

学号 2017302010265

密级

武汉大学本科毕业论文

基于统计的基本面分析

院（系）名 称：经济与管理学院

专 业 名 称：金融工程

学 生 姓 名：鲍余薇

指 导 教 师：李斌 教授

二〇二一年四月

郑 重 声 明

本人呈交的学位论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本学位论文的知识产权归属于培养单位。

本人签名：_____

日期：_____

摘 要

论文的摘要是对论文研究内容和成果的高度概括。摘要应对论文所研究的问题及其研究目的进行描述，对研究方法和过程进行简单介绍，对研究成果和所得结论进行概括。摘要应具有独立性和自明性，其内容应包含与论文全文同等量的主要信息。使读者即使不阅读全文，通过摘要就能了解论文的总体内容和主要成果。

论文摘要的书写应力求精确、简明。切忌写成对论文书写内容进行提要的形式，尤其要避免“第 1 章……；第 2 章……；……”这种或类似的陈述方式。

关键词是为了文献标引工作、用以表示全文主要内容信息的单词或术语。关键词不超过 5 个，每个关键词中间用分号分隔。

关键词：关键词 1；关键词 2；关键词 3；关键词 4；一个非常非常，非常非常长——的关键词 5

ABSTRACT

An abstract of a dissertation is a summary and extraction of research work and contributions. Included in an abstract should be description of research topic and research objective, brief introduction to methodology and research process, and summarization of conclusion and contributions of the research. An abstract should be characterized by independence and clarity and carry identical information with the dissertation. It should be such that the general idea and major contributions of the dissertation are conveyed without reading the dissertation.

An abstract should be concise and to the point. It is a misunderstanding to make an abstract an outline of the dissertation and words “the first chapter”, “the second chapter” and the like should be avoided in the abstract.

Keywords are terms used in a dissertation for indexing, reflecting core information of the dissertation. An abstract may contain a maximum of 5 keywords, with semi-colons used in between to separate one another.

Key words: key word 1; key word 2; key word 3; key word 4; and a very very, very very long key word—the key word 5

目 录

1 引言	1
2 文献综述	5
3 研究设计	7
3.1 数据	7
3.1.1 数据样本	7
3.2 实证方法	7
3.2.1 变量定义与说明	7
3.2.2 模型设定	9
3.2.3 描述性统计	10
4 实证分析	13
4.1 Fama-MacBeth 截面回归	13
4.2 因子模型	13
4.3 稳健性分析	16
5 结论与启示	19
参考文献	21
致谢	23

图片索引

3.1	样本数据的月度有效样本量	8
4.1	五组模型 Q5-Q1 结果（随机森林）	16
4.2	五组模型 Q5-Q1 结果（GBDT）	16
4.3	五组模型多空组合显著性检验	17

表格索引

3.1	变量定义与说明	8
3.2	描述性统计（随机森林方法）	11
3.3	描述性统计（GBDT 方法）	11
4.1	Fama-MacBeth 截面回归	14
4.2	因子模型回归（随机森林）	15
4.3	因子模型回归（GBDT）	15

1 引言

有效市场假说（Efficient Markets Hypothesis, EMH）是现代金融学的理论基石之一。Fama（1965）首次提出了“有效市场”的概念，并定义为：如果在一个证券市场中，价格完全反映了所有可以获得的讯息，那么就称这样的市场为有效市场。他认为市场上存在着大量理性的投资者，在任何时间都有无数人正在搜寻细微的线索去精准预测股票未来的价格，为了自身利益去对股票进行低买高卖，或许对于个人而言仅仅是超额利润的攫取，但在宏观维度上，正是这种高量级的活动快速推进着股票市场价格向正确价格的趋近，任何相关的信息已经完全展现在了股票价格中，所有为了预测付出的时间、金钱和努力都是徒劳无功。进而 Fama（1970）在《Journal of Finance》杂志上发表了最具划时代意义的论文《Efficient Capital Markets: A Review of Theory and Empirical Work》，正式提出了有效市场假说，并对证券市场的信息分为以下三类：一是与交易相关的历史信息，例如历史成交量、价格等；二是当前的公开信息，如公司财报、分红报告等；三是内部信息。Fama 按照以上三种信息把有效市场分为弱式有效、半强式有效和强式有效三类，其中弱式有效证明了技术分析的无效性；半强式有效排除了基本面分析的作用；而强式有效意味着股票价格已经反映了所有信息，投资者不可能战胜市场。至此确立了“有效市场假说”在现代金融领域的基础性地位。

但是现在越来越多的实证研究表明了中国股票市场的非有效性。自 70 年代开始，随着科技的发展，持续识别出能够提供超额收益的异象（因子）。曾有研究发现了股票价格过度波动和可预测性（Shiller, 1981）、股票价格走势逆转（Shieifer、Vishney, 1994）等异常情况；在单只股票或投资组合领域，也发现了规模效应（Bamber 等, 1997）、市盈率效应（Banz, 1981）等，前美国金融学会主席 Cochrane（2011）称其为“因子动物园”（Factor Zoo），这些异象说明投资者能够基于基本面等公开信息去预测股票价格，向有效市场假说发起了挑战。

另外，从行为金融角度看，很多学者指出有效市场假说理论的逻辑存在问题。如果股票价格已经完全反映市场中的信息，那理性投资者在后续的投资活动中就无需花费精力去搜集信息、预测股价了，但这却与市场需要信息才能有效运行相悖。同样，其设定的完美市场假定条件脱离实际。在实际的投资活动中，投资者更多表现为“非理性”，从而导致其决策行为偏离金融理论预测的标准结果，例如羊

群效应、月末效应、规模效应和股权溢价之谜等等金融异象都是投资者“非理性”行为的表现（张元鹏，2015）。虽然行为金融文献并没有完全否定有效市场的界定，但行为金融派的支持者更愿将有效市场看作一种理想状态、不受个人意志（效用和偏好）影响的市场状态（丁志国等，2017）。

从以上几个方面来看，我国股票市场未能做到完全有效，虽然 A 股市场现已发展成为全球第二大股票市场，总市值近 80 万亿元。但由于成立时间晚、相关制度不完善、散户比例大、交易成本高等问题，导致了定价效率的低下。这时基本面分析更能捕捉市场的非有效性，有效预测未来盈余，带来显著的超额回报（汪荣飞和张然，2018）。相对于技术分析依靠图表去预测价格趋势，基本面分析更为关注公司的基本信息，主要包括财务报表或非财务上的信息，并从中评估公司的内在价值。本文将股票价格与内在价格的偏离定义为错误定价，而这种错误定价便是投资者所追捧的超额收益，所以对于错误定价的研究显得极为重要。

由于股票价格是已知的，那么对错误定价的研究关键在于对股票内在价格、公司内在价值的预测。现有研究中常见的计算方法包括现金流贴现法（Free Cash Flow Valuation, FCFF）；相对价值法，即与同行业的相似公司对比，以它们的平均市盈率、市净率及市销率等指标来计算；经济附加值法（Economic Value Added, EVA）；实物期权法等。这些方法虽然起到了一定的预测作用，但由于其选择的指标过少，或太过于依赖对未来的主观预测结果，亦或计算方法过于程式化等原因，很难排除数据窥视（Data snooping）的影响。

基于以上分析，本文创新采用大数据统计方法计算公司的内在价值。利用 2003 年初到 2020 年底的季度财务报表数据，对 A 股所有公司（剔除金融类股票、ST 股票）进行时间维度的截面回归：公司市场价值对于一系列财务指标的线性回归，这些财务指标的组合结果即为公司内在价值，回归残差为公司市场价值与内在价值的偏差，将偏差按照公司市场价值标准化后，成功构建出错误定价因子 M（Mispricing Factor）。其中，线性回归系数按时点分组截面计算，避免了该时点下其他投资因素的影响；纳入 A 股所有公司，回归结果更具有普遍性；且后续模型修订根据客观统计标准，排除了数据窥视的影响。

在成功引入错误定价因子 M 后，接下来便是检验其有效性。问题在于：一、能否通过错误定价因子 M 构建低买高卖的投资组合获取超额收益，进而证明基本面分析的可行性？二、超额利润来自于错误定价还是因子遗漏，股票市场价格是否会向内在价格趋近？本文将针对以上两个问题进行研究。首先从 CSMAR 上搜集 A

股上市公司季度财务数据，计算每个时点公司的内在价值和错误定价因子 M ；后采用 Fama-MacBeth 回归计算错误定价因子 M 系数 β ，检验其是否显著；并且按照错误定价因子 M 将公司分为五组，从价值最被高估到最被低估，运用因子模型得超额收益 α ，观察是否呈现单增趋势；最后观察长期时间内不同分组超额收益的变动趋势，证明超额收益并非来源于风险因子的遗漏。最后结论证明了错误定价因子 M 的有效性，有力支持了基本面分析之于中国股市证券分析的地位。

本文的研究有一定的现实意义与理论贡献：一、丰富了经济学和管理学研究的工具箱；二、丰富了基本面投资的理论和实践研究；三、丰富了中国股票市场有效性的研究。目前对于中国 A 股市场基本面分析的研究过少且不完善，以大数据统计方法的研究更为稀有，统计之于数据分析的重要性不言而喻，其能够有效避免数据窥视的影响，是在此研究方向上的创新。同样，建模方法也有所创新，在线性回归的基础上增加了非线性维度的考量。

后文结构如下：第二部分回顾了中国股市有效性、基本面分析等相关文献，第三部分阐述了本文的研究设计，介绍数据和变量的构建方法；第四部分详细分析了本文的实证结果，第五部分为本文的研究结论。

2 文献综述

对于中国股票市场，已有多位学者研究证明其不具有弱势有效性。贾权和陈章武（2003）对基于市场有效假设的 CAPM 模型以及其他因素与收益率之间的关系进行了实证检验，发现目前我国股市不满足市场有效性的假设，投资者的行为并不是完全理性的；吴振翔和陈敏（2007）通过设计多种投资组合方式，发现在短期（3 个月以内）不能否定市场中无套利的假定；而对于中期和长期（6 个月及 12 个月）统计套利存在，说明我国 A 股股票市场的弱有效性不成立；Lim 和 Brooks（2009）运用非线性依赖检验的方法对 1999 年至 2005 年上证 A 股和 B 股的日数据进行检验，发现 A 股和 B 股均不具备弱势有效性；屈博和庞金凤（2014）利用 Q 统计量法、方差比检验法、广义谱检验、游程检验等多种方法对 2010 年至 2013 年沪深 300 股指期货的当月合约的 5 分钟高频数据进行检验，实证结果表明，随着考察期长度的增加，市场趋于拒绝弱式有效性。

上述研究说明，中国股市能够通过技术分析和基本面分析获取超额收益，实际上很多实证研究也证明了这个结论。对于技术分析，赵国顺（2009）基于时间序列方法，使用 GARCH 模型和 ARIMA 模型对股价波动趋势有着较好的短期预测效果；王劲松（2010）运用技术指标 MA、KDJ 和 MACD 证明在一定时间跨度上技术分析是有效的；石赛男（2011）的实证研究表明 MACD 指标对于 2004 年至 2009 年内的大、中、小盘股均有一定的预测能力；包懿（2015）基于 2010 到 2014 年三只指数标的和一只个股标的，通过持续持有策略和均线穿越法则策略获取了超额收益。对于基本面分析，张然等（2017）采用日历时间组合的方法，证实中国 A 股市场的分析师修正信息具有投资价值，并且这个投资价值主要来源于其能够预测公司基本面信息。汪荣飞和张然（2018）基于季度财务指标构建了六组基本面指标，发现均能够有效预测未来盈余，进而发现分析师和投资者均没有意识到基本面指标的价值。常丹婷和李峰（2020）构建了价值因子、基本面两个维度进行横截面分析，发现只有当估值和基本面预期背离、存在错误定价时，高低估值的对冲组合才能产生显著的超额收益，高达 16.8%。

目前国内有关基本面分析的研究较少，但在成熟的国外资本市场，基本面分析的作用已被大量文献证实。Ou 和 Penman（1989）构建了 68 组基本面指标，发现其能够成功预测未来股票收益涨跌的概率，多空组合在两年内达到 12.5% 的超

额收益。Abarbanell 和 Bushee (1997) 构建了 12 组基本面指标, 研究发现大部分指标能够预测股票未来价值, Abarbanell 和 Bushee (1998) 进一步利用这些基本面指标构建投资组合, 获得了 13.2% 的年化超额收益, 并发现超额收益大部分与未来的业绩公告相关。Piotroski (2000) 对高账面市值比的公司进行了研究, 并用 9 组基本面指标构建出综合指标 FSCORE, 用 FSCORE 挑选出真正的价值股并提升了 7.5% 的收益率。同样, Mohanram (2005) 着眼于低账面市值比的成长股, 利用 8 组基本面指标构建了综合指标 GSCORE。Asness 等 (2019) 根据盈利性、成长性、安全性, 构建了衡量公司质量的 QMJ (quality minus junk) 指标, 发现高质量公司股价较高, 但低质量公司不定, 多头高质量公司空头低质量公司的投资组合信息比率大于 1。

上述研究均根据基本面指标构建了综合指标, 并且构建对应的多空组合获取了超额收益, 证明了基本面分析的有效性。但国内有关基本面分析的研究较少, 运用大数据统计方法的更少, 所以本文将通过构建错误定价因子 M 来进行我国 A 股市场基本面分析可行性的研究。

3 研究设计

3.1 数据

3.1.1 数据样本

本文使用的公司财务报表数据和股票交易数据来自国泰安（CSMAR）数据库，因子模型数据分别来自中央财经大学金融学院、BetaPlus 小组以及 WHUFT 异常因子数据集。由于我国 A 股上市公司的季报数据自 2002 年起才开始公布，为了保障数据的完整性，本文财务指标数据以 2003 年第一季度为数据的起点；而因子模型数据始于 2003 年 10 月。综上，本文数据样本时间区间为 2003 年 12 月至 2020 年 9 月，频次为季度。

对于股票样本的选择，本文以我国 A 股市场 2018 年底上市的公司作为数据样本，规避未来收益数据的缺失。由于 ST 股票存在退市风险、金融行业公司财务报表结构有别于上市公司等原因，本文剔除掉 ST 股票、金融类股票，最后剩余 3300 支。有关数据缺失部分，若某支股票在第 t 期某一指标数据存在缺失，则剔除该股票在 t 期的所有数据。

在完成以上筛选步骤后，2003 年 12 月至 2020 年 9 月的有效样本共 119779 条。图 3.1 展示了 2003 年 12 月至 2020 年 9 月季频月度有效样本量。总体来看，月度样本量随年份呈现上升趋势，由 2003 年 12 月的 872 条有效样本增至 2020 年 9 月的 2616 条。

3.2 实证方法

3.2.1 变量定义与说明

为了计算公司的错误定价因子 M ，首先需要对公司的内在价值进行估计。本文以公司市场价值为因变量，财务指标为自变量进行回归计算，得到的残差即为错误定价；之后通过 Fama-MacBeth 截面回归、因子模型回归等方法验证错误定价因子 M 的有效性。文章中采用的主要变量如表 3.1 所示：

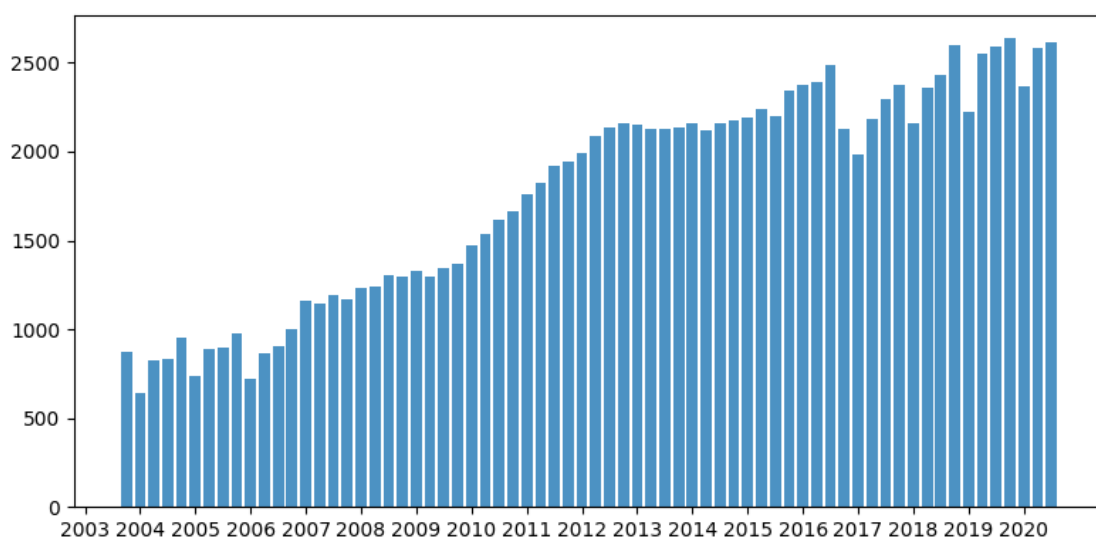


图 3.1 样本数据的月度有效样本量

表 3.1 变量定义与说明

变量名	定义及说明
$mktcap_{it}$	公司 i 于季度 t 时的市场价值
$M_{i,t}$	公司 i 于季度 t 时的错误定价因子
$R_{i,t}$	公司 i 于季度 t 时的月个股收益率
BM	市净率
β	市场投资组合 β 值
$grltnoa$	净经营性资产增长率
PA	毛利率
EP	市盈率
$lagretn$	上一个月的收益率
$mom12$	t-12 月至 t-2 月的累计收益率（共计 11 个月）
$mom36$	t-36 月至 t-13 月累计收益率（共计 24 个月）

3.2.2 模型设定

现有研究中，常见的公司内在价值计算方法包括现金流贴现法、相对价值法、经济附加值法、实物期权法等。这些方法虽然能起到一定的预测作用，但由于其选择的指标过少、过于死板，或太过于依赖使用者对未来的主观预测结果，亦或计算方法过于程式化等原因，很难排除掉数据窥视带来的影响。

根据本文的基本面分析原理，公司内在价值反映在一系列财务指标中，如式 3.1 所示，其中 i 代表不同公司， t 代表不同时间节点。 $mktcap_{i,t}$ 为公司 i 于季度 t 时的市场价值， $f(\cdot)$ 定义了一个参数为 θ 的函数，在本文中为采用机器学习方法中的函数形式， $I_{i,t} = (I_{i,t,1}, I_{i,t,2}, \dots, I_{i,t,N})$ 代表公司 i 在第 t 期时 N 个财务指标向量， $\epsilon_{i,t}$ 为残差项。本文将在后文中介绍采用的机器学习算法。

$$mktcap_{i,t} = f(I_{i,t}; \theta) + \epsilon_{i,t} \quad (3.1)$$

对于财务指标的选择，为避免数据窥视的影响，并非由笔者主观挑选。相反，本文于 CSMAR 数据库下载了所有财务报表数据，包括资产负债表、利润表与直接法和间接法计算的现金流量表，共计 262 个指标。但由于不同公司个体财务情况差异大，过于细分的指标数据未被使用，或是数据库数据录入缺失等原因，为了计算的严谨性，将对 262 个财务指标进行筛选。

基于统计，在所有财务指标中，缺失值比例最大达到 93.6%，最小为 0.01%，均值高达 41.6%，说明有很多指标数据缺失情况严重，故选择缺失值比例小于等于 5% 的所有指标，一共有 51 个。

在确定财务指标后，本文将采用随机森林（Random Forest，简称 RF）和梯度提升树（Gradient Boosting Decision Tree，简称 GBDT）方法进行数据模拟，确定具体函数形式 $f(\cdot)$ 。所有机器学习模型参数的选择均采用网格调参（Grid Search）的方式。由于本文需要逐期进行模型参数的选择，所以在获取每一期的最优参数后，进行式 3.1 的回归，以确定该期的错误定价情况。

$$M_{i,t} = -1 \times \frac{\epsilon_{i,t}}{mktcap_{i,t}} \quad (3.2)$$

回归后残差即为公司市值与内在价值的偏差，将该残差基于市值进行标准化后的负值，定义为错误定价因子 M ，如式 3.2 所示。当公司内在价值低于市场价值时，即残差大于 0，对应 M 为负值，说明股票被高估；当公司内在价值高于市场价值时， M 为正值，说明股票被低估。

其中，考虑到市场以及公司的变化情况，公司的内在价值以及错误定价因子 M 每期都会重新计算一次，对公司进行重新分组调整，避免数据偏误。

总而言之，本文的统计方法不会偏好特定的股票市场、不考虑整个市场在给定时点是否被高估或低估、也不依赖于内在价值的理论模型。相反，本文将公司进行相互比较，使用拟合优度的统计标准来识别公司内在价值如何体现于会计属性。

后续研究中，每期都将根据错误定价因子 M 大小进行公司分组：Q1 到 Q5 五组。其中 Q1 为 M 最小的公司，即股票最被高估；Q5 为 M 最大的公司，即股票最被低估。如果基本面分析有效、财务指标蕴含公司的内在价值，当市场价值最终趋近于内在价值时，最被高估的股票价格会回落，最被低估的股票价格会爬升，以此构建低买高卖的投资组合来获利。

为验证上述猜想，后文将进行 Fama-MacBeth 横截面回归、因子模型回归以及稳健性分析，如式 3.3、3.4 所示。

$$R_{i,t+1} = a_t + b_t M_{i,t} + \sum_{k=1}^K c_{k,t} X_{i,k,t} + \epsilon_{i,t+1} \quad (3.3)$$

$$R_{i,t} = \alpha_i + \sum_{k=1}^K \beta_{i,k} F_{k,t} + \epsilon_{i,t} \quad (3.4)$$

式 3.3 中 $X_{i,k,t}$ 代表公司 i 于季度 t 时第 k 个公司特征，作为控制变量，具体特征见表 3.1 所示；若 b_t 显著，说明错误定价因子 M 能够有效预测股票价格未来走势，证明了其有效性。

式 3.4 中 $F_{k,t}$ 为季度 t 时第 k 个因子值，本文主要采用 Fama-French 三因子模型、Carhart 四因子模型、Fama-French 五因子模型、以及 Hou-Xue-Zhang 四因子模型（后称为 q -因子模型）进行分析；其中 α_i 代表该投资组合的超额收益。如果 Q1 到 Q5 五组超额收益 α 有显著的递增趋势，或是 Q1、Q5 的多空组合超额收益 α 显著大于 0，也能证明错误定价因子 M 的有效性。

3.2.3 描述性统计

表 3.2、3.3 列出了使用随机森林和 GBDT 两种方法计算的相关变量的描述性统计，这里将样本根据错误定价因子 M 分成了 Q1 到 Q5 五组进行考察。其中第一列为所有数据的平均值，第二列为错误定价因子 M 与相应变量的相关系数，第三列为 Q1 组（价值最被高估）的变量平均值，最后一列为 Q5 组（价值最被低估）的变量平均值。

表 3.2 描述性统计 (随机森林方法)

	所有数据	相关系数	错误定价因子 M 分组				
			Q1 (被高估)	Q2	Q3	Q4	Q5 (被低估)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
M	0.2215	1.000	-0.2709	-0.0812	0.0930	0.3301	1.0379
$mktcap$	7.9267	-0.094	13.9565	11.2976	7.4287	4.4057	2.5323
R_t	0.8405	-0.083	2.9935	1.2586	0.4508	-0.0584	-0.4461
R_{t+1}	0.7851	0.039	0.3790	0.5373	0.7598	0.9499	1.3007
$lagretn$	0.0316	-0.035	0.0485	0.0319	0.0280	0.0245	0.0251
BM	0.4498	0.169	0.2908	0.4131	0.4939	0.5333	0.5232
EP	0.0207	0.019	0.0154	0.0217	0.0225	0.0221	0.0217
$beta$	1.1422	0.008	1.1238	1.1396	1.1452	1.1525	1.1508
$grltnoa$	1.7915	0.027	0.3476	0.2675	0.4031	0.2278	8.4563
PA	0.0141	-0.027	0.0175	0.0152	0.0126	0.0116	0.0134

表 3.3 描述性统计 (GBDT 方法)

	所有数据	相关系数	错误定价因子 M 分组				
			Q1 (被高估)	Q2	Q3	Q4	Q5 (被低估)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
M	0.2950	1.000	-0.3103	-0.0609	0.1415	0.4320	1.2745
$mktcap$	7.9267	-0.090	9.6631	14.5515	10.1442	3.4469	1.8191
R_t	0.8405	-0.094	2.8204	1.2833	0.5060	0.0062	-0.4172
R_{t+1}	0.7851	0.036	0.2922	0.5722	0.7975	0.9624	1.3025
$lagretn$	0.0316	-0.030	0.0464	0.0303	0.0248	0.0261	0.0304
BM	0.4498	0.094	0.3057	0.4464	0.5233	0.5142	0.4614
EP	0.0207	-0.043	0.0158	0.0238	0.0256	0.0204	0.0176
$beta$	1.1422	-0.005	1.1293	1.1331	1.1364	1.1523	1.1619
$grltnoa$	1.7915	0.008	0.3489	0.4555	0.1699	2.4593	6.3972
PA	0.0141	-0.058	0.0172	0.0155	0.0137	0.0111	0.0128

在两个描述性统计表格中，从第三列到最后一列，第 $t-1$ 期到第 t 期的收益 R_t 逐渐减少，说明购买被高估的股票收益更高，与现实情况相符，同时第 t 期前一个月的收益 lagretn 也是逐渐减少的趋势。但是，第 t 期到第 $t+1$ 期的收益 R_{t+1} 逐渐增加，说明被高估的股票价格下降，逐渐趋于正常，所以收益变低，而被低估的股票价值上涨到正常价值，收益变高，与本文猜想保持一致。

相比于被高估的 Q1 组股票，被低估的 Q5 组股票平均有着更高的市净率 BM 与市盈率 EP，同样有着更低的市场价值 mktcap ；再加上错误定价因子 M 与市净率 BM 有着正相关关系，说明被高估的股票更多是高市值的价值股，而被低估的股票多是低市值的成长股。

错误定价因子 M 与 β 相关系数低，说明系统性风险并不能解释 M 对于股票未来收益的预测能力。由于 M 与其他异象的相关系数基本都低于 0.05，在后续回归中，将 β 值、过去一个月的收益 lagretn 、市净率 BM、市盈率 EP、净经营性资产增长率 grltnoa 以及毛利率 PA 作为公司层面的控制变量。由于公司市值 mktcap 与错误定价因子 M 高度相关，在后续回归中为避免多重共线性问题，控制变量将不包含公司市值 mktcap 。

4 实证分析

4.1 Fama-MacBeth 截面回归

表 4.1 展现了对于所有公司 Fama-MacBeth 的时间序列截面回归结果，对于两种方法，自左向右三列分别选用了不同的解释变量来表现错误定价因子 M 的预测能力，所有列均控制了行业固定效应，括号内的数字代表 t 值。

其中，第一列仅采用了 M 作为解释变量，结果非常显著；第二、三列新加入了部分公司特征： β 值、市净率 BM 以及三个动量因子 $lagretn$ 、 $mom12$ 、 $mom36$ ；第四、五列加入了更多公司特征变量：市盈率 EP 、净经营性资产增长率 $grltnoa$ 与毛利率 PA ，让回归更为完整，错误定价因子 M 同样显著，说明错误定价因子 M 可以用来预测股票未来收益。

4.2 因子模型

除了对错误定价因子 M 进行 Fama-MacBeth 的截面回归外，本文将依据错误定价因子 M 预测构建 $Q1$ 和 $Q5$ 的多空组合，以月度收益率和风险调节收益 α 作为绩效衡量指标。由于 A 股市场做空机制的限制，本文还构建了 $Q1$ 到 $Q5$ 五组的多头组合，如果从 $Q1$ 组到 $Q5$ 组、从最被高估到最被低估的股票能够呈现超额收益递增的趋势，同样能证明错误定价因子 M 的有效性。

对于因子的选择，使用了常见的 CAPM 模型、Fama-French 三因子模型、Carhart 四因子模型、Fama-French 五因子模型、以及 Hou-Xue-Zhang 四因子模型（后称为 q -因子模型），采用了其中的 6 组因子（CAPM、FF3、FFC、FF5、FF5+MOM、 Q ）进行研究，具体结果见表 4.2、4.3 所示，记录了 $Q1$ 到 $Q5$ 五组多头组合和多空组合的超额收益 α 及 t 值。

可以看出无论使用哪种方法或模型， $Q1$ 到 $Q5$ 的多头组合超额收益 α 均呈现递增趋势，并且非常显著； $Q1$ 、 $Q5$ 的多空组合也有显著的正收益，证明了错误定价因子 M 的作用，以及基本面分析的有效性。

表 4.1 Fama-MacBeth 截面回归

	随机森林			GBDT		
	(1) R_{t+1}	(2) R_{t+1}	(3) R_{t+1}	(4) R_{t+1}	(5) R_{t+1}	(6) R_{t+1}
M	3.028*** (2.79)	4.177*** (3.26)	13.91** (2.55)	1.414*** (2.86)	4.453*** (3.44)	14.94*** (4.56)
β		1.725 (0.68)	15.34*** (2.70)		4.867* (1.94)	21.70*** (7.80)
BM		6.057*** (2.99)	19.30*** (5.28)		10.43* (1.84)	7.151 (0.70)
$lagretn$		-0.788 (-0.57)	-1.129 (-0.08)		2.828* (1.72)	-9.212** (-2.17)
$mom12$		0.825 (0.52)	3.051 (0.65)		-1.855 (-1.13)	10.14*** (2.96)
$mom36$		-0.748 (-0.31)	-0.719 (-0.18)		1.637 (0.72)	2.604 (0.80)
EP			45.38 (0.30)			321.9*** (4.29)
$grltnoa$			-45.67 (-0.51)			9.984 (1.14)
PA			66.95 (0.48)			-161.7** (-2.30)
行业固定效应	控制	控制	控制	控制	控制	控制
N	116527	84462	40542	116527	84462	40542
R^2	0.0600	0.398	0.636	0.0606	0.397	0.638

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

表 4.2 因子模型回归 (随机森林)

模型	系数	Q1	Q2	Q3	Q4	Q5	Q5-Q1
CAPM	mean	0.189	0.347	0.570	0.760	1.111	0.927
	α	0.139	0.274	0.489	0.662	1.007	1.065
	t 值	(1.41)	(2.83)	(5.21)	(6.92)	(7.34)	(2.46)
FF3	α	0.243	0.375	0.598	0.769	1.133	1.065
	t 值	(2.50)	(3.92)	(6.47)	(8.15)	(8.33)	(2.55)
FFC	α	0.420	0.520	0.748	0.910	1.282	0.835
	t 值	(4.33)	(5.43)	(8.09)	(9.64)	(9.39)	(2.40)
FF5	α	-1.304	-1.229	-0.940	-0.870	-0.338	1.036
	t 值	(-11.02)	(-10.58)	(-8.38)	(-7.60)	(-2.03)	(2.38)
FF5+mom	α	-1.113	-1.067	-0.772	-0.712	-0.176	1.006
	t 值	(-9.45)	(-9.21)	(-6.90)	(-6.24)	(-1.06)	(2.30)
Q	α	-0.661	-0.761	-0.657	-0.634	-0.055	0.633
	t 值	(-5.28)	(-6.20)	(-5.55)	(-5.25)	(-0.32)	(1.39)

表 4.3 因子模型回归 (GBDT)

模型	系数	Q1	Q2	Q3	Q4	Q5	Q5-Q1
CAPM	mean	0.102	0.382	0.607	0.772	1.112	1.014
	α	0.052	0.314	0.515	0.676	1.013	0.979
	t 值	(0.54)	(3.29)	(5.48)	(7.00)	(7.34)	(2.78)
FF3	α	0.158	0.407	0.611	0.791	1.151	0.969
	t 值	(1.64)	(4.30)	(6.58)	(8.32)	(8.43)	(2.90)
FFC	α	0.333	0.534	0.751	0.942	1.318	0.957
	t 值	(3.47)	(5.64)	(8.08)	(9.91)	(9.64)	(2.81)
FF5	α	-1.323	-1.246	-0.942	-0.905	-0.267	1.134
	t 值	(-11.27)	(-10.87)	(-8.35)	(-7.84)	(-1.59)	(2.69)
FF5+mom	α	-1.134	-1.098	-0.783	-0.737	-0.087	1.123
	t 值	(-9.71)	(-9.60)	(-6.96)	(-6.40)	(-0.52)	(2.63)
Q	α	-0.704	-0.808	-0.669	-0.631	0.042	0.769
	t 值	(-5.68)	(-6.66)	(-5.61)	(-5.18)	(0.24)	(1.76)

4.3 稳健性分析

上述小节 4.1、4.2 的结果证明了错误定价因子 M 以及基本面分析的有效性，公司内在价值能够完全反映于财务指标，并且市场价值会逐渐趋向内在价值。为此，本文进行以下稳健性分析。

在当前时点 t 计算各公司的错误定价因子 M ，并分为 Q1 到 Q5 五组，构建 Q1 与 Q5 两组的多空组合，基于该投资组合计算未来 36 个季度（9 年）的收益情况，同上文采用 6 组因子（CAPM、FF3、FFC、FF5、FF5+MOM、Q）进行研究，结果如图 4.1、4.2 所示。

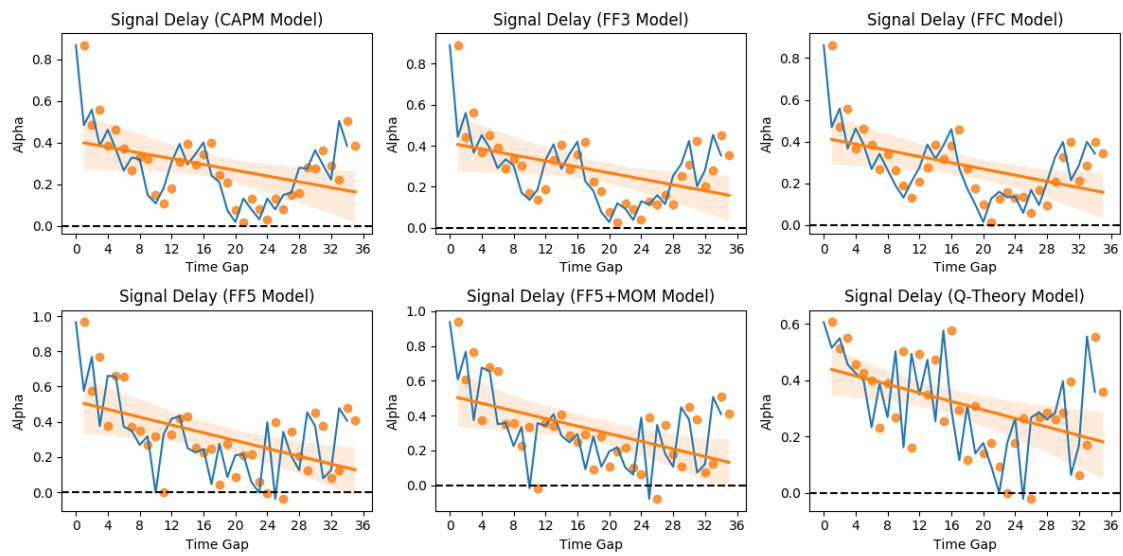


图 4.1 五组模型 Q5-Q1 结果（随机森林）

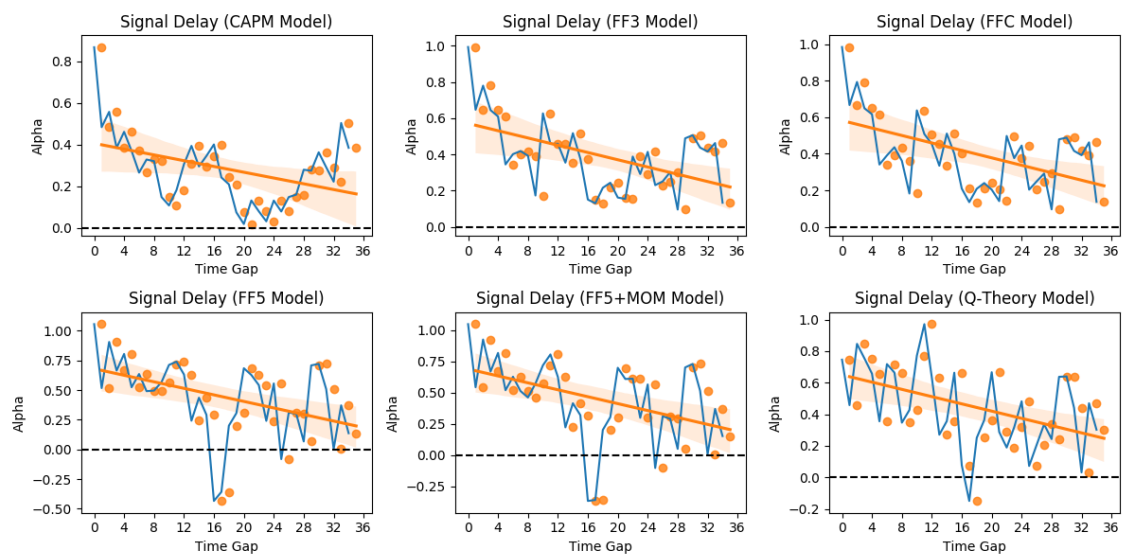


图 4.2 五组模型 Q5-Q1 结果（GBDT）

在五组模型中，多空组合收益均在当前时点 t 拥有最高的超额收益 α ，但后面随着时间增长，收益呈现明显的下降趋势，说明当前时点的错误定价已经随着时间而归正，股票的市场价格能够灵活随着信息进行调整，进而逐渐趋近于内在价值，收益也就越来越小，证明了本文在每一期重新进行内在价值计算、根据错误定价因子 M 对公司进行分组的正确性。若采用过往时点的结果进行未来股票收益的分析，会产生一定的偏误，所以要做到信息的更新换代。

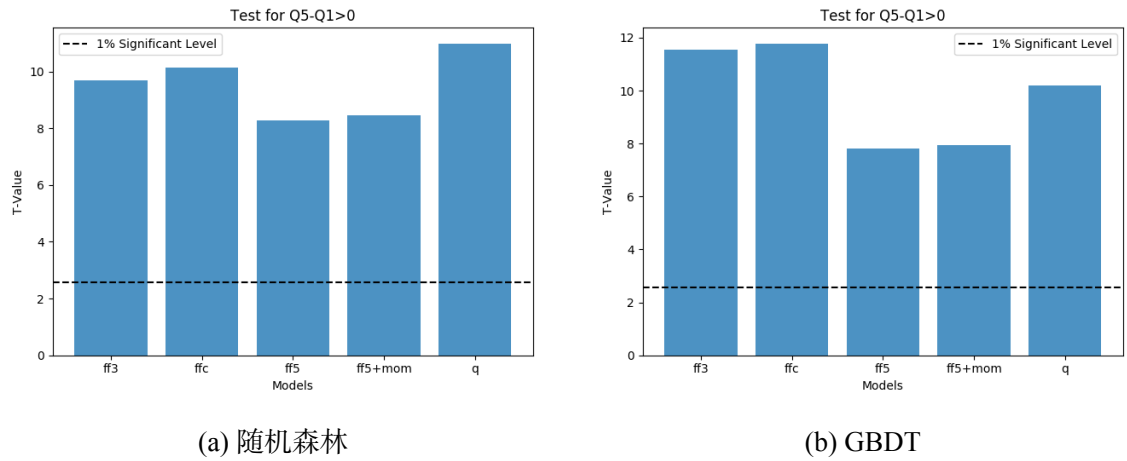


图 4.3 五组模型多空组合显著性检验

图 4.3 是对于五组模型中多空组合收益是否显著大于 0 的检验，所有模型结果均在 1% 的显著性水平下成立，证明了运用财务指标衡量公司内在价值的可行性、根据错误定价因子 M 进行公司分组的有效性。

5 结论与启示

本文基于 2003-2020 年的季度财务报表数据，选取 14 个缺失值比例低、共线性程度小的财务指标，以线性回归的形式完成了对公司内在价值的预测，并以估计的内在价值与市场价值的差异构建了错误定价因子 M ，当 M 为负值时，说明公司价值最被高估，当 M 为正值时，说明公司价值最被低估。根据错误定价因子 M 对公司进行分类，并进行相应的高卖低买策略，研究证明从最被高估到最被低估的投资组合，超额收益 α 呈现显著的递增关系，并且其多空组合能持续 5 年左右的正收益。同样也强调了新旧信息迭代的重要性，需经常性对数据进行重新调整与分析。以上结论在 Fama-MacBeth 截面回归和季度投资组合的五种因子模型检验中均显著符合预期，能够通过未来 28 期的稳健性测试，具有可观的经济意义。综上所述，本文的实证结果表明，基于季度财务报表的统计分析能够有效运用于中国 A 股市场，基本面分析起到了一定的预测作用。

基本面分析的有效性也反映出我国资本市场的效率低下，虽然我国 A 股市场总市值近 80 万亿元，已发展为全球第二大股票市场。但仍然由于成立时间晚、相关制度不完善、散户比例大、交易成本高等问题，导致了定价效率的低下。而市场效率的提高需要从多方面努力。

- (1) 投资者对于市场信息应仔细甄别，避免盲目跟风，根据自身的风险承担水平构建适合自己的投资组合；提高认知水平，学习相关财务知识，能够做到财务报表的简单阅读，市场基本面信息的理解。
- (2) 监管部门应加强建立与完善监管披露机制，让信息传递更具透明性与高效率。在现有科技社会，推进市场数据的信息化，利用大数据技术做到数据的及时披露，及时筛选并纠正市场上的错误信息，引导投资者理性投资；相反，该举措也能做到信息的向上反馈，让政策制定者更为清晰的了解到市场的实际运行情况，为国家宏观政策的制定提供依据。但同时也要做到对用户个人隐私的保护。
- (3) 上市公司需严格遵守会计规则，做到财务报表的及时、准确、公平披露，不得误导投资者，并做好内幕信息的知情人登记工作，减少投资者层面上的信息不对称。

参考文献

致谢

以简短的文字表达作者对完成论文和学业提供帮助的老师、同学、领导、同事及亲属的感激之情。