



高等院校金融学专业系列教材



金融数据分析技术

(基于Excel和Matlab)

元如林 李广明 关莉莉 罗远 主编

赠送
电子课件

清华大学出版社

高等院校金融学专业系列教材

金融数据分析技术 (基于 Excel 和 Matlab)

元如林 李广明 关莉莉 罗 远 主 编

清华大学出版社
北 京

内 容 简 介

本书主要针对金融领域中的问题,介绍如何通过建立数学模型,并运用 Matlab、Excel 等软件工具进行计算的金融数据分析技术,通过金融行业的实际案例,全面介绍数据整理、模型建立、参数确定、计算处理、结果分析的完整过程,并给出详细的上机实验指导,帮助读者亲身体验,以便读者更好地掌握数据分析技术。

本书的主要内容包括金融数据库的基本概念,国内外常用金融数据库、Matlab 和 Excel 等金融数据分析软件工具的使用、金融时间序列分析、金融风险价值计算、资产组合计算、金融衍生品定价计算、固定收益证券计算、信用评分与行为评分等。

本书可以作为应用型高等院校的金融学、金融信息、金融工程、金融数学等专业本科生的教材,也可作为需要金融数据分析技术的其他各专业本、专科生的教材。本书的特点是对每一个金融问题,首先简单地介绍相关金融知识,力求每章自成体系,因此特别适合数学、统计、信息、计算机等非金融类专业的读者。对数据分析技术感兴趣的其他读者,也可将本书作为参考书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

金融数据分析技术(基于 Excel 和 Matlab)/元如林等主编. —北京:清华大学出版社,2016
(高等院校金融学专业系列教材)

ISBN 978-7-302-43525-9

I. ①金… II. ①元… III. ①金融—数据—分析—高等学校—教材 IV. ①F830.41

中国版本图书馆 CIP 数据核字(2016)第 081042 号

责任编辑:孟 攀

装帧设计:刘孝琼

责任校对:王 晖

责任印制:李红英

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62791865

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×230mm

印 张:22.75

字 数:496 千字

版 次:2016 年 6 月第 1 版

印 次:2016 年 6 月第 1 次印刷

印 数:1~2000

定 价:48.00 元

产品编号:067449-01

前 言

随着我国金融信息化的不断推进和金融市场的快速发展,银行、保险公司、证券交易所、证券公司、基金公司、期货交易所、黄金交易所、金融期货交易所等各类金融机构每天都产生大量的金融数据。最近几年,我国互联网金融蓬勃发展,第三方支付、P2P 网贷、众筹融资、大数据金融和金融信息服务等互联网金融企业每天也产生大量的金融数据,这些金融数据如同一座含有丰富信息和知识宝藏的矿山,等待我们去发掘。随着大数据、云计算、移动支付、数据科学、智慧金融等概念和技术的普及,使得人们越来越重视金融数据及其价值。如何从海量的数据中挖掘出有价值的新信息,并发现帮助企业创造价值的新知识,是我国金融行业中信息技术与金融业务深度融合发展面临的主要课题,也是我国金融行业提高国际竞争力的关键。目前我国的金融企业和互联网金融企业都急需大量能够综合运用数学理论、信息技术并精通金融业务的金融数据分析人才和金融数据挖掘人才。

本书主要针对金融领域中的问题,介绍如何通过建立数学模型,并运用 Matlab、Excel 等软件工具进行计算的金融数据分析技术,通过金融行业的实际案例,全面介绍数据整理、模型建立、参数确定、计算处理、结果分析的完整过程,并给出详细的上机实验指导,帮助读者自己亲自体验,以便读者更好地掌握金融数据分析技术。希望本书的出版能为培养金融数据分析人才做出一点贡献,同时为大数据时代各行各业需要数据分析技术的人员提供参考。

本书共分为 8 章。第 1 章主要介绍金融数据库的基本概念,国内外常用金融数据库;第 2 章主要介绍 Matlab、Excel 等金融数据分析软件工具的使用方法;第 3 章主要介绍金融时间序列分析;第 4 章主要介绍金融风险价值计算;第 5 章主要介绍资产组合计算;第 6 章主要介绍金融衍生品定价计算;第 7 章主要介绍固定收益证券计算;第 8 章主要介绍信用评分与行为评分。

目前我国出版的金融计算方面的教材大多只针对已经掌握金融知识的读者,重点介绍如何使用 Excel、SAS、Matlab 等软件进行计算,这类教材对于数学、统计、信息、计算机等非金融类专业的读者,需要花费大量时间补充金融知识。本书的特点是对每一个金融问题,首先简单明了地介绍相关金融知识,力求每章自成体系,不仅方便金融类专业的读者使用,更方便非金融类专业的读者使用。本书的另一个特点是通过金融行业的实际案例,全面介绍数据整理、模型建立、参数确定、计算处理、结果分析的完整过程,并给出详细的上机实验指导,帮助读者亲身体验。

本书的第 1、2、4、6 章以及第 3 章的部分内容由元如林编写,第 3 章的部分内容和第

7 章由李广明编写,第 5 章由罗远编写,第 8 章由关莉莉编写,全书的统稿和 Matlab 计算的内容由元如林完成。

在本书的编写过程中,我们参考了许多经济学和金融学的书籍,特别参考了许多应用数学软件如 Matlab、SAS、SPSS、Excel 等进行金融计算的书籍,还参阅了网上相关内容,也得到许多领导和同事的关心和帮助,在此一并向他们表示衷心的感谢!

由于编者水平有限,特别是本书的内容涉及多学科交叉,疏漏、不足和错误之处在所难免,恳请读者批评指正。

本书获得了中央与地方共建上海金融学院金融信息团队建设项目和上海金融学院教学质量工程(特色教材)项目的资助。

编 者

目 录

第 1 章 金融数据库	1	2.1.2 国产金融数据分析软件介绍 ...	35
1.1 金融数据库的概念	1	2.1.3 金融数据分析软件的选择	36
1.1.1 金融数据库的定义	1	2.2 Matlab 及其金融工具箱	37
1.1.2 金融数据库的起源	1	2.2.1 Matlab 简介	37
1.1.3 金融数据库的作用	2	2.2.2 Matlab 金融工具箱简介	39
1.1.4 金融数据库的分类	4	2.2.3 Matlab 在金融领域的应用	39
1.1.5 金融数据库的选择标准	4	2.3 Matlab 的基础知识	40
1.2 国内外常用金融数据库简介	5	2.3.1 Matlab 的系统开发环境	40
1.2.1 国外金融数据库的概况	5	2.3.2 矩阵及其运算	42
1.2.2 国内金融数据库概况	8	2.3.3 数组及其运算	49
1.2.3 选择合适的金融数据库	12	2.3.4 Matlab 中的常用数学函数	51
1.2.4 免费数据资源的获取渠道	12	2.3.5 图形绘制	55
1.3 锐思数据(RESSET/DB)使用简介	13	2.3.6 编写 M 脚本文件	68
1.3.1 RESSET/DB 的访问途径	13	2.3.7 自定义函数	73
1.3.2 用户类别以及相应的权限	15	2.4 实验二: 金融数据分析软件	
1.3.3 数据查询与下载	15	使用实验	74
1.4 实验一: 金融数据下载实验	27	2.4.1 实验目的	74
1.4.1 实验目的	27	2.4.2 实验原理	74
1.4.2 实验原理	28	2.4.3 实验内容	75
1.4.3 实验内容	28	2.4.4 实验步骤	75
1.4.4 实验步骤	28	2.4.5 实验报告要求	75
1.4.5 实验报告要求	29	本章小结	75
本章小结	30	思考讨论题	76
思考讨论题	31	第 3 章 金融时间序列分析	77
第 2 章 数据分析软件工具	32	3.1 金融时间序列	78
2.1 金融数据分析软件工具简介	32	3.1.1 金融时间序列的概念	78
2.1.1 国外主要金融数据分析		3.1.2 金融时间序列的构成因素	80
软件简介	32	3.1.3 金融时间序列分析	80

3.1.4 金融时间序列的建立.....81	4.1.5 模型的评价方法.....130
3.2 确定性时间序列分析.....82	4.2 使用 Excel 计算风险价值 VaR 的 案例.....130
3.2.1 长期趋势 T_t82	4.2.1 在 Excel 中用参数法的 直接法计算风险价值 VaR.....131
3.2.2 循环变动 C_t82	4.2.2 在 Excel 中用参数法的移动 平均法计算风险价值(VaR).....134
3.2.3 季节变动 S_t85	
3.2.4 确定性时间序列分析小结.....87	4.3 使用 Matlab 软件计算 风险价值(VaR)的案例.....135
3.3 随机性时间序列分析.....88	4.3.1 数据描述.....135
3.3.1 平稳时间序列.....88	4.3.2 采用的模型和方法.....135
3.3.2 自回归移动平均模型.....89	4.3.3 计算结果.....140
3.3.3 模型识别与参数估计.....89	4.3.4 模型评价和比较.....156
3.3.4 金融时间序列的预测.....92	4.3.5 主要结论.....161
3.3.5 模型的检验.....93	4.4 实验四: 金融市场风险的 VaR 计算实验.....162
3.3.6 Matlab 的时间序列工具箱.....94	4.4.1 实验目的.....162
3.4 广义自回归条件异方差模型.....107	4.4.2 实验原理.....162
3.4.1 广义自回归条件异 方差模型.....107	4.4.3 实验内容.....162
3.4.2 GARCH 工具箱.....108	4.4.4 实验步骤.....163
3.5 实验三: 金融时间序列分析实验.....118	4.4.5 实验报告要求.....164
3.5.1 实验目的.....118	本章小结.....164
3.5.2 实验原理.....118	思考讨论题.....165
3.5.3 实验内容.....118	
3.5.4 实验步骤.....119	
3.5.5 实验报告要求.....119	
本章小结.....119	
思考讨论题.....120	
第 4 章 金融风险价值的计算.....121	第 5 章 资产组合的计算.....166
4.1 金融风险价值 VaR 模型.....121	5.1 资产组合基本原理.....166
4.1.1 金融市场风险概述.....122	5.1.1 收益序列与价格序列间的 转换.....167
4.1.2 金融市场风险的 度量与管理.....122	5.1.2 协方差矩阵与相关系数 矩阵间的转换.....169
4.1.3 VaR 模型.....123	5.1.3 资产组合收益率与方差.....174
4.1.4 风险价值 VaR 的计算方法.....125	5.2 资产组合的有效前沿.....176
	5.2.1 两种风险资产组合收益 期望与方差.....176

5.2.2 均值方差的有效前沿	177	6.3.2 欧式期权价格函数	229
5.2.3 带约束条件的资产组合的 有效前沿	178	6.3.3 期货期权定价	231
5.3 用 Excel 进行资产组合计算的 案例	180	6.3.4 隐含波动率	232
5.4 用 Matlab 进行资产组合计算的 案例	194	6.4 Black-Scholes 期权价格的 敏感性分析	233
5.4.1 投资组合常用函数	194	6.5 期权定价的二叉树法	237
5.4.2 投资组合的有效前沿	203	6.5.1 二叉树期权定价模型	237
5.4.3 投资组合的最优资产分配	206	6.5.2 二叉树定价函数	239
5.5 实验五: 投资组合分析计算实验	210	6.6 投资组合套期保值策略	241
5.5.1 实验目的	210	6.6.1 套期保值的基本原理	242
5.5.2 实验原理	210	6.6.2 利用保护性看跌期权策略 进行套期保值	242
5.5.3 实验内容	211	6.6.3 利用期权敏感性参数 进行套期保值	246
5.5.4 实验步骤	212	6.7 实验六: 金融衍生品定价 计算实验	248
5.5.5 实验报告要求	212	6.7.1 实验目的	248
本章小结	212	6.7.2 实验原理	248
思考讨论题	213	6.7.3 实验内容	248
第 6 章 金融衍生品的计算	214	6.7.4 实验步骤	249
6.1 金融衍生品	214	6.7.5 实验报告要求	249
6.1.1 金融衍生品的基本概念	214	本章小结	249
6.1.2 金融衍生品的种类	215	思考讨论题	250
6.1.3 金融衍生品的功能	216	第 7 章 固定收益证券计算	251
6.1.4 金融衍生品的风险管理	217	7.1 固定收益证券的基本概念	251
6.1.5 我国金融衍生品市场的 发展现状	219	7.1.1 固定收益证券	251
6.2 期权	220	7.1.2 美国固定收益证券的种类	253
6.2.1 期权的概念	220	7.1.3 固定收益证券的定价	254
6.2.2 期权的分类	221	7.1.4 固定收益证券的 久期与凸性	258
6.2.3 股票期权的利润函数	221	7.1.5 利率的期限结构	259
6.3 Black-Scholes 期权定价模型	227		
6.3.1 Black-Scholes 方程	227		

7.2 用 Excel 进行固定收益证券分析案例	261	8.2.2 判别分析	284
7.3 用 Matlab 进行固定收益证券计算	266	8.2.3 回归分析	291
7.3.1 现值和终值的计算	266	8.2.4 分类树法	295
7.3.2 计算内部收益率	269	8.2.5 最邻近法	301
7.3.3 固定收益证券产品的定价	270	8.3 信用评分的非统计学方法	303
7.3.4 固定收益证券的 久期与凸性	273	8.3.1 线性规划	304
7.3.5 利率的期限结构	274	8.3.2 非线性规划——整数规划	308
7.4 实验七：固定收益证券计算实验	276	8.3.3 人工神经网络	309
7.4.1 实验目的	276	8.3.4 遗传算法	313
7.4.2 实验原理	277	8.4 行为评分模型及其应用	318
7.4.3 实验内容	277	8.4.1 行为评分简介	318
7.4.4 实验步骤	278	8.4.2 马尔可夫链方法	318
7.4.5 实验报告要求	279	8.4.3 贝叶斯—马尔可夫链方法	324
本章小结	279	8.5 案例	328
思考讨论题	279	8.6 实验八：个人信用综合评分实验	337
第 8 章 信用评分与行为评分	280	8.6.1 实验目的	337
8.1 信用评分与行为评分的基本概念	280	8.6.2 实验原理	337
8.1.1 信用卡与信用卡管理	280	8.6.3 实验内容	347
8.1.2 社会征信体系	281	8.6.4 实验指导	349
8.1.3 信用评分与行为评分	283	8.6.5 实验报告要求	353
8.2 建立信用评分卡的统计学方法	283	本章小结	353
8.2.1 信用评分的统计学 方法简介	283	思考讨论题	354
		参考文献	355

第1章 金融数据库



【学习要点及目标】

- 掌握金融数据库的概念。
- 了解金融数据库的起源、作用及分类。
- 了解国外金融数据库的概况和国内金融数据库的概况。
- 了解金融数据库的选择标准，会选择合适的金融数据库。



【核心概念】

数据 数据库 金融数据 金融数据库 金融数据库应用系统 金融数据库服务产业

1.1 金融数据库的概念

随着我国金融信息化的不断推进和金融市场的快速发展，银行、证券交易所、证券公司、基金公司、期货交易所、黄金交易所、金融期货交易所等各类金融机构每天都产生大量的金融数据，最近几年，我国互联网金融蓬勃发展，第三方支付、P2P网贷、众筹融资、大数据金融服务等互联网金融企业每天也产生大量的金融数据，这些金融数据如同一座含有丰富信息和知识宝藏的矿山，等待我们去发掘。随着大数据、云计算、移动支付、数据科学、智慧金融等概念和技术的普及，使得人们越来越重视金融数据及其价值。

1.1.1 金融数据库的定义

金融数据库就是运用金融理论和计算机数据库技术，将金融机构运营过程、金融市场交易过程和互联网金融企业中产生的各种金融数据进行采集、加工整理，并按一定格式存储便于共享的金融数据集合。它常常与基于该数据库的查询检索、统计分析、模型建立、计算处理等信息技术支持和服务一起，构成金融数据库应用系统，形成金融数据(信息)服务平台，为金融投资、金融研究、金融教学等提供金融数据和相关服务。

1.1.2 金融数据库的起源

随着现代金融学和金融市场的发展，无论是进行金融投资，还是进行金融学研究 and 教学，都离不开金融数据。对于金融机构和互联网金融企业的从业人员来说，金融数据是他们进行金融投资活动和开展研究的必要基础。对于金融数学、金融统计学、金融工程学、



金融信息学、计算金融学、实验金融学等学科领域,金融数据是一个重要的基础,研究者只有收集到全面、准确和完整的数据,才能进行有意义的经济与金融研究。

早期的实证研究者需要花费大量的时间自己进行数据收集整理,不仅影响了研究进度,而且常常因收集的数据不全面、也不一定准确,还会影响研究的结果。另一种情况是,有很多金融数据,普通研究者和普通研究机构难以搜集到,甚至根本无法搜集到,使相关研究无法进行。由于对金融数据的需求越来越迫切,人们越来越注重对金融数据的收集、整理和共享,以便减少重复劳动,降低金融数据收集的成本。于是,专门提供金融数据库的公司出现了,这是社会分工的结果。金融数据库和金融信息服务平台作为一种产品和一种服务,逐渐被人们所接受,并且诞生了一个新兴行业——金融数据服务业。

金融数据服务业是现代金融服务行业的基础和金融创新的源泉。在金融市场比较发达的国家,金融数据服务业的从业人数和资本投入规模都占整个金融服务业很大的一部分。金融数据服务行业在经营范围、安全性、保密性、可靠性和其他监管等方面都有严格的要求,其兼具金融和技术行业特点的性质,决定了其与其他信息服务和技术服务业的区别。由于金融数据和金融技术平台会同时用于银行、证券、保险和基金等不同的金融服务业,在中国金融行业分业经营分业监管的环境下,更显示出第三方独立的金融数据服务行业的必要性。按照目前互联网金融的分类,金融信息服务属于互联网金融六大主要类型之一,金融信息服务主要包括金融数据服务、信用信息服务、金融信息门户网等。我国的金融数据服务是互联网金融中发展最早且最成熟的类型,已有 20 多年的发展历史,不少金融数据服务企业已经是上市公司。

金融数据服务行业的主要业务范围是:对包括个人和企业数据在内的各种金融数据(包括一些用于评估企业和个人信用的非传统金融数据,如企业电子商务网站的浏览量、商品的关注时间、收藏率、交易量等数据,还有财经论坛、博客、微博、微信等文本数据)进行收集、汇总、标准化和发布;提供各种专业的数据定制服务;在原生金融数据的基础上发布相关指数和市场公共基础数据;搭建适用于金融及相关行业的信息技术平台,提供解决方案(包括金融交易系统、清算结算系统、风险控制系统、信息管理系统、数据整合/分析和报告平台、数据挖掘/处理和分析平台、建立和维护投资监控平台等);利用金融数据开展金融分析、资产定价和风险评估;在金融数据基础上设计金融产品和搭建金融模型;研究资产定量管理方法和金融产品量化交易策略;提供与金融数据相关的咨询和外包服务等。

1.1.3 金融数据库的作用

金融数据库是为金融投资者、金融研究者提供的金融信息服务的基础,在金融投资活动和研究、金融实证研究和金融教学中都有十分重要的作用。

1. 有效节约研究者搜集数据、整理数据的时间以提高研究效率

通过金融(研究)数据库提取所需数据可节省金融研究者收集数据、验证数据与计算数据的宝贵时间,提高研究效率。据统计,在没有完整、准确的金融数据库情况下,研究者一

半以上的时间都花在收集、整理数据上,有时甚至会花费80%的时间用于金融数据的收集、整理,不仅效率低且易出错。并且,有很多金融数据,普通研究者和普通研究机构难以搜集到。有了金融数据库后,研究人员可以直接得到研究所需的金融数据,无需花费大量的时间进行相关金融数据的搜集整理,而且专门的金融数据库提供商,能够提供更为全面、更为完整和更为准确的高质量的金融数据,使学术研究尤其是实证研究的效率显著提高,成果明显增加。

2. 降低研究者和研究机构的数据获取成本

与分别从不同的金融数据源机构购买数据比较而言,直接购买金融数据库,可大幅节省研究者,特别是研究机构的数据获取的资金成本,而且并不是所有的金融数据源机构都愿意出售零星小批量的数据。

3. 可以为研究者提供各种专业的数据定制服务

金融研究者不仅需要原始的金融数据,有时也进行数据合并、变换、计算衍生指标等,甚至需要一些特别定制的金融数据,大量数据的整理计算的难度往往很大,需要一些技术含量高的信息处理技术,这些信息处理技术的学习也需要大量的时间投入,而提供金融数据库的专业的金融信息服务企业也提供各种专业的数据定制服务,包括特殊需求的数据整理、模型建立和计算程序设计等服务,支持研究者的研究,让研究者可以专注于金融问题本身的研究。

4. 采用同一金融数据库有利于研究结果的比较

如果每个研究者都使用自己搜集的数据,相关指标的计算处理标准不统一,其研究结果很难进行比较。而采用同一金融数据库,可使标准统一,有利于研究结果的比较。

5. 为金融学教学提供辅助工具并帮助培养急需的相关人才

金融数据库为金融学教学提供了很好的辅助工具,教师可运用金融数据库的数据对各种理论、模型进行演示,使学生能够比较直观地掌握深奥的理论知识和分析方法,从而提高教学效率与质量。运用金融数据库的专业数据,还可以对经典文献、经典模型进行方便的检验,在检验的过程中发现“偏差”和不同市场的特色,从而构建自己的模型,在前人研究的基础上进行新的尝试,拓展研究思路,培养学生的创新能力。

最近几年,大数据已成为大家耳熟能详的名词,金融大数据如同一座含有丰富信息和知识宝藏的矿山,有待去发掘。随着大数据、云计算、移动支付、数据科学、智慧金融等概念和技术的普及,使得人们越来越重视金融数据及其价值。如何从海量的数据中挖掘出有价值的新信息,并发现帮助企业创造价值的新知识,是我国金融行业中信息技术与金融业务深度融合发展面临的主要课题,也是我国金融行业提高国际竞争力的关键。金融数据库有助于培养我国的金融企业和互联网金融企业都急需的能够综合运用数学理论、信息技术并精通金融业务的金融数据分析人才和金融数据挖掘人才。

1.1.4 金融数据库的分类

目前,金融数据库已经发展到比较成熟的阶段,国内外市场上已有各种金融数据库。按照不同的标准,可将其划分为多种类型。

(1) 按金融数据库的内容和用途的不同,分为行情咨询数据库和研究型数据库。

行情咨询数据库主要提供金融市场实时交易行情等咨询类数据,重点在保证金融市场数据的实时性、直接性、准确性,不需要对数据进行加工整理。数据库主要为投资者服务,使用对象以各类投资者为主。

研究型数据库是为研究人员提供服务,它不仅提供原始数据,还为研究者提供一系列对原始数据按专业标准进行深层加工整理得到的各类常用数据指标,包含研究人员需要的更为全面的数据,很多研究性金融数据库提供商还同时提供宏观经济数据,方便研究人员使用。

有的金融数据库提供商同时提供金融市场实时交易行情(包括高频数据)等咨询类数据和研究型数据,形成完整的数据服务平台。

(2) 按金融数据的种类不同,分为单一金融数据库和综合金融数据库。

单一金融数据库主要提供单一种类的金融市场的相关数据,如股票数据库、外汇数据库、债券数据库等,甚至还可以是专门的钢铁交易数据库等。这类数据库虽然数据范围相对狭窄,但是通常更加专业和精细,可以提供更为多样的衍生指标和专题数据。

综合金融数据库通常提供多个金融市场的相关数据。很多投资者特别是机构投资者都在多个金融市场进行投资,研究者也要进行跨市场研究,综合金融数据库正适合了他们的需求。

金融数据库的综合化已经成为一种趋势,越来越多的金融数据库提供商在不断扩大数据范围,同时提供利率、汇率、通货膨胀率等宏观数据,以及各类行业数据,甚至法律信息和与金融有关的政治、社会和科技信息。

1.1.5 金融数据库的选择标准

评价和选择金融数据库的标准主要有以下几个。

- (1) 设计体系科学、合理。
- (2) 数据内容完整、准确。
- (3) 相关衍生指标计算正确。
- (4) 使用方便,服务完善。
- (5) 数据库的结构稳定。
- (6) 数据更新及时。

选择金融数据库时,还要考虑数据的下载方式、对二次开发的支持、与研究和教学的适合度等,选择一个适合自己用途的金融数据库。

1.2 国内外常用金融数据库简介

1.2.1 国外金融数据库的概况

国际著名的金融数据库有美国芝加哥大学商研所金融研究中心的 CRSP 数据库(The Center for Research in Security Prices)、美国著名的信用评级公司标准普尔(Standard & Poor's)的 Compustat 数据库以及纽约交易所 TAQ 数据库等。

1. CRSP 简介

CRSP(Center for Research in Security Prices, 证券价格研究中心)是美国芝加哥大学商研所金融研究中心的产品。CRSP 搜集了市场中最详尽的历史数据,由于它的准确性和权威性,自 1960 年以来,即成为学界及商界的主要数据来源。CRSP 收集的美国股票和指数数据库的来源主要为纽约证券交易所(NYSE)、美国证券交易所(AMEX)及纳斯达克证券交易所(Nasdaq)的上市公司的股票数据,包括每日交易的数据,如股票收盘价等历史数据、美国企业活动信息、企业沿革、并购及联盟状况、资本回收、现金流量等基本财务数据。收录年代情况:纽约证券交易指数,其资料始自 1925 年 12 月 31 日;美国证券交易指数;其资料始自 1962 年 7 月 2 日;纳斯达克证券交易指数,其月资料始自 1972 年 12 月 29 日,日资料始自 1972 年 12 月 14 日。

网址: <http://www.crsp.com/> 或 <http://www.crsp.uchicago.edu>。

2. Compustat

Compustat 数据库是美国著名的信用评级公司标准普尔(Standard & Poor's)的产品。数据库收录有全球 100 多个国家中的 5 万多家上市公司及北美地区公司的详细季度和年度财务报表与财务指标的历史数据,其中包括 7000 多家亚洲的上市公司,还提供 103 个国家的 24000 家公司的国际化信息,收录近 20 年上市公司的财务数据,另外还提供北美回溯版,即 400 家公司自 1950 年以来的财务资料;提供约 180 种模板报表及上市或非上市公司财务数据等信息;并且整合最新或历史性的主要财务数据,利用该数据库提供的 Research Insight 软件可以进行公司及公司财务、行业等分析,制作各种报表及动态图表。应用的研究领域包括资产分析、计量分析、竞争者分析、公司资本结构、财务比率、合并与购并、研究与开发(R&D)、资本及存货投资、股市报酬及资本市场效率等主题。

标准普尔的网站: <http://www.standardandpoors.com/>。

Compustat 网址: <http://www.compustat.com>。

3. NYSE TAQ

NYSE TAQ 是纽约证券交易所(NYSE)的交易和报价(TAQ)高频数据库。NYSE TAQ 数据集是“综合的”(Consolidated),它包括所有在纽约证券交易所、美国证券交易所(AMEX)和纳斯达克(NASDAQ)的全美市场系统(National Market System, NMS)及小盘股(Small Cap)板块挂牌证券的日内交易数据(交易和报价)。TAQ 高频数据覆盖自 1993 年至今,ISSM(另一高频数据库)的数据是从 1983 年至 1992 年。

TAQ 数据集的规模很大。在 2007 年 1 月,报价和交易数据集的大小分别是 121.7G 和 12.3G。如果加上指标数据集(综合报价数据集有 52.1G 的指标数据集,而综合交易数据集有 7.4G 的指标数据集),仅仅一个月的数据量就十分巨大。一般而言,NYSE TAQ 数据库每个月有 4 个数据集,即综合报价数据集、综合交易数据集、红利数据集和主文件数据集(Master File)。

4. PACAP Databases

PACAP Databases 是由美国罗德岛大学环太平洋资本市场研究中心(Pacific-Basin Capital Markets Research Center)建立、维护和发布的研究数据系统。PACAP 的数据库连续和系统地跟踪包括泰国、菲律宾、中国台湾、中国香港、新加坡、马来西亚、韩国、日本等 8 个环太平洋国家和地区的主要证券交易所的广泛信息:每日的股票价格和回报、资本分派、财务报表、市场指数和回报以及经济统计数据。PACAP 的数据库包含从 1975 年开始的历史数据的汇编,每年 9 月进行年度更新。为便于比较,所有国家都使用一个统一的格式。因此,PACAP 的数据库是一个极其宝贵的集学术研究、教学、投资组合管理和资本市场发展于一体的信息源,目前是环太平洋地区资本市场研究的首选数据库,使用其数据的论文广泛见于 Journal of finance(《金融杂志》或《财务杂志》)、Journal of Financial Economics(《金融经济学杂志》)等国际一流期刊。PACAP 多年以来一直以其高品质的数据库产品服务于世界一流大学的研究机构以及全球亚太地区资本市场的研究者。

网址: <http://www.cba.uri.edu/PACAP>。

5. Reuters

路透社(Reuters)是世界前三大多媒体新闻通讯社,有 2300 多名新闻编译人员,有 197 所分社在 130 个国家运行。路透社提供新闻报道和金融数据给报刊、电视台等各式媒体,并向来以迅速、准确享誉国际。另外,路透社提供金融产品及服务,如股票价格和外汇汇率,让交易员可以分析金融数据和管理交易风险;同时路透社的系统让客户可以经由互联网完成买卖,取代电话或是纽约证券交易所的买卖大厅等人工交易方式,它的电子交易服务串连了金融社群。路透社的金融信息服务分为 4 个部分:买卖与交易、研究与资产管理、企业服务和媒体服务,其中超过 90%的收入来自金融服务业务:对股票、外币汇率以及债券等资讯的分析、处理、发送及其相关产品的开发。

路透社提供的主要金融产品及服务有办公及会计支持系统、商品及能源、企业服务、交易决策、Elektron(电子)托管、新兴市场、股票及衍生工具、固定收益、外汇和货币市场、对冲基金解决方案、独立顾问、个人交易者/投资者、投资管理、交易后服务、定价和参考数据、定量研究及贸易、实时解决方案、风险管理、供应方的研究、财富管理等。

2008 年路透集团与汤姆森合并, 更名为汤姆森路透, 成为彭博新闻的主要对手, 两者的市场份额均为 1/3 左右。

网址: <http://thomsonreuters.com>。

6. Bloomberg

彭博通讯社(L.P. Bloomberg)是由迈克·卢本斯·布隆伯格于 1982 年创立的经济性媒体集团, 在短短的 20 多年时间里就超过有着 150 年历史的路透集团, 成为当时全球最大的金融信息服务提供商。彭博通讯社是服务于全球的信息和新闻媒体机构, 总部设在纽约, 全球共有 100 多家办事处、分社或分支机构。彭博新闻是以 1990 年开始的财经新闻发展起来的, 现在已在全球设立 87 个分社, 拥有 1500 名编辑和记者。现今的彭博通讯社提供了许多方式来了解其资讯, 包括电视频道、广播、网络和刊物。

在国际金融界, 彭博资讯的旗舰产品——“彭博专业服务”, 将新闻、数据、分析工具、多媒体报告和“直通式”处理系统前所未有地整合在单一的平台, 通过彭博终端(彭博机)为客户提供服务。彭博终端是一套让专业人士访问“彭博专业服务”(Bloomberg Professional Service)的计算机系统。用户通过“彭博专业服务”可以查阅和分析实时的金融市场数据以及进行金融交易。“彭博专业服务”的目标客户为金融证券投资领域的专业人士。彭博资讯终端不仅是有价值的信息渠道, 也象征着使用者的身份和地位。独立性、权威性、丰富性是彭博资讯的成功之道。大多数大型金融机构使用彭博社提供的服务, 每台彭博终端每个月的租金是 1500 美元。此外, 用户通过彭博终端访问来自交易所的实时报价时需要缴纳额外的费用, 这笔费用是支付给提供报价的交易所的。一些新闻机构也通过彭博终端为客户提供付费新闻。目前彭博社在全球约有 30 万台终端机, 其带来的销售收入占其总营业收入的 85%左右。

网址: <http://www.bloomberg.com/>。

7. WRDS

WRDS(Wharton Research Data Services, 沃顿研究数据服务)是由宾夕法尼亚大学沃顿商学院开发的金融领域的跨库研究工具, 它整合了 Compustat(标准普尔)、CRSP(芝加哥大学)、TFN(Thomson, 汤姆森公司)及 TAQ(纽约交易所)等著名数据库产品。同时, WRDS 也提供包括 CBOE(芝加哥期权交易所)等 10 个公开数据库。该平台是基于 Web 的整合型数据访问服务平台, 可使数据库检索更便利, 并提供验证数据正确性以及数据库的整合等功能。

网址: <https://wrds-web.wharton.upenn.edu/wrds/>。

8. Dow Jones Company

Dow Jones Company(道·琼斯公司)创立于 1882 年,是世界一流的商业财经信息提供商,同时也是重要的新闻媒体出版集团,总部设在美国纽约。道·琼斯公司旗下拥有报纸、杂志、通讯社、电台、电视台和互联网服务,在全球拥有近 1700 名新闻从业人员。道·琼斯编发的股票价格指数更是家喻户晓。2007 年 8 月 1 日,道·琼斯与新闻集团发布联合声明,宣布双方已签署合并协定,新闻集团以 56 亿美元收购道·琼斯。

道·琼斯指数(Dow Jones Index)由道·琼斯公司负责编制并发布,登载在其属下的《华尔街日报》上。历史上第一次公布道·琼斯指数是在 1884 年 7 月 3 日,当时的指数样本包括 11 种股票,由道·琼斯公司的创始人之一、《华尔街日报》首任编辑查尔斯·亨利·道(Charles Henry Dow)编制。1928 年 10 月 1 日起其样本股增加到 30 种并保持至今,但作为样本股的公司已经历过多次调整。道·琼斯指数是算术平均股价指数。道·琼斯 30 种工业股票平均价格指数于 1896 年 5 月 26 日问世。现在,道·琼斯编制发布 4000 多种指数,其中包括道·琼斯工业股票平均价格指数(即闻名于世的道·琼斯指数)、道·琼斯全球股票指数、道·琼斯互联网股票指数以及专门为中国市场编制的道中指数、道中 88 指数、道沪指数及道深指数等。

网址: <http://new.dowjones.com/>。

1.2.2 国内金融数据库概况

1. “新华 08”

“新华 08”于 2007 年 9 月 20 日正式运行,是新华社采用先进的信息与通信技术,自主研发的金融信息服务平台,以终端形式为经济管理部门、金融机构和大中型企业参与国内外债券、外汇、股票、黄金、期货和产权交易,提供交易前的信息收集和分析、交易中的订单递交和风险管理、交易后的清算结算和信息反馈服务,是将实时资讯、行情报价、历史数据、研究工具、分析模型和在线交易融为一体的金融信息综合服务系统。

该平台日发稿过万条,实现了多媒体发布;将国内股票、债券、外汇、期货、黄金、产权等实时行情整合在一个平台上,并实现同一界面展示;自主研发了 40 个金融模型,精确度超过路透社的系统,可以达到小数点后 8 位;建立了覆盖国内所有地市的价格监测采集网络,并在海外 23 个重点城市设立了价格采集点,成为中国大陆首家进驻华尔街进行金融信息采集的机构。“新华 08”的推出,标志着我国建立了主流、权威、系统的经济、金融信息发布渠道,必将大大提升我国在国际金融领域的话语权与竞争力,对我国金融市场发展具有里程碑式的重要意义。国际金融中心建设不仅仅是股票市场和金融机构,更重要的还要有金融信息的收集、发布,这对上海提高金融中心建设中的软实力极为关键。

网址: <http://www.xinhua08.com>。