票市场的渐进交易使得被隐藏的内部消息被揭示的过程, 从而合理解释了无重大外部消息发生的股市崩盘, 并且认为在大多数投资者只能基于当期价格和信息进行投资决策、小部分人占据信息比较优势的市场结构下, 一旦出现外部的价格供给冲击, 投资者不能判定市场指数价格变动是出于宏观基本面因素抑或流动性对冲因素, 投资者持有的信息差异使得股票市场流动性减少, 微弱的供给冲击导致市场指数价格的急剧下降^[3]。Grech and Mazur (1994)^[4]基于美国道琼斯指数以及将局部赫斯特指数用于衡量信息效率, 研究信息效率与股市崩盘之间关联, 发现股票市场的信息效率越低, 市场崩盘发生的可能性越大。

基于行为金融学框架, 研究崩盘风险的两个角度为投资者情绪、异质信念。Hong and Stein (2003)^[5]认为投资者的过度自信会导致投资者异质信念, 并从投资者存在异质信念和市场存在卖空限制这两个角度, 不仅成功解释崩盘的突然爆发性, 还解释了为什么泡沫和崩盘的发生是不对称的, 以及崩盘为什么可以传染。Marin and Oliver (2008)^[6]运用固定效应条件 Logit 模型对美国股票市场的研究发现, 股市交易量与股市崩盘概率成正比, 异质信念程度越

¹ 参考已有文献, 本文主要使用股票市场收益率作为衡量崩盘风险的主要指标: 当股票市场收益率低于历史均值减两个标准差时, 市场当天就发生了股价崩盘风险事件。

大, 股市崩盘可能性越大。近年的研究更多关注行为金融学框架下信息效率与股票崩盘风险之间的关系, 例如, Hutton et al. (2009)^[7] 对公司层面财务信息不透明与崩盘风险的研究; Callen and Fang (2013)^[8] 以及 An and Zhang (2013)^[9] 对机构投资者监督与崩盘风险的研究。

相比于国外, 国内对于股市崩盘现象的研究有待进一步挖掘。潘越等 (2011)^[10] 验证了中国 A 股上市公司财务不透明度与崩盘风险的正向关系。张晓宇和徐龙炳 (2017)^[11] 以限售股解禁为重要事件节点, 研究了在解禁前后大股东的资本运作方式与股价崩盘之间的关系, 他们认为解禁后大股东采取的掏空行为会引发股价崩盘。叶康涛等 (2018)^[1] 对利用股指成分股调整作为一项准自然实验, 考察了股价崩盘风险与成分股调整之间的关系, 他们发现了分析师预测偏差是中介效应这一现象。赵汝为等 (2019)^[12] 则分析了投资者情绪对股价的影响, 他们的结果表明投资者情绪仅对大盘股股价产生明显作用。

2.2 GARCH-MIDAS 模型近年发展和应用

Engle et al. (2013)^[13] 在 Spline-GARCH 模型和波动率长短期分解理论的基础上, 将波动率的长期成分纳入到 MIDAS 框架中, 创造性提出了 GARCH-MIDAS 模型, 其研究发现宏观经济变量确实对股市波动率有着显著影响, 且纳入宏观变量的 GARCH-MIDAS 模型短期及长期的预测效果均优于纳入已实现波动率 GARCH-MIDAS 模型。郑挺国等 (2014)^[14] 将宏观景气指数作为宏观经济代理变量, 在 GARCH-MIDAS 模型中同时纳入宏观景气指数和已实现波动率以构建双因子 GARCH-MIDAS 模型, 结果显示相比单因子模型, 双因子 GARCH-MIDAS 模型预测能力显著改善, 同时发现宏观经济波动和股市波动呈显著相关关系。夏婷和闻岳春 (2018)^[15] 通过 GARCH-MIDAS 模型发现经济不确定性会影响中国股市的波动, 但强度有限, 且对 A 股、B 股影响程度各异。

直接在模型中使用多因子, 随着变量个数增加, 模型的简约性降低, 可能会影响模型的参数估计。近年来, 部分学者运用主成分分析以提取宏观经济变量的关键信息, 并将其纳入到 GARCH-MIDAS 模型中解释长期波动成分。Asgharian et al. (2013)^[16]

与传统 GARCH-MIDAS 模型直接纳入宏观经济变量不同, 其对挑选的一组宏观经济变量进行主成分分析, 并提取出第一主成分以代表宏观经济周期。之后将已实现波动率、第一主成分及其波动率同时将纳入 GARCH-MIDAS 模型中, 结果发现该模型的预测效果优于其他模型。Fang et al. (2018)^[17] 运用 GARCH-MIDAS 模型预测美国黄金期货波动率, 运用主成分分析来提取不同宏观经济变量的关键信息, 结果发现第一和第二主成分均能有效作为宏观经济的代理变量, 宏观经济变量对美国黄金期货波动率影响显著且二者关系在金融危机及之后的更加明显。

2.3 宏观经济因素对股票市场波动率的影响

Errunza et al. (1998)^[18] 运用 VAR 模型研究不同宏观经济变量对美国及欧洲部分国家股票市场收益率波动率的影响, 结果发现不同国家对股市波动率有显著影响的宏观经济变量各异, 且没有宏观经济变量对所有国家的股市波动率影响都显著。Bomfim (2003)^[19] 发现当货币政策未被预期到时, 股票市场波动率显著上升。Engle and Rangel (2008)^[20] 发现通货膨胀预期值对股市长期波动率产生显著作用。Mishkin (2009)^[21] 发现与平常的市场状态相比, 市场处于金融危机时期各个国家的宏观调控政策对股票市场波动率的影响程度更加显著^[21]。对此, Gregoriou et al. (2009)^[22] 认为, 当处于类似于金融危机等市场重大事件时, 宏观经济政策对股票市场波动的作用方向会产生偏移, 使得股票市场波动更剧烈。

在中国不同的政治经济背景之下, 国内的宏观经济变量与股市波动率研究结论与发达国家成熟的股票市场得出的研究结论有所差别, 因此一些学者的研究仍具有很高的借鉴意义。早期的研究如许均华等 (2001)^[23] 发现短期性的宏观经济政策会加大股市的短期波动, 且大于长期连续性宏观经济政策对波动率的影响, 但这种影响随时间衰减。董彩丽等 (2014)^[24] 探索了 GDP 增长率、工业增加值增长率、进出口、货币政策、通货膨胀及经济不确定性等宏观经济变量的水平值对股市波动率的影响。陈守东等 (2014)^[25] 则进一步探究宏观经济变量的波动率对股市波动率的影响并发现了显著的证据。

综上所述, 与国外资本市场相比, 崩盘在国内股市发生的频率更高, 虽然已有文献探讨了影响股

市崩盘的影响因素, 但是如何更好地从定量的角度去预测崩盘风险还有待进一步研究。波动率长期被当作投资者情绪以及股市风险的指标, 从波动率的角度深入研究股市崩盘现象及背后的形成机制我国股票市场良性发展尤为迫切。

3 理论模型和方法

3.1 MIDAS 模型

MIDAS 模型由 Ghysels et al. (2004)^[26] 提出, 本质上是将滞后多项式加入回归模型, 并函数化各频率数据的权重值。MIDAS 模型在近年快速发展, 但基本单变量 MIDAS 模型框架如式:

$$Y_{t+h}^L = \beta_0 + \beta_1 \sum_{j=0}^{N-1} \psi_{N-j}(\omega) \times X_{t-j}^H + \varepsilon_{t+h}^L \quad \#(1)$$

其中 Y_{t+h}^L 为低频因变量, X_{t-j}^H 为高频自变量, $\psi_{N-j}(\omega)$ 为权重函数, 权重函数的引入使得 MIDAS 模型估计参数数量大大减少, 避免了“维度诅咒”问题。MIDAS 方法中最关键部分为权重函数, 权重函数直接决定了滞后各期的权重。常见的权重函数有两种形式, 其一为指数函数形式:

$$\psi_j(\omega) = \frac{\omega^j}{\sum_{j=1}^K \omega^j} \quad \#(2)$$

另外一种为 Ghysels et al. (2007)^[27] 提出的 Beta 权重函数:

$$\psi_j(\omega_1, \omega_2) = \frac{\left(\frac{j}{K}\right)^{\omega_1-1} \left(1 - \frac{j}{K}\right)^{\omega_2-1}}{\sum_{j=1}^K \left(\frac{j}{K}\right)^{\omega_1-1} \left(1 - \frac{j}{K}\right)^{\omega_2-1}} \quad \#(3)$$

通过对 Beta 权重函数中参数 ω_1 、 ω_2 设定, 可以实现 Beta 权重函数不同的形状, 既可以呈现单调递增或单调递减, 也可显示 U 形分布或者驼峰分布, 相比于指数函数的衰减方向的单一性更加灵活。基于此, 本文在 MIDAS 部分主要采用 Beta 权重函数。

3.2 GARCH(1,1)模型

Engle (1982)^[28] 在研究英国通货膨胀率时间序列时发现不同时期内的变量的预测力度差别较大, 预测误差的方差在不同时期各异, 因此提出自回归

$$r_{i,t} - E_{i-1,t}(r_{i,t}) = (E_{i,t} - E_{i-1,t}) \sum_{j=0}^{\infty} \rho^j \Delta d_{i+j,t} - (E_{i,t} - E_{i-1,t}) \sum_{j=0}^{\infty} \rho^j r_{i+j,t} \quad \#(10)$$

其中, $r_{i,t}$ 是 t 月第 i 天对数收益率, $d_{i,t}$ 是同一天的对数股利, $E_{i,t}$ 是同时期下的条件期望值。Engle 和

条件异方差模型 (即 ARCH 模型)。ARCH 模型认为资产收益率序列的扰动项是非线性相关的, 这种非线性相关性可用滞后的简单二次项表示。ARCH 模型能较好处理金融时间序列异方差现象, 表现出了金融时间序列的长记忆性及波动聚集性, 但在实际应用中为了得到更好的拟合效果, 往往需要估计过多的参数, 参数估计准确性很难得到保证。随后 Bollerslev (1986)^[29] 将其拓展为广义自回归条件异方差模型 (即 GARCH 模型)。标准的 GARCH(1,1) 模型如下:

$$r_t = \mu + z_t \quad \#(4)$$

$$z_t = \sqrt{h_t} \times \varepsilon_t \quad \#(5)$$

$$\varepsilon_t | F_{t-1} \sim N(0,1) \quad \#(6)$$

$$h_t = \omega + \alpha z_{t-1}^2 + \beta h_{t-1} \quad \#(7)$$

$$\alpha > 0, \beta > 0, \alpha + \beta < 1 \quad \#(8)$$

其中, r_t 为资产收益率, F_{t-1} 为到 t-1 交易日所获得的全部信息量, z_{t-1}^2 为 ARCH 项, h_{t-1} 为 GARCH 项。GARCH 模型通过用低阶 GARCH 项替代高阶 ARCH 项, 解决了参数过多的问题, 进一步优化了模型识别和拟合过程。

之后学者对 GARCH 模型进行了诸多改进, 其中 Nelson (1991)^[30] 提出的 EGARCH 模型, 即对 GARCH 模型中的条件方差进行对数化^[30]:

$$\ln(h_t) = \omega + \alpha z_{t-1}^2 + \beta h_{t-1} \quad \#(9)$$

条件方差进行对数化保证了条件方差恒为正。基于该条件, 本文的参数估计均在残差正态分布假设下进行, 并采用近似极大似然估计方法得到参数的估计量。

3.3 GARCH-MIDAS 模型

GARCH-MIDAS 模型是基于 GARCH 模型中的 Spline-GARCH 模型及 MIDAS 模型而来的。Spline-GARCH 模型认为不同的信息事件对金融市场的影响是和作用时间长短有关的。传统的分析框架是基于 Campbell (1988)^[31] 提出的对数线性化及 Campbell and Shiller (1991)^[32] 提出的:

Rangel (2008)^[20] 认为收益率的冲击方差具有时变性, 且受到某一信息事件对收益率冲击的作用强度

和作用乘子的影响, 而该信息事件的作用乘子依赖于宏观经济环境。我们将未预期收益率的冲击表示为:

$$r_{i,t} - E_{i-1,t}(r_{i,t}) = \sqrt{\tau_1(z_t)} \times \eta_{i,t} \quad \#(11)$$

$\eta_{i,t}$ 是收益率受到信息事件冲击的作用强度, $\tau_1(z_t)$ 是收益率受到信息事件冲击的作用乘子。因为 $\eta_{i,t}$ 受宏观经济环境影响, $\tau_2(z_t)$ 为 $\eta_{i,t}$ 中受宏观经济因素等长期状态因子影响部分, $g_{i,t}$ 为非负时间序列。信息完全无遗漏时,

$$\tau(z_t) = \tau_1(z_t) \times \tau_2(z_t) \quad \#(12)$$

$$r_{i,t} - E_{i-1,t}(r_{i,t}) = \sqrt{\tau(z_t) \times g_{i,t}} \times \varepsilon_{i,t} \quad \#(13)$$

最终, 可以得出本文使用的 GARCH-MIDAS 模型:

$$r_{it} - \mu = \sqrt{\tau_t \times g_{it}} \times \varepsilon_{it} \quad \#(14)$$

其中 $g_{i,t}$ 为波动率短期成分, τ_t 为波动率长期成分。 $g_{i,t}$ 受日度的市场流动性等短期因素影响; τ_t 受未来的期望现金流和未来折现率影响, 而宏观经济因素通常包含未来的期望现金流和未来折现率的趋势信息。假定短期部分 g_{it} 服从 GARCH(1,1) 过程, 即:

$$g_{it} = (1 - \alpha - \beta) + \alpha \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta g_{i-1,t} \quad \#(15)$$

$$E_{t-1}(r_{it} - \mu)^2 = \tau_t E_{t-1}(g_{it}) = \tau_t \quad \#(16)$$

(1) 单因子 GARCH-MIDAS 模型。

在 MIDAS 模型中用一个低频变量刻画波动率长期成分, 即为单因子 GARCH-MIDAS 模型, 本文 MIDAS 部分均采用 Beta 权重函数形式。

定义已实现波动率如下:

$$\begin{aligned} \log(\tau_t) = & m + \theta_{RV} \sum_{k=1}^K \psi_k(\omega) RV_{t-k} + \theta_l \sum_{k=1}^{K_l} \psi_k(\omega) PC_{l,t-k}^{ml} \\ & + \theta_v \sum_{k=1}^{K_v} \psi_k(\omega) PC_{v,t-k}^{mv} + \varepsilon_t \quad \#(20) \end{aligned}$$

(3) 波动率信息比率。

经典 GARCH(1,1) 模型被投资者广泛应用于金融工程、风险管理等金融领域, 其拟合出的波动率代表投资者“看见的”当前波动率。其局限在于只包含股市的场内数据, 信息熵值较小, 未考虑宏观经济环境对股市波动率的根本性影响。而 GARCH-MIDAS 模型同时运用高频股票市场收益率数据及

$$RV_t = \sum_{i=1}^N r_{i,t}^2 \quad \#(17)$$

对于已实现波动率 RV_t 可以采用固定窗口和滚动窗口两种形式, 本文选取固定窗口形式。将已实现波动率 RV_t 带入公式, 得到 GARCH-MIDAS-RV,

$$\log(\tau_t) = m + \theta_{RV} \sum_{k=1}^K \psi_k(\omega) \times RV_{t-k} + \varepsilon_t \quad \#(18)$$

当用宏观变量因子水平值 $X_{l,t-k}^{ml}$ 或宏观变量因子波动率 $X_{v,t-k}^{mv}$ 代替已实现波动率 RV_{t-k} 时, 模型就变为纳入宏观经济变量的 GARCH-MIDAS-X 模型:

$$\log(\tau_t) = m_v + \theta_v \sum_{k=1}^{K_v} \psi_k(\omega) \times X_{v,t-k}^{mv} + \varepsilon_t \quad \#(19)$$

其中, $X_{l,t-k}^{ml}$ 表示某个宏观经济变量相对当期滞后 k 期的水平值, K_l 表示该变量水平值的最大滞后阶数。 $X_{v,t-k}^{mv}$ 表示某个宏观经济变量相对当期滞后 k 期的波动率, K_v 表示该变量波动率的最大滞后阶数。宏观经济变量波动率的度量, 常见的做法是使用 ARMA(p,q) 模型得到残差项, 然后进行平方运算得到, 本文亦通过该方法度量宏观经济变量的波动率进行 GARCH-MIDAS 建模。

(2) 多因子 GARCH-MIDAS 模型。

将已实现波动率 RV_t , 宏观变量因子水平值 $X_{l,t-k}^{ml}$, 宏观变量因子波动率 $X_{v,t-k}^{mv}$ 同时纳入模型中便可得到包含三因子 $RV + X_l + X_v$ 的 GARCH-MIDAS 模型。本文则基于 Asgharian et al. (2013)^[16] 的方法, 从一组宏观经济变量中提取出主成分作为中国宏观经济的代理变量。将已实现波动率 RV_t , 主成分水平值 PC_l , 主成分波动率 PC_v 同时纳入模型, 形成如下多因子 GARCH-MIDAS 模型:

低频宏观经济变量数据, 信息熵值更大, 且考虑了宏观经济环境对股市波动率的影响, 其拟合的股市波动率可被视为基于宏观经济环境得到的“事实”股市波动率。

本文对市场的非理性狂热定义为, “当前投资者能看见的股市波动率远高于基于宏观经济信息所预示的股市波动率”, 即此时 GARCH(1,1) 模型得到的

股市波动率估计值高于 GARCH-MIDAS 模型得到的股市波动率估计值。由此本文定义波动率信息比率为 GARCH-MIDAS 拟合的波动率与 GARCH(1,1) 拟合的波动率之比:

$$vrratio = \frac{\sigma_{GARCH-MIDAS}^2}{\sigma_{GARCH(1,1)}^2} \quad \#(21)$$

并考察其对股市崩盘风险的预测能力。

$$\begin{aligned} crash_t &= \beta_0 + \beta_1 \times vrratio_{t-1} + \gamma Z_{t-1} + \varepsilon_t \\ crash_t &= \begin{cases} 1 & \text{if } r_t \leq \mu - 2\sigma \\ 0 & \text{if } r_t > \mu - 2\sigma \end{cases} \quad \#(22) \end{aligned}$$

因此可以通过二分类 Logit 模型考察波动率信息比率对股市崩盘风险的预测能力。控制变量包括: 过度自信指标换手率 *turnover*、市场交易额 *volume*、Amihud (2002)^[33] 市场流动性指标 *amihud*、施东晖和孙培源 (2002)^[34] 研究中的市场羊群效应指数 *csad*、前一天的上证综指收益率。为了能有效探测和预测崩盘风险, 本文所有解释变量均选取滞后一天数据。最终模型表达式如下所示:

$$crash_t = \beta_0 + \beta_1 \times vrratio_{t-1} + \gamma Z_{t-1} + \varepsilon_t \quad \#(23)$$

4 数据描述与变量构建

4.1 股票市场指数收益率

本文选取上海证券交易所的上证综合指数 (简称“上证综指”) 作为中国股票市场指数的代理变量, 数据区间为 1996-01-02 至 2019-12-31, 取所有交易日的收盘价指数, 采用对数收益率。在选定的区间共有 5,821 个交易日上证综指对数收益率数据, 均值为 0.0002926, 标准差为 0.01656, 其中最大最小值均接近涨跌幅 10% 的限制, 明显左偏厚尾。为了进行后续的 GARCH(1,1) 及 GARCH-MIDAS 模型实证检验, 需要对上证综指对数收益率 ARCH 效应进行验证。

首先, 对上证综指对数收益率分布进行 Jarque-bera 正态性检验, 结果显示: 偏度显著小于 0, 峰度大于 3, 且 Jarque-bera test 得出的 *p*-value 远小于 0.001, 这说明本文研究区间内的上证综指收益率不服从正态分布; 其次, 对上证综指对数收益率进行平稳性检验, 通过刻画上证综指收益率及其平方的时序发现, 股市收益率明显表现出波动聚集现象, 主要表现在几次金融危机时期, 如 1997 年亚洲金融

(4) Logit 模型。

本文使用的是二分类 Logit 模型。假定股市崩盘发生概率为 π_i , 未发生概率即为 $1 - \pi_i$, 因此服从二项分布。本文将股市崩盘定义为股票市场日度收益率低于历史均值减两个标准差, 即市场当天出现了急速下跌的崩盘风险^{[1][6]}。具体形式如下:

危机、2008 年全球金融危机、2015 年国内股市泡沫; 最后, 在检验 ARCH 效应之前, 采用 ADF 单位根检验对股票市场收益率序列的平稳性进行检验, 从检验结果可知, 单位根检验结果显示在 1% 的显著性水平下拒绝不平稳原假设, 即股市对数收益率序列平稳, 通过 LM-test 的检验结果可知, ARCH 效应检验结果卡方统计量值为 636.92, 对应的 *p* 值几乎为 0, 在 1% 的显著性水平上拒绝原假设, 上证综指对数收益率序列存在 ARCH 效应, 可以进行 GARCH 模型的拟合。

4.2 其它宏观经济变量

国内外学者的研究已充分表明, 大多数宏观经济变量对股票市场波动率有显著影响。综合国内资本市场现状及过往的研究, 本文选取七个宏观经济变量: 通货膨胀率 *cpi*、工业增加值月同比增速 *industry*、货币供应量月同比增速 *m2*、社会消费零售总额月同比增速 *sales*、一年期央行定期存款利率 *depositrate*、经济政策不确定指数 *epu*、人民币美元汇率 *exchangerate*。宏观经济变量波动率的常见计算方法为, 对宏观经济变量时间序列进行 ARMA(*p*,*q*) 模型拟合, 以其残差序列平方值作为波动率的度量, 本文采用该方法计算宏观经济变量波动率:

$$X_t = \beta_0 + \sum_{i=1}^p \beta_i X_{t-i} + \sum_{j=1}^q \alpha_j \varepsilon_{t-j} \quad \#(24)$$

X_t 为宏观经济变量的水平值, X_{t-i} 为宏观经济变量的滞后项, ε_t 为残差项, *p*、*q* 分别为解释变量和残差项的最大滞后阶数。宏观经济变量波动率描述性统计如表 1 所示。

表 1 宏观经济变量波动率描述性统计

变量	<i>cpi</i>	<i>m2</i>	<i>industry</i>	<i>sales</i>	<i>deposirate</i>	<i>epu</i>	<i>exchangerate</i>
频率	月度	月度	月度	月度	月度	月度	月度
样本数	288	288	288	288	288	288	288
均值	0.00003	0.00008	0.0008	0.00023	0.00001	0.50026	0.00208
最大值	0.00061	0.00253	0.01732	0.00731	0.00032	9.63907	0.0604
最小值	0	0	0	0	0	0	0
标准差	0.00006	0.0002	0.00213	0.0007	0.00004	1.13038	0.0063
偏度	4.6637	7.9287	4.2397	6.2181	7.5723	4.3912	5.7368
峰度	31.7440	86.6072	20.3606	45.4625	58.0302	23.5272	39.8481

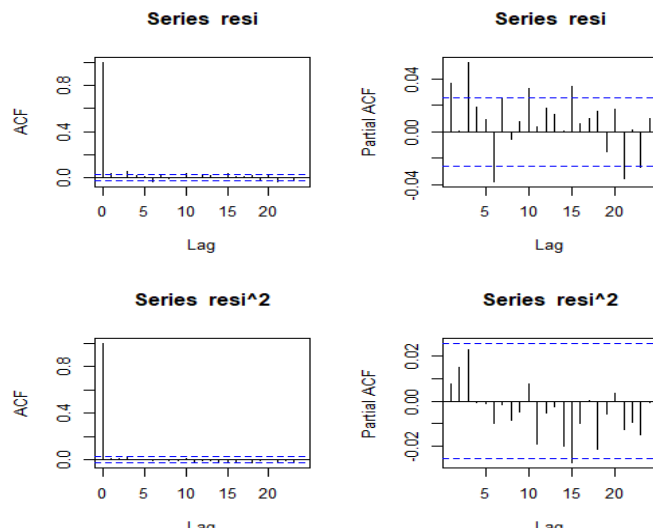


图 1 GARCH(1,1)标准化残差 ACF/PACF 函数值

5 实证结果与分析

5.1 GARCH(1,1)模型结果分析

从表 2 结果可知,除了 μ 之外所有系数在 1% 的显著水平下显著地异于零,说明上证指数收益率过去时刻的波动大小对当前波动大小有明显的影响,具有波动聚集效应。此外 $\alpha + \beta$ 小于 1 满足约束条件,约等于 1 则表明上证指数收益率条件方差序列具有长记忆性,收益率波动性具有较高的持续性,收益率时间序列总体风险较大。

为了验证 GARCH(1,1)模型拟合后的残差序列不存在波动聚集效应及自相关性,通过画出 GARCH(1,1)标准化残差的时间序列可知,残差序列没有明显的波动聚集效应。图 1 刻画了 GARCH(1,1)标准化残差 ACF、PACF,以观察残差序列的自相关性,从图中可知,残差序列的 ACF 及 PACF 函数值在置信区间内,这表明残差序列无自相关性或者相

关性很弱。残差平方序列的 ACF 和 PACF 函数值无截尾或拖尾现象,大多数函数值位于置信区间内,残差平方序列无自相关性。进一步对 GARCH(1,1)标准化残差自相关性进行 Ljung-Box 自相关检验,结果显示,在 5% 的显著水平下,滞后项为 10、15、20 时得到的 P 值分别为 0.7962、0.3987、0.4701,不能拒绝原假设,即 GARCH(1,1)标准化残差平方不存在序列相关性。

表 2 GARCH(1,1)模型回归结果

参数名	系数值	标准误	t 值	Pr(> t)
μ	2.265e-04	1.550e-04	1.462	0.144
α	0.07848	3.392e-07	12.461	< 2e-16
β	0.920	6.298e-03	155.492	< 2e-16
ω	1.603e-06	5.916e-03	4.725	2.31e-06

5.2 GARCH-MIDAS 模型结果与分析

(1) 基准模型 GARCH-MIDAS-RV 估计结果。

本文的基准模型和 Engle 和 Rangel (2008) 类似, 选取股票市场的低频已实现波动率 (RV) 做单因子 GARCH-MIDAS 回归, MIDAS 部分如下式:

$$\log(\tau_t) = m + \theta \sum_{k=1}^K \varphi_k(\omega) \times RV_{t-k} + \varepsilon_t \quad \#(25)$$

所使用的权重函数 $\varphi_k(\omega)$ 是 Beta 单 ω 权重函数, 最大滞后期数根据 AIC、BIC 准则及经济规律选择 12, 之后的 GARCH-MIDAS 模型权重函数及最大滞后期均与基准模型保持一致。表 3 报告了模型的结果, 参数 α 和 β 显著为正, 且二者之和小于 1, 表明上证综指日度对数收益率波动率高度集中, 股市波动率的短期成分服从 GARCH(1,1) 过程。已实现波动率 RV 对股市波动率长期成分影响参数 θ 显著为正, 说明了已实现波动率对股市波动率长期成分有显著的正向作用, 而权重函数参数 ω 不显著, 表明滞后各期已实现波动率对股票市场波动率的影响程度差异不大。

图 2 则刻画了股票市场的年度化的条件方差及波动率的低频长期成分。从图中可以看到, 波动率的低频部分过滤了短期日内噪音信息的影响, 相比于条件方差序列更加平缓。中国股票市场至今几次的暴跌崩盘风险都在图中体现: 1997 年亚洲金融危机, 2008 年全球金融危机及 2015 年国内股市极速下跌。

(2) 包含宏观经济变量水平值及波动率的 GARCH-MIDAS-X 模型估计结果。

将宏观经济变量水平值替代已实现波动率 RV_t 代入模型, MIDAS 部分将变为如下形式:

$$\log(\tau_t) = m + \theta \sum_{k=1}^K \varphi_k(\omega) \times X_{l,t-k} + \varepsilon_t \quad \#(26)$$

$$\log(\tau_t) = m + \theta \sum_{k=1}^K \varphi_k(\omega) \times X_{v,t-k} + \varepsilon_t \quad \#(27)$$

宏观经济变量水平值和波动率 GARCH-MIDAS-X 模型估计结果如表 4 和表 5 所示。首先, 对宏观经济变量水平值模型分析, 可以看到对于所有宏观经济变量水平值的 GARCH 参数 α 和 β 均显著为正。除了通货膨胀率 cpi 水平值及汇率 $exchangerate$ 水平值之外, 各宏观经济变量水平值对股市波动率长期成分影响系数 θ 均显著。通货膨胀率 cpi 水平值系数 θ 的不显著性与国内外的经济理论及

相关实证结果有些偏差, 究其原因可能是本文选取的数据时间区间较长, 而其中 90 年代国内通货膨胀率脱离实体经济异常偏高而影响了 cpi 的显著性。汇率系数 θ 的不显著性可能是由于我国实行的国外资本管制制度, 国外资本进入国内股票市场的门槛较高, 汇率水平对股市的影响不大。

大多数宏观经济变量系数 θ 显著但符号并不一致, 除了经济政策不确定性指数 $e pu$ 系数 θ 为负外, 其他宏观经济变量系数 θ 均显著为正, 与国内外实证结果一致。经济政策不确定性指数 $e pu$ 系数 θ 为负, 说明经济政策不确定性程度的上升会降低股市波动率, 可能原因在于当政策不确定性较高, 投资者避险情绪更浓而趋于保守。由于股票市场相比于其他市场的风险性更高, 股市对投资者吸引力下降而导致投资者涌向其他金融市场, 股市活跃度下降, 股市的波动率随之降低。

表 3 基准模型 GARCH-MIDAS-RV 估计结果

参数名	系数值	标准误	p 值
μ	0.02488	0.01774	0.1609
α	0.09009	0.07548	0.0000
β	0.8692	0.1433	0.0000
θ	0.6947	0.1810	0.0001
ω	2.7735	3.3184	0.4033
m	0.8030	0.1317	0.0000
AIC		19,392.22	
BIC		19,431.97	
MSE		1.0019e-04	
LLH		-9,690.11	

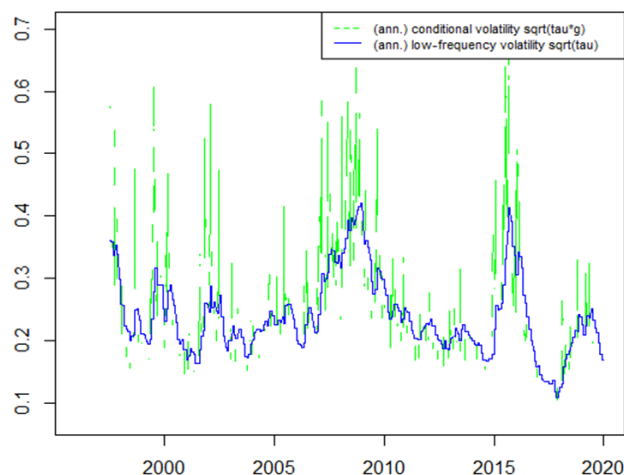


图 2 GARCH-MIDAS-RV 条件方差和低频成分

权重函数参数 ω 决定了滞后各期宏观经济信息的权重,各宏观经济变量参数 ω 均不显著,表明滞后各期宏观经济信息对股票市场波动率的影响程度差异不大。从 AIC、BIC、LLH 值来看,各宏观经济变量水平值模型中 *m2* 水平值模型拟合程度最好;从 MSE 值看, *industry* 水平值模型拟合程度最好。

其次,对宏观经济变量波动率模型分析,可以看到与宏观经济变量水平值结果一致,宏观经济变量波动率的参数 α 和 β 均显著为正。系数 θ 的显著性和符号正负情况则与宏观经济变量水平值结果有所差异。除了工业增加值增长率 *industry* 波动率、一年期央行定期存款利率 *deposirate* 波动率及汇率 *exchangerate* 波动率之外,宏观经济变量波动率对股市波动率长期成分影响系数 θ 均显著为负,说明宏观经济变量波动率的升高会降低股市波动率,究其缘由可能是因为宏观经济环境剧烈波动使得投资者对宏观基本面信心下降,风险偏好降低,减少对股票市场的资产配置。大部分参数 ω 显著且大于 1,预示着滞后的各期宏观经济变量的波动率对股市波动率的作用程度随滞后阶数增加而递减。对比各宏观经济变量波动率模型的 AIC、BIC、LLH 值可知,经济政策不确定性指数波动率模型拟合程度最好,对应的 AIC、BIC、LLH 值最小。

最后,上述对宏观经济变量水平值和波动率 GARCH-MIDAS-X 模型的分析可知,除了汇率之外,本文所选取的大部分宏观经济变量水平值和波动率对股市波动率的长期成分均有显著的影响,说明了国内宏观经济环境对股市波动率有显著的影响。

(3) 主成分分析及 GARCH-MIDAS-RV+PC+Pcvol 模型估计结果。

由于 GARCH-MIDAS 的计算复杂性,当在模型中加入过多宏观变量时会产生过度识别及结果非收敛等问题,而主成分分析能从一组宏观经济变量信息中提取关键信息。本文将对以上七个宏观经济变量进行主成分分析,并将提取出的主成分运用于 GARCH-MIDAS 模型。

首先,对各宏观经济变量之间的相关性进行分析,结果如表 6 所示。可以看到部分宏观经济变量之间的相关系数较大,其中消费者价格指数 *cpi* 与社会零售消费总额增速 *sales* 之间的相关系数达 0.715,究其原因可能是二者所覆盖的范围较为重合,强相关性在情理之中。经济政策不确定性指数和其他的宏观经济变量均呈负相关性,原因在于当经济形势较好时,政府采取的经济政策不确定性较小且在投资者预期内,经济平稳有序增长,即经济政策不确定性与宏观经济环境背道而驰。

表 4 包含宏观经济变量水平值的 GARCH-MIDAS-X 模型估计结果

	<i>cpi</i>	<i>m2</i>	<i>industry</i>	<i>sales</i>	<i>deposit_rate</i>	<i>epu</i>	<i>exchangerate</i>
μ	0.0214 (0.01694)	0.0230 (0.01672)	0.0220 (0.01657)	0.0227 (0.01696)	0.0221 (0.01682)	0.0232 (0.01658)	0.0216 (0.01675)
α	0.0756*** (2.590e-3)	0.0789*** (2.909e-3)	0.0774*** (1.262e-3)	0.0770*** (4.355e-3)	0.0763*** (3.300e-3)	0.0743*** (4.171e-3)	0.0764*** (3.305e-3)
β	0.920*** (5.192e-5)	0.912*** (8.881e-6)	0.917*** (2.126e-5)	0.916*** (2.195e-5)	0.916*** (1.457e-5)	0.921*** (5.871e-5)	0.920*** (7.419e-5)
θ	0.0494 (0.0120)	0.0771*** (0.0271)	0.0811** (0.0379)	0.1054** (0.0159)	0.2224** (0.0933)	-0.0020*** (6.948e-4)	0.00227 (0.00166)
ω	1.000 (1.000)	1.000 (0.361)	1.380 (0.922)	5.987 (3.830)	26.172 (29.274)	17.769 (1.180)	1.000 (1.703)
<i>m</i>	0.01387** (0.00634)	-0.00091 (0.00430)	-0.00502 (0.00356)	-0.00130 (0.00673)	0.00468** (0.00399)	0.0177 (0.00752)	-0.00138 (0.0131)
AIC	19,412.86	19,396.65	19,401.66	19,399.42	19,400.51	19,395.80	19,409.52
BIC	19,452.61	19,436.40	19,441.42	19,439.17	19,440.26	19,435.55	19,449.27
MSE	1.0025e-4	1.0021e-4	9.9297e-5	1.0043e-4	1.0104e-4	9.9577e-5	9.9912e-5
LLH	-9,700.428	-9,692.324	-9,694.830	-9,693.708	-9,694.255	-9,691.897	-9,698.757

注: *, **, ***分别代表 10%, 5%, 1%的显著性水平;括号内为参数估计值标准误;LLH 代表最大似然函数值, AIC 代表赤池信息准则, BIC 代表贝叶斯信息准则, MSE 为均方误差。

表 5 包含宏观经济变量波动率的 GARCH-MIDAS-X 模型估计结果

	<i>cpi</i>	<i>m2</i>	<i>industry</i>	<i>sales</i>	<i>deposit_rate</i>	<i>epu</i>	<i>exchangerate</i>
μ	0.02145 (0.01691)	0.02142 (0.01690)	0.02130 (0.01693)	0.02141 (0.01687)	0.02146 (0.01691)	0.02174 (0.01645)	0.02176 (0.01687)
α	0.0771*** (2.248e-3)	0.0789*** (2.276e-3)	0.0762*** (2.004e-3)	0.0758*** (1.618e-3)	0.0771*** (2.216e-3)	0.0754*** (5.744e-6)	0.0781*** (2.690e-3)
β	0.920*** (9.284e-5)	0.920*** (9.342e-5)	0.921*** (8.614e-5)	0.921*** (7.703e-5)	0.919*** (9.159e-5)	0.919*** (1.304e-6)	0.919*** (9.767e-5)
θ	-0.01288** (0.00507)	-0.01287** (0.00631)	-0.3423 (0.0324)	-0.8669* (0.0159)	-0.01267 (0.0118)	-0.0049*** (4.255e-3)	-0.1039 (0.0255)
ω	2.561** (0.0100)	2.562** (0.0227)	5.806** (0.0244)	17.495*** (0.0943)	2.523* (0.0134)	1.000 (0.00614)	1.001 (0.00976)
m	0.01728** (0.00674)	0.01726** (0.00681)	0.01748*** (0.00624)	0.01756*** (0.00529)	0.01726*** (0.00666)	-0.0141*** (0.00172)	0.01731*** (0.00753)
AIC	19,413.16	19,413.16	19,411.88	19,407.22	19,413.16	19,391.46	19,411.79
BIC	19,452.91	19,452.91	19,451.63	19,446.97	19,452.92	19,431.21	19,451.54
MSE	9.9790e-5	9.9791e-5	9.9963e-5	9.9907e-5	9.9815e-5	9.9417e-5	9.9731e-5
LLH	-9,700.579	-9,700.579	-9,699.936	-9,697.609	-9,700.580	-9,689.728	-9,699.892

注: *, **, ***分别代表 10%, 5%, 1%的显著性水平; 括号内为参数估计值标准误; LLH 代表最大似然函数值, AIC 代表赤池信息准则, BIC 代表贝叶斯信息准则, MSE 为均方差。

表 6 宏观变量相关系数矩阵

	<i>cpi</i>	<i>m2</i>	<i>industry</i>	<i>sales</i>	<i>deposirate</i>	<i>epu</i>	<i>exchangerate</i>
<i>cpi</i>	1.000	0.198	0.233	0.715	0.488	0.045	-0.166
<i>m2</i>	0.198	1.000	0.512	0.496	0.529	-0.556	0.356
<i>industry</i>	0.233	0.512	1.000	0.351	0.135	-0.468	0.358
<i>sales</i>	0.715	0.496	0.351	1.000	0.402	-0.239	-0.215
<i>deposirate</i>	0.488	0.529	0.135	0.402	1.000	-0.329	0.295
<i>epu</i>	0.045	-0.556	-0.468	-0.239	-0.329	1.000	-0.383
<i>exchangerate</i>	-0.164	0.356	0.358	-0.215	0.295	-0.383	1.000

表 7 主成分与宏观经济变量相关系数及方差贡献比

	PC1	PC2		PC1	PC2
<i>cpi</i>	0.53	0.73	SS loadings	3.00	1.75
<i>m2</i>	0.84	-0.18	Proportion Var	0.43	0.25
<i>industry</i>	0.69	-0.24	Cumulative Var	0.43	0.68
<i>sales</i>	0.70	0.60	Proportion Explained	0.53	0.31
<i>deposirate</i>	0.70	0.12	Cumulative Proportion	0.53	0.84
<i>epu</i>	-0.66	0.46			
<i>exchangerate</i>	0.38	-0.73			

其次碎石图的结果显示, 由于前两个主成分特

征值大于 1, 可提取出两个主成分。从表 7 可知, 第一主成分 PC1 和各宏观经济变量的相关系数均较高, 大部分在 0.5 之上, 第二主成分 PC2 与部分宏观经济变量相关系数较高, 且与 *m2*、*deposirate* 的相关系数较低。同时第一主成分 PC1 方差贡献比约为第二主成分 PC2 方差贡献比的两倍, 因此本文选取第一主成分 PC1 作为国内宏观环境的代理变量。

选取选取第一主成分 PC1 作为国内宏观环境的代理变量之后, 通过 ARMA 模型拟合求出其波动率。将已实现波动率、第一主成分水平值及其波动率同时纳入 GARCH-MIDAS 模型之中, 得到最终三因子 GARCH-MIDAS-RV+PC+PCvol 模型。从表 8 结果

可以看到, 短期波动率成分的 GARCH 过程系数 μ 、 α 、 β 均显著, 短期成分 GARCH 效应显著。三个系数 θ 中, 已实现波动率系数 θ_{rv} 、主成分水平值系数 θ_{pc} 显著为正, 而主成分波动率系数 θ_{pcvol} 不显著, 这说明在长期波动率 MIDAS 部分中纳入宏观经济变量水平值及已实现波动率会降低宏观经济变量波动率对股市波动率的影响。与系数 θ 情况类似, 三个权重函数系数 ω 中, 已实现波动率系数 ω_{rv} 、主成分水平值系数 ω_{pc} 显著大于 1, 说明权重函数值随着滞后阶数增加逐渐递减, 而主成分波动率系数 ω_{pcvol} 却不显著。

表 8 GARCH-MIDAS-RV + PC + PCvol 结果

参数名	系数值	标准误	p 值
μ	0.02159*	0.01678	0.09906
α	0.0767***	0.01165	0.0000
β	0.9124***	0.01463	0.0000
θ_{rv}	0.0350***	0.01183	0.001544
θ_{pc}	0.3267***	0.1177	0.002747
θ_{pcvol}	-0.3268	0.001206	0.1052
ω_{rv}	3.3404**	1.8089	0.03242
ω_{pc}	4.6873**	2.4953	0.03018
ω_{pcvol}	4.9977	9.3571	0.2966
m	0.00922***	0.001262	0.0000
AIC		19398.08	
BIC		19417.95	
MSE		1.0183e-04	
LLH		-9696.039	

注: *, **, ***分别代表 10%, 5%, 1%的显著性水平; LLH 代表最大似然函数值, AIC 代表赤池信息准则, BIC 代表贝叶斯信息准则, MSE 为均方误差。

最后, 为了对比 GARCH-MIDAS-RV 与 GARCH-MIDAS-RV+PC+PCvol 模型的拟合能力, 本文一方面从 AIC、BIC、LLH、MSE 值对二者进行对比, 从表 3 和表 8 的对比可以看到, GARCH-MIDAS-RV 模型的赤池信息准则 AIC 更小, 而 GARCH-MIDAS-RV+PC+PCvol 模型的贝叶斯信息准则 BIC、最大似然函数值 LLH 及均方误差 MSE 更小, 总体来看二者拟合能力相差不大, GARCH-MIDAS-RV+PC+PCvol 模型略微更好。

另一方面, 使用 Diebold-Mariano 检验对比 GARCH-MIDAS-RV 模型与 GARCH-MIDAS-RV+

PC+PCvol 模型预测精度, Diebold-Mariano Test 的备择假设是基准模型的预测精度低于基于宏观变量主成分的因子模型。Diebold-Mariano 检验 DM 值为 -0.25948, p 值为 0.3976, 不能拒绝原假设, 即 GARCH-MIDAS-RV+PC+PCvol 模型预精度上没有显著优于 GARCH-MIDAS-RV。

5.3 Logit 模型估计结果及预测能力检验

(1) Logit 模型估计结果。

Logit 模型估计结果如表 9 所示, 可以看到, $vrratio_1$ 的系数 p 值远小于 0.01, 在 1%水平下显著为负, 与本文的主要假设一致。当 $\sigma_{GARCH(1,1)}^2$ 比 $\sigma_{GARCH-MIDAS}^2$ 越大, 发生崩盘的可能性越大, 即市场投资者所“看见的”的波动率与基于综合宏观经济环境及股票市场金融数据所得到的“真实”波动率比值越大时, 预示着当前股票市场波动存在着较大的非理性因素, 股票市场发生崩盘风险的可能性越高, 与本文的主要假设一致。除了市场流动性指标, 其他控制变量系数均显著。市场流动性指标 $amihud$ 不显著的原因可能为 $amihud$ 指标本是基于市场收益率和市场交易额数据而得到, 而本文中已将市场前一天的收益率和市场前一天交易额作为控制变量纳入到模型中导致流动性指标解释力度下降而系数不显著。

表 9 Logit 模型估计结果

变量名	系数值	标准误	z 值	p 值
截距项	5.726	2.810	2.038	0.041595
$vrratio_1$	-1.076	2.861	-3.760	0.000170
$turnover_1$	1.497	3.905	3.835	0.000126
$csadl_1$	3.344	8.656	3.864	0.000112
$amihud_1$	0.1144	0.2370	0.483	0.629320
$volume_1$	1.199e-05	5.547e-06	2.162	0.030641
$return_1$	-1.049	4.165	-2.519	0.011775

(2) Logit 模型预测能力检验。

为了检验模型的预测效果, 将样本分为训练集和测试集以测试模型的预测股市崩盘风险的能力。样本的分类方法有两种, 第一种是根据国内股市历史数据发现 15 年、18 年期间发生较多的股票价格极端下跌, 由于股市极端性下跌为小样本事件, 因此设定训练集样本时间区间为 1996 年初至 2014 年末, 测试集时间区间为 2015 年初至 2019 年末。第二种是更换测试集和训练集的选取方式, 采用 k 折交叉

验证进一步评估模型的预测性能, 由于本文中数据量较大, 取 k 为 5。

为了测试模型的预测能力, 本文采用不同状态下 ROC 曲线面积 AUC 值进行度量。ROC (Receiver Operating Characteristic curve) 接收者操作特征曲线横坐标是伪阳性率 (False Positive Rate), 纵坐标是真阳性率 (True Positive Rate)。AUC (Area under the Curve of ROC) 是 ROC 曲线下方的面积, 表示正例排在负例前面的概率, 是判断二分类预测模型优劣的标准。由于本文中样本分布不平衡, $crash$ 为 1 的样本占比较低, 而 AUC 的计算方法同时考虑了分类器对于正例和负例的分类能力, 在样本不平衡的情况下, 依然能够对分类器做出合理的评价。

精确度、准确率、召回率是基本且常见的衡量分类模型的指标, 但在二分类且正反例不平衡的情况下, 尤其是对小样本时, 精确度等指标基本没有价值。另外, 在分类模型中, 预测结果一般都是以概率的形式表现, 如果要计算准确率, 通常都会手动设置一个阈值来将对应的概率转化成类别, 这个阈值也就很大程度上影响了模型准确率的计算, AUC 能避免把预测概率转换为类别而很好描述模型整体性能的高低。因此, 本文将采用 AUC 值来评价

采用的 Logit 模型分类能力。

当训练集样本时间区间为 1996 年初至 2014 年末, 测试集时间区间为 2015 年初至 2019 年末时, 训练集模型结果如表 10 所示, 训练集 $vrratio_1$ 的系数为负且显著, 仍与假设一致。训练集的 AUC 值为 0.722, 大于 0.5, 表明训练集中模型的分类能力较好。

表 10 Logit 模型训练集估计结果(1996-2014)

	系数值	标准误	z 值	p 值
截距项	4.327	3.063	1.413	0.157774
$vrratio_1$	-9.269	3.123	-2.968	0.00300
$turnover_1$	1.259	4.306	2.923	0.00346
$csadl_1$	3.561	9.104	3.912	9.17e-05
$amihud_1$	0.0933	0.2519	0.370	0.71107
$volume_1$	1.883e-05	1.175e-05	1.603	0.10899
$return_1$	-8.473	4.993	-1.697	0.08970

当采用 5 折交叉验证变换训练集测试集时, 训练集 Logit 模型估计结果如表 11 所示, 5 折交叉 Logit 模型训练集 $vrratio_1$ 的系数为负且显著, 仍与假设一致。在表中, 5 次测试集 AUC 值及平均值为 0.77, 再一次印证模型的预测能力较好且相对稳定。

表 11 5 折交叉 Logit 模型训练集估计结果及 AUC 值

	系数值	标准误	z 值	p 值		
截距项	5.465	2.925	1.868	0.061737		
$vrratio_1$	-1.046	2.978	-3.514	0.000441		
$turnover_1$	1.575	4.007	3.930	8.49e-05		
$csadl_1$	3.274	8.902	3.677	0.000236		
$amihud_1$	0.09611	0.2498	0.385	0.700449		
$volume_1$	8.878e-06	6.034e-06	1.471	0.141198		
$return_1$	-1.035	4.369	-2.368	0.017861		
	1	2	3	4	5	平均值
AUC	0.7692	0.7330	0.7943	0.7432	0.8042	0.7688

5.4 Logit 模型稳健性检验

(1) 更换股票市场 $crash$ 变量定义。

上述实证结果中对股票市场崩盘风险定义为当股票市场收益率低于历史均值减两个标准差时, 市场当天发生崩盘风险。现将股票市场发生崩盘风险新定义如下: 当天以及之后连续四天 (一周五个交

易日) 股票市场收益率之和小于 -10%, 股票市场发生崩盘风险。由表 12 可知, $vrratio_1$ 的系数仍显著为负, 与本文假设及实证结果相符。

(2) 更换模型样本区间。

由于 2005 年中国股票市场股权分置改革实现了同股、同权、同价, 改革对中国股票市场影响深远。

本文剔除股权改革之前的股票市场数据, 选取 2006 年初至 2019 年末为样本区间, 进行稳健性检验。此时, Logistic 模型结果如表 13 所示, 可以看到, *vrratio_1* 的系数仍显著为负, 与本文假设及实证结果相符。

(3) 更换计量方法

将 Logit 模型更换为 Probit 模型进行稳健性检验, 结果如表 14 所示, 可以看到, *vrratio_1* 的系数仍显著为负, 与本文假设及实证结果相符。

表 12 稳健性测试 1 结果

	系数值	标准误	z 值	p 值
截距项	1.519	3.942	3.854	0.000116
<i>vrratio_1</i>	-2.162	4.064	-5.322	0.000000
<i>turnover_1</i>	1.294	5.785	2.238	0.025250
<i>csadl_1</i>	4.818	1.062	4.538	5.68e-06
<i>amihud_1</i>	1.495e-05	7.576e-06	1.973	0.048456
<i>volume_1</i>	-3.620	0.5151	-0.703	0.482234
<i>return_1</i>	-4.853	6.177	-0.786	0.432068

表 13 稳健性测试 2 结果

	系数值	标准误	z 值	p 值
截距项	1.261	4.077	3.092	0.00199
<i>vrratio_1</i>	-1.778	4.128	-4.308	1.65e-05
<i>turnover_1</i>	1.693	7.609	2.224	0.02612
<i>csadl_1</i>	5.479	1.808	3.031	0.00244
<i>amihud_1</i>	-7.403e-05	7.849e-06	-0.943	0.34556
<i>volume_1</i>	0.2610	0.2531	1.031	0.30255
<i>return_1</i>	-8.277	4.776	-1.733	0.08305

表 14 稳健性测试 3 结果

	系数值	标准误	z 值	p 值
截距项	5.086	1.820	2.795	0.00519
<i>vrratio_1</i>	-7.808	1.829	-4.268	1.97e-05
<i>turnover_1</i>	7.701	3.780	2.037	0.04160
<i>csadl_1</i>	2.998	9.170	3.270	0.00108
<i>amihud_1</i>	-3.892e-06	3.927	-0.991	0.32159
<i>volume_1</i>	0.1351	0.1405	0.962	0.33604
<i>return_1</i>	-3.577	2.355	-1.519	0.12874

6 结论

本文构建了波动率信息比率指标预测股票市场崩盘风险现象。具体步骤为在 GARCH-MIDAS 模型中检验了国内宏观经济变量信息对股票市场波动率影响的显著性, 并通过结合 GARCH(1,1) 及 GARCH-MIDAS 模型各自得出的波动率构建出国内股票市场崩盘预测指标波动率信息比率, 通过 Logit 回归验证了波动率信息比率与国内股票市场崩盘风险的显著负相关性, 并检验了 Logit 模型的预测能力, 本文具体的结论有如下几点:

第一, GARCH(1,1)模型能有效拟合上证综指日度对数收益率数据, 股票市场日度波动率服从 GARCH(1,1)过程;

第二, 中国股票市场波动率与中国宏观经济变量水平值及波动率显著相关, 单因子 GARCH-MIDAS-X 模型结果显示, 除了通货膨胀率和汇率之外, 其他五个宏观经济变量水平值对国内股市波动率均有显著影响; 波动率方面, 除了工业增加值增速波动率与汇率波动率外, 其他五个宏观经济变量波动率对国内股市波动率均有显著影响, 同时, 权重函数参数值大部分高于 1, 宏观经济变量对股市波动率的影响随着滞后阶数增加而下降, 符合现实中经济规律;

第三, 三因子 GARCH-MIDAS-RV+PC+PCvol 模型拟合能力优于 GARCH-MIDAS-RV 模型, 本文采用主成分分析从七个宏观变量中提取出主成分作为整体宏观经济环境的代理变量, 从 BIC、LLH、MSE 指标上看, 三因子 GARCH-MIDAS-RV+PC+PCvol 模型拟合能力优于 GARCH-MIDAS-RV 模型, 另外, 使用 Diebold-Mariano 检验对比 GARCH-MIDAS-RV 模型与 GARCH-MIDAS-RV+PC+PCvo 模型预测精度, 结果显示三因子 GARCH-MIDAS 模型的预测精度与基准模型 GARCH-MIDAS-RV 没有显著区别。

第四, 波动率信息因子与国内股票市场崩盘风险显著负相关, Logit 模型结果中波动率信息因子系数显著为负, 表明波动率信息因子值与国内股票市场崩盘概率呈负相关性, 另外, 稳健性检验结果中波动率信息因子系数也显著为负, 波动率信息因子与股票市场崩盘也呈现出显著的负相关性。

第五, 包含波动率信息因子的 Logit 模型能有效预测国内股市崩盘风险, 基于历史事实取 1996 年初

至 2014 年末期间数据为训练集, 2015 年至 2019 年期间数据为测试集, 同时采用 5 折交叉验证表明包含波动率信息因子的 Logit 模型的预测能力较高。

参考文献

- [1] 叶康涛, 刘芳, 李帆. 股指成份股调整与股价崩盘风险: 基于一项准自然实验的证据[J]. 金融研究, 2018(3): 172-189.
- [2] Pindyck, R S. Risk, Inflation, and the Stock Market[J]. The American Economic Review, 1983, Vol.74, No.3: 335-351.
- [3] Romer, D H. Rational Asset Price Movements Without News[J]. The American Economic Review, 1992, Vol.83, No.5: 1112-1130.
- [4] Grech, D, Zygmunt M. Can One Make Any Crash Prediction in Finance Using the Local Hurst Exponent Idea[J]. Physica A-statistical Mechanics and Its Applications, 2004, Vol.336, No.1: 133-145.
- [5] Hong H, Stein J C. Differences of Opinion, Short-Sales Constraints and Market Crashes[J]. Review of Financial Studies, 2003, Vol.16, No.2: 487-525.
- [6] Marin J M, Olivier J. The Dog That Did Not Bark: Insider Trading and Crashes. Journal of Finance, 2008, Vol.63, No.5: 2429-2476.
- [7] Hutton A P, Marcus A J, Tehranian H. Opaque Financial Reports, R2, and Crash Risk[J]. Journal of Financial Economics, 2009, Vol.94, No.1: 67-86.
- [8] Callen J L, Fang X H. Institutional Investor Stability and Crash Risk: Monitoring Versus Short-Termism?[J]. Journal of Banking and Finance, 2013, Vol.37, No.8: 3047-3063.
- [9] An H, Ting Z. Stock Price Synchronicity, Crash Risk, and Institutional Investors[J]. Journal of Corporate Finance, 2013, Vol.21, No.1: 1-15.
- [10] 潘越, 戴亦一, 林超群. 信息不透明, 分析师关注与个股暴跌风险[J]. 金融研究, 2011(9): 138-151.
- [11] 张晓宇, 徐龙炳. 限售股解禁、资本运作与股价崩盘风险[J]. 金融研究, 2017(11): 158-174.
- [12] 赵汝为, 熊熊, 沈德华. 投资者情绪与股价崩盘风险: 来自中国市场的经验证据[J]. 管理评论, 2019(3): 50-60.
- [13] Engle R F, Ghysels E, Sohn B. Stock Market Volatility and Macroeconomic Fundamentals[J]. The Review of Economics and Statistics, 2013, Vol.95, No.3: 776-797.
- [14] 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(12): 118-139.
- [15] 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.
- [16] Asgharian H, Ai J H, Farrukh J. The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH-MIDAS Approach[J]. Journal of Forecasting, 2013, Vol.32, No.7: 600-612.
- [17] Fang L B, Yu H H, Wen X. Forecasting Gold Futures Market Volatility Using Macroeconomic Variables In The United States[J]. Economic Modelling, 2018, Vol.72: 249-259.
- [18] Errunza V, Hogan K. Macroeconomic Determinants of European Stock Market Volatility[J]. European Financial Management, 1998, Vol.4, No.3: 361-377.
- [19] Bomfim A N. Pre-announcement Effects, News Effects, and Volatility: Monetary Policy and the Stock Market[J]. Journal of Banking & Finance, 2003, Vol.27, No.1: 133-151.
- [20] Engle R F, Rangel J G. The Spline-GARCH Model for Low Frequency Volatility and its Global Macroeconomic Causes[J]. Review of Financial Studies, 2008, Vol.21, No.3: 1187-1222.
- [21] Mishkin F S. Is Monetary Policy Effective During Financial Crises[J]. The American Economic Review, 2009, Vol.99, No.2: 573-577.
- [22] Gregoriou A, Kontonikas A, MacDonald R, Montagnoli A. Monetary Policy Shocks and Stock Returns: Evidence from the British Market. Financial Markets And Portfolio Management[J]. Swiss Society for Financial Market Research, 2009, Vol.23, No.4: 401-410.
- [23] 许均华, 李启亚. 宏观政策对我国股市影响的实证研究[J]. 经济研究, 2001(9): 12-21+95.
- [24] 董彩丽, 陈强, 郑旭. 基于混频模型的 CPI 短期预测研究[J]. 统计研究, 2014 (12): 25-31.
- [25] 陈守东, 易晓激, 刘洋. 政策不确定下中国股市与宏观经济相关性的非对称效应研究[J]. 当代财经, 2014(1): 45-55.
- [26] Ghysels E, Santa-Clara R, Valkanov R I. The MIDAS Touch: Mixed Data Sampling Regression Models[R]. CIRANO Working Papers, 2004, 2004s-20, CIRANO.

- [27] Ghysels E, Sinko A, Valkanov R I. MIDAS Regressions: Further Results and New Directions[J]. *Econometric Reviews*, 2007, Vol.26, No.1: 53-90.
- [28] Engle R F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation[J]. *Econometrica*, 1982, Vol.50, No.4: 987-1007.
- [29] Bollerslev T. Generalized Autoregressive Conditional Heteroskedasticity[J]. *Journal of Econometrics*, 1986, Vol.31, No.3: 307-327.
- [30] Nelson D B. Conditional Heteroscedasticity in Asset Returns: A New Approach. *Econometrica*, 1991, Vol.59: 347-70.
- [31] Campbell J Y. A Variance Decomposition for Stock Returns[J]. *The Economic Journal*, 1991, Vol.101, No.405: 157-179.
- [32] Campbell J Y, Shiller R J. The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors[J]. *Review of Financial Studies*, 1988, Vol.1, No.3: 195-228.
- [33] Amihud Y. Illiquidity and Stock Returns: Cross-Section and Time-Series Effects[J]. *Journal of Financial Markets*, 2002, Vol.5, No.1: 31-56.
- [34] 施东晖, 孙培源. 基于 CAPM 的中国股市羊群行为研究——兼与宋军、吴冲锋先生商榷[J]. *经济研究*, 2002(2): 64-70.

收稿日期: 2021 年 2 月 9 日

出刊日期: 2021 年 5 月 10 日

引用本文: 陈倩, 王锦烽, 高翔, 基于 GARCH-MIDAS 模型的中国股票市场崩盘风险预测[J]. *国际金融进展*, 2021, 3(1):1-15

DOI: 10.12208/j.aif.20210001

检索信息: RCCSE 权威核心学术期刊数据库、中国知网 (CNKI Scholar)、万方数据 (WANFANG DATA)、Google Scholar 等数据库收录期刊

版权声明: ©2021 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS