

征信大数据

理论与实践

CREDIT REFERENCE
& BIG DATA



通过征信大数据
对中国产业结构变迁进行分析

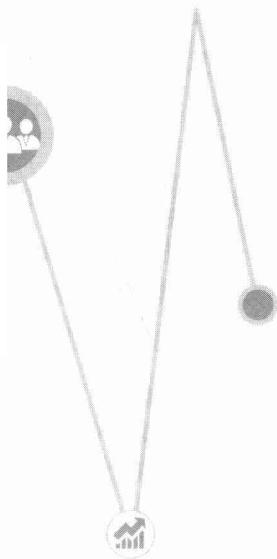
从各产业的微观表象
探究中国经济现代化转型的深层次脉搏

征信大数据

理论与实践

CREDIT REFERENCE
& BIG DATA

常州大学图书馆
藏书章



通过征信大数据
对中国产业结构变迁进行分析

从各产业的微观表象
探究中国经济现代化转型的深层次脉搏

责任编辑：肖 炜

责任校对：李俊英

责任印制：赵燕红

图书在版编目 (CIP) 数据

征信大数据：理论与实践 (Zhengxin Dashuju: Lilun yu Shijian) / 姚前等著. —北京：中国金融出版社，2018. 3

ISBN 978 - 7 - 5049 - 9412 - 7

I. ①征… II. ①姚… III. ①信用制度—研究—中国 IV. ①F832. 4

中国版本图书馆 CIP 数据核字 (2018) 第 017141 号

出版 **中国金融出版社**

发行

社址 北京市丰台区益泽路 2 号

市场开发部 (010)63266347, 63805472, 63439533 (传真)

网上书店 <http://www.chinafph.com>

(010)63286832, 63365686 (传真)

读者服务部 (010)66070833, 62568380

邮编 100071

经销 新华书店

印刷 北京市松源印刷有限公司

尺寸 155 毫米 × 230 毫米

印张 13

字数 195 千

版次 2018 年 3 月第 1 版

印次 2018 年 3 月第 1 次印刷

定价 50.00 元

ISBN 978 - 7 - 5049 - 9412 - 7

如出现印装错误本社负责调换 联系电话 (010) 63263947



中国的金融发展史就是一部“新金融”的历史，金融业的版图无时无刻不在演变、重塑。不断革新的金融工具、运行机制和参与主体塑造了不断变化的金融业态和格局。理念与技术的创新在推动金融结构演进、金融改革深化的同时，也为整个金融业的发展带来了机遇与挑战。

“新金融书系”是由上海新金融研究院（Shanghai Finance Institute, SFI）创设的书系，立足于创新的理念、前瞻的视角，追踪新金融发展足迹，探索金融发展新趋势，求解金融发展新问题，力图打造高端、权威、新锐的书系品牌，传递思想，启迪新知。

上海新金融研究院是一家非官方、非营利性的专业智库，致力于新金融领域的政策研究。研究院成立于2011年7月14日，由中国金融四十人论坛（China Finance 40 Forum, CF40）举办，与上海市黄浦区人民政府战略合作。研究院的宗旨是：探索国际金融发展新趋势，求解国内金融发展新问题，支持上海国际金融中心建设。

上海新金融研究院努力提供一流的研究产品和高层次、有实效的研讨活动，包括举办闭门研讨会、上海新金融年会、金融科技外滩峰会，开展课题研究，出版《新金融评论》、“新金融书系”等。

“中国金融四十人论坛”（CF40）是中国最具影响力的非官方、非营利性金融专业智库平台，专注于经济金融领域的政策研究与交流。论坛正式成员由40位40岁上下的金融精锐组成，即“40×40俱乐部”。CF40的宗旨是：以前瞻视野和探索精神，致力于夯实金融学术基础，研究金融领域前沿课题，推动中国金融业改革与发展。

序 言

信息是客观世界各种事物的特征的反映，数据则是信息的具体表现形式，或通过记录，或通过采集，或通过其他方式“捕捉”而得到。有人类的时代，即有数据。上古社会“结绳记事”，殷商时期“唯殷先人，有典有册”，均是鲜明写照。

数据之重要不言而喻。于企业，数据是运行基础。有了不同来源和渠道的各种数据，如人力资源数据、财务数据、客户数据、市场数据、经济数据等，企业才可能进行合理的生产经营、投资、销售和管理决策。数据是生产要素，对数据进行加工、分析和挖掘，可以得到有价值的知识，不仅能为企业的经营决策提供数据化的科学参考，还能让企业快速响应市场变化，提高管理效率，优化企业流程。信息化和数字化，已成为企业经营的基本战略。

于国家，就像《商君书·去强》所言：“强国知十三数：竟内仓府之数，壮男壮女之数，老弱之数，官士之数，以言说取食者之数，利民之数，马、牛、刍藁之数”。秦国之所以能统一六国，在于秦律，而秦律成功运行的基础则是绵密细致的国情统计体系。出土的秦简显示，秦国各级官府设有“计”的职位，专门负责数据统计，人口、田地、畜牧、天气等均记录在案，定期更新，并及时送往咸阳留存，于是中央集权足不出户，即可知天下。因此，“佚而治，约而详，不烦而功，治之至也。秦类之矣”（荀子·强国）。后续的王朝更替似乎也印证了《商君书》的警示：“欲强国，不知国十三数，地虽利，民虽众，国愈弱至削”。历史学家黄仁宇将近代以来中国的落后归咎于缺乏“数目字管理”，认为这是阻碍中国现代化转型的重



要症结。

数据重要，数据分析更重要。“有借必有贷，借贷必相等”的复式簿记法本质上即是一种数据分析方法。它创造了一种科学的从会计凭证中获取有关经济往来和经营成果重要信息的方法，为现代企业和商业社会的形成准备了一个完美的经济信息系统。德国哲学家歌德（Goethe）把它赞誉为“人类智慧的绝妙创造之一，每一个精明的商人从事经营活动都必须利用它”。不仅如此，经济学家和统计学家们还建立了计量经济学与实验经济学，提出各种计量分析框架、思想和理论，如面板数据计量经济学、微观计量经济学、金融计量经济学、宏观计量经济学等，为经济数据的实证分析提供重要的方法论基础。

时至今日，随着信息技术和人类生产生活交汇融合，互联网快速普及和云计算的发展，数据呈现出全新的表象：大数据（Big Data）。大数据具有数据规模海量（Volume）、数据流转快速（Velocity）、数据类型多样（Variety）、数据价值巨大（Value）等四大特征。数据规模海量和流转快速意味着无法在有限时间内用传统IT技术和硬件工具对其进行感知、获取、管理、处理和服务，需要新的特殊技术和硬件系统，以有效地存储和处理大量的经过容忍时间内的数据，比如大规模并行处理（MPP）数据库、分布式文件系统、分布式数据库、云计算平台和可扩展的存储系统等；数据类型多样意味着数据变量高维，传统的基于低维变量的数据分析技术不再适用，而基于高维变量的大数据分析技术将成为重点。数据价值巨大意味着从数据中获取价值是大数据分析的核心，需要专家知识和理论建模的进一步提高，还需分类分析、聚类分析、关联分析、神经网络、机器学习等数据挖掘技术的进一步成熟和完善。

大数据时代的来临将对经济发展、社会生活和国家治理产生重大影响。习近平总书记指出：“大数据发展日新月异，我们应该审时

度势、精心谋划、超前布局、力争主动，深入了解大数据发展现状和趋势及其对经济社会发展的影响，分析我国大数据发展取得的成绩和存在的问题，推动实施国家大数据战略，加快完善数字基础设施，推进数据资源整合和开放共享，保障数据安全，加快建设数字中国，更好服务我国经济社会发展和人民生活改善”，并强调“建设现代化经济体系离不开大数据发展和应用”，“要运用大数据提升国家治理现代化水平。要建立健全大数据辅助科学决策和社会治理的机制，推进政府管理和社会治理模式创新，实现政府决策科学化、社会治理精准化、公共服务高效化”。

而征信与大数据有着天然的联系，征信的业务都是围绕着数据展开。根据党中央、国务院关于加快社会信用体系建设、加快建设征信系统的重要战略部署，2004年初，人民银行开始组织商业银行建设全国集中统一的个人征信系统。2005年8月，个人征信系统完成与全国所有商业银行和部分有条件的农村信用社的联网运行，并于2006年1月在全国联网运行。企业征信方面，2004年，人民银行启动银行信贷登记咨询系统升级工作，将原来分布在337个城市的三级分布式数据库升级为全国集中统一的企业征信系统，实现数据在全国的大集中，统一对外服务。2005年12月，企业征信系统实现主要商业银行的全国联网运行。2006年6月末，企业征信系统实现所有中资、外资商业银行和有条件的农村信用社的全国联网运行，并于2006年7月底完成全国范围内与银行信贷登记系统的切换工作。根据《征信业管理条例》，央行征信系统依法采集征信数据。经过10多年来的建设和完善，央行征信系统的数据具有全样本、大覆盖、跨周期的特点，不仅数据量大，而且最重要的是包含丰富的价值，是典型的金融大数据。

征信大数据可以服务于三个目的：一是征信服务高效化。征信大数据有助于进一步提升征信在金融机构信用风险管理中的作用，



有效解决信贷市场的信息不对称问题，提高社会融资便利，缓解实体企业“融资难、融资贵”问题。二是社会治理精准化。比如在个人征信权益保护方面，基于征信大数据分析的异常查询实时监测模型，可以对个人征信系统的违规查询行为产生威慑作用，倒逼查询机构加强内部管理，合法使用信用信息，以保障信息主体的权益，促进征信市场健康发展。三是政府决策科学化。征信大数据可以深度地应用在分析和预测信贷市场整体发展状况和趋势、信贷市场结构和资产质量变化情况、重点行业的信贷业务、区域间信用经济发展和信贷活跃人口变迁情况等方面，从而为货币政策、产业政策和金融监管提供更加全面、实时和更有政策涵义的决策依据。

一言蔽之，征信大数据是一个数据金矿，它的“挖掘”将有助于新时代下现代化经济体系的建设以及国家治理现代化水平的提升。然而，遗憾的是，当前征信大数据方面的研究工作尚处于空白，鉴此，我们不揣浅陋，进行了这方面的尝试。

在理论篇，本书全面地研究了大数据概念、国外发展动态、我国发展现状和未来发展趋势，并从基础架构支持、数据采集、数据存储、数据计算、展现与交互、数据算法等多层次多角度对大数据技术展开分析，继而聚焦于征信大数据，揭示了征信大数据的特征、在经济分析的重要意义以及可探索的方向。

在实践篇，本书对征信大数据的应用进行了多方面探索，一是在个人征信权益保护方面，通过数据挖掘技术，从海量查询记录中，分析查询用户的行为模式，归纳出平稳型、增长型、小幅跳跃型、周期跳跃型四种异常查询实时监测模型（见本书第四章）；二是在研究人口迁徙方面，对 3.9 亿信息主体、48.8 亿条征信记录进行分析挖掘，全面刻画了改革开放以来全国劳动力人口迁徙的轨迹、特征及未来人口迁徙的趋势（本书第五章）；三是在评估城市信用方面，综合借贷双方重点关注贷款规模、获取信贷便捷度及信贷风险等三

个因素，对 50 亿条征信记录进行分析挖掘，采用层次分析法建立城市信用指数模型，对中国各城市的信用情况进行了综合评价，从而得出近 10 年来中国城市信用特点及发展趋势（本书第六章）；四是在产业经济分析方面，基于征信大数据分析，从信贷资源流动与配置入手，研究我国房地产业、制造业、批发零售业、住宿餐饮业、建筑业、交通运输、仓储和邮政业、农林牧渔业、采矿业、电力生产供应业、水利、环境和公共设施管理业等各细分行业的信贷结构，由此基于信贷视角，揭示了近十年来我国产业结构变迁路径，并分析了产业结构调整 and 信贷供给之间的关联性（本书第七章至第十八章）。

值得一提的是，本书的研究同时还受到了一本名著《美国产业结构》（詹姆斯·W·布罗克编著）的启发，该书对美国 12 个重要产业的历史、市场结构、市场行为、市场绩效和公共政策问题展开了详细分析。笔者研读该书之后，深感我们中国也需要这样的一本书，能够详细地分析中国各个产业的发展历史、结构特征与市场行为，从各产业的微观表象，去探究中国经济现代化转型的深层次脉搏，因为笔者深信，大江大河的气势磅礴是由一个个浪花的起起伏伏而构成，一个国家的现代化，必然会体现为各个微观层面的结构变迁。而恰好，实时海量、几乎覆盖各个行业、接近于全体样本的征信大数据，为研究我国产业结构变迁提供了可能。鉴此，本书尝试性地以信贷结构为切面，利用征信大数据对我国产业结构变迁进行了分析，以图“窥一斑而知全豹”。

希望本书的出版对我国征信大数据研究有所助益。当然，我们也清醒认识到，因学识所限，本书的研究还有待深入，比如在产业经济分析方面，可以立足于当前供给侧结构性改革背景，进一步深入挖掘征信数据中的经济信息，从而更好地服务新时代的经济建设与发展。比如研究“僵尸企业”问题（行业分布、地区分布、信贷



机构分布、时序变化)；针对农民工“离土不离乡”现象，研究人口迁移回流后地区间的财政转移支付问题；还有中美产业结构对比等。诸多遗憾，我们暂且留待下一步研究。

本书的研究数据有多个来源，一部分是征信系统的基本统计，已经经过合规要求进行了脱敏处理，不包含具体金融信息。另一部分来源是公开的数据，比如 Wind 数据库、国家统计局部门的数据等。本书的研究仅代表个人观点，不代表所在机构的意见。

周炳、景志刚、高健、李靖、徐方林、楼骁涵、卫旭骏、邵伟凯、刘威、乐国锋、司恩哲、胡青青、邓林慧、杨志、李状君、熊欣等同事、中国人民银行金融研究所陈华和中国金融出版社肖炜老师对本书给予了大量帮助，特此感谢。

书中的其他不足与需要改进的地方，亦请方家指正。

是为序。

2018 年 1 月 6 日

目 录

理 论 篇

第一章 大数据概述	3
一、大数据概括	3
二、国内外大数据发展探究	4
第二章 大数据技术	11
一、大数据技术层	11
二、传统大数据算法	18
三、大数据算法新趋势	30
第三章 征信大数据探索	39
一、征信大数据简介	40
二、征信大数据对产业经济分析的重要意义	42
三、征信大数据的实践	44

实 践 篇

第四章 基于数据挖掘的个人征信系统异常查询实时监测 模型	49
一、引言	49
二、以业务经验驱动异常查询监测	50



三、以数据挖掘驱动异常查询实时监测模型	51
四、模型应用及讨论	61
五、结束语	61
第五章 基于层次分析法的中国城市信用指数模型研究	63
一、引言	63
二、国内外信用指数发展状况	63
三、城市信用指数模型	64
四、研究结果及数据展示	69
五、结束语	72
第六章 基于征信大数据分析的中国劳动力人口迁徙研究	74
一、引言	74
二、研究简介	75
三、数据描述和研究方法	75
四、研究结果及数据展示	78
五、利用 Logistic 模型预测未来青壮年劳动力人口迁徙趋势	84
六、结束语	88
第七章 基于征信数据观中国近 10 年产业间信贷资源的调整 路径	90
一、引言	90
二、征信数据简介	91
三、我国信贷结构和产业结构调整关联分析	91
四、基于信贷视角的区域产业结构调整情况分析	95
五、结论及政策建议	98
第八章 基于大数据的中国房企信贷分析：现状与趋势	101
一、研究背景	101
二、房地产企业信贷市场分析	102

三、结语	109
第九章 基于大数据的个人房贷分析：现状与趋势	111
一、研究背景	111
二、消费者房屋贷款市场分析	112
三、结语	121
第十章 基于征信大数据的批发零售业信贷结构分析	123
一、引言	123
二、批发零售业贷款总体发放情况	123
三、批发零售业贷款的流向情况	126
四、需要关注的问题	129
第十一章 基于征信大数据的住宿餐饮业信贷结构分析	131
一、引言	131
二、住宿餐饮业贷款总体发放情况	131
三、住宿餐饮业贷款的流向情况	133
四、需要关注的问题	136
第十二章 基于征信大数据的建筑业信贷结构分析	139
一、引言	139
二、行业信贷情况	139
三、结语	144
第十三章 基于征信大数据的交通运输、仓储和邮政业信贷	
分析	145
一、研究背景	145
二、相关产业的信贷市场分析	145
三、结语	150



第十四章 基于征信大数据的采矿业信贷结构分析	152
一、引言	152
二、行业信贷情况	152
三、结语	159
第十五章 基于征信大数据的电力生产供应业信贷结构分析 ...	160
一、引言	160
二、行业信贷总体情况	160
三、结语	165
第十六章 基于征信大数据的水利、环境和公共设施管理业 分析	167
一、研究背景	167
二、相关基础设施行业的信贷市场分析	167
三、结语	172
第十七章 供给侧改革下我国制造业信贷投向结构分析	174
一、引言	174
二、信贷投放总体情况	174
三、基于供给侧改革的制造业贷款结构新特点	177
四、制造业信贷结构调整过程中仍需关注的问题	181
第十八章 “三农” 信贷的痛点	183
一、导语	183
二、贷款不良率上升	184
三、分项差异显著	184
四、农信社成中坚	185
五、结语	186

结语.....	187
一、内容回顾.....	187
二、未来展望.....	189

理 论 篇

