

学校编码: 10384

分类号_____密级_____

学号: X2013156318

UDC_____

厦门大学

硕士学位论文

基于财务与估值指标的面板 Logit 动态
选股模型研究

Research on Dynamic Stock Selection by Panel Logit Model
Based on Financial Data and Valuation Index

杨松

指导教师姓名: 李常青教授

专业名称: 工商管理(EMBA)

论文提交日期: 2015年6月

论文答辩日期: 2015年月

学位授予日期: 2015年月

答辩委员会主席:

评阅人:

2015年6月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。
2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士学位论文摘要库

摘要

本文通过主成分分析方法以及面板 Logit 模型建立了一个基于财务与估值指标的动态选股模型，使其既可为一个长期的选股策略被单独使用；亦可为事件型选股以及市场技术面选股提供一个先期的初选股票池，以提高量化选股的科学性和严谨性。针对财务数据和市场数据更新频率不同的难题，提供了一个两阶段的股票筛选方法；同时在进行个股投资时，也能为个股的财务基本面进行综合评分，以辅助投资者进行投资决策。

在模型构建的过程中，本文运用了更符合实际的数据处理方式。同时专注于有效信息挖掘，避免在因子规模上进行粗放式竞争，未将过多的精力浪费在增加一些仅仅为了可以淘汰的无用的因子。相反，而是论文力图用科学严谨的数量方法去挖掘每个重要财务因子的有效信息，为证券投资提供更有价值的参考。此外，在实证过程中对比多种选股策略收益，保障结果的稳定性。

经过相关的实证检验，作者认为：

首先，简单的因子排序模型效果较差，不具备实际的操作价值。对于企业财报数据而言，传统的因子排序模型因其简单、粗糙的处理方式，无法充分提取财务数据中的有效信息，并不适用作为选股投资策略的参考。

其次，本文所使用量化模型对于纯财务数据处理效果优秀，收益良好，具有实际的投资价值。

再次，本文量化模型具有良好的后期拓展性。在引入市场化因子后，可使基本模型收益有一个明显提高，同时风险有效降低，说明本文的量化模型具有较好的拓展性和潜力，可以对后续的选股策略提供精简、优质的企业样本。

最后，本文量化模型同样能运用于非周期行业的长期投资。通过基于企业年报的实证分析可知，对于非周期行业而言，本文的量化选股模型对于投资周期的波动具有极高的稳健性，对于以年为周期的投资策略，可以做到相近的年化收益率、超额年化收益率和胜率。

关键词：财务数据；估值指标；主成分分析；面板 Logit 模型；选股模型

厦门大学博硕士学位论文摘要库

Abstract

This paper established a dynamic stock selection model by a principal component analysis method and panel Logit Model based on pure financial data. It is a long-term stock selection strategy. Moreover, it can provide a basis for stock pool for the event type stock selection and market technical stock selection. At the same time, in view of the problems, which financial data and market data update frequency is different, it provides a two-stage stock filtering method. It also can give a rating of the company's financial data, to help investors make investment decisions.

In actual application process, this paper uses a more practical way of data processing. This article focuses on effective information mining. Avoid vulgar competition on the scale factor in order to save the efficiency. On the contrary, this article attempts to use scientific and rigorous method to explore the effective information for each of the key financial factors in order to provide more valuable reference for the securities investment. In addition, in the empirical process, we compared incomes of strategies; safeguard the stability of the results.

After the relevant empirical test, we get the following conclusion:

First of all, the effect of simple factor sorting model is poorer, does not have the actual operation value. For corporate earnings data, the traditional factor sorting model method is very simple and rough, thus it is unable to fully extract the useful information in the financial data, can not apply for stock investment strategy reference.

Secondly, for the pure financial data, effect of quantitative model of this paper is good, has the actual investment value. After the introduction of market factors, income of the basic model has an obvious improve, at the same time, risk is effectively decreased

Late again, in this paper, quantitative model has good expansibility. That means in this paper, the quantitative model has a good potential, is able to provide concise and high quality enterprise samples for subsequent strategy.

Finally, this paper quantitative model can also be applied to the long-term investment of non-cycle industry. The empirical analysis shows that for non-cycle industry, in this paper, the quantitative stock selection model has high robustness.

Key words: Financial data、 Valuation index、 Principal component analysis 、 panel Logit Model、 Stock selection model

厦门大学博硕士学位论文摘要库

目录

第一章导论	1
第一节研究背景	1
第二节研究目的	2
第三节研究意义	3
第四节文献综述	4
第五节研究内容和框架	8
第六节研究创新	10
第二章量化投资概述	12
第一节量化投资的基本概念	12
第二节量化投资的发展历程	13
第三节量化投资在中国发展的现状和机遇	14
第四节 Logit 模型概述	15
第三章传统多因子模型下的财务筛选模型验证.....	17
第一节数据说明	17
第二节数据处理	19
第三节传统多因子排序模型的选股方法：是优化还是徒劳	21
第四节本章小结	24
第四章基于面板 Logit 模型的动态选股模型验证.....	25
第一节模型选股策略流程	25
第二节基于纯财务数据的非周期行业组合收益	28
第三节基于纯财务数据的周期行业组合收益	33
第四节本章小结	36
第五章扩展模型：估值指标优化下的 Logit 模型分析.....	37
第一节估值指标优化组合收益：非周期行业	37
第二节估值指标优化组合收益：周期行业	40
第三节稳健性检验：基于年报数据的量化选股结果	42
第四节本章小结	43

第六章结论与建议	45
第一节结论	45
第二节模型的一些改进设想	46
第三节实务中运用多因子模型需要注意的问题	46
参考文献	48

厦门大学博硕士论文摘要库

第一章导论

第一节研究背景

随着中国市场经济进程的不断完善，我国资本市场建设得到了充分的发展。其中，股票市场作为资本市场的重要组成部分，对我国经济起着积极的推动作用。自 1990 年上交所成立起至今，我国证券市场已经历了 25 年的发展，在 25 年的探索实践中，我国的证券市场在国内改革的春风中，制度不断完善，内涵不断丰富，在国际资本市场的快速发展中有着举足轻重的影响作用。不过，相对于欧美发达市场而言，我国证券市场的整体发展仍处于初级阶段，仍具有极大的发展潜力。在这一基础上，对于各类证券投资策略方法的分析和实践，能有效地促进证券市场投资方式的多样化；不仅有助于评估证券投资价值，提高投资者投资决策的科学性，还有助于降低市场的整体风险，促进证券市场稳定，具有一定的研究价值。

从理论上来说，市场的证券价格受多种因素的共同影响，其中包括交易参与方的心理和行为状态以及国家的宏观经济等因素的改变，而上市公司披露的财务报表数据更是交易参与者判断该公司投资价值的重要依据，会直接影响该公司在市场上的定价。在分析市场的股票价格走势的预测方法中，这种基于上市公司财报信息的分析方法被称作基本面分析。它与关注股票价格历史行情，走势图形的技术面分析方法一起，是我国证券投资分析的两种主要流派。不过，无论是技术面分析还是基本面分析，都认为市场中证券价格的影响因素都是多元化的。而在形形色色庞杂的影响因素中，如何辨识整理并加以应用，转而为证券价值评估和建立投资策略提供依据，一直是国内外学者研究的热点，从资产选择理论到资本资产定价模型，进而到罗斯的套利定价模型，投资分析理论的范式从单一影响因素标准逐渐进化为多因素模型分析，这使得证券市场的投资策略更加理性，资源配置功能更加完善，带动了欧美资本市场的急速发展，也使得证券投资分析研究进入空前的繁荣时期。而各种证券投资分析模型的研发和投入实用，也为量化投资提供了施展拳脚的舞台以及广阔的发展前景。

量化分析不同于基本面分析和技术面分析，却可以将两者的优点相结合。通

过将技术面的图形趋势分析定量化,同时加上基本面数据本身属于定量分析范畴,量化分析可以将两种分析方法有效结合,达到证券投资分析过程中对于选股和择时的双重要求。在众多的量化投资模型中,多因子模型被多数研究认为是契合我国当前证券市场运行的一个经典模型,而其中 Fama 和 French 提出的三因子模型上更是成为国内券商和投资基金研究的热点问题,以此模型为基础进行延伸和拓展的各种多因子模型也正被广泛应用于各项量化金融产品之中。

第二节 研究目的

本文目的是构建一个纯财务的量化选股模型,即可作为一个长期的选股策略被单独使用;亦可为事件型选股以及市场技术面选股提供一个先期的初选股票池,以提高量化选股的科学性和严谨性。针对财务数据和市场数据更新频率不同的难题,提供了一个两阶段的股票筛选方法;同时在进行个股投资时,也能为个股的财务基本面进行综合评分,以辅助投资者进行投资决策。而本文以纯财务数据为基础,利用特定的量化方法主要是基于以下两个方面的考量。

首先,从本文的研究方法上来说,近几年量化投资策略在国内受到越来越多的关注。2004年,国内推出第一支量化基金“光大保德信量化核心基金”,2011年国泰君安资产管理公司的“君享量化”则标志着中国证券市场开始有了第一支量化对冲基金,随后,国泰君安资产管理公司陆续推出了另外的五只量化产品。在当年,这6只量化产品尽数排进了国内集合理财产品收益排行榜的前20位,良好的收益率激发了市场对量化产品的活跃度,也使得量化产品在近几年逐渐开辟了自己的天空,成为了各投资机构中不可或缺的投资产品种类。不过,相比发达国家量化投资历经40年的实践发展,我国量化投资仍处于起步阶段,总体上仍显稚嫩。这主要表现在两个方面:首先,从发行规模上来看,量化产品在市场上的占比仍然较低。据WIND数据统计,截止2014年底,国内市场共有104只量化基金,总规模为1358.24亿元,相比公募基金4.47万亿的总净值规模,量化基金仅占比3.04%。其次,从量化产品的原理和市场表现来看,现阶段国内量化产品原理简单,数据处理粗糙,同质化程度高,在遇到市场风格转变的时候缺乏灵活的应对,回撤较大。鉴于此,量化投资分析在国内证券市场上仍具有较大的发展空间,有继续深入研究的需求。

其次,从本文研究内容上来说,上市企业披露的财务数据是市场判断证券价格的主要信息来源,同时也是市场参与者进行证券投资分析的重要依据。股票的价值之所以存在,是因为股票代表股利的资本化,当持续的现金流折现后仍存在正的现值,股票才存在价值。上市公司披露的财务数据在上述过程中所起到的作用,就是改变市场参与者对公司未来持续现金流的预期,继而影响股票当前的价值。换句话说,上市公司披露的财务数据能影响人们对它未来运营状况的预期,未来运营状况直接影响了未来的股利水平,上市公司未来的股利水平预期影响现在的股价。就这样,上市公司当前披露的财务数据与其证券价格的变动紧密地联系在了一起。不过单一财务信息对证券价格波动的解释力度有限,难以发掘出上市公司财务指标与证券价格之间的内在联系,全面探讨各种财务数据对证券价格走势的影响有其必要性,鉴于财务数据的种类繁多,项目庞杂,这势必涉及到多因子数据的处理。

第三节 研究意义

量化分析在逐渐得到国内投资者青睐的同时,多因子模型也在国内资本市场的实务应用中得到了快速的发展。在国内资本市场有效性不断提高的基础上,上市公司披露的财务数据越来越完善,使得证券价格与企业基本面的联系愈发紧密。因此,如何利用统计学、金融学等原理,通过计算机的高效处理优势,大概率地捕捉上市公司的财务动态,识别上市公司的经营状况发生了哪些变化,成为各类投资者越来越关注的问题。因此,一个科学、系统的多因子量化模型在当前我国证券市场就显得有着一定的理论和现实意义。

1、理论意义

其一、本文模型是建立在传统的 Fama-French 模型基础上的一种拓展和优化,不仅可以通过模型的分析了解我国股票市场的价格动向,也有利于相关的多因子模型在我国资本市场进行后续的深入研究。

其二、本文的量化模型为验证我国资本市场中的财务信息有效性提供了良好的工具。除此之外,本文所应用的模型还具有较好的兼容性,不仅可以处理财务信息,形成先期股票池,还可以在后续分析中加入市场技术面信息以及量化的行为金融学因素,同时保证了不同种类的数据信息不会相互出现干扰。这为后续的

相关研究具有一定的借鉴意义。

其三、多因子模型是量化投资分析的基础模型，许多复杂量化模型都是以多因子模型为基础不断改善和拓展演化而来。可以说，多因子模型是量化分析的基础，本文所使用的量化模型是对现阶段常用的基础多因子模型的拓展和优化，在一定程度上拓展了该量化模型的外延和内涵，这对之后建立更加完备的量化分析理论体系具有一定的意义。

2、现实意义

多因子模型是当前我国证券市场运用最广泛的选股模型，研究这一模型及其拓展形式对于各类证券研究或投资机构更好地认识国内股市，提高投资者的整体收益都具有十分重要的价值。此外，量化模型的持续深入研究也可以促进投资理论体系的整体推进，有利于推动证券投资机构的发展，提高市场的有效性，对于增强市场的整体稳定性，降低投资风险有着一定的现实意义。

第四节文献综述

量化投资方法早在 70 年代就被实务界人士广泛运用于资本市场获利，其理论基础来源于 1952 年马科维茨 (Markowitz) 建立的均值方程理论和 1964 年夏普 (Sharpe) 等人提出的资本资产定价模型 (CAPM)。资本资产定价模型认为资产的系统性风险是唯一定价因子，系统性风险越大，要求获得的收益补偿就越多。进入 80 年代，学者发现系统性风险不是资产定价的唯一影响因素 (Ross, 1976; Fama 和 French, 1993; Carhart, 1997; Hou 等, 2012)，把更多因子纳入定价模型，丰富和发展了资产定价理论。

一、多因子模型

(一) 国外相关研究

随着理论和计算机技术的发展，学者发现股票收益不仅与市场系统性风险相关，还与其他风险因子有关，研究发现会计信息、市盈率倒数等一系列指标都可以预测股票收益。因此到了 80 年代，实务界已经不再局限于资本资产定价模型进行投资，把众多风险因子纳入模型，其中包括了行业因子和各项宏观因子。

因子选股模型已经成为目前量化投资策略中应用最为广泛的选股模型之一。优秀的因子能够满足如下条件：在时间维度上表现稳定，即因子对股票收益率影响是一致的；在不同市场和样本中有一定的区分度，能够解释个体之间的差异；有相同因子的股票其行为路径是相同的。

在各类多因子模型中，影响最大且运用最为广泛的是 Fama 和 French(1993) 提出的三因子模型。基于跨期资本资产模型和套利定价模型，他们通过分组统计检验股票收益率差异，对公司共同特征进行筛选，实证检验发现市场因子、市净率和公司规模可以解释公司大部分的收益率，从而构建得到了三因子模型。这个多因子均衡定价模型可以表示为：

$$E(R_{it}) - R_{ft} = \beta_i[E(R_{mt}) - R_{ft}] + S_i E(SMB_t) + h_i E(HML_t)$$

其中 R_{ft} 表示时间 t 的无风险收益率， R_{mt} 表示时间 t 的市场收益率； R_{it} 表示资产 i 在时间 t 的收益率； $E(R_{mt}) - R_{ft}$ 是市场风险溢价， SMB_t 为时间 t 的市值 (Size) 因子的模拟组合收益率， HML_t 为时间 t 的账面市值比 (book-to-market) 因子的模拟组合收益率。 β_i 、 S_i 、 h_i 分别表示三个因子的系数。

Fama-French 三因子模型最大的缺陷是它完全来自于统计检验得到的结果，缺乏理论基础。但这个缺陷对实证结果影响不大，只要新的因子与市场因子相关性较低，并且新的因子能够解释剩余的收益率变动，那么这个模型就能得到跨期资本资产模型和套利定价模型的支持。

Carhart (1997) 对三因子模型进行了改进，把市场中流行的动量因子纳入模型，构成了 4 因子模型，模型对股票收益率的预测能力高于 Fama 和 French 的三因子模型。而 Hou、Xue 和 Zhang (2012) 把动量因子和价值因子进行成替换投资因子和 ROE 因子，他们实证结果发现该模型预测能力优于 Carhart 的四因子模型。在 Hou 等 (2012) 研究基础上，Fama 和 French (2014) 最新的研究把盈利能力和公司投资模型因子纳入三因子模型，得到了五因子模型。

(二) 国内相关研究

国内学者对多因子模型的研究工作主要是应用国外成熟的模型检验在中国资本市场的有效性，如吴世农、许年行 (2004) 采用 A 股上市公司数据，实证检验了 CAPM 和 Fama-French 三因子模型有效性，他们发现价值因子和规模因

子在中国同样具有较强的预测效果，三因子模型的解释能力优于 CAPM 模型。沈浩（2009）采用价值成长（GARP）量化投资方法在中国股市进行模拟投资，实证结果显示该量化方法在我国 A 股市场 1996 年至 2008 年期间也有着较好的表现；于栋华（2009）采用加拿大皇家银行（RBC）的选股模型在 2007 年至 2009 年对 A 股市场进行的实证检验表明，该量化模型可以取得较为稳定的超额收益。

还有学者在经典模型基础上进行改进和完善，如王源昌等（2010）针对我国资本市场特征，在 Fama-French 三因子模型基础上加入了市盈率和换手率，实证结果表明改进的模型预测能力优于原模型。王茵田和朱英姿（2011）采用 Fama-French 模型筛选因子的方法，从众多因子中提取了八个表现较为优秀的因子，包括了期限利差、投资资本比、工业增加值变化率、账面市值比、盈利股价比、现金流股价比、市场风险溢价和回购利率，他们的模型对我国 A 股市场风险溢价的解释能力更强。

二、财务指标选股模型

在多因子模型中，财务指标是投资者运用得最多的因子，其主要投资逻辑是：在资本市场上，股票价格反映了上市公司的会计盈利信息，而公司财务状况可以通过构建财务指标模型进行量化打分，因而投资者投资那些盈利预期最好的公司可以获得超额收益。

（一）国外相关研究

自二十世纪六十年代，鲍尔（Ball）和布朗（Brown）开创性地对会计盈利问题与股票收益率关系进行研究以来，会计信息与资本市场领域已经成为会计学研究的重点。他们研究发现投资于有超额盈利能力公司的投资者可以取得超额收益，这说明股票价格反映了上市公司的会计盈利信息。

至此之后，大量学者从财务指标入手，选取财务健康的公司构建投资组合，检验组合是否具有超额收益。在众多研究文献中，Piotroski 教授在 2002 年发表的论文对投资界影响深远，2008 年美国个人投资者协会（AAII）跟踪的 56 个选股策略中只有 Piotroski 教授选股策略的收益率为正，当年收益率高达 32.6%。

在此之前，投资界奉行采用 PB 指标构建反转组合，即选择那些低 PB 公司

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.