

元数据技术在金融业综合统计中的应用研究

陈 燕, 廖燕平, 王世醒, 周 弘

(中国人民银行南昌中心支行调查统计处, 江西 南昌 330008)

摘要:2018年4月,国务院办公厅发布《关于全面推进金融业综合统计工作的意见》,提出对基础统计要素制定统一标准,建立集中共享的国家金融基础数据库,解决数据割裂、关联性不强的问题。元数据作为一种与数据结构本身无关的编码技术,能有效实现各类不同标准数据的集中共享,对实现金融业综合统计有重要作用。试图通过研究借鉴国内外应用研究及实践经验,探究元数据技术在我国金融业综合统计中的应用方案,并以目前成功试点的江西金融业综合统计工作案例来评估应用方案,从相关性、准确性、时效性和安全性方面论证其可行性,最后得出结论。

关键词:元数据技术;集中共享;金融业综合统计;金融基础数据库

中图分类号:F831.2 **文献标识码:**A **文章编号:**2095-0098(2019)04-0062-07

一、研究背景

2018年4月,国务院办公厅发布《关于全面推进金融业综合统计工作的意见》(以下简称《意见》),正式拉开了全面开展金融业综合统计的序幕。《意见》明确了金融业综合统计在服务实体经济、防范金融风险、深化金融改革方面的重要性,要求建立覆盖所有金融机构、金融基础设施和金融活动全生命周期的统计监测体系,着力解决“有效集中、充分共享”两大核心问题,最终建立国家金融基础数据库。然而,在长期分业监管的环境下,银行、证券、保险及其他金融领域的统计制度、统计标准及统计系统均自成体系,要建立集中共享的国家金融基础数据库面临着众多困难。近年来,我们在统计数据与元数据交换标准国际经验借鉴^[1]、金融统计标准下元数据信息框架构建^[2]等方面进行了探索性研究。在《意见》正式落地后,在总结前期研究成果的基础上,结合我省金融业综合统计工作经验及成效,进一步提出将元数据技术应用于金融业综合统计工作的重要性,并对元数据应用的方案和可行性进行了研究及评估。

二、元数据在统计领域的应用研究现状

(一)国内应用研究现状

国内研究方面,国内学者多是从统一标准方面研究了数据的集中共享,而未深入到元数据技术层次。于一超和何琳(2012)^[3]分析了将信息标准化管理运用于商业银行的优势,并构建相应的标准化管理框架和机制。徐燕(2015)^[4]总结互联网金融特征,从统计内容、统计对象、统计范围等方面构建一套互联网金融统计指标。柳成洋和王东升(2015)^[5]认为“十三五”时期,我国金融统计标准化政策导向应包括全面创新金融标准化管理体制机制,积极以标准化支持投资审批“简政放权”改革,力求加快研制和优化调整并重,努力提升我国金融标准国际化水平。

国内实践方面,中国标准化研究院已发布多项元数据标准或征求意见稿,如《政务信息资源目录体系第3部分:核心元数据》《地理信息元数据》《全国主要产品分类产品类别核心元数据》《产品标签内容核心元数

收稿日期:2019-01-15

基金项目:中国人民银行重点课题“金融统计元数据研究”

作者简介:陈 燕(1964-),女,江苏盐城人,经济师,研究方向为公共管理、金融。

据》等等。中国人民银行也一直积极开展相关研究工作,自2010年起开始跟踪统计数据与元数据交换技术动态,通过借鉴欧盟、美国等国外先进经验,于2014年制定并发布了我国《统计数据和元数据交换标准》,为我国确立了统计数据和元数据交换的行业标准。人民银行理财与资金信托专项统计、江西和安徽推进的金融业综合统计试点工作中,虽未形成专门的元数据标准体系,但都积极应用了元数据技术的思想,通过制定统一的基础标准,实现了不同领域、不同标准下的数据集中采集、校验和存储。

(二) 国际应用研究现状

联合国统计司、国际货币基金组织及世界银行等国际组织已经将元数据应用于统计领域,并取得实质性成果,通过与世界各地国家或地区统计机构的协调配合,建立多种形式的数据平台。联合国统计司在其千年发展目标项目^[6]中,制定了专门的元数据结构,在此基础上,通过国家数据信息渠道与15个国家之间交换社会经济发展指标的信息,以及共享来自100个国家的指标,为项目参与国家提供更好的信息服务。欧盟统计局较早开始应用元数据技术,在成员国之间建立了统计数据与元数据交换的基础设施,用于简化数据和元数据的交换过程。该基础设施中构建了众多的区域块,旨在根据对每个统计机构或组织的数据仓库的映射来提供数据和结构元数据。世界银行、国际货币基金组织协助非洲开发银行于2012年启动了非洲信息高速公路项目,以支持非洲开发银行的成员国统计能力建设。该项目在54个非洲国家和16个区域组织安装了开放数据平台,该平台应用了元数据技术,不仅实现了成员国之间的数据交换和共享,还可实现无缝地向国际货币基金组织传输宏观数据。

(三) 应用元数据的意义

从国内外应用研究情况看,元数据作为一种与数据结构本身无关的编码技术,它能有效实现各类不同标准数据的集中共享,对实现金融业综合统计有重要作用。元数据由参考元数据和结构元数据组成,参考元数据用于标记数据的存储位置、交换方式、字段属性及类型等要素,而结构元数据是将各类不同标准的原始数据转换为格式统一的无差别数据。元数据在金融业综合统计中重要性有以下三点:

一是元数据可以加强对金融数据全生命周期的治理,应对数据更新需求。它提供了数据从哪里来、如何到达此处以及相关的数据转换规则等信息,能清晰地反应数据的流向,通过分析元数据,可以显著提升数据治理水平。二是通过元数据加强数据间映射,可减少数据存储压力,强化数据交流。一行两会、中债登、中证登、各地股权交易中心、清算中心等都已建立较为完善的数据库,通过元数据共享的方式能有效降低数据交流的成本,避免重复建设和数据存储资源的浪费,缩短建立统一的国家金融基础数据库的时间。三是元数据自身也是一种标准,它规定了数据的属性,因此应用元数据的过程也是推动标准化的过程。近年来,统计数据维度呈几何级增长,而元数据符合维度结构化数据管理的要求,元数据制定者根据金融业务的实践定义数据属性,使数据采集、数据传输、数据处理等各环节的使用者对数据的理解统一起来,可减少对数据的误解。

综上,元数据在数据采集、传输、处理等各方面都有巨大的发挥空间,接下来本文结合我国实际情况对元数据技术在金融基础数据库应用展开进一步详细阐述,并结合江西省金融业综合统计工作对元数据技术的应用进行可行性论证。

三、元数据在金融业综合统计中的应用方案

(一) 元数据的应用原则

《意见》由国务院办公厅印发,明确了工作的重大意义、总体要求、工作机制、关键支撑、主要任务及保障措施,可见此项工作已经上升到国家高度,这为元数据的应用奠定了基础。根据《意见》精神,元数据技术在金融业综合统计中的应用要遵循“统一布局、分步实施、局部突破”的原则。统一布局是由人民银行牵头制定完善金融业综合统计元数据的基础标准。分步实施是指在基础标准确立后,各行业在基础标准上根据需要制定本行业内部的元数据标准。局部突破是在统计基础相对较好的金融领域或地区率先开展试点,推进元数据技术的落地应用,以检验应用方案的可行性。

(二) 元数据的应用流程

元数据应用流程贯穿整个金融业综合统计过程,从标准制定、数据采集及处理、数据传输,直至提交到国

家金融基础数据库中,都需要它发挥作用。反之,当从国家金融基础数据库中查询分析数据时,也需要利用元数据技术来传输和解析。

1. 制定元数据标准

元数据技术是借助同一的编码体系,在数据底层建立的无差异的共享机制。元数据分为结构元数据和参考元数据。结构元数据用于统计数据的说明和识别,参考元数据更为普遍地描述和定性了统计数据归集、传输、存储流程,且通常不与数据的特定观测结果或序列相关,但是与整个数据的集合或者数据的提供者有关。目前我国各金融领域统计标准及数据结构差异较大,为数据集中共享带来较大的难题。为解决这一问题,需先制定金融行业基础统计的元数据标准,统一规范金融机构、金融工具、金融交易对手方所属经济部门、金融基础设施等基础统计要素的定义、口径、分类和编码等规则。然后,各金融监管部门制定所辖行业的元数据标准,行业元数据标准根据基础统计标准的分类要求,进一步细化分类,特别是金融工具、金融交易对手方所属经济部门等属性。细化后的行业分类标准可直接或重新加工组合对应基础统计标准的分类,确保统计口径、分类的一致性。

2. 数据归集

建立基础及行业的元数据标准后,便可着手开展数据的归集工作。对于可通过对标方式归集的数据,利用元数据交换技术建立金融业综合统计基础数据归集平台,在数据源上依托金融管理部门现有信息系统资源库,构建分层次的数据组织架构(元数据基础设施的一部分)。对于现有统计能够支持近期工作目标的内容,由该平台从金融管理部门现有数据库归集相关数据,如各金融行业及机构的资产负债表及损益情况。对于现有统计不能支持近期工作目标的内容,可由该平台根据新建统计制度,直接从金融机构采集数据,如资产管理产品统计(如图1)。

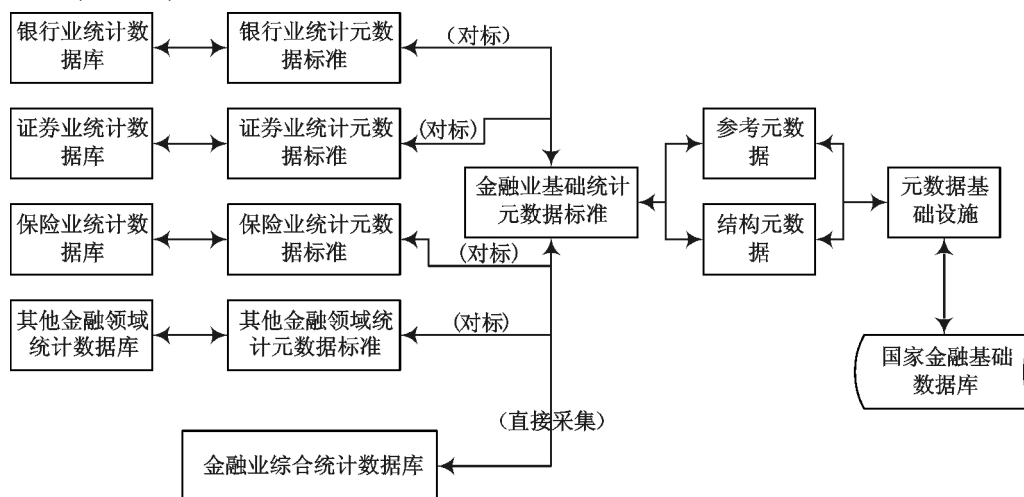


图1 元数据技术在金融业综合统计中的应用流程

3. 集中校验

《意见》中提到需要对数据集中校验,同样能通过元数据技术实现。如图1中,所有金融数据经过对标或直接采集后,都能够转换成以基础元数据标准表示的数据,这就排除了因为标准不统一而带来的校验难题。只需要将各行业数据与基础统计标准对标,建立数据间的逻辑对应关系,就可进行跨主题关联校验。从整体上实现对数据质量的自动化管控和全面治理,支持数据应用层面的统一汇总、信息整合、关联穿透。

4. 数据汇总共享及查询

集中校验通过后,可根据基础统计标准汇总加工数据,统一制定金融业综合统计管理报表体系。然后逐步建立金融业综合统计数据共享机制,在符合国家保密规定的前提下,提高统计透明度和数据使用效率,实现对基础数据源、汇总指标以及统计报表的充分共享。在查询过程中,用户首先通过相关查询系统输入需求,平台将查询需求解析为元数据,传输到元数据基础设施,元数据基础设施再从国家金融基础数据库获取相应的结构元数据,再通过配套的参考元数据解析转换为可识别的内容。

四、实例论证元数据技术应用的可行性

在《意见》下发之前,为策应全国金融业综合统计工作,江西省积极推进辖内试点工作,历经四年成功打造出金融业综合统计工作“江西样板”。金融业综合统计对象为银行业(含资产管理公司)、证券业(含期货公司)、保险业、典当、融资担保、小贷、融资租赁、商业保理、住房公积金中心及第三方支付十大类机构,覆盖江西省所有金融或类金融业务类型,按季度采集各类机构的资产负债表和损益表数据。在江西省金融业综合统计工作试点过程中,元数据技术的思想得到初步应用。本文正是基于江西省金融业综合统计的实践,来评估元数据技术应用的可行性。

(一) 建立元数据标准

一是建立统计数据基础标准。人行南昌中支编制了江西省金融业综合统计核心指标体系,并根据各行业机构业务变化适时调整指标体系,以确保统计的科学性。同时定期向省统计局申报立项,保障统计工作的合法性和权威性。

二是建立了元数据基础标准和数据采集系统。根据核心指标,制定了参考元数据和结构元数据标准,方便数据采集。同时建设了金融业综合统计信息管理平台,用于归集各类机构的数据。直报机构通过该平台提交数据,数据提交时会根据元数据标准,转换为结构元数据,同时产生参考元数据,用于记录数据存储位置、提交方式、提交时间等信息。

三是通过统计数据及元数据基础标准实现了与各行业现有统计标准的对标。具体分为三个方面,一是银行业、小贷公司的资产负债和损益数据,这两类机构现有统计报表与金融业综合统计指标体系相似度较大,因而通过现有统计数据直接对标转换;二是实现除银行、小贷公司外其他机构现有报表与核心指标体系的对标,制定了分类填报指引,对采集到的原始数据进行转换归集;三是不能通过现有统计报表对标转换的指标,如损益表中业务及管理费项下的具体科目,由各机构根据金融业综合统计核心指标体系,直接从其他业务报表或系统中加工提取。

四是开发了江西省金融业综合统计数据处理系统,建立了江西省金融业综合统计数据库,实现各行业统计数据的集中校验、汇总共享和统一存储。数据在信息管理平台与数据处理系统之间通过结构元数据报文进行传输,然后根据参考元数据解析,在相应的系统界面上显示为可阅读的内容。

(二) 元数据应用流程示例

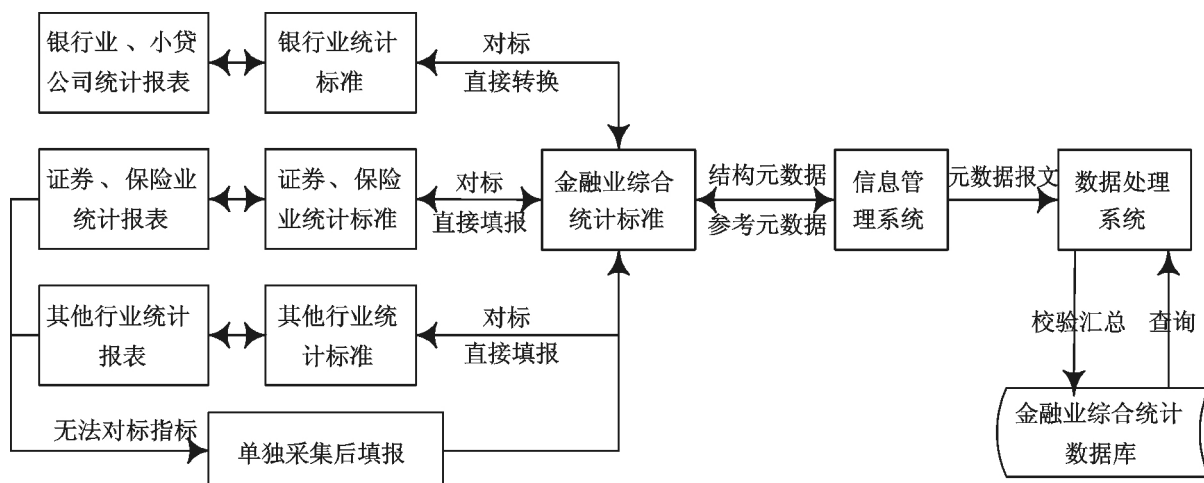


图2 江西省金融业综合统计工作中元数据应用流程

无论行业的数据是以上述流程中哪种方式采集,但基本原理是一致的,即将原始报表指标与金融业综合统计指标进行对标,根据对标关系填报数据,或根据金融业综合统计指标体系单独采集一些无法对标的指标。

1. 元数据表制定

为规范数据采集,我们定义了5张参考元数据表、1张结构元数据表及1张逻辑校验表。参考元数据表

分别为金融机构用户表、人民银行用户表、机构信息表、地区信息表、指标表,结构元数据表为数据表,集中校验则是通过逻辑校验关系表,逻辑校验表实际也是通过元数据标准来校验的。表 1 显示了结构元数据表,它是记录每条数据的标准编码,包含本条记录的 ID、指标编码、指标值、所属机构编码、所属地区编码、业务类型、数据日期、提交时间等属性。

表 1 数据表

字段名	属性	类型	约束	说明
Record_ID	本条记录的 ID	Char(7)	非空 & 唯一 & Key	自动生成
Index_Code	指标编码	Char(20)	非空	外键
Index_Value	指标值	Char(20)	非空	
Index_Dep	所属机构编码	Char(14)	非空	
Index_Area	所属地区编码	Char(7)	非空	
Index_Type	业务类型	Int(2)	非空	人民币/外币
DateTime	数据日期	Date	非空	
SubTime	提交时间	Date	非空	

下面举例说明,以下分别是金融业综合统计中证券公司 A 和保险公司 B 的数据信息,它们是根据表 1 的编码规则进行处理后,转换成了结构元数据,如表 2 所示。在不知道元数据标准的情况下,用户无法解读这条数据。

表 2 结构元数据示例

本条记录的 ID	指标编码	指标值	所属机构编码	所属地区编码	业务类型	数据日期	提交时间
1000001	101	2.3	301	360400	1	2018-06-30	2018-07-16
1000002	102	100	686	360100	1	2018-06-30	2018-07-12

传输时上述结构元数据以下面这种报文形式呈现:

1000001,101,2.3,301,360400,1,2018-06-30,2018-07-16

1000002,102,100,686,360100,1,2018-06-30,2018-07-12

通过结构元数据标准,便可以知道逗号分隔符号隔开的每个字段指代的含义,及指标值的单位等。但指标编码和机构编码、地区编码还必须查找参考元数据中的指标信息、机构信息和地区信息的标准来识别。以机构信息为例,见表 3。

表 3 机构信息表

字段名	属性	类型	约束	说明
Dep_Code	机构编码	Char(14)	非空 & 唯一 & Key	
Dep_Name	机构名称	Char(40)	非空	
Area	所属地区编码	Char(7)	非空 & 外键	
Dep_Type	机构类别	Int(2)	非空	

通过参考元数据解析,可得到以下统计信息。此时,用户便可知晓该统计信息所代表的含义。例如,第一条记录表达了所属九江市的证券公司 A 在 2018 年 6 月末资产负债表中现金项人民币余额为 2.3 万元,该数据于 2018 年 7 月 16 日上报,在系统中记录编号为 1000001。

表4 结构元数据解析后示例

本条记录的ID	指标	指标值	所属机构	所属地区	业务类型	数据日期	提交时间
1000001	现金	2.3 万元	证券公司 A	九江市	人民币	2018-06-30	2018-07-16
1000002	存放除货币当局外金融机构款项	100 万元	保险公司 B	南昌市	人民币	2018-06-30	2018-07-12

2. 数据归集

数据归集是利用对标关系,将各行业数据统一转换为金融业综合统计数据。对标后,各行业数据在金融业综合统计中呈现形式是结构元数据。例如,证券公司 A 和保险公司 B 报金融监管部门的资产负债表中一般没有“101 现金”和“104 存放除货币当局外金融机构款项”这些指标,而是通货、货币资金、自有资金存款、客户资金存款、结算备付金、清算存款、其他货币资金等,因此需要进行对标。其中通货、货币资金中的现金需填入“101 现金”项,货币资金中存款、自有资金存款、客户资金存款、结算备付金、清算存款、其他货币资金等填入 104 “存放除货币当局外金融机构款项”及其子项。

3. 集中校验、汇总

各行业数据集中校验和汇总均是在金融业综合统计数据处理系统中完成,同时银行业、小贷公司数据也是以对标形式,转换后导入数据处理系统,最终数据均存储在金融业综合统计数据库。为进行集中数据校验,系统设置逻辑校验关系表,对采集到的数据进行统一校验,校验规则也是以元数据的形式表现。数据采集完备且校验无误,系统会进行汇总,汇总数据同样以结构元数据的形式存储在金融业综合统计数据库中。相反地,数据处理系统可以对结构元数据进行分析处理,并结合参考元数据对结构元数据进行解析,生成用户可以直接使用的报表。

(三) 元数据应用评估结果

从上述评估过程看,虽然江西省金融业综合统计中暂未形成独立的元数据技术标准,但实际上它已经将元数据技术应用于标准制定、数据采集及处理、数据传输和数据存储整个过程,基本符合了本文第三部分提出的元数据应用方案,从而论证了该应用方案的可行性。

五、结论

本文根据前期研究成果及国内外应用现状,提出了元数据技术在我国金融业综合统计中的应用方案,并以江西省金融业综合统计经验实例对该方案进行了论证。从论证结果看,本文提出的应用方案是可行的,而且它完全符合了相关性、准确性、高效性和安全性四个特性。相关性方面,制定完善金融业综合统计的基础元数据标准和分行业元数据标准,将可对标的的数据直接完成对接,不可对标的的数据以基础元数据标准进行采集,逐步可实现了各行业金融业综合统计元数据的归集。准确性方面,利用从各行业采集的结构元数据,进行行业内逻辑校验、总分校验和跨行业校验,实现了集中统一校验,保证了数据的准确性。时效性方面,统一了元数据标准和实行对标后,各报数机构数据采集效率会大幅提高,且后期数据进行加工处理也更为高效,相较于直接采集原始报表再做处理,效率大幅提高。安全性方面,通过使用元数据标准的实施,数据从采集端到存储端的中间过程,都是以编码形式的元数据呈现,而解析结构元数据的参考元数据一般只存放在采集端和存储端,传输过程中数据实际是加密的,从而保证了数据的安全性。

参考文献:

- [1]陈燕,李贞. 欧盟实施统计数据与元数据交换(SDMX)标准的经验与启示——以荷兰为例[J]. 金融教育研究, 2016, 29(2): 65-72.
- [2]金融统计元数据研究课题组. 金融统计标准下元数据信息框架构建研究[J]. 金融教育研究, 2018, 31(3): 31-39.
- [3]于一超,何琳. 商业银行信息标准化管理的理论与实践[J]. 国际金融, 2012(10): 33-36.
- [4]徐燕. 互联网金融统计方法探索[J]. 调研世界, 2015(7): 44-46.
- [5]柳成洋,王东升. 金融标准化发展及“十三五”政策导向[J]. 开放导报, 2015(1): 39-42.
- [6]联合国统计司. 千年发展目标[EB/OL]. <http://unstats.un.org/unsd/mdg/>.

Research on the Application of Metadata Technology in Financial Industry Comprehensive Statistics

CHEN Yan , LIAO Yanping , WANG Shixing , ZHOU Hong

(Survey and Statistics Office of Nanchang Central Branch , People's Bank of China ,
Nanchang , Jiangxi 330008 , China)

Abstract: In April 2018 , the general office of the state council issued the opinions on comprehensively promoting comprehensive statistical work in the financial sector , in order to solve the problem of data fragmentation and weak correlation , a unified standard for basic statistical elements and a centralized and shared national financial basic database are proposed. As a coding technology unrelated to the data structure itself , metadata can effectively realize the centralized sharing of various standard data , which plays an important role in the realization of financial industry comprehensive statistics. This paper tries to explore the application scheme of metadata technology in China's financial industry comprehensive statistics by referring to domestic and foreign applied research and practical experience; at the same time , the application scheme is evaluated by a case study of Jiangxi financial industry comprehensive statistics , the feasibility is demonstrated from the aspects of relevance , accuracy , timeliness and safety.

Key words: metadata technology; centralized shared; financial industry integrated statistics; basic financial database

(责任编辑: 沈 五)

(上接第 61 页)

Research on Measurement and Management of Bank Liquidity Risk under the New Trend

WANG Ce¹ , WEN Xianming¹ , ZHOU Yilun²

(1. School of Economics and Management , Changsha University of Science and Technology , Changsha ,
Hunan 410114 , China; 2. Henan Yufa Group , Zhengzhou , Henan 450014 , China)

Abstract: Using the theory of liquidity risk management , this paper studies the actual situation of commercial banks' liquidity risk in China's economic operation from the perspectives of monetary policy , asset securitization , foreign exchange and interest rate marketization , and analyzes the new challenges faced by the new normal. Through the establishment of a time series model , the partial liquidity index is used to empirically measure the liquidity of commercial banks. The empirical results show that the capital adequacy ratio has a weak influence on liquidity and the non - performing loan ratio is negatively correlated with the liquidity of commercial banks. Commercial banks should focus on the control of non - performing loan ratio.

Key words: commercial banks; liquidity risk; time series regression model; capital adequacy ratio; non - performing loan ratio

(责任编辑: 黎 芳)