



中南财经政法大学学术文库

多元统计

在组合投资中的应用研究

APPLIED RESEARCH OF MULTIVARIABLE STATISTICS IN PORTFOLIO INVESTMENT

吕盛鸽 著

 中国财政经济出版社

文库

中南财经政法大学学术文库

多元统计在组合投资中的应用研究

吕盛鸽 著

中国财政经济出版社

图书在版编目(CIP)数据

多元统计在组合投资中的应用研究/吕盛鸽著. —北京:中国财政经济出版社, 2002. 8

(中南财经政法大学学术文库)

ISBN 7-5005-6017-6

I. 多... II. 吕... III. 多元分析—应用—股票—证券投资—研究—中国 IV. F832.51

中国版本图书馆 CIP 数据核字(2002)第 062894 号

中国财政经济出版社出版

URL: <http://www.cfeph.com>

E-mail: cfeph@drc.gov.cn

(版权所有 翻印必究)

社址:北京海淀区阜成路甲 28 号 邮政编码:100036

发行处电话:(010)88190406 财经书店电话:(010)64033436

湖北南财文化发展有限公司电话:(027)88391585 88391589

信阳长城印刷厂印刷 各地新华书店经销

850×1168 毫米 32 开 7.5 印张 181 千字

2002 年 8 月第 1 版 2002 年 8 月信阳第 1 次印刷

印数:1—1500 册 定价:15.50 元

ISBN 7-5005-6017-6/F·5269

(图书出现印装问题,南财公司负责调换)

中南财经政法大学学术文库

编辑委员会

主 任:吴汉东

副主任:郭道扬 张中华 赵凌云 覃有土

委 员:刘可风 卢现祥 熊胜绪 杨灿明

范忠信 罗 飞 朱新蓉 陈池波

齐文远 张新国 杨云彦 夏成才

姚 莉 陈景良 杨宗辉 朱延福

主 编:赵凌云

编辑部成员:姚 莉 陈敦贤

刘普生 朱冬生

序

一个没有思想活动和缺乏学术氛围的大学校园，哪怕它在物质上再美丽、再现代，在精神上也是荒凉、冷清和贫瘠的。欧洲历史上最早的大学就是源于学术。大学与学术的关联不仅体现在字面上，更重要的是，思想与学术，可谓大学的生命力与活力之源。

我校是一所学术气氛浓郁的财经政法高等学府。范文澜、嵇文甫、潘梓年、马哲民等一代学术宗师播撒的学术火种，五十多年来一代代薪火相传。因此，在世纪之交，在合并组建新校而揭开学校发展新的历史篇章的时候，学校确立“学术兴校，科研强校”的发展战略。这不仅是对学校五十多年学术文化与学术传统的历史性传承，而且将成为谱写新世纪学校发展新篇章的战略性手笔。

“学术兴校，科研强校”的“兴”与“强”，是奋斗目标，更是奋斗过程。我们是目的论与过程论的统一论者。我们将对宏伟目标的追求过程寓于脚踏实地的奋斗过程之中。由学校斥资资助出版《中南财经政法大学学术文库》，就是学校采取的具体举措之一。

本文库的指导思想或学术旨趣,首先在于推出学术精品。通过资助出版学术精品,形成精品学术成果的园地,培育精品意识和精品氛围,提高学术成果的质量和水平,为繁荣国家财经、政法、管理以及人文科学研究,解决党和国家面临的重大经济、社会问题,作出我校应有的贡献。其次,培养学术队伍,特别是通过对一批处在“成长期”的中青年学术骨干的成果予以资助推出,促进学术梯队的建设,提高学术队伍的实力与水平。第三,培育学术特色。通过资助在学术思想、学术方法以及学术见解等方面有独到和创新之处的成果,培育科研特色,力争通过努力,形成有我校特色的学术流派与学术思想体系。因此,本文库重点面向中青年,重点面向精品,重点面向原创性学术专著。

春华秋实。让我们共同来精心耕种文库这块学术园地,让学术果实挂满枝头,让思想之花满园飘香。



2001年11月28日

前 言

在股票市场里有成百万的投资者、商人和投机家,他们做出成百上千万个决策,有些决策是独立于他人做出的,有些是彼此相互影响做出的,这些决策形成一个合力左右着股票价格。我们可以把投资人的决策看成是成百上千万个偶然因素,这些偶然因素无时不影响着股票价格,因而股票价格具有不确定性。针对股价的不确定性,本书引进概率统计中随机向量概念,用多元统计方法分析和揭示中国股票市场的内在规律,构建适合机构投资者操作的证券组合投资模型。

本书在第二、三和四章中,介绍了如何用多元统计方法分析我国证券市场的一般规律(由于涉及到的多元统计方法可以使用国内外通用的 SPSS 和 SAS 软件去实现,书中不再附计算程序);第五章根据我国开放式基金随时面临投资者赎回要求这一特点,研究了如何用聚类分析和主成分分析优选流动性强且具有投资价值的股票,建立适合开放式基金管理人操作的组合投资方法;第六章根据我国 A 股市场存在着暂时不能流通的国有股和法人股这一

国情,用聚类分析、判别分析、因子分析和回归分析建立了《A股平均股价流通量模型》,并用建立的模型研究国有股减持定价方法;第七章介绍了如何用《A股平均股价流通量模型》寻找预期收益率的股票,用概率统计方法建立了以较大概率达到预期收益率下的证券组合投资模型,并由极值原理确定组合投资比例分配方法;第八章介绍了逐步回归分析,并用逐步回归分析法建立了收益预测模型。

本书的主要内容是作者近几年发表的六篇有关多元统计、证券组合投资论文的核心内容以及在这些论文基础上进一步的研究成果,这六篇论文是:《聚类判别法》,《统计研究》1993年第四期;《A股平均价格对流通量双曲线模型》,《统计研究》2000年增刊;《预期收益率下的证券组合投资比例分配》,《数量经济技术经济研究》2001年第9期;《国有股减持定价方式探讨》,《中南财经大学学报》2001年第3期;《分层抽样单位数的一种配置方法》,《统计与决策》2000年第1期;《新股申购收益率统计分析》,《统计与决策》2001年增刊。用多元统计分析方法研究中国股市的内在规律,还处在尝试阶段,再加上笔者学识所限,书中可能有不少疏漏乃至错误之处,诚恳地欢迎读者批评指正。

吕盛鸽

2002年6月于武汉

内 容 提 要

在股票市场里有成百万的投资者、商人和投机家,他们做出成百上千万个决策,有些决策是独立于他人做出的,有些是彼此相互影响做出的,这些决策形成一个合力左右着股票价格。我们可以把投资人的决策看成是成百上千万个偶然因素,这些偶然因素无时不影响着股票价格,因而股票价格具有不确定性。针对这种不确定性,本书引进概率统计中随机向量概念,用多元统计方法分析和揭示中国股票市场的内在规律,构建适合机构投资者操作的证券组合投资模型。本书共有八章,各章的主要内容如下:

为了便于后面各章的学习,第一章中介绍了本书涉及到的矩阵知识以及多维随机向量概念。矩阵方面的内容多数未给出证明,仅证明了实对称矩阵谱分解式及特征根极值性质。随机向量部分介绍了多维随机向量的分布、密度函数、数字特征矩阵等概

念,给出了数字特征矩阵的估计方法,证明了随机向量协方差矩阵、相关系数矩阵的性质。

在经济活动中,众多经济指标(视为随机向量)无时不在变化,常常需要寻找这些经济指标变动合力引起的变化比较大的几个方向,主成分分析能够做到这一点,它是一种把多个指标化为少数几个综合性指标的统计方法。第二章中,给出了主成分的推导过程以及主成分的方差、方差贡献和累积方差贡献率等概念,并且给出了主成分的几何解释、常用的主成分分析方法,最后介绍了如何根据股票收盘价数据,用主成分分析方法筛选活跃的股票。

因子分析是研究如何从多个指标中提炼出公共因素的多元统计方法,它是通过建立因子模型来实现的。到目前为止,因子分析的模型和理论还不完善,但它还是在包括经济学在内的许多学科领域中取得了成功应用。在第三章中介绍了因子模型、载荷矩阵的统计意义、方差最大正交旋转、因子得分等概念,以及用观察数据进行因子分析的两种方法。介绍了如何用因子分析研究股价之间的联动关系、对股份公司的分类、分析上证(或深圳)综合指数的走势等证券投资方面的问题。

聚类方法是数理统计中研究“物以类聚”的一种统计方法。数值分类方面有两大类问题:一类问题是不存在事前分类的情况下,而进行数据结构的分类,这是聚类分析所要解决的问题。另一类是已知研究对象的分类情况,将某些未知个体归属其中某一类,这是判别分析所要解决的问题。聚类分析的方法有很多,第四章仅研究了目前国内外应用最广泛的系统聚类分析法,介绍了如何规定样本(或变量)点之间的距离以及类与类之间距离、系统聚类分析法的基本思想、八种系统聚类方法,并以最短距离法为例,说明了系统聚类分析法的一般步骤以及如何分析聚类图。判别分析的内容十分丰富,第四章仅介绍了距离判别和费歇(Fisher)判别,以

及判别函数的两种检验方法(F 检验法、统计拟合检验法)。

我国开放式证券投资基金的启动,给证券组合投资提出了新的要求,因为开放式基金随时面临投资者的赎回,为了满足投资者的赎回要求,基金管理人应该寻找资金进出方便且具有投资价值的证券进行组合投资。在第五章中,用第四章的系统聚类分析和第二章的主成分分析,研究如何从市场中筛选具有投资价值且资金进出方便的股票进行组合投资,并且给出了一种组合投资比例的确定方法。最后通过实证研究,从沪市的某 53 只股票中筛选出 10 只适合开放式基金组合投资的股票。

我国上市公司 A 股价格受诸多因素影响,其中重要的因素有两个:一是流通量的大小,二是上市公司盈利的增长率。在第六章中,用第四章的聚类分析和判别分析、第三章的因子分析,整理沪、深股市 1100 家上市公司的有关资料,然后用回归分析方法建立了《平均股价流通量模型》,该模型揭示了我国证券市场平均股价与流通量之间的关系。我们知道国有股减持的关键之一在于定价,而定价的根本问题在于寻找国家股股东与流通股股东利益的均衡点。减持某只股票的国有股时,用《平均股价流通量模型》可以找到一种定价方式,这种定价方式可以确定国家股股东与流通股股东利益的均衡点,它既能最大限度地收回国有资产的升值效应,又不会对股票二级市场形成冲击。

从经济学中我们知道,价格总是围绕价值作上下波动的,证券市场同样遵循这条规律。在第七章中我们介绍了如何用《平均股价流通量模型》寻找价值被低估的股票,用概率统计方法建立了以较大概率达到预期收益率下的证券组合投资模型,并由极值原理确定组合投资比例分配方法。通过实证研究,从 2000 年每股收益在 0.210—0.254 元之间的 26 只股票中,找到了 7 只可能达到预期收益率的股票进行组合投资,确定了组合投资比例,并计算出能

达到预期收益率的概率。

在实际问题中,往往碰到影响因变量 Y 的因素有许多,而这些因素之间存在着多重共线性,特别是在经济和贸易的观察数据中,各自变量之间有高度的相互依赖性,如果用这些数据建立回归方程,就会给回归系数估计值带来不合理的解释。为了得到一个稳健的可靠的回归模型,就应该把所有对因变量 Y 影响显著的变量选入模型,剔除对 Y 影响不显著的。第八章介绍了如何利用观察数据,从大量可共选择的变量中,选择对回归方程重要变量的方法——逐步回归分析法,并用该方法从上市公司众多财务指标中,筛选出对收益影响大的,建立了平均收益预测模型。

本书的主要内容是作者近几年发表的六篇有关多元统计、证券组合投资论文的核心内容,这些论文是:《聚类判别法》,《统计研究》1993 年第四期;《A 股平均价格对流通量双曲线模型》,《统计研究》2000 年增刊;《预期收益率下的证券组合投资比例分配》,《数量经济技术经济研究》2001 年第 9 期;《国有股减持定价方式探讨》,《中南财经大学学报》2001 年第 3 期;《分层抽样单位数的一种配置方法》,《统计与决策》2000 年第 1 期;《新股申购收益率统计分析》,《统计与决策》2001 年增刊。用多元统计分析方法研究中国股市的内在规律,还处在尝试阶段,还有很多问题值得研究,例如,可以考虑用典型相关分析来优选上市公司,用主成分分析法分配投资于各板块的资金比例,用聚类分析、对比分析来分析上市公司财务指标之间的关系等等,作者将与同行对这些问题进行深入探讨。

Applied research of multivariable statistics in portfolio investment

Abstract

Millions of investors, businessmen and speculator make millions of decisions in the stock market, which could be independent decisions or interactive decisions so as to form a general power of fluctuating stock prices. We can regard investors' decisions as millions of episodic factors with never ending effect to stock prices, so uncertainty we have to get in stock prices. Specially, we introduce a concept of stochastic vector in probability & statistics and multivariable statistical methods to analyze and explore the inner rules of China stock market, and to build an applicable securities investment model readily for instructional investors. There are eight chapters and main contents as follows respectively:

In order to our get going easily of the following study,

chapter 1 includes referenced matrix knowledge and multidimensional vector concept. Few proof to knowledge pertaining to matrix but only to general decomposition of real symmetric matrix and the properties of eigen-root limit value. In the part of stochastic vector concepts of multidimensional vector distribution, probability distribution function, numerical eigenmatrix, etc are included and the estimate method of numerical eigenmatrix is also given, then properties of covariance matrix of random vector and correlation coefficient matrix proves.

In the economic activities lots of economic indexes will never end changing and we are always in need of seeking the major directions caused by the general power of these economic indexes. Principal component analysis is ready for it, which transforms a large scale of indexes into several synthetic indexes. In chapter 2, deductive process of principal component analysis and concepts of variance of principal component, variance contribution and cumulative variance contribution are given. And its geometric explanation and pop method of principal component analysis are also given, how to sieve and get volatile stocks with the method of principal component analysis according to the data of stock end price.

Factor analysis is a multivariable statistical method with regard to how to get a public factor from several indexes with the factor model. Till now, though neither the factor analysis model nor its theory is perfect, it does matter much in many other subjects and fields, surely economics included. In chapter 3, concepts of factor model, statistical meaning of loading ma-

trix, maximum orthogonal rotation of variance, factor score and the two methods to get factor analysis going with data are introduced, research about the mutual relation among stock — prices with factor analysis, stock—company divisions and tendency of the synthetic index from Shang SE. or Shen SE. and something the like also included.

Cluster analysis is a statistical method to perceive a problem of "kinds of the same gather together" in mathematical statistics. There are two kinds of problems in numerical divisions; one is data—structure—division under the condition of no pre—division, which is that the cluster analysis should do; the other is to put unknown individuals into some kind under the condition of determined divisions, which is what the differentiate analysis should do. Among lots of methods of cluster analysis, chapter 4 only gives the most pop systematic method of cluster analysis in the world for the moment. It introduces how to get the distance between samples (or variables) and between types, the basal theory of systematic cluster analysis, eight systematic method of cluster analysis, the general steps of systematic method of cluster analysis with the example of method of shortest distance and how to analyze cluster diagram. Chapter 4 also introduces distance differentiation and Fisher test and the two estimate methods of differentiate function. (F test, Statistical fit test)

With the launch of open securities — investment fund in China, new problem we have to face about securities portfolio investment 'cz we will be ready to face the redemption from investors. So a fund manager should seek securities easily turn-

over and worthy to invest to make investors satiable. In chapter 5 with the systematic cluster analysis of chapter 4 and principal component analysis of chapter 2, we perceive how to get the valuable securities with easily passing in — and — out capital to undergo portfolio investment and we also give a method to get a portfolio investment proportion. In practice, we can get ten proper stocks out of random — selected 53 stocks from Shang SE.

In China, prices of A stock are interactive with many factors, of which two are the most important: one is flow, the other is the growth rate of stock — company's profit. In chapter 6, with the cluster analysis of chapter 4 and factor analysis of chapter 3 and relevant materials of 1100 firms in Shang&Shen SE., we get "Model of average stock — price flow" with regression analysis method, showing us the relationship between the average price of our securities and the flow. As we all know, one of the keys of less — holding national stocks lies on how to price, and the inner problem is to seek the profit balance between national stock shareholders and flow stock shareholders. Then we can get a pricing method with "Model of average stock price flow" as we want to hold less the national part of some stock. This method can fix the profit balance between those two types of shareholders, which can get back the effect of the national capital at the maximum level and never crash the second stock market.

It is well known that price fluctuates around value and it also means something in the securities market. In chapter 7 we

talk about how to seek the down—estimated—valued stocks with "Model of average stock price flow", that is to say, we build a securities portfolio investment model with regard to larger probability below the expected yield rate and we can ascertain the proportional distribution method of the portfolio investment with the extremum theory. In practice, from 26 stocks with return per share between 0.210—0.254 Yuan in 2000 we only get 7 stocks which are beyond expected yield rate and then fix on the portfolio proportion. Finally, probability of the gettable expected yield rate can be right computed.

In practice, we may come across many factors make dependent variable Y volatile and multilinear lies among those factors, especially among data from economy and trade. According to the highly mutual dependence among the variables, we will get a non—proper explanation toward regression coefficient estimate value if we put those data into use to build a regression equation. To get a robust and reliable regression model, we need to choose those typical variables to Y but not those non—typical. In chapter 8 stepwise regression analysis method can help us choose those important and typical variables from piles of selectable variables. Still, we can select what matters much to yield rate form piles of fiscal indexes to build an average yield forecasting model.

This book holds the quintessent contents from the six essays relating to multivariable statistics and securities portfolio investment published in recent years and those essays as follows are: "Cluster differentiate analysis", "Statistical research"