

中国银行数据库的学术生长 ——基于CBD建设的探讨

纪洋
厦门大学

一、银行数据：理解中国金融、中国经济的一把钥匙

- 银行业在我国金融体系中发展早、比重大
- 是金融市场化改革的重要着力点
 - 利率市场化、存款保险、破产条例
- 与数字金融、金融科技有密切联系

- 没有显性危机
- 数据缺失、变量缺失

数据缺失举例

➤ 隐藏的风险？

在挪动的风险

——不良贷款率

——关注类贷款率

(手动+复核)

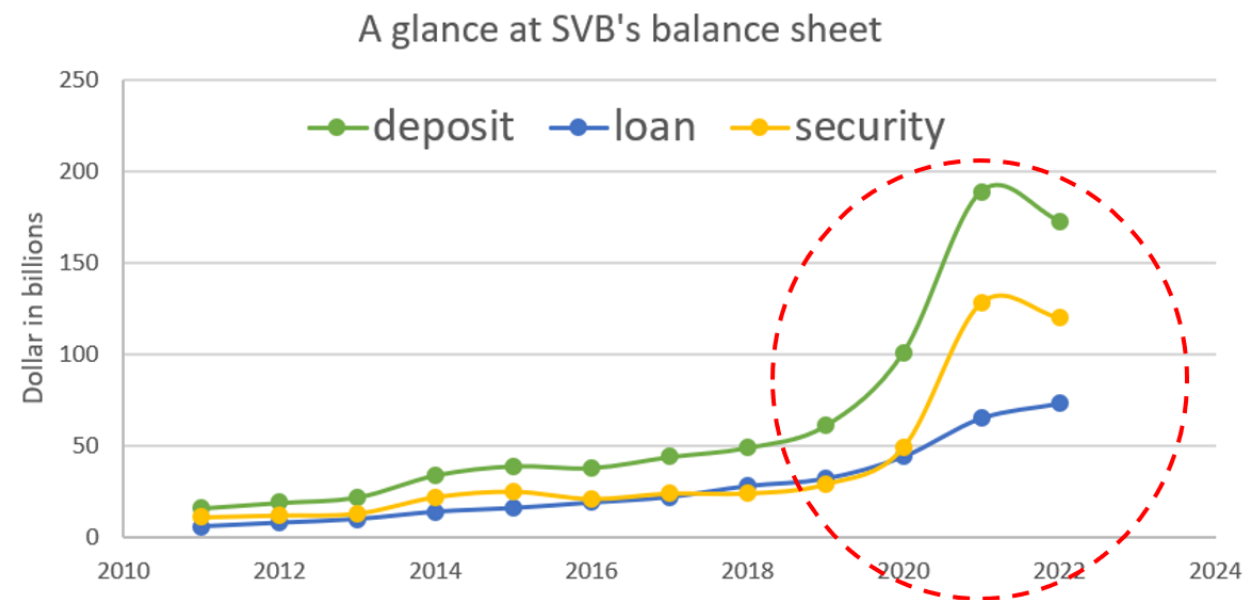
	<i>r_npl</i>		<i>r_special_mention</i>	
	Non-connected	Connected	Non-connected	Connected
	(1)	(2)	(3)	(4)
<i>r_loan_deposits</i>	0.002 (0.007)	0.081*** (0.028)	0.042 (0.033)	0.570** (0.289)
<i>lnasset</i>	-0.108 (0.066)	0.452** (0.192)	-0.957*** (0.364)	4.247* (2.327)
<i>r_equity</i>	0.012 (0.009)	0.029 (0.031)	-0.078 (0.048)	0.190 (0.197)
<i>duality</i>	0.059 (0.160)	0.183 (0.235)	1.008 (0.839)	0.670 (1.575)

数据缺失举例

➤ 隐藏的风险？

在挪动的风险——不良贷款率、关注类贷款率

在积累的风险——包商银行破产



Data source:
<https://ir.svb.com/financials/annual-reports-and-proxies/default.aspx>

10

数据缺失举例

➤ 隐藏的风险？

在挪动的风险——不良贷款率、关注类贷款率

在积累的风险——包商银行破产

➤ 基层银行真正的故事？

数字化转型滞后的农商行

存款消失的河南村镇银行

数据库基本结构与资料来源

❖ CBD三大数据模块：经营，风险，治理，共有6,458字段

- 政策依据：银监[2007]7号《商业银行信息披露办法》

❖ 基础资料来源：公开的信息披露资料

年报	编号	资料存档命名	类型代码
银行官网	1	年报	无标记
金融时报	2	年报摘要-有年报时使用	a
上海金融报	3	商业银行统计年鉴	s
地方报纸	4	经审计财务报表	t
中国货币网	5	大额/同业存单发行计划	c
中国债券网	6	信用风险评估报告	r
中国证券网	7	银监会开业批复/招股说明书	p
网络	8	社会责任报告	e
交易所/新三板	9/x		

数据库概览：截止2023年初

	银行总数 (含停业)	营业 银行数	CBD涵盖 机构数	CBD涵盖 营业机构数	CBD资料 银行-年份样本数	CBD数据 银行-年份样本数
政策性及国有大型银行	9	9	9	9	161	280
股份制银行	12	12	12	12	227	317
城商行（含住房储蓄银行）	156	126	155	125	2021	2167
农商行（含农合行）	1665	1629	1088	1059	6266	7439
农信社（含省、市联社）	544	544	123	114	576	650
村镇银行	1645	1645	388	388	1081	1430
民营银行	19	19	19	19	93	102
外资银行	43	41	43	41	464	504
总计	4093	4025	1837	1767	10889	12889

❖ 原始资料：15,917份资料，101GB

二、本土数据库建设：智慧+知识+信息→数据

- ❖ 数据⇒信息⇒知识⇒智慧
 - Data⇒Information⇒Knowledge⇒Wisdom
- ❖ 信息是组织过的数据
 - Information is organized data
- ❖ 知识是提炼过的信息
 - Knowledge is refined information
- ❖ 智慧是结构化的知识
 - Wisdom is structured knowledge
- ❖ ——数据库是经过提炼的数据？

有效数据来自于对现实的了解

➤ 文化：

支付宝发展早期，是如何识别用户身份的？

中国企业高管的管理决策，受哪些独特因素的影响？

➤ 市场现状：

信息与市场波动来自于哪里？——报纸？官网？

➤ 有效变量：

城镇登记失业率？贷款利率？存款利率？

基于国情的智慧+知识+信息，是中国本土数据库的基础

- 籍贯——方言、宗族、晋商\闽商\粤商
- 属相：2015年1月——属马，2015年9月——属羊
- 信息来源：公众号、B站、新浪微博

数据库构建的深层次问题之三：资料获取

❖ 银行原始信息披露资料的搜索、登记、存档

■ 核心问题：缺乏集中信息披露平台

- ✓ 文件允许银行以三种主要形式进行信息披露：1.网络；2.报纸；3.营业场所张贴
- ✓ CBD利用前两种信息披露渠道，但进一步的困难在于：什么网络？什么报纸？
- ✓ 长期积累，摸索出银行信息披露的规律，如公众号披露等

三、从数据到研究，从数据搜集者到优秀研究者

- “做理论的多看看数据现实，做实证的多想想理论框架”
- “如果没有深度触摸过一套数据，博士生涯是不完整的。”

——研究生培养：19名硕士、8名博士毕业论文依托该数据库

——工作组机制：本科生为主，每批次30-50人

- 累计参与人员417人次、320人

总结：八年磨一剑

- 一把钥匙，打开中国金融研究的谜题
- 一个模板，探究中国本土数据库的构建之路
- 一个福袋，为数百名学子备下了一份厚礼