CSMAR

新数据 新科研

——CSMAR数据库的创新应用

主讲人：杨曼莎

www.csmar.com

深圳希施玛数据科技有限公司

Contents

01

CSMAR简介

- 数据库概况
- 基础操作指引

数据库概况

全　　称：China Stock Market & Accounting Research Database

中国经济金融研究数据库

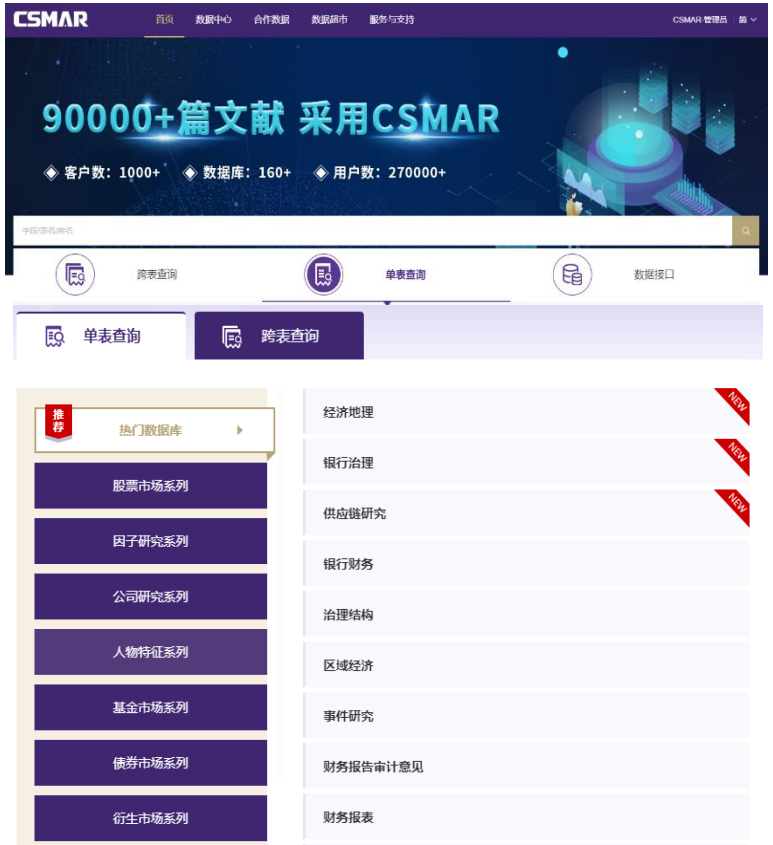
定　　位：**研究型精准**数据库

标　　准：CSMAR数据库参照CRSP、COMPUSTAT等权威数据库的标准。

服务对象：以研究和量化投资分析为目的的学术高校和金融机构。

内　　容：将数据库分为股票、公司、基金、债券、衍生、经济、行业、海外、资讯系列数据库。涵盖中国证券、期货、外汇、宏观、行业等经济金融主要领域的高精准研究型数据库，是投资和实证研究的基础工具。

官　　网：http://cn.gtadata.com/或http://www.gtarsc.com



数据库概况

18
系列

160+
数据库

50000+
字段

- 公司研究系列
- 股票市场系列
- 因子研究系列
- 专题研究系列
- 商品市场研究系列
- 科技金融研究系列
- 基金市场系列
- 海外研究系列
- 债券市场系列
- 人物特征系列
- 市场资讯系列
- 行业研究系列
- 经济研究系列
- 衍生市场系列
- 绿色经济系列
- 货币市场系列
- 银行研究系列
- 板块研究系列

数据库概况

序号	数据库
1	中国股票市场交易数据库
2	中国证券市场指数研究数据库
3	中国上市公司财务报表数据库
4	中国上市公司财务指标分析数据库
5	中国上市公司财务报告审计意见数据库
6	中国银行财务研究数据库
7	中国上市公司首次公开发行研究数据库（A股）
8	中国上市公司增发配股研究数据库
9	中国上市公司股东研究数据库
10	中国上市公司红利分配研究数据库
11	中国上市公司治理结构研究数据库
12	中国上市公司并购重组研究数据库
13	中国银行间交易研究数据库
14	中国证券市场基金评价研究数据库
15	中国债券市场研究数据库
16	中国商品期货市场研究数据库
17	中国股票市场收益波动研究数据库
18	中国股票市场基本分析研究数据库
19	中国股票市场实践研究数据库
20	中国黄金市场交易研究数据库
21	中国上市公司专利数据库

数据库概况

CSMAR数据库的研发，始终确保：

数据来源权威、采集流程专业、质检流程规范

为广大用户提供高质量数据，助力学术研究。

01

权威

- **数据源**包括上交所、深交所、上期所、港交易所、中金所等，国家外汇管理局，中证信息,中国年鉴信息网等

02

基础

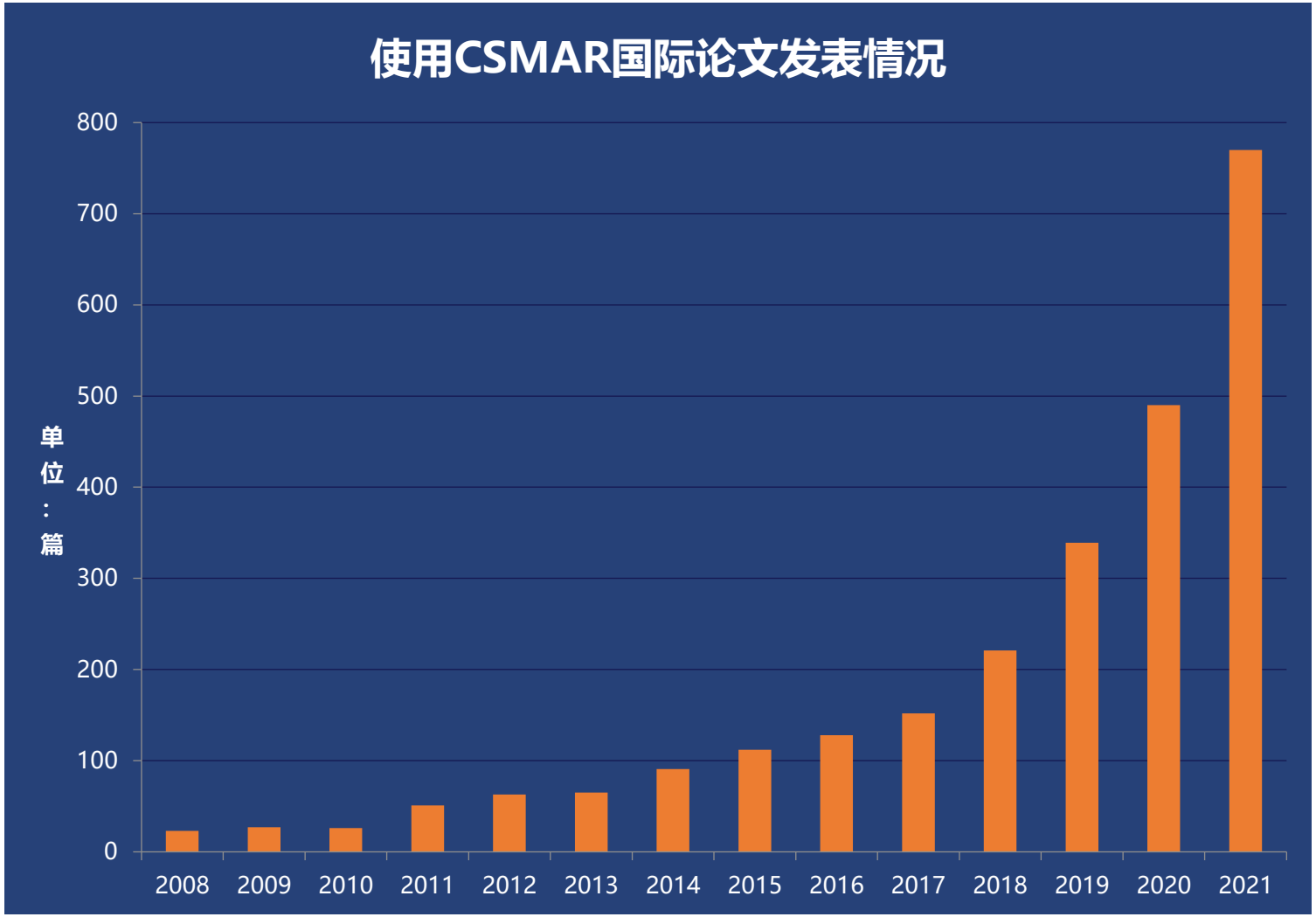
- CSMAR数据被**摩根斯坦利**选用，作为编制MSCI-A股指数的**基础**。

03

引用

- 截至2021年12月31日，国内外使用**CSMAR**数据的论文多达96000多篇。

数据库概况



数据来源于Wiley Online Library 、 ScienceDirect，截止时间2021.12.31

数据库概况

中国科学技术大学引用CSMAR情况

根据中国知网的检索结果，2020年至今，中国科学技术大学引用CSMAR在权威国内期刊上发表了12篇文章。

主要研究主题

- 企业经营绩效波动
- 战略差异
- 全要素生产率
- 职业化管理
- 商业银行效率
- 管理者过度自信
- 金融科技
- 对外直接投资
- 政府补助
- 企业全要素生产率
- 企业创新
- 资本深化
- 绩效关系
- 企业融资约束
- 环境不确定性

[1]郑明贵,董娟,钟昌标.资本深化对中国资源型企业全要素生产率的影响[J].资源科学,2022.

[2]巴曙松,吴丽利,熊培瀚.政府补助、研发投入与企业创新绩效[J].统计与决策,2022.

[3]胡志亮,郑明贵.企业战略差异、环境不确定性与企业经营绩效波动[J].华东经济管理,2021.

[4]王相宁,刘肖.金融科技对中小企业融资约束的影响[J].统计与决策,2021.

[5]肖宵,马骏,李新春,李书娴,姚振玖.家族企业的对外直接投资与职业化管理[J].管理学报,2021.

[6]楚雪芹,李勇军,崔峰,梁樑.基于两阶段非期望DEA模型的商业银行效率评估[J].系统工程理论与实践,2021.

[7]刘志迎,支援援,吴瑞瑞.开放获取资源能够促进二元创新吗？[J].东北大学学报(社会科学版),2021.

[8]鲁小凡,窦钱斌,宋伟,葛章志.海归高管与企业创新效率：助力还是阻力？[J].科技管理研究,2021.

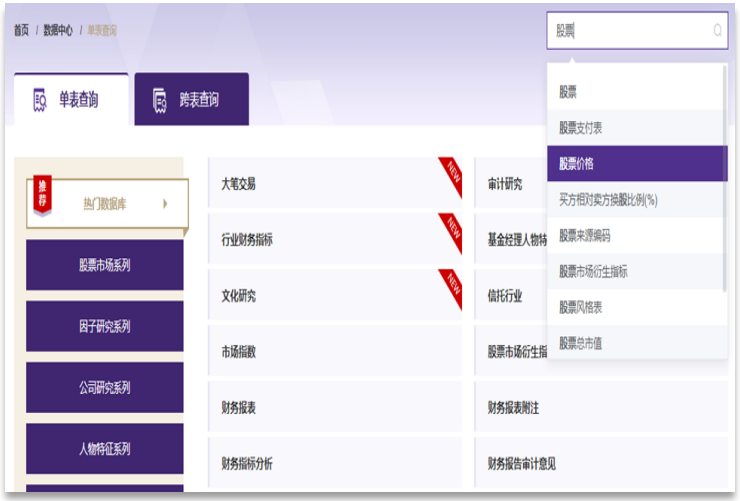
[9]冯锋,张燕南.企业社会责任与公司绩效关系再讨论——基于上市公司企业社会责任评级数据的实证分析[J].吉林大学社会科学学报,2020.

[10]郑明贵,潘咏雪,胡志亮.战略差异对企业经营绩效波动影响的路径研究[J].华东经济管理,2020.

[11]曹崇延,翟青梅.管理者过度自信对企业扩张影响的统计检验[J].统计与决策,2020.

[12]魏玖长,丁奚.重特大安全事故震慑效应的影响因素研究[J].中国行政管理,2020.

数据库概况



数据服务

- 数据合作
- 数据接口
- 数据定制
- ……

数据查询

- 关键字搜索
- 单表查询
- 跨表查询
- ……

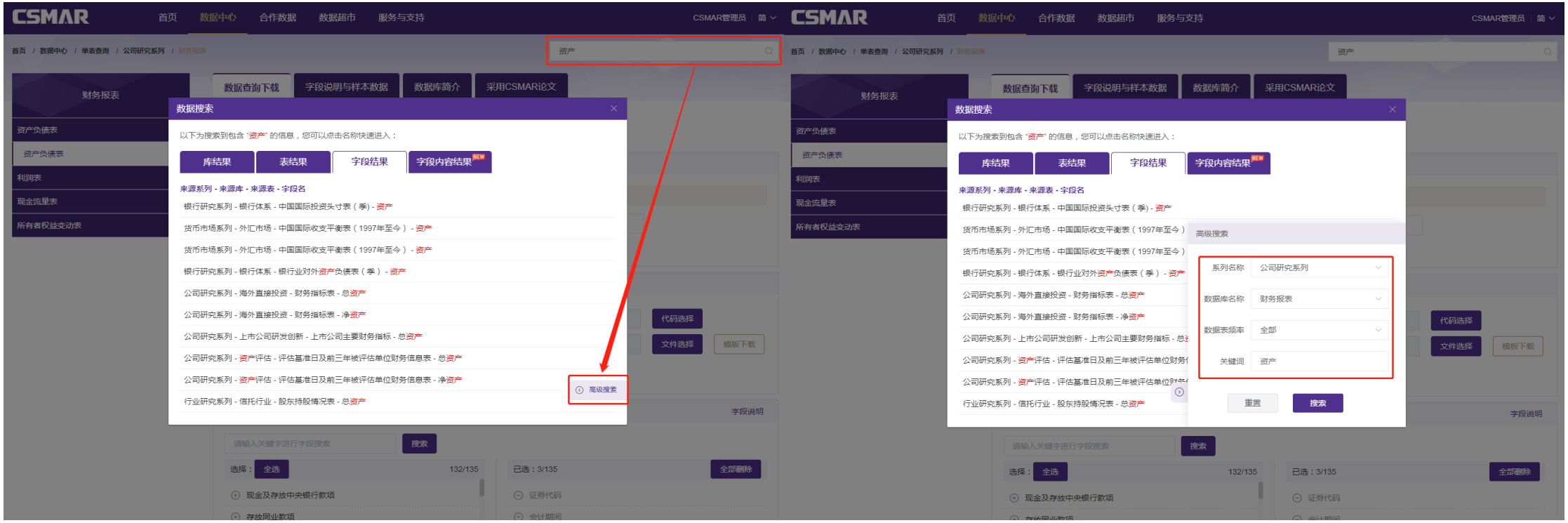
数研通平台

- 术语检索
- 文献检索
- ……

基础操作指引

1. 关键字搜索（支持模糊搜索）

在搜索框中输入关键字，搜索框下显示包含关键字的相关字段、数据表、数据库信息，点击搜索，即可展示包含此关键字的数据库/表/字段名称/字段内容。增加了高级搜索功能，用户可自定义设置系列名称、数据库名称、数据表频率（日、周、月、季、年）及搜索关键词，更高效地搜索数据内容。



基础操作指引

2. 数据查询

如果您希望查看某个数据库的数据，只需点击【数据中心】-“单表查询”指定系列界面的某个数据库名称，将会进入当前数据库的数据查询页面。有权限的数据库为黑色字体显示，无权限的数据库为灰色字体显示。

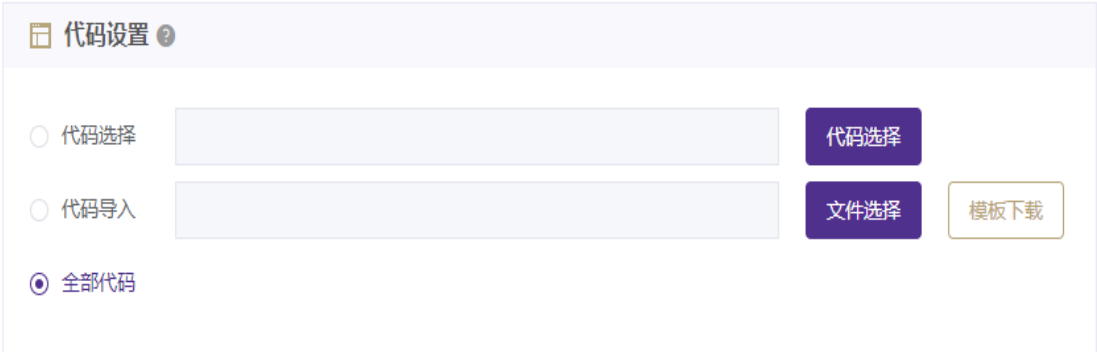
数据库的数据查询页面展示当前数据库的数据库信息，其中：黑色字体显示有权限的数据表，灰色字体显示无权限的数据表。

基础操作指引

3. 时间设置与代码设置

设置所需查询的数据时间区间，无权限的数据表不可进行时间设置；有权限的数据表时间设置范围是权限内的时间区间。

可以通过以下三种方式进行代码设置：全部代码、代码选择、代码导入（请严格按照代码模板进行导入）。



基础操作指引

4. 字段选择与条件筛选

选择需要查询下载的字段，可以通过单击字段项或者点击【全选】按钮进行字段选择。如果当前数据表包含字段太多，您可以通过在输入框内输入您想查找的字段关键字进行实时搜索。如需了解当前数据表的字段说明，可点击【字段说明】进行查看。

如果您想对查询数据设定筛选条件，可以通过设置条件方式进行条件限定，以满足符合一个或者多个条件组合的数据结果。

基础操作指引

5. 联表查询

单表查询模块，针对用户经常使用的数据表提供【联表查询】功能，实现多个相关数据表的数据自动提取、合并。

功能对比	单表查询（联表查询）	跨表查询
组合数据表的频率	相同的数据频率	年、季、月、日数据频率自由组合
组合数据表的范围	针对某个数据主题相关的数据表进行联表查询，较跨表查询，可以针对区域经济、县域经济等数据进行组合查询	可进行多数据主题的数据表组合查询，如同时查询宏观经济+公司行情+公司财务数据
组合数据表的数据筛选条件设置	除了时间、代码外，还可以针对主表进行更多数据筛选条件设置	仅针对时间和代码进行设置

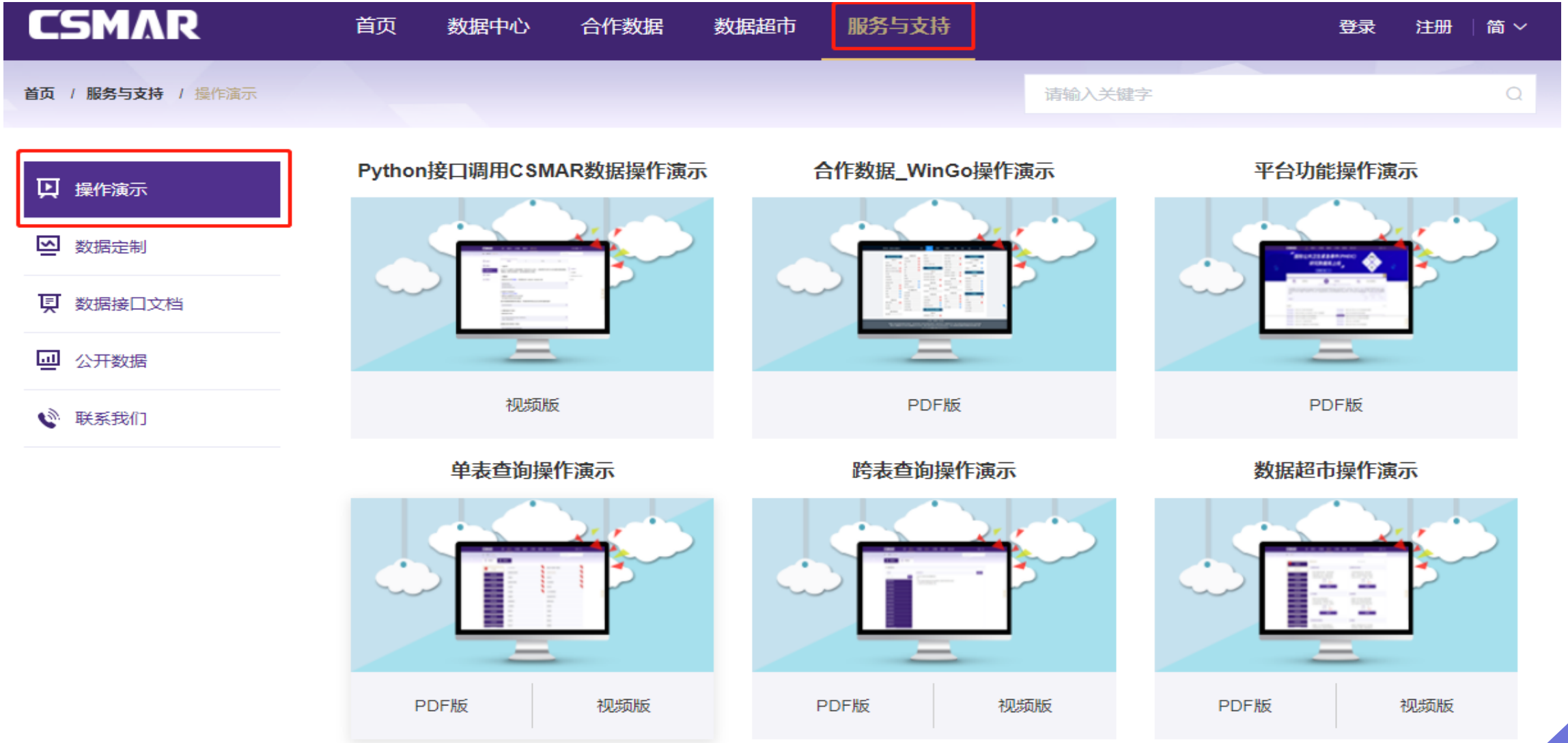
基础操作指引

6. 数据下载与数据库资料

设置好查询条件后，点击【下载数据】。新页面打开数据下载概要，您将看到具体的条件设置信息，同时，下方显示查询数据的下载压缩包。点击压缩包名称，可直接下载数据到本机。

如果您想查看某个数据表的样本数据或字段说明，请点击【字段说明和样本数据】。或者，您可以通过点击【数据库说明书】查看当前数据库的简介和下载数据库说明书。当有此数据库的表权限时，才能下载数据库说明书。如果您想了解采用数据库发表的文献情况，可点击【相关文献】查看，目前主要展示《会计研究》、《经济研究》、《金融研究》等期刊中采用CSMAR数据的文献列表。

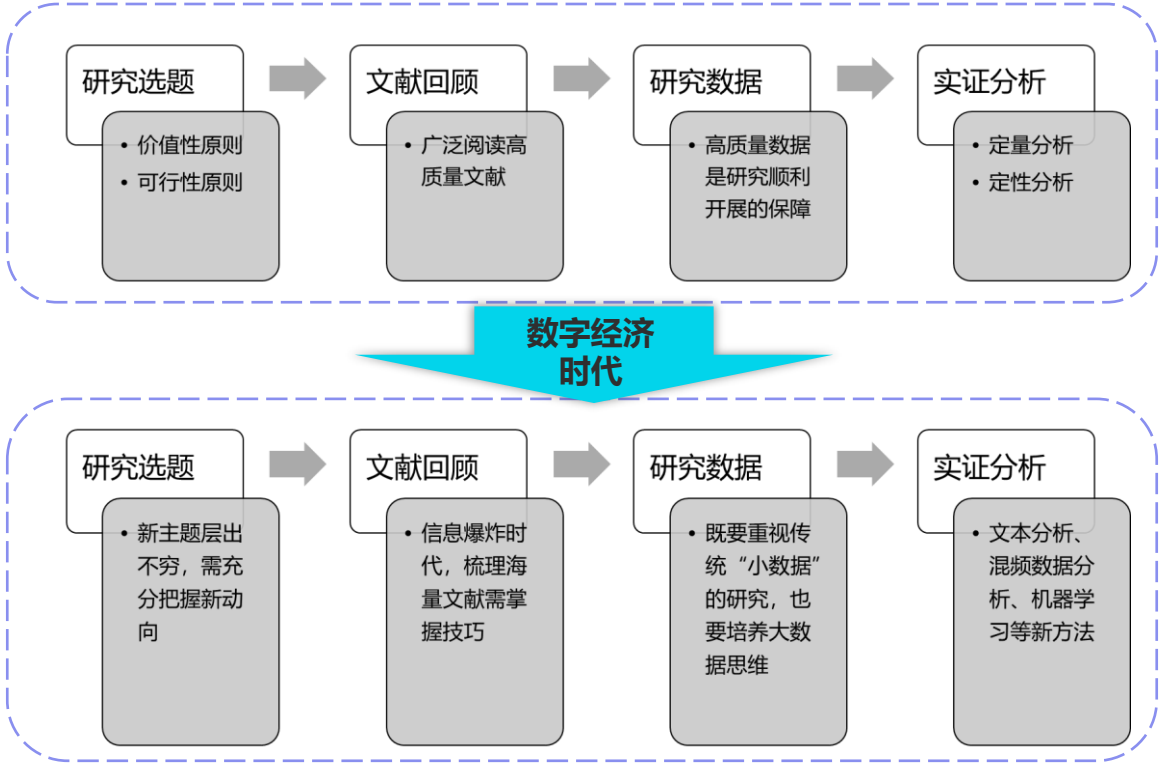
基础操作指引


CSMAR
首页
数据中心
合作数据
数据超市
服务与支持
登录
注册
简
首页 / 服务与支持 / 操作演示
请输入关键字
操作演示
数据定制
数据接口文档
公开数据
联系我们
Python接口调用CSMAR数据操作演示
视频版
合作数据_WinGo操作演示
PDF版
平台功能操作演示
PDF版
单表查询操作演示
PDF版
视频版
跨表查询操作演示
PDF版
视频版
数据超市操作演示
PDF版
视频版

02

CSMAR与实证研究创新

- 研究选题
- 文献回顾
- 研究数据
- 实证分析



研究选题	✧ CSMAR数据库：热门研究主题与CSMAR数据库；前沿研究主题与CSMAR数据库		
文献回顾	✧ CSMAR数据库：“关联CSMAR论文“板块，实现数据与文献联动 ✧ 数研通：聚焦经济金融权威文献的检索		
研究数据	✧ CSMAR数据库：数据全面，部分特色库体现科研大数据的应用		
实证分析	定量分析	文本分析	✧ CSMAR数据库：部分特色库包含文本分析指标，可直接用于建模 ✧ WinGO财经文本分析平台：可进行文本分析
		混频分析	✧ CSMAR数据库：可获取混频数据用于建模
		机器学习	✧ CSMAR数据库：数据全面，为机器学习分析方法提供数据基础

研究选题——重要原则

01 价值性原则

（1）选题应有社会价值、实践意义

选题必须与国家经济建设和社会发展的需求与总目标相一致。

（2）选题应有科学价值、学术意义

选取具有学术价值的课题，最根本的是选择具有**学术创新意义**的课题。

创新在学术论文中**主要体现**为：发现、提出新问题；完善、补充前人的观点、理论；否定纠正前人某一结论、成说；发掘、提供新的资料；采用新的角度方法，作出新的论证；运用已有理论研究解决社会和学科发展中的迫切问题。

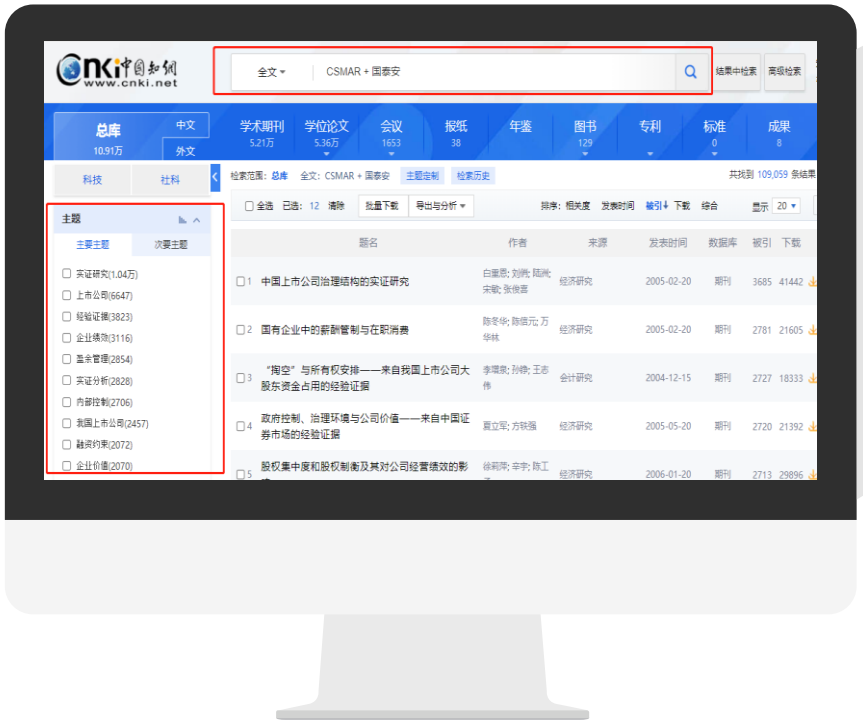
02 可行性原则

（1）选题要难易、大小适中

（2）选择有条件完成的课题

需要充分考虑获取必要文献资料的条件、完成课题的时间条件、导师的指导条件等。

研究选题——热门研究主题与CSMAR数据库



在**知网**检索“*全文 - CSMAR + 国泰安*”，可获取采用了CSMAR数据发表的论文，了解研究主题情况。

使用CSMAR数据的一些**热门研究主题**包括：

- 融资约束
- 企业创新
- 公司治理
- 股权质押
- 盈余管理
- 内部控制
- 信息披露
- 企业价值
- 企业社会责任
- 投资效率
- 审计质量
- 家族企业
- 企业金融化
- 分析师
- 股价崩盘风险
- 股票流动性
- 环境规制
- 产业政策
- 经济政策不确定性
- 混合所有制改革

在**知网**进行**高级检索**，如：

- *主题 - 融资约束*
- *全文 – CSMAR + 国泰安*

可获取该主题下使用CSMAR数据发表的论文，了解具体研究内容。

针对上述热门主题，讲义里列出了研究中常用的CSMAR子库，供学习者参考。

研究选题——热门研究主题与CSMAR数据库

表 热门研究主题常用的CSMAR子库

融资约束	财务报告审计意见、区域经济、上市公司与子公司专利、内部控制、股票市场交易、经营困境、财务指标分析、违规处理、社会责任、行业财务指标、上市公司基本信息、财务报表附注、财务报表、股东、上市公司研发创新、并购重组、股权性质、治理结构、审计研究、融资融券
企业创新	财务报告审计意见、上市公司及子公司专利、上市公司与子公司专利、财务指标分析、上市公司人物特征、机构投资者、行业财务指标、上市公司基本信息、财务报表附注、财务报表、海外直接投资、股东、资质认定、民营上市公司、上市公司研发创新、股权性质、治理结构、审计研究、分析师预测
公司治理	财务指标分析、上市公司基本信息、并购重组、上市公司人物特征、股票市场衍生指标、违规处理、治理结构、审计研究、股票市场交易、财务报表、股东、会计信息质量
股权质押	财务报告审计意见、区域经济、股票市场交易、债券、财务指标分析、上市公司人物特征、机构投资者、股权质押、上市公司贷款、上市公司基本信息、财务报表附注、财务报表、资源、股东、并购重组、股权性质、全球暖化、治理结构、分析师预测、关联交易、新闻
盈余管理	股票市场收益、财务指标分析、上市公司基本信息、并购重组、股权性质、财务报告审计意见、财务报表附注、治理结构、审计研究、财务报表、机构投资者、股东、分析师预测、首次公开发行（A股）、会计信息质量、股票市场交易
内部控制	财务指标、诉讼仲裁、财务指标分析、上市公司基本信息、股权性质、财务报告审计意见、违规处理、治理结构、EVA专题、审计研究、财务报表、机构投资者、股东、内部控制
信息披露	投资者情绪、财务指标分析、股权性质、治理结构、财务报表、机构投资者、股东、分析师预测、首次公开发行（A股）、股票市场交易
企业价值	财务指标分析、上市公司基本信息、治理结构、区域经济、财务报表、机构投资者、股东、分析师预测、上市公司研发创新
企业社会责任	财务指标分析、上市公司基本信息、上市公司人物特征、股权性质、治理结构、财务报表、行业财务指标、股东、分析师预测、上市公司与子公司专利、上市公司研发创新、股票市场交易
投资效率	财务指标分析、上市公司基本信息、股权性质、财务报告审计意见、治理结构、社会责任、审计研究、财务报表、股东、分析师预测、内部控制、会计信息质量、股票市场交易

研究选题——热门研究主题与CSMAR数据库

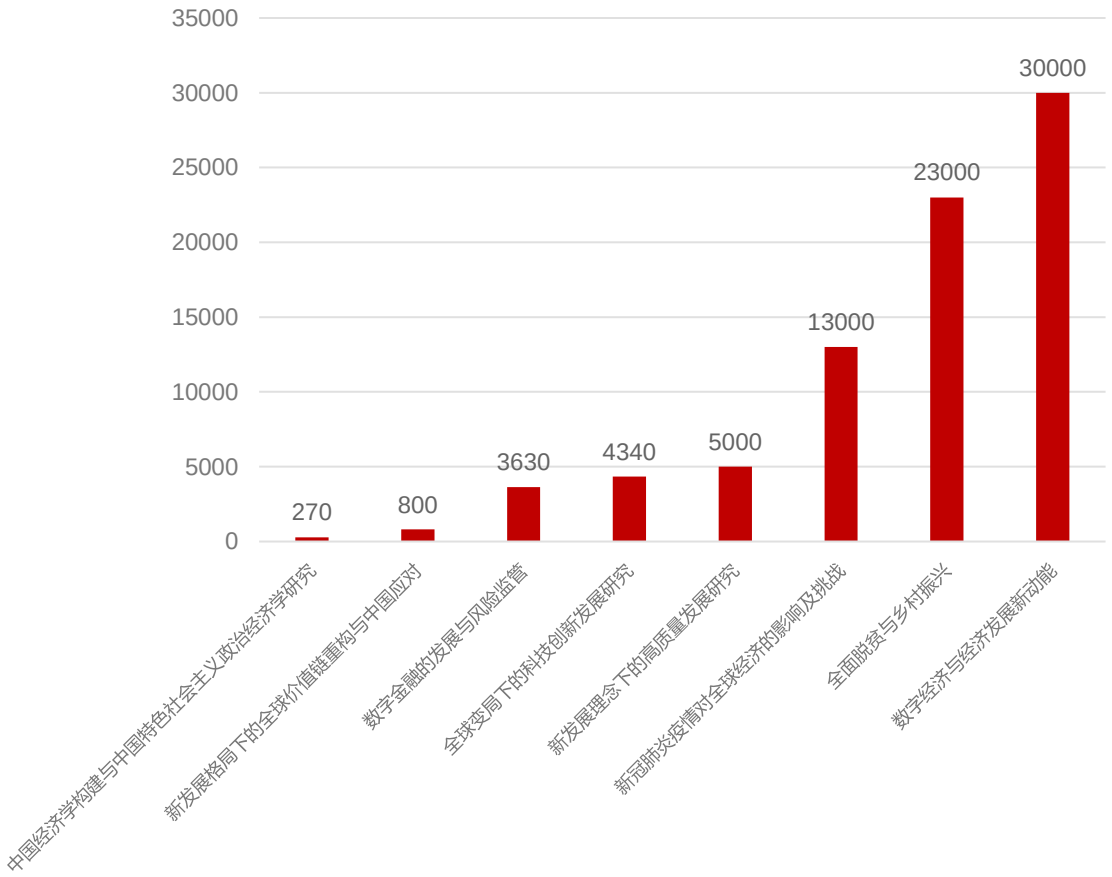
表 热门研究主题常用的CSMAR子库

审计质量	股权分置改革、财务指标分析、上市公司基本信息、上市公司人物特征、财务报告审计意见、股权性质、治理结构、财务报表附注、审计研究、财务报表、机构投资者、股东、融资融券、会计信息质量、股票市场交易
家族企业	财务指标分析、股权性质、上市公司人物特征、财务报表附注、治理结构、首次公开发行（A股）、财务报表、海外直接投资、家族企业、股东、民营上市公司、行业财务指标、宏观经济、上市公司研发创新
企业金融化	财务报告审计意见、区域经济、行为金融、内部控制、股票市场交易、财务指标分析、上市公司人物特征、机构投资者、行业财务指标、股权质押、上市公司基本信息、财务报表附注、财务报表、股东、宏观经济、上市公司研发创新、股权性质、治理结构、审计研究
分析师	财务指标分析、上市公司基本信息、行为金融、财务报表、机构投资者、分析师预测、首次公开发行（A股）、股票市场交易
股价崩盘风险	财务报告审计意见、行为金融、家族企业、市场指数、会计信息质量、股票市场交易、财务指标分析、股票市场衍生指标、机构投资者、财务报表附注、财务报表、股票流动性、股东、民营上市公司、宏观经济、上市公司研发创新、股权性质、治理结构、分析师预测
股票流动性	财务指标分析、治理结构、财务报表、股票流动性、行业财务指标、关联交易、股票市场交易
环境规制	财务报表、财务报表附注、财务指标分析、股权性质
产业政策	上司公司研发创新、财务指标分析、上市公司基本信息、股权性质、财务报告审计意见、治理结构、审计研究、财务报表、机构投资者、股东、上市公司与子公司专利、首次公开发行（A股）、宏观经济、股票市场交易
经济政策不确定性	财务指标分析、股权性质、治理结构、区域经济、财务报表、股东、分析师预测、市场指数、业绩预告、宏观经济、股票市场交易
混合所有制改革	财务指标分析、国有股拍卖与转让、治理结构、审计研究、财务报表、股东、行业财务指标、宏观经济

研究选题——前沿研究主题与CSMAR数据库

2020年经济管理学前沿研究主题

(1) 学术性 (2) 社会性 (3) 全面性(4) 平衡性 (5) 综合性 (6) 包容性 (7) 关注度



1. 新冠肺炎疫情对全球经济的影响及挑战
2. 中国经济学构建与中国特色社会主义政治经济学研究
3. 新发展理念下的高质量发展研究
4. 数字经济与经济发展新动能
5. 全面脱贫与乡村振兴
6. 新发展格局下的全球价值链重构与中国应对
7. 数字金融的发展与风险监管
8. 全球变局下的科技创新发展研究

参考资料：中国人民大学书报资料中心经济编辑部. 2020年中国经济学与管理学研究热点分析[J]. 经济学动态, 2021.

研究选题——前沿研究主题与CSMAR数据库

研究热点	主要研究内容	关联数据库
1.新冠肺炎疫情对经济的影响	新冠肺炎疫情的发展演变及其经济社会影响，包括宏观、微观及国际关系层面的冲击，以及对我国贸易、产业链、就业、地方财政等的影响；应对新冠肺炎疫情的对策和措施。	**新冠疫情与经济研究**、**公共卫生事件PHEIC**、新闻、宏观经济、区域经济、**国际宏观** 行业研究系列、公司研究系列、股票市场系列
2.中国经济学构建	突出发展战略与规划研究和宏观经济调控研究两大命题，打造彰显中国特色的国民经济学；“三农”领域的研究；金融学本土化等。	**文化研究**、**经济内循环**、宏观经济、区域经济、农村金融经济、**精准扶贫**、民营企业、家族企业
3.新发展理念下的高质量发展研究	包括国家治理、经济、社会、文化和生态五个维度，相关研究分析了高质量发展的内涵、动力以及相关测度指标体系的构建等。	**文化研究**、**碳中和**、**环境研究**、**绿色专利**、**润灵环球ESG**、**商道融绿ESG**、绿色金融、全球暖化、资源研究、社会责任
4.数字经济与经济发展新动能	对数字经济本质和运行规律的理论性探索；围绕数字经济助推经济发展变革并成为经济发展新引擎开展的理论和应用研究；对数据这一新型生产要素在社会经济运行与创新发展变革中的作用机理的研究；围绕数字产业化和产业数字化开展的理论、实证与对策性研究；数字经济治理研究；数字经济领域的统计与核算研究。	**数字经济**、**金融科技**、**金融机构分支机构**、一带一路、**国际宏观**、东盟十国宏观经济、企业创新 公司研究系列
5. 全面脱贫与乡村振兴	脱贫攻坚、相对贫困治理、脱贫攻坚与乡村振兴的有效衔接、后小康社会时期乡村振兴等。	**精准扶贫**、农村金融经济、**金融机构分支机构**、宏观经济、区域经济、县域经济
6. 新发展格局下的全球价值链重构与中国应对	新冠肺炎疫情影响下全球价值链的发展方向；全球价值链与国际经济问题；中国攀升全球价值链的路径和模式。	**供应链**、**经济内循环**、一带一路、**数字经济**、上市公司研发创新、专利、**专利被引用**
7. 数字金融的发展与风险监管	数字金融的价值与影响；数字金融的风险与监管。	**金融科技**、专利、区域经济、县域经济、企业创新、上市公司研发创新
8. 全球变局下的科技创新发展研究	我国的科技创新优势和劣势、我国创新体系构建与创新环境、创新政策的制定实施与作用效果、我国参与全球产业链创新链分工、关键核心技术方面的突破、推进企业创新与平台创新、促进科技成果转化与知识产权管理等。	宏观经济、**国际宏观**、**经济内循环**、**专利被引用**、**数字经济**、**绿色专利**、区域经济、专利、企业创新、上市公司研发创新

CSMAR与研究选题创新——前沿研究主题与CSMAR数据库

数字经济发展与企业价值提升[J].经济问题,2021.

内容提要

有效地发挥数字经济的作用，引导和帮助企业进行数字化转型，对促进企业价值提升，进而驱动经济高质量发展有重要意义。文章以我国2011—2019年A股上市公司为研究样本，将“宽带中国”战略的实施视为一场准自然实验，以是否实施“宽带中国”战略试点来衡量数字经济发展水平，采用双重差分模型实证检验数字经济对企业价值的影响以及企业生命周期不同阶段数字经济对企业价值影响关系的差异。研究发现，数字经济发展有利于促进企业价值提升。对不同生命周期阶段的企业而言，数字经济发展对企业价值的影响效应差异明显。

研究方法

本文采用**双重差分模型**进行实证分析。

$$tobinq_{it} = \alpha_0 + \alpha_1 data_{it} + \varphi controls_{it} + h_{it} + \mu_i + v_t + \varepsilon_{it}$$

研究数据

企业价值、数字经济发展、企业生命周期以及控制变量：企业成长性、企业规模、独立董事比例、资产负债率、股权集中度、董事会规模、监事会规模、审计意见等。（2011-2019年沪深上市公司）。

变量	定义
企业价值（tobinq）	选取托宾Q作为企业价值的测度指标
数字经济发展（data）	以是否实施“宽带中国”战略试点来衡量数字经济发展水平
企业生命周期	初创期（1~6 年）、成长期（7~11 年）和成熟期（12年及以上）
企业成长性（growth）	采用主营业务收入增长率测度
企业规模（ln asset）	采用总资产的自然对数测度
独立董事比例（indep）	采用独立董事人数与董事会总人数的比值测度
资产负债率（lev）	采用总负债与总资产的比值测度
股权集中度（first）	采用第一大股东持股比例之和测度
董事会规模（ln dsh）	采用董事会规模的自然对数衡量
监事会规模（ln jsh）	采用监事会规模的自然对数衡量
审计意见（opinion）	当出具标准无保留意见时取值为 1，否则取值为 0
管理层持股比例（manhold）	采用管理层持股数量占股本总数的比值测度

数据获取

CSMAR公司研究系列

- 财务报表
- 财务指标分析
- 股东
- 股权性质
- 治理结构
- ……

另一种测算方式：
分别从数字基础设施、数字产业化、产业数字化等方面构建了数字经济发展水平量化指标体系，基于熵值法测度。
参考文献：李英杰,韩平.中国数字经济发展综合评价与预测[J].统计与决策,2022.

综合指标	指标类别	指标名称及计量单位
数字经济发展水平	数字基础设施	互联网普及率(%)
		光缆线路长度(万公里)
		IPV4地址数(十块)
	数字产业化	电子信息制造业增速(%)
		软件业务收入(万元)
		信息通信产业就业人数(万人)
		ICT行业固定投资占全社会总投资比例(%)
	产业数字化	每百家企业拥有网站数(个)
		拥有电子商务交易活动的企业数(十个)
		电子商务交易额(万亿元)

指标来源：CSMAR数字经济研究数据库

CSMAR与研究选题创新——前沿研究主题与CSMAR数据库

大数据应用对中国企业市场价值的影响[J].经济研究,2021.

内容提要

文章通过对A股上市公司的年报进行文本分析，构建了衡量公司层面“大数据”应用程度的指标，探讨了企业大数据应用的发展状况及决定因素，检验了大数据应用对公司市场价值的影响。研究发现：第一，规模较大、有形资产比例较低、盈利能力较强，以及所在地区市场化程度较高的公司更可能在生产经营过程中应用大数据；第二，大数据的应用可以显著提高公司的市场价值；第三，主要的影响机制在于大数据的应用显著提高了公司的生产效率和研发投入，而相关技术和人才供给的不足可能会阻碍大数据对市场价值的积极影响。文章结论对中国未来大数据相关的政策设计具有参考价值，为推动实体企业生产经营与大数据的高效融合提供了经验证据和指导建议。

研究方法

I. 采用**Probit和OLS模型**分析大数据应用程度的决定因素；

II. 基于如下基准模型分析大数据应用对公司市场价值的影响。

$$Y_{ijpt} = \gamma_0 + \gamma_1 BigData_{ijpt} + \gamma_2 Controls_{ijpt} + \delta_{pt} + \gamma_{jt} + \mu_i + \xi_{ijpt}$$

变量	变量定义
lnBigdata	表1中大数据相关关键词在年报中出现的次数加一后取对数
Tobin's Q	公司总市值与总负债之和除以公司总资产
lnAssets	总资产取自然对数
Lev	总资产除以股东权益
PPE_TA	固定资产除以总资产
lnAge	当年年份减去上市年份加1，再取自然对数
SOE	按公司实际控制人性质确定
SalesGrowth	当年营业收入除以上一年营业收入后减一
ROA	净利润除以总资产

关键词	定义
大数据	企业收集、处理与利用的海量、高速、多样化的数据要素或资产。
海量数据	根据高德纳公司对大数据的定义，海量规模是大数据的重要特征之一。
数据中心	安置计算机系统及相关部件的设施，用于在网络基础设施上传递、加速、展示、计算、存储数据信息。信息时代下，大数据需要安全可靠、高效率的数据中心进行存储、计算和交换。
信息资产	指由企业拥有或者控制的能够为企业带来未来经济利益的信息资源。根据高德纳公司的报告，大数据本质上是一种信息资产。
数据化	将均匀、连续的数字比特结构化和颗粒化，形成标准化的、开放的、非线性的、通用的数据对象，并基于不同形态与类别的数据对象，实现大数据的应用。
算力	也称哈希率，指比特币网络处理能力的度量单位，也是计算机计算哈希函数时输出的速度。

研究数据

大数据相关关键词在年报中出现的次数，公司估值指标，公司规模、杠杆率、固定资产比率、公司年龄等（2006-2017沪深上市公司）。

相关指标获取：CSMAR数字经济研究数据库-上市公司数字化

上市公司数字化	上市公司基本信息（年）、上市公司数字化转型指标（年）、上市公司数字化转型程度（年）、数字经济上市公司主要财务指标（年）、数字经济上市公司专利申请获得情况（年）、数字经济上市公司研发投入情况表（年）	2000-	年

指标分类	指标名称
人工智能技术	人工智能、商业智能、图像理解、投资决策辅助系统、智能数据分析、智能机器人、机器学习、深度学习、语义搜索、生物识别技术、人脸识别、语音识别、身份验证、自动驾驶、自然语言处理
区块链技术	数字货币、智能合约、分布式计算、去中心化、比特币、联盟链、差分隐私技术、共识机制
云计算技术	内存计算、云计算、流计算、图计算、物联网、多方安全计算、类脑计算、绿色计算、认知计算、融合架构、亿级并发、EB级存储、信息物理系统
大数据技术	大数据、数据控掘、文本控掘、数据可视化、异构数据、征信、增强现实、混合现实、虚拟现实
数字技术应用	移动互联网、工业互联网、移动互联、互联网医疗、电子商务、移动支付、第三方支付、NFC支付、B2B、B2C、C2B、C2C、O2O、网联、智能穿戴、智慧农业、智能交通、智能医疗、智能客服、智能家居、智能投顾、智能文旅、智能环保、智能电网、智能能源、智能营销、数字营销、无人零售、互联网金融、数字金融、Fintech、金融科技、量化金融、开放

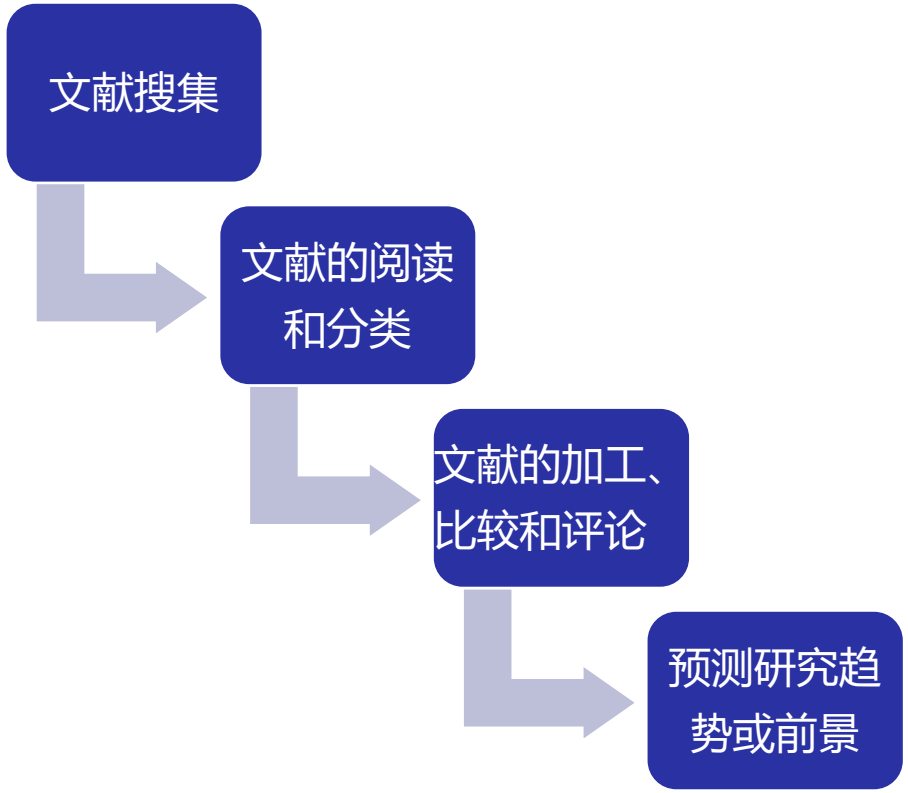
数据获取

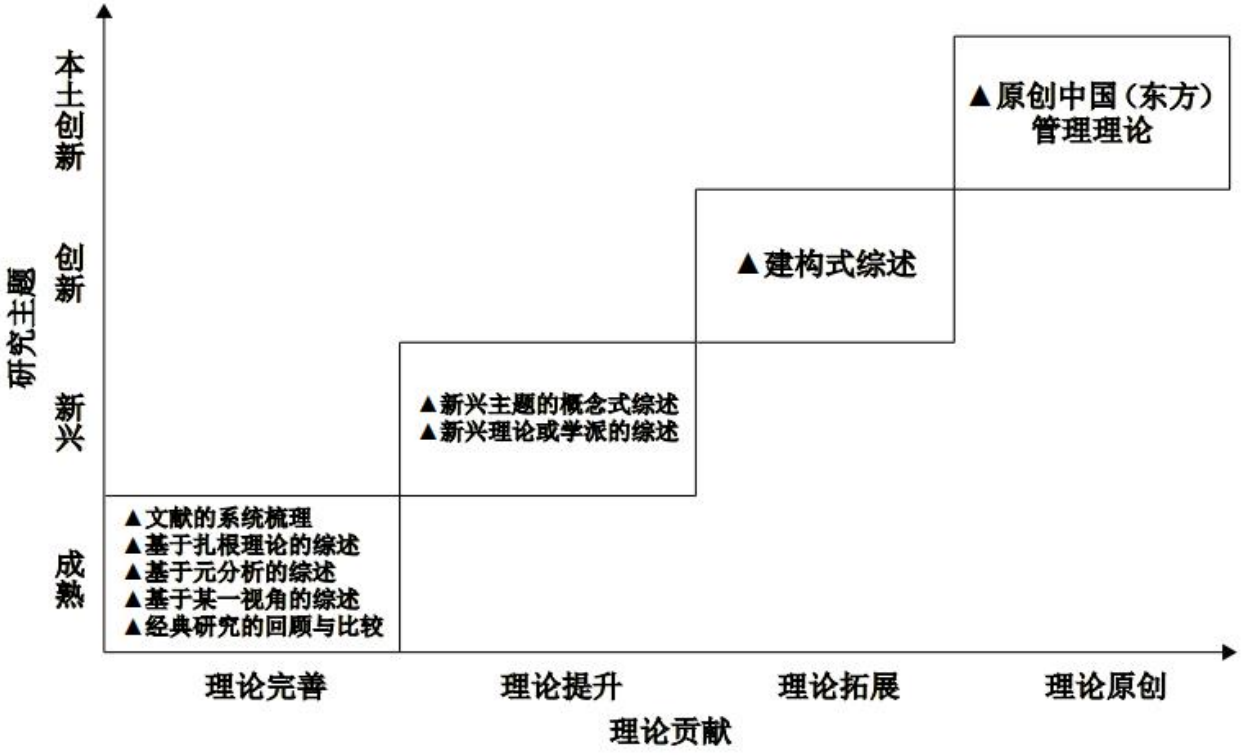
CSMAR公司研究系列

- 财务报表
- 财务指标分析
- 股权性质
- …

文献回顾——文献综述

文献综述是实证论文的重要组成部分，作用在于揭示研究的现状，阐明选题依据、研究目的与意义，提出文章的创新之处。好的文献综述必须客观详实、有条有理地述评、分析国内外研究现状，以述带论，以论为主，说明拟研究问题与已有研究之间的差距，展现作者拟做出的贡献。

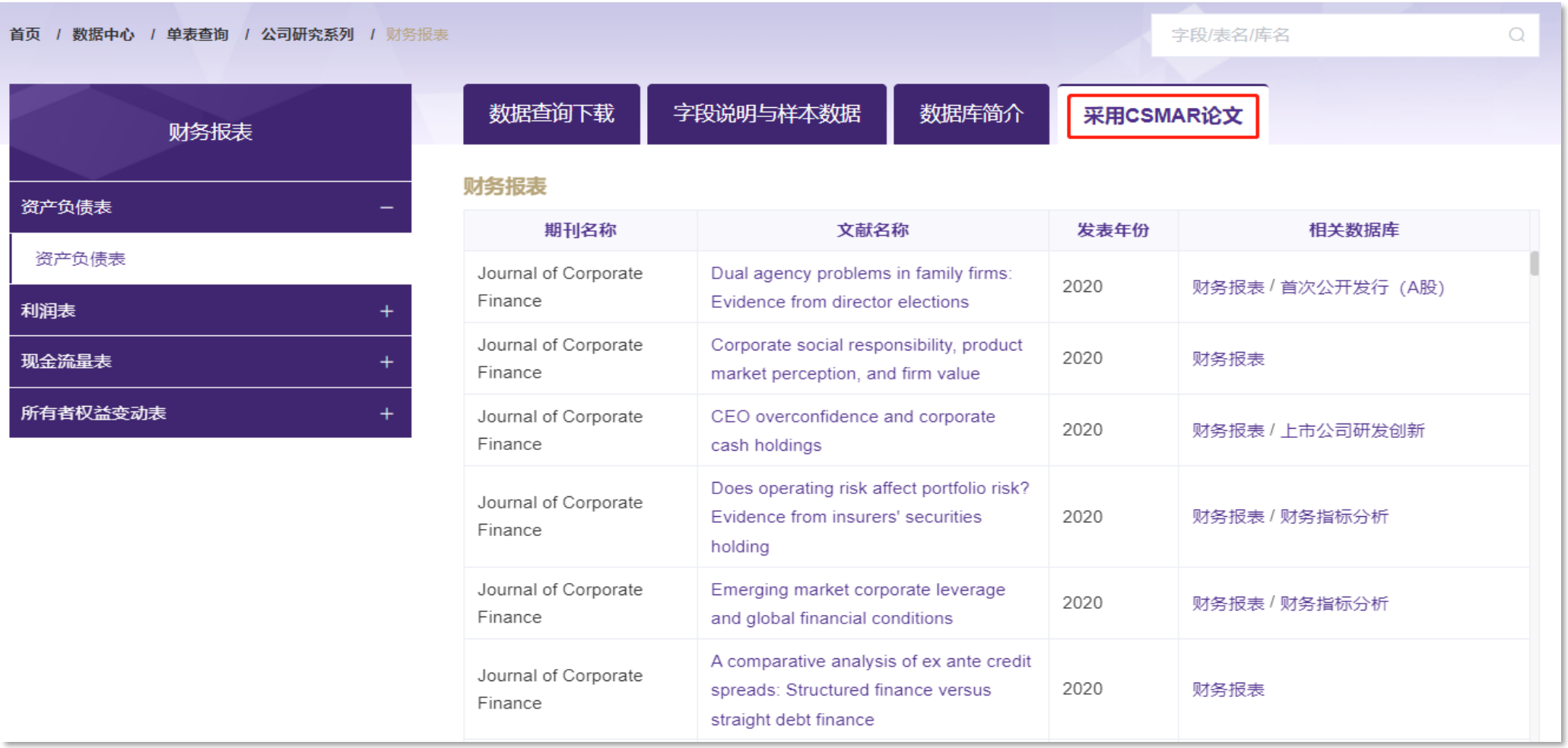




参考资料：朱秀梅,杨姗.管理类综述文章的撰写范式[J].外国经济与管理,2019.

CSMAR与权威文献——数据库与关联文献

CSMAR数据库中，进入各子库，【采用CSMAR论文】板块可查看近几年国内外权威期刊上所发表的与该子库关联的论文。



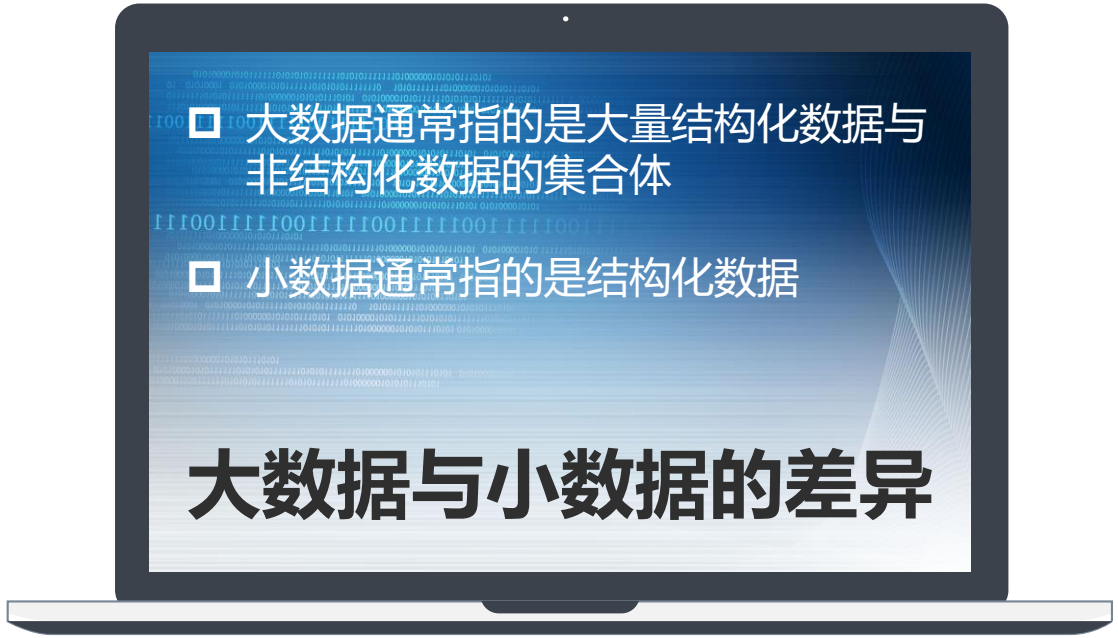
期刊名称	文献名称	发表年份	相关数据库
Journal of Corporate Finance	Dual agency problems in family firms: Evidence from director elections	2020	财务报表 / 首次公开发行（A股）
Journal of Corporate Finance	Corporate social responsibility, product market perception, and firm value	2020	财务报表
Journal of Corporate Finance	CEO overconfidence and corporate cash holdings	2020	财务报表 / 上市公司研发创新
Journal of Corporate Finance	Does operating risk affect portfolio risk? Evidence from insurers' securities holding	2020	财务报表 / 财务指标分析
Journal of Corporate Finance	Emerging market corporate leverage and global financial conditions	2020	财务报表 / 财务指标分析
Journal of Corporate Finance	A comparative analysis of ex ante credit spreads: Structured finance versus straight debt finance	2020	财务报表

CSMAR与权威文献——数研通文献检索功能

数研通文献检索模块展示经济金融相关的最新国际顶级期刊及国内权威期刊的文献信息，可通过关键词搜索快速查询相关文献，并可以进行期刊及发表年份筛选。每个文献展示对应的期刊、摘要、关键词和文献链接。

研究数据——大数据与小数据

数据是进行实证研究最基本、最重要的要素。随着经济社会的发展，数据形态和来源日益多样化，研究数据的选取大体上经历了“只能收集到少量的数据——尽量多地收集数据——科学利用样本数据——综合利用各类数据——选择使用大数据”这样的发展过程。



01.样本的差异

样本容量；样本来源；样本数据类型

02.精确性的差异

大数据对数据收集和分析的精确性要求低于小数据

03.关注的因素关系差异

大数据更关注相关关系，而小数据更关注因果关系

04.价值发现的维度差异

大数据的价值发现主要在于广度，小数据则主要在于深度

研究数据——大数据与小数据

大数据的局限性

- 大数据并非所有时候都是“全数据”
- 大数据并非大家都可以用
- 大数据并不意味着数据的多样化
- 大数据重相关而轻因果
- 大数据不能准确反映人的社会政治行为

小数据的必要性

01.

大数据只能被动地挖掘、收集已经客观发生了的行为信息，而抽样调查和实验研究则可以“制造”数据。

02.

抽样调查的样本在特定情况下比某些“大数据”更具有代表性。

03.

小数据研究在因果关系的分析上别有特点。

04.

小数据能更好地规避学术伦理的问题。

研究数据——大数据思维与小数据研究

尽管大数据和相关的挖掘分析快速增长，小数据仍将继续成为研究领域的重要组成部分。在不久的将来，不大可能会出现大数据研究取代小数据的范式转变，小数据和大数据将相互补充，我们要基于大数据思维开展小数据研究，将机器学习、深度学习和人工智能算法等大数据技术类比应用于小数据上，可以更加充分的释放小数据中的大价值。

大数据思维

1. 广义数据
2. 动态数据
3. 全体数据
4. 先有数据后有总体
5. 有中选优
6. 挖掘价值

界定大数据总体，即根据所拥有的大数据资源，进行时间、空间和数据三个维度的界定，形成特定的大数据总体

对大数据总体进行由粗到细的梳理，确定数据的存在类型，锁定数据所依附的个体身份，根据不同类型的数据定义小数据（变量），确定各种类型数据的分组方法和测度方法

根据一定标准对小数据进行清洗和筛选，剔除无用、虚假（错误）的小数据，对可用的小数据进行编码和分组

开展基于统计规则的数据运算，即依据统计方法和计算技术对大量的小数据进行各种运算，形成各类可供进一步分析的综合性数据，初步确定可进行深入探讨的问题或主题

借助各种统计模型，寻找各种可能存在的分布规律、相关关系（因果关系）和动态趋势

以统计或可视化的方式给出数据结论，合成可供决策的明确信息

参考资料：李金昌.基于大数据思维的统计学若干理论问题[J].统计研究,2016.

CSMAR与大数据的科研运用——特色数据库

近年来随着对市场微观结构数据的深入探索，越来越多研究人员关注日内股票实际的交易情况特征，收盘数据无法满足对日内数据的探索，但是日内数据普遍存在的问题是数据量太大，难以进行统计建模等工作，所以学者们耗费了大量时间精力在进行日内数据转换频率转变成日频数据。基于实际的客户需求，CSMAR团队设计并研发了已实现指标研究数据库，该库采用了大量的日内数据，并进行频率转换，统计计算，最终形成了部分日内转日频指标数据表，此库的推出旨在为客户提供更加全面丰富的市场微观结构数据。本数据库主要包括七张数据表，涵盖了买卖价差、已实现指标、信息不对称、买卖不平衡等指标。

已实现指标研究数据库

- 个股买卖价差表（日）
- 个股已实现指标表（日）
- 指数已实现指标表（日）
- 个股跳跃指标表（日）
- 指数跳跃指标表（日）
- 个股买卖不平衡指标表（日）
- 个股知情交易概率指标表（日）

数据库概况

个股买卖价差 指数已实现指标	个股已实现指标 股跳跃指标
指数跳跃指标 个股买卖不平衡指标	个股知情交易 概率指标

- 上线时间：2020.09
- 更新频率：月
- 起始时间：2003-
- 表字段数：7张表，101个字段

- 数据来源：沪深交易所提供的海量日内交易数据

- 特色指标：相对有效价差、已实现波动率、已实现偏度、跳跃波动率、主动性买入成交金额、个股知情交易概率（VPIN）等

CSMAR与大数据的科研运用——特色数据库

中国股票市场的已实现偏度与收益率预测

金融研究，2018.

风险与收益的关系一直是金融学的核心问题之一。股票市场投资人除了承担市场风险，同时还承担了偏度风险。本文构造了中国股票市场的已实现偏度，并检验了其对中国股票市场收益率的预测能力。实证结果显示，当前较低的已实现偏度可以显著预测下个月中国股票市场较高的超额收益率。在控制了一系列的其它股票预测变量之后，该结论依然成立。此外，基于四种不同的构造方法，已实现偏度对上海和深圳两个股票市场都具有显著的预测能力。从经济解释上，本文发现已实现偏度对股票收益率的预测能力是通过影响股票市场的交易活跃程度，从而传导到股票市场收益率。

1

构造中国股市的偏度风险

① 利用 5 分钟的日内高频指数价格构造日度股票收益率

$$r_{t,i} = p_{t,i} - p_{t,(i-1)}$$

② 计算日度已实现偏度

$$RDVar_t = \sum_{i=1}^{N} r_{t,i}^2 \qquad RDSkew_t = \frac{\sqrt{N}\sum_{i=1}^{N} r_{t,i}^3}{RDVar_t^{3/2}}$$

③ 月度已实现偏度为月内所有交易日的日度已实现偏度之和

$$RSkew_t = \frac{1}{22}\sum_{i=0}^{21} RDSkew_{t-i}$$

2

检验样本内预测能力

$$R_{t+1} = \alpha + \beta RSkew_t + \varepsilon_{t+1}$$

t+1时刻股市超额收益率　　t时刻已实现偏离度

$$R_{t+1} = \alpha + \beta X_{i,t} + \varepsilon_{t+1}$$

第i个中国经济变量

$$R_{t+1} = \alpha + \beta RSkew_t + \sum \phi_i X_{i,t} + \varepsilon_{t+1}$$

3

检验样本外预测能力

$$\hat{R}_{n_1+1} = \hat{\alpha}_{n_1} + \hat{\beta}_{n_1} RSkew_{n_1}$$

$$\hat{R}_{n_1+2} = \hat{\alpha}_{n_1+1} + \hat{\beta}_{n_1+1} RSkew_{n_1+1}$$

$$\bar{R}_{t+1} = \frac{1}{t}\sum_{j=1}^{t} R_j$$

$$R_{OS}^2 = 1 - \frac{\sum_{k=1}^{n_2}(R_{n_1+k} - \hat{R}_{n_1+k})^2}{\sum_{k=1}^{n_2}(R_{n_1+k} - \bar{R}_{n_1+k})^2}$$

相关指标查询路径

因子研究系列 → 已实现指标 → 指数已实现指标表

- 已实现偏度

实证分析——定量分析

由于社会经济数据的非实验性质，利用定量分析和计量建模的科学研究方法显得尤其重要。大数据时代带来的“数据革命”，更是凸显了定量分析的必要性和重要性。定量分析的研究逻辑通常为“找变量 — 建模型 — 假设检验”。

找变量

- 选好变量
- 借助理论选择变量

建模型

- 问题导向
- 注意适用条件
- 简约性原则
- 模型可解释性
- 与现实的关联度
- 数据质量

假设检验

- 模型证据与数据证据、统计假设和经济假说的差别
- 统计关系和经济学因果关系的差别
- 正确使用统计方法

参考资料：洪永淼,汪寿阳.数学、模型与经济思想[J].管理世界,2020.

实证分析——定量分析

实证分析——定量分析的创新变化

创新案例一：文本分析

在不断增长的社会数据中，文本数据扮演着重要的角色。得益于数据技术的迅猛发展以及数字设备的广泛应用，政策文献、社交媒体、法律文书、档案史料、访谈资料、宣传文案、消费者评论等多样化的文本数据逐渐得到发掘，为研究者提供了更加丰富的实证素材和更为多元的研究视角。

文本数据的特征

数据来源多样化

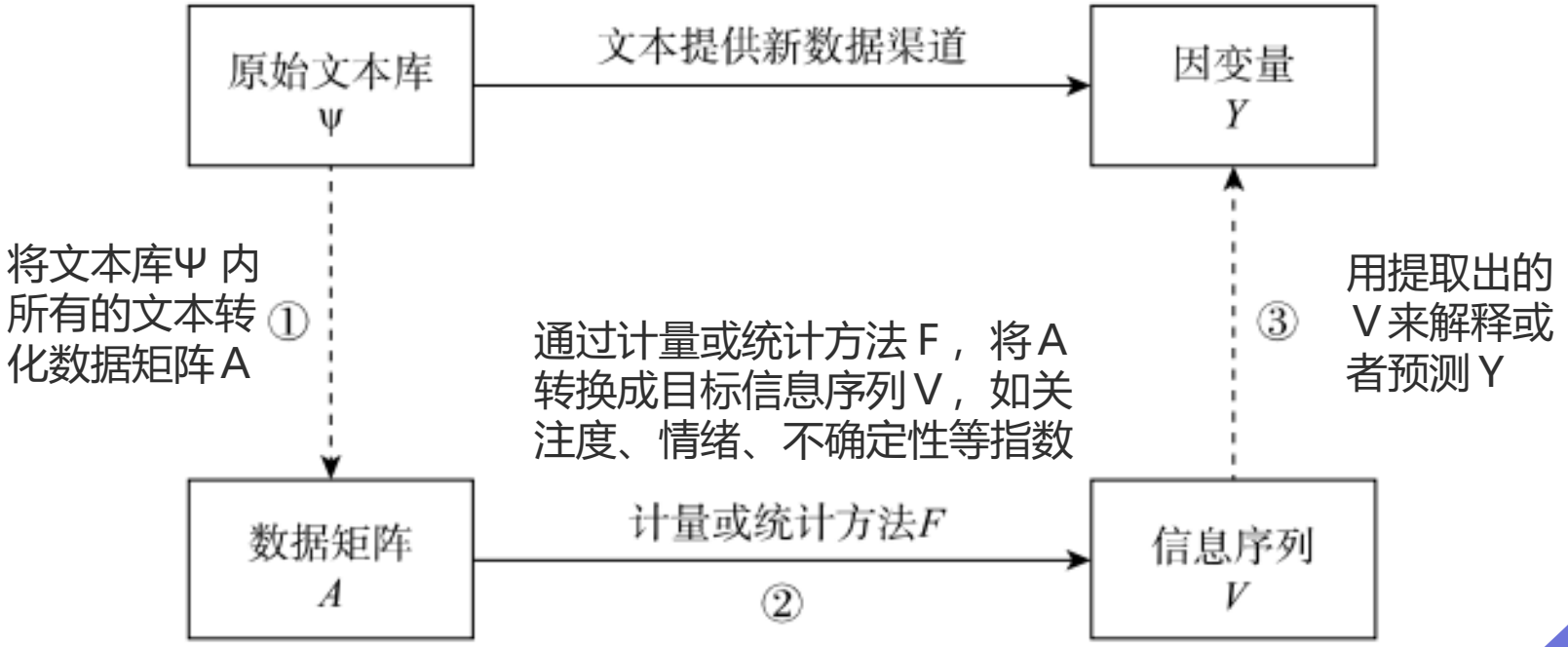
推特，微博，论坛帖子，微信公众号，上市公司年报，电话录音文稿，招聘广告，公司年报、季报、公告，IPO招股说明书，分析师研究报告，会议纪要，有影响力的政治、经济、金融领域人物的演讲，央行等政府机构定期和不定期发布的各类信息，等等。

数据体量几何级增长

随着文本信息从纸质媒介向以互联网为媒介的方式转移，文本数据收集和传输成本大幅度降低，为计算机领域的自然语言处理方法（NLP）提供了应用场景。

时频高

经济和金融领域数据多为年度、季度、月度、周度数据，而文本数据的频率可以高达秒级（如网民在网络平台上发布的消息和观点的时间颗粒度），为高频研究提供了数据基础。



参考资料：沈艳,陈赟,黄卓.文本大数据分析在经济学和金融学中的应用:一个文献综述[J].经济学(季刊),2019.

实证分析——定量分析的创新变化

创新案例一：文本分析

在经济学中的应用

- 经济政策不确定指数
- 行业分类
- 度量和预测经济周期
- 媒体报道偏差
- 量化央行政策沟通

在金融学中的应用

- 关注度指数（投资者关注度、媒体关注度）
- 文本情绪（媒体情绪（语调）、管理层语调、投资者情绪）
- 文本可读性
- 新闻隐含波动指数
- 投资者分歧

类　型	具体方法	适用的特征	优　点	缺　点	主要代表文献
字典法	基于词汇分类表的词频统计方法	语调；可读性；管理者特征；风险；竞争；前瞻性；研发信息	1.方法简单，应用极为广泛；2. 可以度量大多数的文本特征，尤其是语调、可读性及前瞻性等文本特征；3. 相关研究文献极多，可研究的主题和范围很大	1.受到语言种类和专业领域的限制；2. 词语存在一词多义、语义模糊和上下文语境问题，引起度量偏差；3. 不具有自适应的学习能力；4. 难以识别虚假性特征	Henry（2008）；Li（2008）；Kothari等（2009）；Feldman等（2010）；Loughran和Mcdonald（2011）；Li等（2013）；Hoberg和Phillips（2016）
无监督的机器学习方法	贝叶斯模型；向量机方法；K-近邻算法；决策树算法；随机森林算法；其他的人工智能学习方法	语调；可读性；重复性；风险；竞争；虚假性；融资约束	1.可以基于不同的训练文本进行模型构建，具有动态和适应性强的特点；2. 适宜度量竞争、虚假性、语调、相似性等特征	1.相比字典法，该方法难度较大；2. 需要预先分类，分类过程中可能存在偏误	Antweiler和Frank（2004）；Li（2010b）；Cecchini等（2010）；Huang和Li（2011）；Humpherys等（2011）；Zhou和Kapoor（2011）；Purda和Skillicorn（2015）
有监督的机器学习方法	LDA文档主题生成模型	风险	1.无须预设分类集和培训样本，能够减少人工识别误差；2. 能够自我学习，归纳出文本的内在特征	1.方法十分复杂、难度大；2. 目前处于应用初期，应用面较窄；3. 目前只限于度量“风险”特征	Bao和Datta（2014）；Campbell等（2014）

参考资料：沈艳,陈赟,黄卓.文本大数据分析在经济学和金融学中的应用:一个文献综述[J].经济学(季刊),2019.

实证分析——定量分析的创新变化

CSMAR与文本分析——中国股吧舆情研究数据库

中国股吧舆情研究数据库利用深度学习模型对网络股票贴吧的股评文本进行判断，整理出各上市公司股评的情绪和观点态度，并对此进行筛选、量化和统计，为您提供按上市公司、时间和发帖者特征分类统计的量化舆情数据。中国股吧舆情研究数据库的推出，旨在为研究者进行量化舆情分析提供帮助，促进金融领域量化研究的发展。

- 投资者情绪总统计表
- 文本情绪总统计表
- 投资者情绪分类筛选统计
  - 投资者情绪筛选统计表（吧龄）
  - 投资者情绪筛选统计表（发帖终端）
  - 投资者情绪筛选统计表（发帖时间）
  - 投资者情绪筛选统计表（净评论）
- 文本情绪分类筛选统计
  - 文本情绪筛选统计表（吧龄）
  - 文本情绪筛选统计表（发帖终端）
  - 文本情绪筛选统计表（发帖时间）
  - 文本情绪筛选统计表（净评论）

中国股吧舆情研究数据库

- 上线时间：2019.01
- 更新频率：日
- 起始时间：2010-
- 表字段数：10张表，438个字段

- 数据来源：东方财富股吧的评论文本信息

- 特色指标：投资者情绪和文本情绪信息，发帖者的吧龄、活跃度、影响力等方面的指标

看涨情绪指数a	看涨情绪指数b	情绪一致性指数
0.733333	1.540445	0.320131
-1	-0.693147	1
1	1.098612	1
1	0.693147	1

实证分析——定量分析的创新变化

CSMAR与文本分析——中国股吧舆情研究数据库

范文：网络情绪能够影响股市羊群效应吗?[J].财经问题研究,2019.

内容提要

为检验网络情绪对股市羊群效应的影响，本文选取创业板指数成分股收益率数据和东方财富网股吧发帖文本对中小投资者进行经验研究。本文首先利用文本挖掘技术建立一个适用于中小投资者的金融情感词库，构建三项网络情绪指标，然后应用分位数回归模型分析网络情绪对股市羊群效应的影响。研究发现，第一，中小投资者的网络分歧情绪能够整体上减弱过度自信从而减轻逆向羊群效应，参与热情能够促进信息交流从而减轻正向羊群效应，但看涨情绪对羊群效应没有显著影响。第二，相比于股市下跌时期，股市上涨时期的羊群效应受参与热情的影响更大，相比于股市上涨时期，股市下跌时期的羊群效应受分歧情绪的影响更大。第三，网络信息互动并不一定总是降低投资者的有限理性程度。因此，有必要利用网络情绪信息监测中小投资者的羊群行为，给予其适当的理性引导，从而促进中国金融市场有效运行和健康发展。

中小投资者网络情绪指标

文本情绪指标。基于词典分类法构造**分歧指数**。

分歧指数反映的是股民发帖的不一致程度，是每个交易日中情绪值的方差。分歧指数越高，代表股民发帖中对看涨、中立、看跌情绪的分散程度越大。

相关指标获取：CSMAR中国股吧舆情研究数据库

投资者情绪表	证券代码、证券简称、发帖日期、发帖人影响力、发帖人吧龄、帖子数量、阅读量、点赞量、看涨帖子数量、中立帖子数量、看跌帖子数量、投资者情绪指数、情绪一致指数等统计指标	2010～	日

羊群效应的度量

羊群效应的相对大小通常由**个股收益率的分散度**来度量。文章利用横截面收益率绝对偏差(CSAD) 来衡量羊群效应的相对大小。

相关指标获取：CSMAR行为金融研究数据库

股票市场羊群效应指标表（日）	参考宋军（2001）、孙培源（2002）构建中国股票市场羊群效应测度指标，包括：市场类型、CSSD（考虑再投资等权平均法）、CSAD（考虑再投资等权平均法）等	1990~	日

实证分析——定量分析的创新变化

CSMAR与文本分析——数字经济研究数据库

数字经济研究数据库提供经济发展总体情况（国民经济核算、人力投入、科技创新、固定资产投资）、数字产业化（通信业、互联网和相关服务、软件和信息技术服务业、电子信息制造业、共享经济）、产业数字化（农业数字化、工业数字化、数字金融、电子商务）、政策与资讯及上市公司数字化数据。

中国数字经济研究数据库

- 上线时间：2021.11
- 更新频率：年、季、月、日
- 起始时间：1949-
- 表字段数：104张表，1115个字段

- 数据来源：中国统计年鉴、中国电子信息产业统计年鉴、中国科技统计年鉴、中华人民共和国工业和信息化部、上市公司定期报告等。

- 特色指标：上市公司数字化转型程度（人工智能技术、区块链技术、云计算技术、大数据技术、数字技术应用）

证券代码	指标分类	指标名称	
002210	云计算技术	物联网	1
002210	数字技术应用	电子商务	4
002212	云计算技术	云计算	23
002212	云计算技术	信息物理系统	2
002212	大数据技术	数据挖掘	1

实证分析——定量分析的创新变化

CSMAR与文本分析——数字经济研究数据库

范文：基于文本挖掘的数字化水平与运营绩效研究[J].统计与信息论坛,2021.

内容提要

从微观角度研究企业数字化发展水平和企业运营绩效的动态关系，进而为中国数字产业发展和结构升级寻求微观经验基础。在对数字化水平影响企业运营绩效的传导机制分析基础上，选用2015-2019年上市企业的经营数据，并结合文本挖掘方法和中介效应模型实证分析了数字化水平对企业运营绩效的影响。

研究发现：企业数字化水平显著提升了企业运营绩效；在数字化水平影响企业运营绩效过程中，信息对称性水平表现出完全中介效应，商业模式创新水平起到部分中介效应，管理成本降低水平并未表现出中介效应，即数字化水平对运营绩效的提升作用主要是通过提升企业内部信息对称性水平和促进企业商业模式创新来实现的。因此，为提升数字化水平对企业运营绩效的促进作用，中国企业应努力实现数字产业转型升级，构建高效的动态数字信息分享平台，创新基于数字产业的商业模式并注重数字化进程中的成本控制。

变量性质	变量名称	变量符号	计算方法
被解释变量：运营绩效 Per	总资产收益率	Roa	$Roa=NR/0.5(TA_1+TA_2)$，其中 NR、TA_1、TA_2 分别表示净利润、年初总资产和年末总资产
	净资产收益率	Roe	$Roe=NR/0.5(TE_1+TE_2)$，其中 NR、TA_1、TA_2 分别表示净利润、年初净资产和年末净资产
核心解释变量：Dig	数字化水平	Diga	来源于文本挖掘
		Digb	财报附注中无形资产明细中与数字经济相关部分占无形资产总额的百分比
中介变量：Med	信息对称性水平	inf	$inf_{i,t}=Ar_{i,t}/TA_{i,t-1}=(Tar_{i,t}-Nar_{i,t})/TA_{i,t-1}$，其中 Tar、Nar 和 TA 分别表示企业总应计项目、应计项目和企业总资产
	商业模式创新水平	Sty	来源于文本挖掘
	管理成本降低水平	Cst	财务报表中管理费用总额和营业收入的比值
控制变量：Control	总资产增长水平	Toa	$Goa_{it}=(TA_{it}-TA_{i,t-1})/TA_{i,t-1}$，TA 表示总资产规模
	经营性现金流量	Nc	$Nc=\ln(Ncf)$，Ncf 表示企业的经营性现金净流量
	销售成本率	Sot	$Sot=Toc/sale$，Toc 表示企业营业成本，sale 表示企业销售收入
	资产负债水平	Deb	$Deb=Debt/TA$，Debt 和 TA 分别表示企业总负债和总资产
	固定资产百分比	Asr	$Asr=Fix/TA$，Fix 和 TA 分别表示企业固定资产净额和总资产
	第一股东持股比例	Fir	$Fir=N_1/TN$，N_1 和 TN 分别指第一大股东持有的流通股股数和该公司流通在外的总流通股股数
	审计报告性质	Aud	审计意见为标准无保留意见取值为 1，否则为 0
	股权性质	Stc	企业为国有控股时取值为 1；否则为 0
	两职合一性	Cob	总经理和董事长为一人担任取值为 1，否则为 0
	董事会独立性	Ine	独立董事人数和董事会总人数的比例

变量	数字化水平 Diga	商业模式创新水平 Sty
关键词 keyword	AI、GIS、ERP、OA、U9、RMR、互联网+、区块链、商务智能、智能办公、云计算、云储存、物联网、工业化 4.0、数据可视化、深度学习	O2O、生态协同、模式创新、PaaS、线上线下、创新模式、市场定位、盈利模式、销售模式、智能成本控制
合计	16	10

实证分析——定量分析的创新变化

CSMAR与文本分析——中国上市公司经营困境研究数据库

中国上市公司经营困境研究数据库从八个方面介绍公司经营困境，包括公司基本信息、特殊处理与特别转让情况、管理层治理分析（管理层信息披露情感分析、管理层治理能力、失信被执行人信息）、财务指标分析、融资约束、非效率投资、过度负债、财务困境。

中国上市公司经营困境研究数据库

- 上线时间：2021.11
- 更新频率：年、半年、日
- 起始时间：1990-
- 表字段数：16张表，299个字段

- 数据来源：主要为衍生数据，收录的数据来源包括上市公司发布的定期报告、临时公告以及中国执行信息公开网等。

- 特色指标：预警提示（融资约束、非效率投资、过度负债、财务困境）、文本分析（管理层语调、文本相似性）等

负面词汇数量	词汇总量	句子数量	文字数量	情感语调1
169	3595	99	7187	0.037
194	4598	88	9310	0.0333
539	8336	195	16465	0.0131
426	7205	177	14184	0.0447
213	4578	108	9317	0.0227

实证分析——定量分析的创新变化

CSMAR与文本分析——中国上市公司经营困境研究数据库

范文：管理层讨论与分析的语调操纵及其债券市场反应[J].管理世界,2022.

内容提要

本文研究了管理层讨论与分析（MD&A）语调的操纵行为及其债券市场反应。

研究发现，MD&A 异常积极语调与预警 Z值负相关，债务重组正相关，这表明 MD&A 异常积极语调暗示了企业较高的未来风险，这与语调的信息增量解释相悖，因此MD&A异常积极语调更可能是操纵的结果。

进一步研究发现，MD&A异常积极语调越大，债券信用评级越高，且该正向关系在与评级机构利益冲突大、信息透明度低的公司子样本中更显著；此外，债券投资者能够识别语调操纵行为，但随着债券市场公众投资者的参与，MD&A 异常积极语调与债券信用利差之间呈现出一定的负向关系，且这种负向关系在信息透明度低的企业组中更加显著。

◆ 对语调的定义

Tone：（积极词汇数-消极词汇数）/（积极词汇数+消极词汇数）

Tone1：（积极词汇数-消极词汇数）/（词汇总数）

相关指标获取：CSMAR上市公司经营困境研究数据库

管理层信息披露情感分析	证券代码、行业代码、行业名称、管理层讨论与分析内容、与前一年相比文本相似度、正面词汇数量、负面词汇数量、词汇总量、句子数量、文字数量、情感语调1、情感语调2、管理层盈利预测类型编码、管理层盈利预测类型等	2017-	半年

◆ 分解异常语调

NTone表示有关基本面的中性语调，即正常语调；ABTone表示异常语调，代表管理层对语调的战略选择，可能是增量信息抑或操纵结果。

NTone与ABTone 分别是下方回归模型的预测值与残差。

$$Tone_{it} = \alpha + \beta_0 Roa_{it} + \beta_1 Ret_{it} + \beta_2 Size_{it} + \beta_3 MV_{it} + \beta_4 Std\_Ret_{it} + \beta_5 Std\_Roa_{it} + \beta_6 Age_{it} + \beta_7 Loss_{it} + \beta_8 \Delta Roa_{it} + \delta_{ht} + \gamma_{pt} + \varepsilon_{it}$$

◆ 检验异常语调的影响

异常语调与企业破产风险与债务重组的关系、与信用评级的关系、与债券利差的关系。

实证分析——定量分析的创新变化

CSMAR与文本分析——WinGo财经文本数据平台

中国上市公司文本数据库

词频子库
- 财务报告 热
- 董事会报告章节
- 管理层讨论与分析章节(全文) 热
- 管理层讨论与分析章节(未来展望) 热
- 审计报告
- 财务报表附注
- IPO招股说明书
- 内部控制评价报告
- 业绩说明会 热
- 社会责任报告 热
- 互动易 新
- 问询函 新
- 公告总库 新

相似词工具
- 深度学习相似词 热
- 同义词词林
- 语义相似词

自定义特征工具
- 自定义特征

文本特征子库
- 文本相似性 热
- 文本相似性网络 热
- 语调 热
- 可读性 热
- 创新 热
- 风险
- 前瞻性
- 竞争战略
- 问询函问答相似度 新
- 区块链研究 新
- 政府-上市公司采购合同 新

会计金融指标子库
- 应计盈余管理
- 真实盈余管理
- 会计稳健性
- 会计信息可比性
- 分析师跟踪数量
- 分析师预测误差
- 分析师预测分歧度
- 事件研究
- 股价同步性
- 股价崩盘风险
- 中美上市公司估值对比系统

中国政府文本数据库

词频子库
- 政府工作报告（国务院）
- 政府工作报告（省级行政区）
- 政府工作报告（地级行政区） 热

相似词工具
- 深度学习相似词 热

自定义特征工具
- 自定义特征

文本特征子库
- 文本相似性 热
- 文本相似性网络 热
- 政府官员特征 新
- 中国地方政府采购合同 新

美国上市公司文本数据库

词频子库
- 年度财务报告（10-K等） 热
- 季度财务报告（10-Q等）
- MD&A（年报） 热
- MD&A（季报）
- 风险章节（年报） 热
- 风险章节（季报）
- IPO招股说明书
- 盈余电话会议（文档层面） 热
- 盈余电话会议（段落层面） 热

相似词工具
- 深度学习相似词 热
- 传统知识库同义词

自定义特征工具
- 自定义特征

文本特征子库
- 语调 热
- 可读性 热
- 文本相似性 热
- LIWC 新
- 姓氏-文化集群 新

新冠疫情数据库

上市公司疫情新闻
- 上市公司疫情新闻 新

全球疫情
- 全球疫情 新

专利数据库

专利数据库
- 专利文本指标
- 专利基本信息 新
- 公司专利个数 新
- 专利质量指标 新
- 公司专利质量指标 新

在线服务
- 中文分词 热
- PDF解析
- LDA主题模型 热
- STM主题模型
- Word2vec模型
- Doc2vec模型

WinGo是中国首家基于中美上市公司披露文本的人工智能财经数据平台。平台从学术研究需求出发，聚焦于中美海量财经文本数据。针对两国截然不同的文本披露规则和财经文本特点，平台应用**自然语言处理、深度学习和人工智能技术对财经文本进行深度加工**，给学者们提供了多种类型文本的**词频、相似词、文本特征**等现成的数据，从而大幅降低广大研究和分析人员的研究成本。

操作视频

实证分析——定量分析的创新变化

创新案例二：混频数据模型的应用

大数据使得构建高频化的重要经济指数或经济变量成为可能，这些高频宏观经济变量对研究宏观经济的档期或近期的运行状况、进行经济预测，具有重要的现实意义。

基础混频（MIDAS）模型

Ghysels et al.（2004）提出的基础 MIDAS(m,K) 能直接将低频数据 y_t 和高频数据 $x_t^{(m)}$ 通过参数化的多项式权重 $B(L^{1/m};\theta)$ 整合成如下的单方程的回归模型：

$$y_t = \beta_0 + \beta_1 B(L^{1/m};\theta)x_t^{(m)} + \varepsilon_t \quad (1)$$

其中，m 表示混频数据的倍差，$B(L^{1/m};\theta) = \sum_{k=1}^{K} w(k;\theta)L^{(k-1)/m}$，$L^{i/m}x_t^{(m)} = x_{t-i/m}^{(m)}$，$i = 0,1,\cdots,K-1$，$K$ 为用于参数化权重函数的高频数据滞后阶数。本文所有 MIDAS 类预测模型的估计都选取了两参数的指数 Almon 多项式权重函数，且对权重函数中的参数都进行了 $\theta_1 \leqslant 300$，$\theta_2 < 0$ 的约束处理，以满足宏观经济分析与预测所需的权重形式，且能保证权重函数为正，使方程获得零逼近误差的良好性质（Ghysels & Valkanov，2006；Clements & Galvão，2008），其具体形式为：①

$$w(k;\theta) = \frac{e^{(\theta_1 k+\theta_2 k^2)}}{\sum_{k=1}^{K} e^{(\theta_1 k+\theta_2 k^2)}} \quad (2)$$

参考资料：刘汉,刘金全.中国宏观经济总量的实时预报与短期预测——基于混频数据预测模型的实证研究[J].经济研究,2011.

应用场景

O1 经济预测与影响分析：宏观、行业、股票市场等

O2 指数编制：金融景气指数、金融稳定指数等

传统的计量经济模型都是基于同频数据进行建模分析，否则将面临模型无法识别的情况。然而，在日常经济环境中，受统计口径和方式的影响，不同类型的数据其频率存在一定差异。

- ✓ 股票、期货、商品价格、客流量等大多为日度甚至时度数据；
- ✓ 价格指数、供应量、存栏量等大多为月度数据；
- ✓ 国民生产总值、产业增加值等大多为季度数据；
- ✓ 人口数、固定资产投资流量、种植面积等大多为年度数据。

传统分析中，一般将涉及的混频数据先转换为同频数据，但存在高频数据损失有效信息或由于人为操作增加无效信息从而增加误差等问题，进而可能会对模型估计、策略选择带来影响。

传统分析中，混频数据转换为同频数据常用方法：

- ✓ 通过计算均值或取离散点替代等算法将高频数据降频；
- ✓ 通过拟合、插值法或桥接模型法等算法将低频数据升频。

如何基于混频数据开展建模分析，成为了学术界的一个研究热点。混频数据模型则能完美避免上述传统分析中的问题。

混频模型的优势：

- ✓ 一方面，其能够将不同频率的变量纳入同一模型，充分挖掘高频数据的有效信息，综合考虑高低频变量的滞后阶数和权重函数进行建模；
- ✓ 另一方面，其能避免传统计量模型在预测方面的不足和假设要求，可以根据解释变量的超前发生性进行提前预测。

参考资料：吴培, 李哲敏. 混频数据模型应用研究现状及展望[J]. 统计与决策, 2021.

实证分析——定量分析的创新变化

创新案例二：混频数据模型的应用

经济预测与影响分析

基于混频数据的宏观经济组合预测模型与实证. 统计与决策, 2020.

文章利用我国日度金融数据和月/季度宏观经济数据，从伪样本外预测的角度，构建混频数据抽样模型（MIDAS），并加入金融、经济领先因子，对比四类组合预测模型对宏观经济的预测精度。结果显示：组合预测模型能减少对宏观经济预测的系统误差，提高预测精度。

日度数据	商品指标、公司风险指标、股票指标、政府债券指标、外汇汇率指标
月度数据	生产资料订单指数、耐用品采购订单指数、PMI、非制造业PMI、生产资料库存指数、出口价格指数、PPT、CPI
季度数据	城镇失业率、城镇居民人均可支配收入、城镇居民人均消费性支出、中小企业发展指数、GDP累计同比贡献率、固定资产投资价格指数、消费者信心指数、当期收入感受指数、未来收入信心指数

基于混频数据的社会物流成本预测. 统计与决策, 2020.

降低社会物流成本对提升国民经济运行效率至关重要。文章从宏观层面出发，基于若干宏观指标对社会物流总费用及其在GDP中的占比进行预测，为解决不同频率数据建模的问题，选用ADL-MIDAS模型方法进行预测分析。从模型应用效果看，ADL-MIDAS模型能够有效提高估计精度。

被解释变量	1992年-2018年年度社会物流总费用
	1992年-2018年年度社会物流总费用在GDP中占比
解释变量	1992M1-2019M11月度全社会用电量
	1992Q1-2019Q3季度GDP
	1992Q1-2019Q3季度第三产业在GDP中占比
外生变量	1992年-2018年年度公路运输量在总货运量中占比

混频投资者情绪与股票价格行为. 管理科学学报, 2018.

文章采用混频数据抽样模型(MIDAS)研究了混频投资者情绪对中国股市收益率及其波动的影响。通过构建日度、周度及月度这三种不同频率的投资者情绪，实证结果发现，混频情绪对当期收益率及其波动都存在显著的正向影响，并且与传统回归模型相比，MIDAS 模型具有更强的解释能力。

被解释变量	2010年5月-2016年6月上证综指股市收益率与已实现波动
投资者情绪	市盈率、换手率、腾落比率、融资融券比的日度、周度、月度指标

实证分析——定量分析的创新变化

创新案例二：混频数据模型的应用

指数编制

中国混频金融状况指数的构建. 统计与决策, 2017.

文章从货币政策经济增长目标出发，构建了新的混频金融状况指数(MFFCI)编制公式，使用MF-VAR模型，测算了金融状况变量的混频权重系数，实证编制和应用了中国MFFCI，同时与同频金融状况指数(SFFCI)进行了比较分析。结果表明，MFFCI无论与GR的相关性、因果关系，还是对GR的领先性和预测能力，都比SFFCI好，说明MFFCI更适合中国。

季度数据	1999Q1-2015Q3国内生产总值(GDP)
月度数据	1999M1-2015M9货币供应量、利率、人民币汇率、股票价格、房地产价格

金融稳定是否应纳入中国货币政策目标. 南方经济, 2019.

文章首先使用混频动态因子模型（MF-DFM）构建中国首个混频金融稳定指数（MF-FSI），接着把MF-FSI作为金融稳定的代理变量，比较分析纳入与不纳入金融稳定的中国货币政策损失函数差异。结果表明：中国混频金融稳定指数是金融稳定的一个实时性有效测度指标；中国货币政策目标应纳入金融稳定,以减少货币政策福利损失。

季度数据	不良贷款、证券化率、住房贷款、外汇储备量、M2、财政赤字
月度数据	存贷款比例、社会融资规模、上证综合指数、沪深两市股价、实际利率

中国电力景气指数的混频非对称测度. 管理科学学报, 2018.

文章构建了能够同时分析季、月 2 种频率的 MF-MS-SW 模型，选择 21 个指标组成的先行、一致和滞后混频样本数据，构建中国混频电力景气指数及预警信号系统。结果表明：构建的 MF-MS-SW 计量模型较好地刻画了中国电力景气指数的波动特征，具有多频率和非线性的特征，且与中国总体经济发展状态具有高度的一致性，可以用来预警与预测。

季度数据	固定资产投资价格指数、国内生产总值、第二产业增加值、平板玻璃产销率、建筑业增加值
月度数据	全社会用电量、发电量、火力发电量、工业用电量、城乡居民生活用电量、商品房销售面积、电力生产主营业务收入

实证分析——定量分析的创新变化

CSMAR与混频数据——相关数据系列

宏观经济数据

统计频率较低，多包含月/季/年度频率数据，其数据稳定性较强。

经济研究系列——中国宏观经济研究数据库

反映整个国民经济与社会发展状况的统计指标，数据内容包括国内生产总值、固定资产投资、居民收入与消费、财政收支、价格指数、国内贸易、对外贸易、就业与工资、能源生产与消费、环境保护、农业、工业、建筑业、金融业、邮电和运输业、旅游、国际收支、保险业共18个领域的数据。

经济研究系列——国际宏观综合数据库

国际宏观综合数据库包括世界经济总览，世界经济展望预测，景气指数，金融，制造业，军事，全球性问题，物流绩效八大模块，11个二级节点，设计了33张表，覆盖全球180多个国家的宏观数据，收录的数据来源于联合国统计署、IFM，全球清算银行等。数据最早为1896年。

行业研究系列

包括能源、房地产、通信、汽车、交通运输、保险、钢铁、有色金属、医药、新能源、石油化工、农林牧渔、物流、旅游、土地交易、信托行业等16个库

金融数据

统计频率较高，多包含日/小时频率数据，其数据信息量较多。

专题研究系列——中国投资者情绪指标研究数据库

甄选能够代表中国投资者情绪指标，收录了中国证券登记结算公司对投资者账户的周统计信息、及投资者保证金变化信息，中国证券保护网统计的投资者信心，及中国景气检测中心绘制的消费者、企业家等的信心，BW综合情绪指数主要因子、国内主流学者研究的情绪指数及大量代理情绪因子等。

股票市场系列

包括股票市场交易、融资融券、大笔交易、大宗交易、市场指数、股权分置改革、停复牌、特殊处理与特别转让、股票市场衍生指标、转融通、沪港通与深港通等11个数据库。收集了自1990年上海证券交易所和深圳证券交易所成立以来中国上市公司的资料、全部交易数据。

货币市场系列

包括外汇市场、黄金市场交易、货币市场与政策工具等3个数据库。

衍生市场系列

包括商品期货、权证市场、股指期货、国债期货、个股期权、股指基权、商品期权等数据库。

实证分析——定量分析的创新变化

创新案例三：机器学习与经济管理学研究

机器学习作为大数据研究的标志性产物，在处理多变量问题时有着较好的表现，因此越来越广泛地用于经济管理学研究中。机器学习使得研究者获得了以前通过人工投入无法获得的海量数据，检验了一些依靠传统方法无法有效的假设，这在一定程度上拓展了经济管理学研究的边界。近年来，研究者将机器学习方法尝试性地应用到探讨收入差距演变、金融市场波动、异质性员工的个性行为特征与企业人力资源管理和绩效提升、全球气候保护、公共政策分析等现实复杂的问题中，取得了比传统计量实证方法等更为深入的分析效果。

真实经济/管理问题涉及许多变量

变量的影响是高度非线性的，或存在变量之间的交互项

经济预测比统计推断更重要

如果研究满足（但不限于）上述三个条件，那么机器学习技术就是有价值的

实证分析——定量分析的创新变化

创新案例三：机器学习与经济管理学研究

机器学习技术在经济管理学中的应用

01
数据生成

机器学习可以帮助学者获得以前很难或无法获得的数据，进而对一些更具挑战性的假设进行检验。
通过机器学习获得数据的主要方式是**文本挖掘**及**图像识别**。

02
预测

机器学习可以更有效地探索变量之间的相关性，进而做出较为精准的预测。
计量经济学的目的不仅是预测，更在于解释现实中的现象以找到背后规律，用来预测的函数形式越简单越好。机器学习则恰恰相反，只要这个函数能够很好地模拟现实，哪怕函数形式再复杂也无所谓。在这一点上，机器学习不拘泥于“可解释性”，灵活地选择函数形式进行拟合数据，这使得其预测能力强过了计量经济学传统方法。

03
因果识别

社会科学特别是经济学实证研究的核心是因果识别。由于机器学习在预测方面的优势，它可以被用来预测反事实进而获得因果效应。

实证分析——定量分析的创新变化

CSMAR与机器学习方法——经典数据与创新方法

范文：高管个人特征与公司业绩——基于机器学习的经验证据[J].管理科学学报,2020.

内容提要

文章首次采用机器学习算法中的 Boosting 回归树，全面考察了多维度高管特征对公司业绩的预测性，以我国2008年 ~ 2016年的上市公司为样本，研究了高管的多维个人特征是否能预测公司业绩，并进一步分析了对公司业绩预测能力较强的高管个人特征及其预测模式。

研究发现: 1) 整体而言，在我国公司 CEO 和董事长的特征对公司业绩的预测能力较弱; 2) 在众多高管个人特征之中，高管持股比例和年龄对公司业绩的预测能力较强; 3) 高管持股比例和年龄与公司业绩之间的关联都呈现出非线性的特点，与以往的理论较为吻合。本研究不仅利用机器学习方法从一个更为全面的视角对中国的高管特征进行了研究，也为公司高管聘任和激励机制设计等方面提供了有益的启发。

公司绩效	
adEBITDAper	经过行业中值调整的 EBITDA 占总资产的比例
adTobinQ	经过行业中值调整的托宾 Q 值

CEO 个人特征	
Duality	CEO 和董事长是否两职分离，1 为是
Gender．c	是否为女性，1 为是
Age．c	年龄的对数
Share．c	年末持股比重
Parttime．c	是否在其他公司兼职，1 为是
ProFun．c	是否有生产、技术、设计职能经验，1 为是
MgtFun．c	是否有市场、战略、人力管理职能经验，1 为是
SkiFun．c	是否有财务、法律职能经验，1 为是
Oversea．c	是否有海外工作、求学经历，1 为是
GovBack．c	是否有政府工作经历，1 为是
Academic．c	是否有学术研究经历，1 为是
FinBack．c	是否有金融行业工作经历，1 为是
董事长个人特征	
Duality	CEO 和董事长是否两职分离，1 为是
Gender．b	是否为女性，1 为是
Age．b	年龄的对数
Share．b	年末持股比重
Parttime．b	是否在其他公司兼职，1 为是
ProFun．b	是否有生产、技术、设计职能经验，1 为是
MgtFun．b	是否有市场、战略、人力管理职能经验，1 为是
SkiFun．b	是否有财务、法律职能经验，1 为是
Oversea．b	是否有海外工作、求学经历，1 为是
GovBack．b	是否有政府工作经历，1 为是
Academic．b	是否有学术研究经历，1 为是
FinBack．b	是否有金融行业工作经历，1 为是

公司层面变量	
Ln Asset	总资产的对数
State．share	国有股份占比
Leverage	资产负债率
PPE	固定资产占总资产的比例
Esty	公司寿命的对数

公司研究系列

上市公司人物特征

公司研究系列

实证分析——定量分析的创新变化

CSMAR与机器学习方法——经典数据与创新方法

范文：“倾斜性”政策、生产部门变迁与南北地区发展差异-来自机器学习的因果推断[J].财经研究,2022.

内容提要

“倾斜性”政策对各区域形成优势互补、共同发展的新格局发挥了重要作用。借助于这 些区域政策,地方政府会更积极地介入经济活动,其对生产部门结构变迁的冲击具有时空异质性，这为讨论南北发展分化提供了一种新的线索。因此,文章构建了“倾斜性”政策、两部门结构与区域 发展的均衡模型 ，使用机器学习算法对 2003-2018 年 277 座地级及以上城市数据开展统计分析，尤其采用**因果森林（causal forest）算法**对“倾斜性”政策的阶段性效果进行了有效估计，并有效地解释了南北发展分化新原因。

机器学习方法的优势

与传统政策评估模型需要进行参数化设置不同，机器学习分析非参数模型有巨大优势。作为有监督机器学习（如随机森林）往经济可解释性方向突破的最新进展，**因果森林**可大幅提高政策评估的可信度。这种以数据驱动方式获得因果推断结果，其可信度来源于两方面：一方面，因果森林通过大量的子样本产生平均处理效应（ATE）的平滑估计，其预测值以真实条件平均处理效应（CATE）为中心；另一方面，有学者证明了因果森林算法估计结果满足渐近正态性。

研究数据

实证样本选自2003-2018年277座地级及以上的城市，具体数据来源于《中国城市统计年鉴》、CSMAR数据库。

变量名称	具体指标
地理区位	南北区位，以秦岭—淮河为界划分南北方城市[1]
	板块区域，按照年鉴划分四大区域
	港口距离，这里基于可利用港口的便利性来衡量
经济发展禀赋	2003 年人均 *GDP*
金融状况	信贷增速
人力资本	人力资本
运输条件	货运量
投资规模	固定资产投资
政府行为指数	建设用地面积增速
	建设用地面积变异系数
	财政收支比
城市规模	辖区人口
两部门结构	可贸易部门与不可贸易部门[2]之比
城市引力流	2004—2011 年引力流
	2012—2018 年引力流

研究结果

①旨在协调区域平衡发展的政策在南方地区的效果优于北方，尤其在2012 年以后，效果更明显。

②“倾斜性”政策会提高欠发达地区 不可贸易部门的份额,引致长期生产率提升速度比可贸易部门更缓慢，尽管这种现象在我国普遍存在，但北方城市相对更严重，这是南北区域发展差距扩大的重要原因。

因此，考虑到传统区域政策的异质性和不可持续特征，未来区域政策的重心应侧重于如何提升欠发达地区人力资本积累以及企业的贸易参与度。

03

CSMAR最新数据资源

- 最新数据库简介

最新数据库简介

2019

- 海外直接投资
- 股吧舆情
- 信托行业
- 行业财务指标
- 基金经理人物特征
- 审计研究
- 文化研究
- 行为金融

2020

- 对赌协议
- 金融机构分支机构
- 新冠疫情主题
  - -新冠疫情与经济研究
  - -PHEIC
- 会计信息质量
- 已实现指标
- 国际宏观

2021

- 上市公司环境研究
- 供应链研究
- 银行治理研究
- 绿色专利研究
- 经济内循环研究
- 经济地理研究
- 润灵环球ESG评级
- 商道融绿ESG评级
- 金融科技
- 数字经济
- 经营困境

2022

- 碳中和

更多新库即将上线

敬请关注！

最新数据库简介

2021新库发布会集锦

01

供应链研究
上市公司环境研究

02

银行治理
绿色专利

03

内循环研究
经济地理

04

润灵环球ESG
商道融绿ESG

05

金融科技

06

数字经济

07

经营困境

04

CSMAR近期动态

- 学术活动
- 科研资讯

学术活动

CSMAR实证研究主题工作坊

为帮助广大学子与科研工作者系统解决在实证研究与论文写作当中的各种难题，深圳希施玛数据科技有限公司联合各大高校，广邀名师，定期开展实证研究主题线上工作坊，以精彩纷呈的主题讲座，与广大学子共攀学术高峰。

往期精彩课程

数据加油站| CSMAR工作坊实证论文复刻讲座

(2022.03.19-2022.04.09)

由武汉大学联合深圳希施玛数据科技有限公司举办的系列讲座，每期讲座均邀请在国内外核心期刊发表过文章的作者，复刻权威论文，帮助研究者深入掌握经济管理学研究论文写作方法。

日期	时间	分享论文题目	主讲人
3月19日	09:30-11:30	Does short selling affect a firm's financial constraints，Journal of Corporate Finance，2020. 卖空会影响公司融资约束吗?	中国人民大学商学院教授 孟庆斌
3月26日	09:30-11:30	Differences of opinion, institutional bids, and IPO underpricing，Journal of Corporate Finance，2020. 意见分歧、机构竞价与IPO抑价	西南财经大学会计学院教授 高升好
4月9日	09:30-11:30	技术并购、市场反应与创新产出，南开管理评论，2021	南开大学商学院教授 姚颐

科研资讯

CSMAR
官方公众号

数据资源上新

权威文献解析

科研活动资讯

热门文章推荐

- CSMAR前沿探索：环境经济问题研究
- CSMAR前沿探索：把脉疫情之下的经济新态势
- CSMAR前沿探索：香港疫情数据感知报告
- 研究“经济内循环”该从何下手?
- 助推经济高质量发展？经济地理研究势在必行!
- 如何利用ESG评级数据做研究?
- 实证研究论文选题、文献综述撰写技巧，看这一篇就够了!
- 纯干货！实证分析中的数据与模型
- 在核心期刊发表论文是一种怎样的体验

Thank you

感谢参与

www.csmar.com

深圳希施玛数据科技有限公司