

A 股市场收益率波动的长记忆性

摘要

A 股市场收益率波动的长记忆性研究对于有效市场假说的检验具有重要的意义。本文采用 GPH 估计法对 A 股市场所有个股的日收益率平方进行探究,通过滚动估计得到面板的长记忆指标,并从横截面的角度研究了股价波动的长记忆指标与预期收益之间的关系。实证结果显示,样本内 98.92% 股票的收益率波动都具有显著的长记忆性。通过单变量分组实验,本文发现股价波动的长期记忆程度与其未来的超额收益之间存在显著的负相关关系。针对这一负相关关系本文提出了解释机制,即长记忆程度高的股票具有更高的波动可预测性,因此投资者所要求的风险补偿也更低。最后本文从行业门类的角度验证了长记忆性存在的稳健性。本文是少数关注长记忆性在横截面和资产定价方面的研究,也是该方向上少数引入行业门类对比的研究。

关键词: 长记忆性, 资产定价, GPH 估计法, 波动性

LONG MEMORY IN STOCK RETURN VOLATILITY

IN CHINESE STOCK MARKET

ABSTRACT

The study on the long memory of return volatility in the Chinese stock market is of significant importance for testing the Efficient Market Hypothesis. This paper explores the daily squared returns of all individual stocks in the Chinese stock market using the GPH estimation method. By employing rolling estimation, panel long memory indicators are obtained, and the relationship between stock price volatility's long memory indicators and expected returns is examined from a cross-sectional perspective. Empirical results reveal that 98.92% of the sampled stocks exhibit significant long memory in return volatility. Through univariate grouping experiments, it is found that there exists a significant negative correlation between the long-term memory of stock price volatility and its future excess returns. In response to this negative correlation, this paper proposes an explanatory mechanism, suggesting that stocks with higher long memory exhibit greater volatility predictability, thereby leading to lower risk compensation demanded by investors. Finally, the robustness of the existence of long memory is verified from the perspective of industry categories. This paper represents one of the few studies focusing on the long memory in cross-sectional and asset pricing aspects and is among the minority that introduces industry categories for comparison in this direction.

KEYWORDS: long memory, asset pricing, GPH estimator, volatility

目 录

一、引言	1
（一）研究背景	1
（二）研究意义	2
（三）研究方法 & 框架	2
（四）研究创新点	3
二、文献综述	4
（一）长记忆性	4
（二）长记忆指标的估计方法	5
（三）文献述评	5
三、样本与变量	7
（一）样本选择	7
（二）长记忆指标的估计	7
四、实证分析	9
（一）长记忆指标的描述性统计和显著性检验	9
1. 描述性统计	9
2. 假设检验	10
（二）股价波动长记忆性与预期收益	11
1. 单变量分组	11
2. 均值检验	11
3. CAPM 模型检验	12
（三）机理分析	13
1. 可预测性的度量	13
2. 结果分析	14
五、稳健性检验	16
六、结语	19
（一）研究结论	19
（二）研究不足	20
参考文献	21
致谢	23

一、引言

（一）研究背景

“有效市场假说”作为金融领域研究中的核心概念，对资产定价、风险管理等领域的理论和实证研究产生了重要影响。该理论认为在一个有效市场中，所有可获取的信息都已被充分地反映在资产的价格中。根据这一假设，投资者对信息的反应呈线性关系，资产收益率遵循随机漫步过程，并服从正态分布。因此，在有效市场中，不存在可以预测未来价格走势的方法，投资者无法利用市场上的任何信息来获得超额收益。

尽管许多金融理论以有效市场假说为基础，并在实践中得到了广泛应用，但对于有效市场假说是否成立一直存在争议。根据该理论，资产收益率应该服从正态分布，但实证研究中发现的“尖峰厚尾负偏”现象与此相悖。为了解决这一矛盾，Mandelbrot（1960）提出了股票收益应服从稳态帕雷托分布的观点。这种分布不仅具有尖峰厚尾的特性，还能够描述趋势性和周期性，以及突发性变化。行为金融学的奠基人 Shiller（1992）则从投资者行为的角度提出，市场不有效是由于噪音交易者的存在，这与投资者理性的假设相矛盾。噪音交易者对价格波动过度反应，导致市场波动加剧，从而产生价格的可预测趋势。类似地，Peters（1996）认为投资者对信息的反应是非线性的，可能存在滞后性，或者只有当信息量达到一定阈值时才会爆发性显现。因此，Peters（1996）指出资本市场不仅不满足马尔科夫性质，甚至存在长记忆性，即过去的信息可能反馈在当前甚至未来很长一段时间的市場表現中。

在理论研究中，长记忆性（long memory）指的是时间序列数据中存在的一种持久性依赖关系，表明过去的观测值对当前和未来的观测值产生长期的影响。相对应地，短期记忆性表示过去的观测值对当前和未来的观测值的影响会以指数形式迅速衰减。具体而言，长记忆性时间序列是一类特殊的平稳时间序列，其自相关函数（ACF）虽然也随滞后 $k \rightarrow \infty$ 趋于零，但其收敛到零的速度仅为负幂次 $k^{-\alpha}$ （Tsay, 2013）。长记忆性的存在表明时间序列数据具有一定变化趋势和周期性波动，而不是完全的随机波动。

作为以个人投资者为主导的新兴市场，我国 A 股市场普遍存在非随机的价格波动行为，如趋势延续性、价格反转、周期性波动等。这些行为暗示着长记忆性的存在，也表明 A 股市场具有一定的可预测性。事实上，大部分实证研究结果都表明，我国 A 股市场存在显著的长记忆性特征（余俊等，2008；余立威等，2017）。然而，目前的研究主要集中于时间序列的维度，对横截面维度关注较少；对指数的实证分析较多，而对个股数据的探究不足。此外，对长记忆指标的估计方法也在不断更新迭代中。因此，本文将我国 A 股市场全部股票为研究对象，采取优化的 GPH 估计法，检验收益率波动的长记忆性是否在个股层面普遍存在；同时进行横截面维度的研究，探究长记忆性与预期收益

率间的联系，并提出一种可能的作用机理。

（二）研究意义

本文的研究可能具有以下研究意义：

首先，验证长记忆性的存在性可以对经典有效市场假说的成立提出质疑。有效市场假说认为市场对信息反馈迅速，价格变动是随机的，但长记忆性的存在暗示了股价具有一定的可预测性，挑战了这一假说。其次，长记忆性的存在性能够支持行为金融学对于投资者非理性行为和市场波动的理论观点。这一发现可以提供证据，表明投资者在决策过程中可能会受到历史信息的影响，从而导致股价的长期趋势和波动性。另外，对长记忆性的横截面研究可以帮助更好地理解股票市场的行为特征。通过深入研究长记忆性，可以揭示市场参与者的交易行为和决策模式，有助于预测市场未来的发展趋势。股价波动长记忆性的研究还可以为金融理论的进一步发展提供新的思路和证据支持，从而在丰富现有理论的同时，也为后续的研究者提供了新的思路和视角。

在实践中，长记忆性的研究有助于改进风险管理模型，提高风险管理的准确性和有效性。通过对股价波动长期趋势的理解，可以更好地评估资产的风险水平，从而更有效地进行资产配置和风险控制。同时，长记忆性与预期收益率的联系可以为投资者提供更有效的投资策略。投资者可以在资产定价时囊括长记忆性相关的因子，更准确地制定交易策略，降低投资风险，提高收益率。最后，对长记忆性的研究也对金融政策的制定具有一定的指导意义。了解股价波动的特征可以帮助监管部门更好地制定监管政策，维护市场稳定和健康发展。

（三）研究方法及框架

本文主要采用了理论分析及实证研究的方法，验证股价波动长记忆性的存在性及其与预期收益间的关系。本文主要结构如下：

第一章为引言，主要介绍了本文的研究背景、研究意义与创新点、行文结构。

第二章为文献综述，介绍了长记忆性的提出历史和现有研究结果，对比了长记忆指标的多种估计方法，并给出对应述评。

第三章为样本与变量，列举了数据的获取途径和具体描述，并记录了使用 Python 来进行清洗和转换原始数据的详细流程。本章也介绍了长记忆时间序列的典型模型 $ARFIMA(0, d, 0)$ 和长记忆指标 d 的理论与实际含义，随后根据 GPH 估计法的理论研究结论确定滚动窗口的长度和有关参数，并使用 Python 程序计算得到长记忆指标的面板数据。

第四章为实证分析，包含三个主要研究问题的假设检验及结果分析。首先长记忆指标进行了描述性统计和假设检验，以验证我国 A 股市场长记忆性的存在性和普遍性，并简单探究了其时变特性。然后在对长记忆性与预期收益的关系的探究中，本文使用了单变量分组方法，并对分组的均值和 CAPM 模型下的 α 值进行了对比分析和显著性检验。最后在机理分析中，使用异质自回归 (HAR-RV)

模型来度量可预测性，得到长记忆性通过可预测性影响预期收益的作用机制。

第五章为稳健性检验，检验了行业门类对本文结果无明显影响。本章再次使用了描述性统计的方法，对比了行业间长记忆指标统计特征的异同，验证了第四章的结论对大多数行业门类都成立。

第六章为结语，总结了本文研究结论，同时反思了本文的不足和局限之处，并对未来的研究方向进行了展望。

（四）研究创新点

本文的研究可能在以下方面具有创新性：

首先，就研究视角而言，现有研究大多是对股票市场长记忆的存在性进行检验，本文则在此基础上进一步探究了其与预期收益率的关系。现有研究往往基于时间序列的维度，本文则更多基于横截面维度来揭示 A 股市场长记忆性的特征，尤其是行业间对比提供了创新的视角。

其次，就研究对象而言，现有文献主要集中在美国等成熟市场，而本文选择了我国 A 股市场作为研究对象。一方面，它拓展了对长记忆性的研究范围，为该领域的研究增添了新的维度和深度；另一方面，它也为我们更深入地探讨新兴市场和成熟市场之间的差异提供了重要的补充。本文还囊括了 A 股市场全部个股，并按行业门类划分，与现有文献主要以指数为研究对象相比，提供了更细致和多元的结论。

最后，本文使用了长达 27 年的 A 股市场日频数据，与目前同类研究相比样本量更大。本文还首次使用 Python 实现了 GPH 算法，扩展了该算法的应用环境。

二、文献综述

（一）长记忆性

对长记忆性的研究最早可以追溯到 Hurst (1951)，他通过对尼罗河的水文数据进行分析，提出了长记忆性用以描述流量数据的持续性，并在 40 多年水位控制研究的基础上创建了经典 R/S 分析（重标极差分析）。此后对长记忆性的研究广泛出现在了科学领域，如物理学、气候学、生物学等。Mandelbrot (1971) 首次将长记忆性的理论应用于资本市场，并进一步发展了 R/S 分析法。

长记忆性理论的引入带来了大量新的时间序列模型和方法论，因为用于短记忆时间序列的旧方法不再适用于分析、估计和预测这些过程。最常见的模型是由 Granger & Joyeux (1980) 和 Hosking (1981) 提出的自回归分数积分移动平均 (ARFIMA) 模型，以及由 Baillie 等人 (1996) 提出的分数积分广义自回归条件异方差 (FIGARCH) 模型。这些模型分别是短期记忆 ARMA 和 GARCH 模型的扩展。

目前国内外学者关于长记忆性的实证研究主要集中在对不同金融市场的长记忆性的存在性检验和对比上。余俊等 (2008) 使用修正 R/S 分析和 V/S 分析，对世界上 31 个国家或地区的股票指数进行了长记忆性的比较研究，得到以美国为代表的大多数发达国家（德国等一些欧洲国家例外）股市一般不存在长记忆性，而中国等发展中国家（巴西等南美洲国家例外）大多存在显著的长记忆性，尤其中国股市的长记忆性最强。余立威和陈涛 (2017) 则结合平均小波系数法和自回归原始自助法，对中国内地、美国、中国香港、德国的股市进行了实证分析，发现仅中国内地的股市存在显著的长记忆性。苑莹和庄新田 (2008) 通过多种统计分析方法得出中国沪深两市均具有长记忆性。但李云红等 (2015) 的研究则认为中国沪深两市指数、日经 225 指数和标准普尔 500 指数在多种半参数估计法下都未表现出显著的长记忆性，只在极端事件发生时表现出收益率序列的相关性。

而对金融市场中长记忆性的形成机制，目前存在多种解释。Lillo 和 Farmer (2004) 认为，市场本身存在信息成本，且投资者接受信息与理解信息间存在延迟，这导致了长记忆性的出现。赵桂芹和曾振宇 (2003) 则认为投资者对信息的非线性反应，导致信息量超过阈值时会爆发性地体现在价格中，从而展现出证券市场的长记忆性特征。他们以上海证券交易所的交易量作为研究对象，指出长记忆性的主要原因是市场中存在较多的噪声交易者，而信息披露和金融工具的缺乏加剧了这一现象。

此外，近年来也有学者从资产定价的角度指出，横截面维度的长记忆性研究存在重要价值。Nguyen 等 (2020) 以美国股票为研究对象，首次分析了波动率长记忆性与资产定价的关系。他们验证了长记忆性在个股收益率的波动中普遍存在，并与公司规模、历史业绩和跳跃强度有关。他们的研究发现，美国股市中 95% 的股票具有显著的长记忆性，平均长记忆指标为 0.22；更高的长记忆性

与较大的公司规模、较差的历史业绩和较少的价格跳跃相关。陈淼鑫和黄振伟（2023）则聚焦中国市场，发现长记忆指标中包含公司特征因素以外的重要定价信息，在不同市场状态和不同持有期下，长记忆指标的风险溢价始终显著为负。

（二）长记忆指标的估计方法

根据已有文献，长记忆参数的估计方法主要包括三类：非参数估计法、参数估计法以及半参数估计法。

非参数估计法中出现最早、应用最广的为经典 R/S 分析法（即重标极差分析法），由 Hurst（1951）提出，并由 Peters（1994）首先应用于资本市场。该方法通过对时间序列的累积求和，并计算不同时间尺度下的极差与标准差比值的平均值，然后对这些值进行对数变换和线性拟合，最终得到 Hurst 指数的估计值。但该方法在序列长度较短时通常会存在有偏性，且由 Hurst 指数描述的记忆性仅对线性过程有效，因此需要结合各类改进和修正方法才能得到较准确的结果。另一种经典的非参数估计法是由 Giraitis（2003）提出的重标方差法（V/S 法）。V/S 统计量不易受到离群点的影响，且能更好地处理均值附近的波动，因此解决了 R/S 方法对 Hurst 指数估计有偏的问题。尽管如此，该方法依旧对数据长度敏感，实际样本量往往不足以使该方法达到最准确。

参数估计法包括 Engle 和 Bollerslev（1986）提出的 IGARCH 模型，Nelson（1990）提出的 EGARCH 模型，Baillie（1996）提出的 FIGARCH 模型，和 Davidson（2004）提出的 HYGARCH 模型等。这类方法对模型设定的准确性要求较高，即使很小的结构调整都会引起参数估计结果的剧变，因此往往无法通过稳健性检验。如果想要得到一致、无偏、有效的估计量，需要同时估计模型中的所有参数，并满足严格的假设条件，因此近年来往往被半参数估计法所替代。

而半参数估计法主要包括 Geweke 和 Porter-Hudak（1983）提出的 GPH 估计法和 Robinson（1995a）提出的 LW (local whittle) 估计法。由于这类方法无需估计模型的所有参数，也不强调随机扰动项的正态性，因此拥有更大的适应性和更强的稳健性，近年来得到了更多学者的青睐。两种方法均较为常用，但 GPH 估计法除了计算简便快捷外，还具有如下优势。Robinson（1995b）证明了一定条件下，GPH 得到的估计量渐进服从正态分布，相比于 LW 估计法，其估计量的极限分布形式更为简单，易于检验。GPH 估计法的稳健性问题主要存在于带宽参数 m 的设置范围，但 Hurvich 和 Deo（1999）的研究给出了带宽参数 m 的设置范围，能够极大程度减小这一影响。最后，Kunold 等（2009）通过横向比较 GPH 估计法和 LW 估计法，表明一定样本量下 GPH 方法具有更高的准确率，明确给出了两种方法各自更为适用的范围。

（三）文献述评

综上所述，在长记忆性方面，目前研究已经介绍了理论研究长记忆性所需的几种时间序列模型，也对长记忆性的存在性做出了非常全面的检验。尽管检验结果并不统一，但在针对我国股市的探讨

上，大多数研究者认为沪深市场存在长记忆性，且这一现象能够被一些行为金融学理论所解释。在此基础上，近年有学者从横截面维度探讨了长记忆性在资产定价上的作用，但这方面的研究较为稀少。因此，本文将首先检验我国 A 股市场的长记忆性存在性，然后探究其与预期收益间的关联，并给出这一关联的一种可能作用机理，最后对长记忆性的表现进行行业间比对，拓展该方面的研究深度。

而在长记忆指标的估计方法的选择上，非参数估计法的准确性和参数估计法的稳健性均较低，因此实证研究中往往采取半参数估计法。而半参数估计法中的 GPH 估计法由于其简捷性，以及在本文研究的样本量下更具优势，因此本文采用 GPH 估计法来实现长记忆指标的估计。

三、样本与变量

（一）样本选择

本文以我国 A 股市场全部股票为研究对象，日股票综合数据来源为锐思金融研究数据库（www.resset.com）。另外，实证分析中使用的 Fama-French 因子数据来源于中央财经大学中国资产管理研究中心（sf.cufe.edu.cn/info/1198/13526.htm）和国泰安数据库，稳健性检验中使用的证监会行业门类代码（Csrciccd1）采取 2023 年上半年版。

考虑到 1996 年 12 月 16 日起深交所和上交所实行涨跌幅限制，对股价波动特性造成巨大影响，为保证样本的可比性，选取样本期为 1997 年 1 月至 2023 年 12 月。为避免股票停复牌的影响，并考虑到长记忆指标的滚动估计窗口为 5 年，估计频率为每月，本文根据如下规则筛选样本：

- （1）剔除 ST 或 PT 的股票；
- （2）剔除上市不满 3 年的股票；
- （3）剔除过去 1 年内不足 150 个交易日，或过去 1 个月不足 10 个交易日的股票；
- （4）每次估计时，要求过去 5 年中的交易天数大于区间总交易天数的 70%（以 875 天记），且过去 1 个月的交易天数大于 10 天。

本文共计从锐思金融研究数据库获取 198 份文件的数据，每份文件含有不超过 100001 条样本，样本字段及描述如表 3-1。

表3-1 样本字段及描述

样本字段	字段描述
上市状态	Norm（此为筛选条件）
股票代码_Stkcd	最新 6 位股票代码
证监会行业门类代码_Csrciccd1	字符型，A~S 不等
日期_Date	字符型，19970101~20231231 不等
日收益率_Dret	浮点型，精度为小数点后 4 位

（二）长记忆指标的估计

根据 Tsay（2013），长记忆时间序列研究中的典型模型是自回归分数积分移动平均模型，即 $ARFIMA(0, d, 0)$ ，其中参数 d 即被称为长记忆指标。当 $-0.5 < d < 0.5$ 时， $ARFIMA(0, d, 0)$ 模型的具体表达式为

$$(1 - L)^d X_t = \xi_t \quad (3-1)$$

其中 $\{\xi_t\}$ 是均值为零的独立同分布白噪声， $\{X_t\}$ 为时间序列。事实上，当 $d < 0$ 时，时间序列表现为平稳的反记忆过程，即 ACF 在滞后阶数 $k = 1$ 时是负值，股价波动会带来负相关的影响；当 $d = 0$ 时，时间序列退化为 $ARMA$ 模型，表现为短记忆过程，即 ACF 随 $k \rightarrow \infty$ 快速趋于 0，股价波动的影响会迅速消失，这与有效市场假说相符合；当 $0 < d < 0.5$ 时，时间序列表现出缓慢的收敛，即 ACF 以 $k^{-\alpha}$ 的速度趋于 0，但时间序列依旧是弱平稳的，此时股价波动带来长记忆的影响，其影响效果随时间缓慢降低；当 $0.5 \leq d < 1$ 时，时间序列不再平稳，但仍具有均值回复的特性，表明股价波动产生的影响会高度持久；当 $d \geq 1$ 时，时间序列成为永久记忆过程，表现为发散，此时任何波动都会对未来股价产生巨大影响，且不会回复到均值。

另外，当 d 处于 0~0.5 区间时，可以解得平稳解 $\{X_t\}$ 的 ACF 为

$$\rho_k = \frac{d(d+1)\dots(d+k-1)}{(-d+1)(-d+2)\dots(-d+k)}, k = 1, 2, \dots \quad (3-2)$$

因此长记忆指标 d 越大表明长记忆性越强，也表明信息对股价造成的影响时间越长。但在实证研究中，如果仅关心时间序列长记忆性，而不关心其稳定性，则无需考虑 0~0.5 区间的限制， d 值大小均可表明长记忆性的强弱。

结合理论与本文的样本选取，使用 GPH 方法实现对长记忆指标 d 的估计。根据 Geweke 和 Porter-Hudak (1983)，GPH 方法通过对周期图与谱密度形状进行对数线性回归来估计长记忆指标 d 。具体而言，首先设定带宽参数 m ，并结合样本量 n 计算对应的 m 项傅里叶频率

$$\lambda_j = \frac{2\pi j}{n}, j = 1, 2, \dots, m. \quad (3-3)$$

然后计算样本值（即日收益率的平方） X_t 在每一项傅里叶频率 λ_j 处的周期图

$$I(\lambda_j) = \frac{1}{2\pi n} |\sum_{t=1}^n X_t e^{it\lambda_j}|^2 \quad (3-4)$$

并以

$$4 \sin^2\left(\frac{\lambda_j}{2}\right) \quad (3-5)$$

代表谱密度的形状。通过线性回归模型

$$\log(I(\lambda_j)) = \beta_0 + \beta_1 \log\left[4 \sin^2\left(\frac{\lambda_j}{2}\right)\right] + \epsilon_j \quad (3-6)$$

得到 β_1 的估计值 $\widehat{\beta_1}$ ，则长记忆指标 $d = -\beta_1$ 。

在参数的选择上，根据 Stroe-Kunold 等 (2009) 的测试，1024~2048 为样本量 n 的最佳取值区间，且此时 GPH 方法具有所有半参数估计方法中最好的估计效果。考虑到一年约有 250 个交易日，取 5 年滚动窗口内将得到约 1250 个样本点，能够在符合最佳区间的同时得到尽可能多的长记忆指标估计数据。而对于带宽参数 m 的选择，根据 Hurvich 和 Deo (1999) 的研究，采用 $m = n^\delta$ 的设置形式。由于 $\delta = 0.5$ 是理论最佳取值和标准设置，本文即采取 $m = \sqrt{n}$ 。

四、实证分析

（一）长记忆指标的描述性统计和显著性检验

1. 描述性统计

对样本期内所有股票进行筛选，并对满足要求的股票依滚动窗口使用 GPH 估计法，得到长记忆指标的面板数据。计算所得长记忆指标的描述性统计如表 4-1 所示。

表4-1 长记忆指标的描述性统计

估计总数	均值	标准差	偏度	峰度
360407	0.3619	0.1773	0.0431	-0.2104

（续）表4-1 长记忆指标的描述性统计

25% 分位数	50% 分位数	75% 分位数	d<0 占比	0<d<0.5 占比	0.5<d 占比
0.2364	0.3594	0.4858	1.62%	75.73%	22.65%

由表 4-1 可知，样本期内共得到 360407 个长记忆指标估计值，75.73%的估计值落在 0~0.5 区间，表明 A 股市场的股票日收益率波动序列大多具有长记忆性，且是弱平稳的；1.62%的估计值小于 0，表明有极小部分时间序列具有负记忆或反持久；22.65%的估计值高于 0.5，表明有一部分股票的日收益率波动时间序列并不稳定。长记忆指标的平均值为 0.3619，标准差为 0.1773，表明在平均水平上，日收益率波动的时间序列具有显著的长记忆性。由 25%分位数为 0.2364 和 75%分位数为 0.4858 可知，长记忆指标的估计值有相当一部分都落在 0.2~0.5 区间内。根据 Stroe-Kunold 等（2009），这也是 GPH 估计法最为准确的区间。结合滚动窗口所决定的样本量，上述估计从理论上满足高准确度的要求，因此应具有较高可信度。

对每个时间点上有效的长记忆指标取算数平均值，得到长记忆指标的时间序列，以折线图表示；同时对每个时间点的估计数量进行计数，得到长记忆指标个数的时间序列，以柱状图表示。结果如图 4-1 所示。

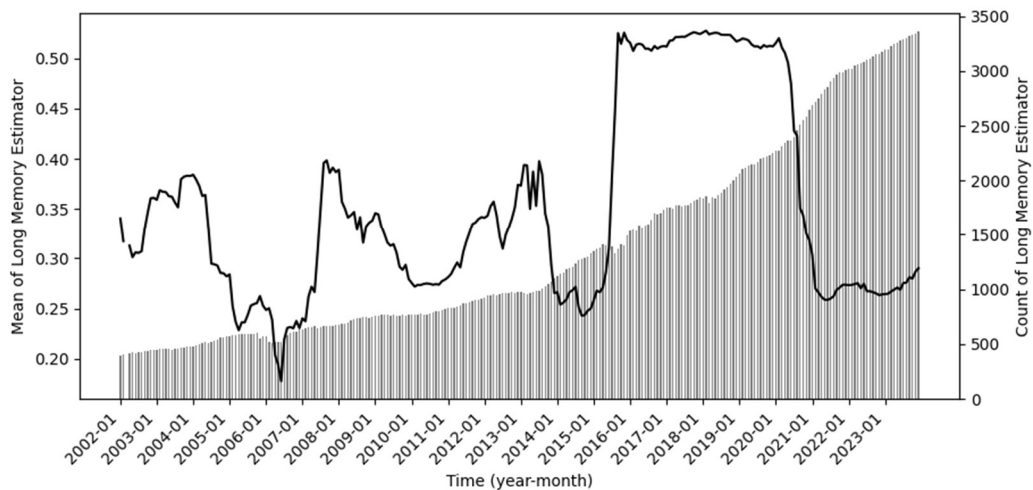


图 4-1 平均长记忆指标的时变描述

由图 4-1 可知，长记忆指标的平均值始终大于 0，且除 2015~2020 年间的特殊时段外始终小于 0.5，这一特殊时段可能是由 2015 年 A 股股灾导致的。长记忆指标随时间存在较大波动，且没有明显的趋势性，这表明 A 股市场的股票日收益率波动在样本期内始终存在长记忆性，但长记忆性并未随时间减少。从这个角度来说，我国股市的有效性，尤其是信息反馈的有效性并没有明显的提高。

此外，由于 2002 年 2 月的交易日数量少于 10 天，按筛选规则该月无法计算有效的长记忆指标，因此上图曲线存在断点。为保证时间序列的连续性，本章第二、三节使用的 GPH 数据自 2002 年 3 月开始。

2. 假设检验

假设 1: A 股市场股票日收益率波动具有长记忆性。

使用单侧 t 检验的方法，对每支股票的长记忆指标计算 t 值，并计算得 A 股市场长记忆指标的平均 t 值为 27.65，表明从平均意义上说，长记忆指标极显著大于 0，无法拒绝原假设。

另外，对个股长记忆指标的 t 值分布进行统计，得到图 4-2。样本内 98.92% 的股票都具有显著大于 0 的长记忆指标，这体现出长记忆性不仅仅是平均意义上的，更是普遍意义上的。

因此本文得出与陈淼鑫等（2023）相同的结论，即我国股票日收益率波动存在非常显著的长记忆性。根据赵桂芹（2003）的解释，A 股市场普遍存在的长记忆性与个人投资者居多、投资者不理性、信息披露不完整等问题有关。中国证券市场上个人投资者占比较高，且个人投资者往往了解经济学和证券投资知识较少，并伴有新股民的涌入，加上存在信息披露的不规范的问题，许多投资者只能依托于噪声信息进行交易，以“跟风”和“跟庄”作为主要投资策略，呈现出对信息的非线性反应。这些原因都造成了 A 股市场显著、普遍存在长记忆性。

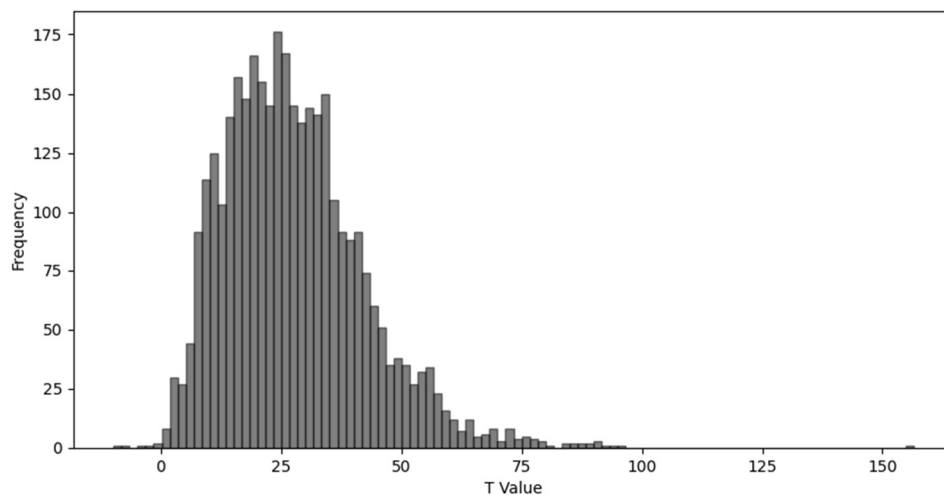


图 4-2 个股长记忆指标的 t 值频数直方图

（二）股价波动长记忆性与预期收益

1. 单变量分组

为了研究股价波动的长记忆性与横截面股票收益率间的关系，本文采用了一种基于长记忆指标 d 的排序方法。具体而言，本文将所有股票按照每个月的长记忆指标升序排列，并将它们等分成 10 组，即将股票分为 10 个子样本。其中，第一组（L 组合）包含了长记忆性最低的股票，即长记忆指标排序前 10% 的股票；而第十组（H 组合）则包含了长记忆性最高的股票。另外，本节还考虑了做多 L 组合并做空 H 组合投资策略，称为 L-H 组合。在形成组合时，考虑等权重加权，并记录每个组合在接下来一个月内的收益率。每个月重新根据新的长记忆指标对股票进行排序分组，且长记忆指标的数据与组合持有期的收益率数据之间不存在重叠。

2. 均值检验

计算十一个组合每月收益率的平均值，并计算其相对于 0 的 t 值，结果如表 4-2 所示。

表4-2 单变量分组的月收益率均值

	L	Q2	Q3	Q4	Q5
Count	260	260	260	260	260
Mean	0.0127	0.0127	0.0130	0.0126	0.0126
Sd	0.0855	0.0869	0.0872	0.0887	0.0880
T	2.3881	2.3564	2.3984	2.2916	2.3025
P	0.0088	0.0096	0.0086	0.0114	0.0110

(续) 表4-2 单变量分组的月收益率均值

Q6	Q7	Q8	Q9	H	L-H
260	260	260	260	260	260
0.0132	0.0124	0.0123	0.0108	0.0019	0.0023
0.0890	0.0895	0.0905	0.0894	0.0242	0.0222
2.3840	2.2254	2.1841	1.9440	1.2614	1.7005
0.0089	0.0135	0.0149	0.0265	0.1041	0.0451

假设 2: L-H 组合具有大于 0 的平均收益。

由表 4-2 可知, L-H 组的组合平均收益为 0.23%, 在 $p=0.05$ 的水平下显著大于 0。这表明做空 H 组合做多 L 组合的投资策略具有显著的正收益。

另外, 从 L 组到 H 组的 10 个投资组合展现出大致逐渐降低的平均收益率, 具体而言, L 组的平均收益率为 1.27%, 随着长记忆指标的增加降低到了 H 组的 0.19%。这表明在横截面的维度上, 长记忆性越高的股票预期收益越低, 也即股价波动的长记忆性和预期收益之间存在负向关系,。

3. CAPM 模型检验

为了排除市场因子对收益率的影响, 我们继续考虑了 CAPM 模型。回归模型如下

$$R_i - R_f = \alpha_i + \beta_i \times (R_m - R_f) + \epsilon_j \quad (4-1)$$

对十一个组合分别进行上述回归, 得到表 4-3 的结果。

表4-3 单变量分组的 α 值

	L	Q2	Q3	Q4	Q5
α	0.3731%	0.3397%	0.3823%	0.5434%	0.3213%
Sd	0.0016	0.0017	0.0015	0.0016	0.0016
t Stat	2.3494	2.0203	2.4757	3.3006	2.0378
P-value	0.0196	0.0444	0.0139	0.0011	0.0426

(续) 表4-3 单变量分组的 α 值

Q6	Q7	Q8	Q9	H	L-H
0.4492%	0.4033%	0.3617%	0.3315%	0.1137%	0.2594%
0.0017	0.0017	0.0018	0.0017	0.0016	0.0013
2.6990	2.3119	1.9994	1.9008	0.7244	1.9438
0.0074	0.0216	0.0466	0.0584	0.4695	0.0530

假设 3: L-H 组合具有大于 0 的 α 。

由表 4-3 可知, L-H 组具有 0.2594%的超额收益率, 在 10%置信度下显著大于 0。这表明做空 H 组合做多 L 组合的投资策略具有显著的超额收益。

类似地, L 组到 H 组的 α 也呈现大致降低的趋势, α 从 L 组的 0.3731%减少到 H 组的 0.1137%。

以上分析均表明股价波动的长记忆性高低很可能是一个市场因子以外的定价因素。

（三）机理分析

从前文的分析可知，股价波动的长记忆性越高意味着股票的预期收益相对更低。一种可能的解释是，更高的长记忆性与波动率更高的可预测性相关联，而更高的可预测性意味着更低的风险。根据 CAPM 理论，如果投资者能够更准确地预测未来的波动率水平，那么他们所面临的风险将会降低，这是因为投资者能够更有效地规避波动率增加所带来的风险。因此，投资者对于风险的感知降低，他们对风险的需求也随之下降，导致了对风险的补偿更低。这进一步意味着，对于具有更高可预测性波动率的资产，投资者所要求的预期收益率更低。这就能解释为什么长记忆程度越高的股票预期收益率越低。为了验证这一影响机制，本文研究了长记忆指标与波动率的可预测性之间的关系。

1. 可预测性的度量

根据 Corsi (2009)，可以使用异质自回归 (HAR-RV) 的拟合程度（通常是调整后 R^2 ）来衡量波动率的可预测性。具体而言，以某月的波动率作为因变量，以滞后若干期的波动率作为自变量，构建一元或多元线性回归。回归模型的拟合程度表明滞后波动率能够解释当期波动率的程度，也就是通过历史波动率能够在多大程度上预测当期波动率。

本节采取的异质自回归对每只股票分别进行，自回归项从滞后一个月 (M)，三个月 (3M)，六个月 (6M)，一年 (1Y)，两年 (2Y)，三年 (3Y)，五年 (5Y) 依次递增，共计七个模型。模型的具体设定形式如公式 (4-2) ~ (4-8)。

$$HAR(1): RV_{t+1}^M = \alpha + \beta RV_t^M + \epsilon_t \quad (4-2)$$

$$HAR(2): RV_{t+1}^M = \alpha + \beta_1 RV_t^M + \beta_2 RV_t^{3M} + \epsilon_t \quad (4-3)$$

$$HAR(3): RV_{t+1}^M = \alpha + \beta_1 RV_t^M + \beta_2 RV_t^{3M} + \beta_3 RV_t^{6M} + \epsilon_t \quad (4-4)$$

$$HAR(4): RV_{t+1}^M = \alpha + \beta_1 RV_t^M + \beta_2 RV_t^{3M} + \beta_3 RV_t^{6M} + \beta_4 RV_t^{1Y} + \epsilon_t \quad (4-5)$$

$$HAR(5): RV_{t+1}^M = \alpha + \beta_1 RV_t^M + \beta_2 RV_t^{3M} + \beta_3 RV_t^{6M} + \beta_4 RV_t^{1Y} + \beta_5 RV_t^{2Y} + \epsilon_t \quad (4-6)$$

$$HAR(6): RV_{t+1}^M = \alpha + \beta_1 RV_t^M + \beta_2 RV_t^{3M} + \beta_3 RV_t^{6M} + \beta_4 RV_t^{1Y} + \beta_5 RV_t^{2Y} + \beta_6 RV_t^{3Y} + \epsilon_t \quad (4-7)$$

$$HAR(7): RV_{t+1}^M = \alpha + \beta_1 RV_t^M + \beta_2 RV_t^{3M} + \beta_3 RV_t^{6M} + \beta_4 RV_t^{1Y} + \beta_5 RV_t^{2Y} + \beta_6 RV_t^{3Y} + \beta_7 RV_t^{5Y} + \epsilon_t \quad (4-8)$$

其中， RV_t^M 指第 t 期的月度已实现波动率，通过将该股票当月每一天的收益率平方求和得到。如公式 (4-9)， $R_{t,i}$ 为第 t 期当月的第 i 个日收益率。

$$RV_t^M = \sum_i R_{t,i}^2 \quad (4-9)$$

多期的已实现波动率定义为期限内月度已实现波动率的算术平均值。以第 t 期的三个月已实现波动率为例，其计算公式如 (4-10)。

$$RV_t^{3M} = \frac{1}{3}(RV_t^M + RV_{t-1}^M + RV_{t-2}^M) \quad (4-10)$$

同样可以计算得到第 t 期的六个月、一年、两年、三年和五年已实现波动率 RV_t^{6M} , RV_t^{1Y} , RV_t^{2Y} , RV_t^{3Y} , RV_t^{5Y} 。

对样本中所有个股计算各期的月度已实现波动率和多期已实现波动率，得到时间序列数据，并根据模型 $HAR(1) \sim HAR(7)$ 进行异质自回归，得到模型的调整后的 R^2 。然后按照长记忆指标升序排列并等分为十组，依次记为 L, Q2~Q9, H。最后对各组内股票在不同异质自回归模型下的平均调整后 R^2 ，那么该值可以度量不同分组的平均波动率可预测性。

2. 结果分析

表 4-4 和图 4-3 展示的是各回归模型在不同分组下的平均调整 R^2 。

表4-4 不同分组下的 $adj. R^2$

	L	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	H
HAR(1)	0.0131	0.0635	0.1187	0.1709	0.1922	0.2075	0.2286	0.2220	0.1982	0.2072
HAR(2)	0.0203	0.0614	0.1374	0.1977	0.2131	0.2228	0.2331	0.2326	0.2005	0.2061
HAR(3)	0.0213	0.0987	0.1685	0.2092	0.2227	0.2305	0.2395	0.2437	0.2149	0.2127
HAR(4)	0.0896	0.1514	0.2044	0.2464	0.2397	0.2344	0.2549	0.2506	0.2269	0.2365
HAR(5)	0.1927	0.2173	0.2337	0.2685	0.2578	0.2399	0.2770	0.2563	0.2457	0.2548
HAR(6)	0.2317	0.2403	0.2641	0.2826	0.2718	0.2602	0.2848	0.2700	0.2557	0.2594
HAR(7)	0.2021	0.1851	0.2256	0.2746	0.2758	0.2678	0.2785	0.2660	0.2695	0.2666

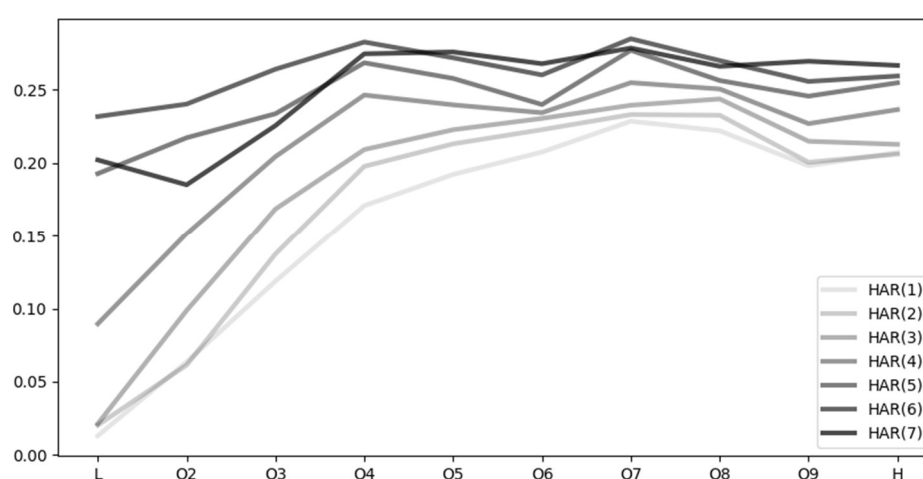


图 4-3 不同分组下的 $adj. R^2$ 对比图

每一个模型下调整 R^2 都大致随长记忆指标的增加而提高，这表明股价波动长记忆性更强的股票确实具有更高的波动率可预测性。尽管 Q9 与 H 组的 $adj. R^2$ 略微不符合上升趋势，但本文认为这是

由于长记忆指标大于 0.5，收益率波动序列不再平稳，导致 HAR-RV 模型的预测效果不佳。总体而言，上述结果表明所假设的作用机制成立：更高的长记忆性与波动率更高的可预测性相关联，而更高的可预测性意味着更低的风险，因此根据 CAPM 理论，投资者对风险补偿的需求也更低，从而形成了更低的预期收益率。

五、稳健性检验

本节旨在探究是否某些行业的公司具有较高或较低的长记忆指标。由于不同行业之间存在显著的特征差异，因此可能呈现出截然不同的长记忆性，这将导致本文的研究结果不适用于某些行业门类。具体而言，不同行业的盈利模式、市场结构、商业周期、法规政策均存在差异。高科技行业可能具有高增长性和高风险，而公用事业行业可能具有稳定的现金流和低风险；某些行业可能具有垄断或寡头市场，而另一些行业可能具有竞争激烈的市场；周期性行业（如建筑和制造业）可能更受经济周期的影响，而非周期性行业（如公用事业和医疗保健）可能相对稳定；政府的政策措施可能对某些行业的经营环境和盈利能力产生重大影响。这些因素都可能显著影响行业股价波动的长记忆性。因此，对比行业间的长记忆指标的表现对验证本文结果的稳健性是十分必要的。

本节使用了 2023 年上半年版证监会行业门类代码（Csrciccd1），具体见表 5-1。简洁起见，后续将用字母指代对应行业。

表5-1 2023年上半年版证监会行业门类代码（Csrciccd1）

Csrciccd1	对应行业门类	Csrciccd1	对应行业门类
A	农、林、牧、渔业	K	房地产业
B	采矿业	L	租赁和商务服务业
C	制造业	M	科学研究和技术服务业
D	电力、热力、燃气及水生产和供应业	N	水利、环境和公共设施管理业
E	建筑业	O	居民服务、修理和其他服务业
F	批发和零售业	P	教育
G	交通运输、仓储和邮政业	Q	卫生和社会工作
H	住宿和餐饮业	R	文化、体育和娱乐业
I	信息传输、软件和信息技术服务业	S	综合
J	金融业		

对不同行业门类的长记忆指标做描述性统计，得到结果如表 5-1 所示。其中 O 由于估计总数远小于其他行业门类，故在后续分析中不作考虑。所有门类（除 O 外）均具有显著大于 0 的长记忆指标，表明长记忆性存在于每个行业中。绝大多数行业的长记忆指标平均值都非常接近全市场的均值，说明行业代码无法很好解释长记忆指标的横截面变化，也不会对本文结果造成很大影响。

尽管如此，行业 P 的长记忆指标均值低于全市场水平，且 $d < 0$ 占比显著高于全市场，表明该行业体现出反记忆性的股票比例较高。这可能是因为教育市场的不确定性较高，政策调整频繁、教育

需求波动等因素使得投资者更倾向于在股价剧烈波动时抛售股票，以规避风险，从而加剧了反记忆性。教育行业面临的宏观经济因素也可能在其中起到作用，如人口结构变化的不确定性可能影响投资者对行业的信心，导致他们更倾向于短期投机而非长期持有，从而降低了长记忆性。

行业 J 也体现出低于平均水平的长记忆性，并且 $0.5 < d$ 占比显著更低，表明该行业的股票波动较少出现高度长记忆性，即存在更少的长期趋势。这可能是因为金融业具有高度灵活性和快速调整能力，导致股票价格的长期依赖性受到限制。同时，金融业受到严格的监管和政策限制，这可能会影响金融机构的业务策略和风险管理，从而限制了长期趋势的形成。

另外，本节考察了不同门类的长记忆指标平均水平随时间变化的趋势，得到图 5-1 的结果。大多数行业的时变趋势都很一致，与全市场的变化趋势较为统一，表明从横截面来看长记忆指标在行业间的差异的确并不非常突出。

表5-2 不同门类长记忆指标的描述性统计

	估计总数	均值	标准差	平均 t 值	$d < 0$ 占比	$0 < d < 0.5$ 占比	$0.5 < d$ 占比
A	4685	0.3691	0.1626	33.85	1.52%	76.82%	21.66%
B	8285	0.3817	0.1432	37.43	0.54%	80.51%	18.95%
C	238299	0.3609	0.1769	27.67	1.61%	75.96%	22.43%
D	14313	0.3589	0.1823	29.94	2.10%	74.99%	22.92%
E	7173	0.3694	0.1748	27.31	1.23%	74.03%	24.75%
F	19307	0.3671	0.1807	29.21	1.39%	73.32%	25.29%
G	13949	0.3610	0.1886	28.29	1.48%	74.81%	23.71%
H	1317	0.3375	0.1526	35.62	1.59%	84.28%	14.12%
I	18294	0.3685	0.1929	23.72	2.03%	69.93%	28.04%
J	8033	0.3275	0.1527	28.53	2.03%	88.21%	9.76%
K	14228	0.3786	0.1659	32.49	1.12%	74.91%	23.97%
L	2795	0.3574	0.1845	21.47	1.43%	76.17%	22.40%
M	1823	0.3291	0.1821	21.88	2.36%	81.35%	16.29%
N	2669	0.3505	0.2036	22.32	4.05%	71.82%	24.13%
O	28	-0.1212	0.0841	-7.63	92.86%	7.14%	0.00%
P	211	0.2663	0.2025	13.52	10.43%	72.99%	16.59%
Q	482	0.3809	0.1923	21.97	1.24%	64.32%	34.44%
R	3245	0.3688	0.1827	23.68	1.23%	74.58%	24.19%
S	1271	0.3683	0.1564	33.34	1.89%	76.00%	22.11%
全市场	360407	0.3619	0.1773	27.65	1.62%	75.73%	22.65%

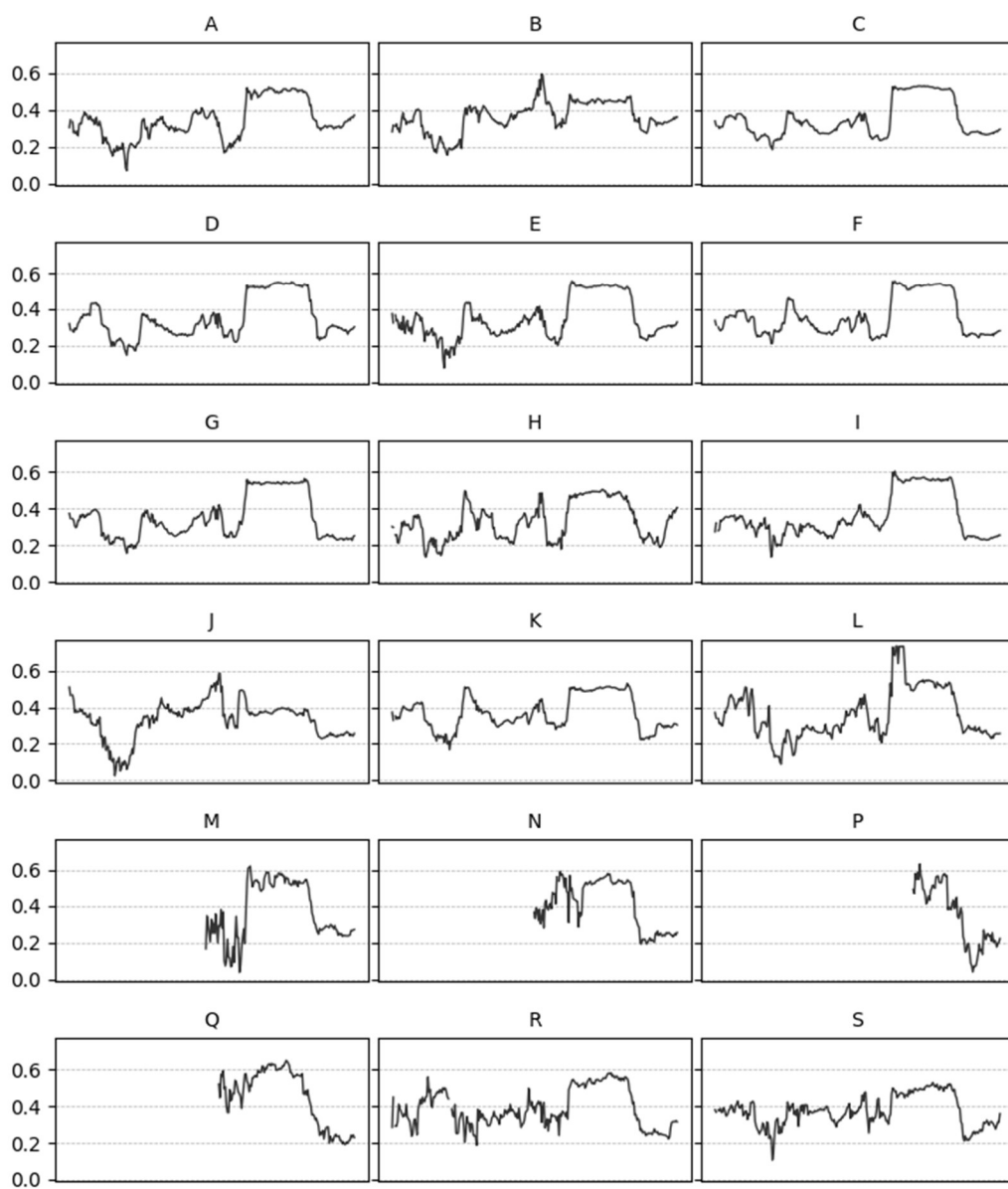


图 5-1 不同门类长记忆指标的时变对比图

六、结语

（一）研究结论

本文以 A 股市场个股的股价波动为研究对象，通过对滚动窗口使用 GPH 估计法得到长记忆指标的面板数据，并以横截面的视角研究了长记忆指标与预期收益率间的关系，以及该关系的可能解释机制。本文得出了一系列重要结论如下：

首先，实证结果表明，我国 A 股市场绝大部分个股收益率波动都具有显著的长记忆性。在本文样本所取的 1997 年 1 月到 2023 年 12 月间，98.92% 的股票都得到了显著大于 0 的长记忆指标，且全市场的平均长记忆指标为 0.3615，同样显著大于 0。这一发现不仅验证了 A 股市场存在长记忆性的普遍性，也挑战了传统的有效市场假说。长记忆性的存在暗示了股价具有一定的可预测性，这与有效市场假说所主张的市场信息效率存在一定的冲突。然而，这种冲突也为市场参与者提供了机会，他们可以通过深入研究和有效利用长记忆性信息来获取超额收益。

其次，本文通过单变量分组实验，发现股价波动的长记忆程度与未来的超额收益之间存在着显著的负相关关系。等权重买入长记忆程度最低的组合，同时卖出长记忆程度最高的组合，这样的投资策略可以取得 0.23% 的平均收益，在 $p=0.05$ 的水平下显著大于 0。相比于长记忆程度最低的组合具有 0.37% 的超额收益率，长记忆程度最高的组合的 α 仅有 0.11%，两者也存在显著差异。这意味着长记忆程度低的股票存在更高的未来超额收益，说明长记忆指标中确实包含一定的定价信息。这一发现对于投资者制定有效的投资策略具有重要的指导意义。例如，投资者可以在资产配置时更倾向于选择长记忆性低的资产，以提高预期收益。

随后，本文对这一负相关关系提出了一种解释机制。通过实证分析发现，在单变量分组中，长记忆指标越大，异质自回归的调整 R^2 越大。由于异质自回归的拟合程度（调整 R^2 ）可以用来衡量股票波动率的可预测性，说明长记忆程度越高的股票分组具有更高的波动率可预测性。而根据 CAPM 理论，波动率可预测意味着风险降低，投资者对风险的需求下降，导致风险补偿更低，预期收益率也更低。因此长记忆指标越大的股票会具有相对更低的预期收益率。

此外，本文从行业门类的角度对长记忆性存在的稳健性进行了验证。结果显示，长记忆指标在几乎所有门类中均显著大于 0，表明长记忆性的存在性并未受到不同行业门类的明显影响。绝大多数行业（除教育业、金融业）的长记忆指标平均值都非常接近全市场的平均水平，说明行业代码不是影响长记忆指标的横截面变化的重要原因。另外，各行业长记忆指标的时变特征也与全市场变化趋势较为一致。这进一步确立了我们对于 A 股市场长记忆性的研究结论的可靠性。这也意味着长记忆性可能是 A 股市场整体性质的一部分，而非某些特定行业的特征。

综上所述，本文不仅对 A 股市场收益率波动的长记忆性进行了全面深入的研究，而且在研究方

法和视角上具有一定的创新性。本文的研究不仅对学术界提供了有价值的研究成果，也为投资者、监管机构以及金融市场的相关参与者提供了实践指导和决策支持。

（二）研究不足

本研究存在一些局限性。首先，尽管本文所采用的 GPH 法是估计长记忆指标的常用算法，且具有良好的性质，但近年来随着数学领域的不断发展，新的数学方法不断出现。这些新方法也许可以给出更为准确的估计，从而提高研究的严谨度。其次，本文没有涉及对伪长记忆性的检查和剔除，这可能会导致长记忆性结果的失真，从而影响我们的结论和推断。另外，本文在单变量分组时仅采取了等权重加权的方法，实际上，市值加权也是非常重要且常用的方法，能够更加贴近现实中投资组合的构造方法。因此，在未来的研究中，可以考虑使用其他新兴数学方法来估计长记忆指标，并增加对伪长记忆性的处理，并在单变量分组时加入市值加权构造投资组合的步骤，研究其与等权重加权得到的结果是否一致。

此外，未来的研究还可以考虑在以下方向做出拓展深入：

在资产定价方面，可以在 CAPM 的基础上使用 Fama-French 三因子或五因子模型来研究长记忆指标带来的超额收益。进一步，可以通过 Fama-Macbeth 回归等方法进一步探究长记忆指标是否是当前模型未涉及的定价因子。如果能够在控制市值、账面市值比、动量指标、特质波动率、彩票指标、非流动性指标等因素的情况下，观测到长记忆指标具有显著不为零的风险溢价，则能更为严谨地证明长记忆指标的资产定价能力。

在机理分析方面，可以通过研究投资者结构对长记忆指标的影响，解释其与预期收益的负相关关系。按长记忆程度得到的不同分组，可能具有不同的短期投机者与长期持有者比例，而投资者结构的差异可能导致长记忆性与预期收益之间的负相关关系。这也许能够从行为金融学的角度为长记忆指标的定价能力提供另一种解释。

总之，长记忆性的存在性和横截面特性将继续成为金融市场研究的重要课题。期待未来能有更多的学者和从业者投入到这一领域的研究中，共同推动金融理论和实践的不断发展和完善，更全面地理解市场的行为特征，为投资者提供更有效的风险管理和投资决策支持。

参考文献

- [1] 陈淼鑫, 黄振伟. 股价波动的长记忆性与横截面股票收益——基于中国市场的实证研究[J]. 中国管理科学, 2023, 31(4): 1-10.
- [2] 李云红, 魏宇, 张帮正. 股票市场历史信息的长记忆性特征研究[J]. 中国管理科学, 2015, 23(9): 37-45.
- [3] 余俊, 方爱丽, 熊文海. 国际股票市场收益的长记忆性比较研究[J]. 中国管理科学, 2008, 16(4): 24-29.
- [4] 余立威, 陈涛. 股票市场的长记忆及其动态结构[J]. 数理统计与管理, 2017, 36(6): 1106-1118.
- [5] 苑莹, 庄新田. 中国股票市场的长记忆性与市场发展状态[J]. 数理统计与管理, 2008, 27(1): 156-163.
- [6] 赵桂芹, 曾振宇. 证券市场长期记忆特征的实证分析[J]. 管理科学, 2003 (2): 40-44.
- [7] Tsay R S. 金融数据分析导论: 基于 R 语言[J]. 2013.
- [8] Baillie R T, Bollerslev T, Mikkelsen H O. Fractionally integrated generalized autoregressive conditional heteroskedasticity[J]. Journal of econometrics, 1996, 74(1): 3-30.
- [9] Corsi F. A simple approximate long-memory model of realized volatility[J]. Journal of Financial Econometrics, 2009, 7(2): 174-196.
- [10] Davidson J. Moment and memory properties of linear conditional heteroscedasticity models, and a new model[J]. Journal of Business & Economic Statistics, 2004, 22(1): 16-29.
- [11] Engle R F, Bollerslev T. Modelling the persistence of conditional variances[J]. Econometric reviews, 1986, 5(1): 1-50.
- [12] Geweke J, Porter-Hudak S. The estimation and application of long memory time series models[J]. Journal of time series analysis, 1983, 4(4): 221-238.
- [13] Giraitis L, Kokoszka P, Leipus R, et al. Rescaled variance and related tests for long memory in volatility and levels[J]. Journal of econometrics, 2003, 112(2): 265-294.
- [14] Granger C W J, Joyeux R. An introduction to long-memory time series models and fractional differencing[J]. Journal of time series analysis, 1980, 1(1): 15-29.
- [15] Hosking, J R M. Fractional differencing[J]. Biometrika, 1981, 68(1): 165-176.
- [16] Hurst H E. Long-term storage capacity of reservoirs[J]. Transactions of the American society of civil engineers, 1951, 116(1): 770-799.
- [17] Hurvich C M, Deo R S. Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series[J]. Journal of Time Series Analysis, 1999, 20(3): 331-341.
- [18] Lillo F, Farmer J D. The long memory of the efficient market[J]. Studies in nonlinear dynamics & econometrics, 2004, 8(3).
- [19] Mandelbrot B. The Pareto-Levy law and the distribution of income[J]. International economic review, 1960, 1(2): 79-106.
- [20] Mandelbrot B. When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models[J]. The Review of Economics and Statistics, 1971: 225-236.
- [21] Nelson T O. Metamemory: A theoretical framework and new findings[M]//Psychology of learning and

motivation. Academic Press, 1990, 26: 125-173.

[22] Nguyen D B B, Prokopczuk M, Sibbertsen P. The memory of stock return volatility: Asset pricing implications[J]. Journal of Financial Markets, 2020, 47: 100487.

[23] Peters E E. Fractal market analysis: applying chaos theory to investment and economics[M]. John Wiley & Sons, 1994.

[24] Peters E E. Chaos and order in the capital markets: a new view of cycles, prices, and market volatility[M]. John Wiley & Sons, 1996.

[25] Robinson P M. Gaussian semiparametric estimation of long range dependence[J]. The Annals of statistics, 1995a: 1630-1661.

[26] Robinson P M. Log-periodogram regression of time series with long range dependence[J]. The annals of Statistics, 1995b: 1048-1072.

[27] Shiller R J. Market volatility[M]. MIT press, 1992.

[28] Stroe-Kunold E, Stadnytska T, Werner J, et al. Estimating long-range dependence in time series: An evaluation of estimators implemented in R[J]. Behavior research methods, 2009, 41(3): 909-923.

致谢

感谢指导老师和卢紫怡学姐在本文的书写过程中提供了巨大帮助，他们细致的修改和严谨的建议使我的文章最终得以成型。也感谢答辩组老师对本文提出了关键性的批评，让我尽早弥补了本文结构上的问题。

感谢走过这四年的自己，走出了小镇做题家心态，学会了平衡生活，培养了兴趣爱好，也磨砺了性格。感谢钟子健在全方面为我提供的支持，尤其是情绪价值。感谢父母的支持，无论是出国访学时的经济支出，还是对我转专业的理解；尤其感谢母上大人时不时的快递投喂，让我在上海也不必“休说鲈鱼甚脍”。

感谢金融学院的老师们，让我领略到这个专业的趣味与深刻；同时感谢教务和辅导员老师们的热心负责，让我的大学生活顺利而愉快。感谢数学学院对我的巨大帮助和鼓励，尤其是刘泓志老师，对我的学习、申请和幽默感提供了极大帮助。也感谢信息与管理学院几位老师对我学业规划上的指导。

感谢校足球队的朋友们和教练以及郭学长，我们不仅一起训练、比赛、聚餐，在平时也维系着极有凝聚力的关系。一起通宵看世界杯的日子，是疫情期间最疯狂最快乐的事情；每年的年夜饭，也是大学中最有归属感的时刻。感谢金融实验班的同学们，大家都极为优秀，无论学习还是为人处世都带着我一同进步。感谢我的饭搭子们，没有你们我将错过无数的美食。

感谢我的 ThinkPad T14 Gen1，这是我用过体验感最棒的电脑，不仅联网质量优秀，电池续航持久，而且在四年中任劳任怨，极其靠谱。感谢 Python，不仅让我的工作效率扶摇直上，更是改变了我的思维方式。感谢 Forza Horizon 带我领略不列颠的晚霞与墨西哥的火山；感谢 Stardew Valley 让我体验“但使愿无违”的田园生活。

最后，由衷地感谢所有帮助过我的人，祝你们生活幸福，万事如意。