

Python 数据分析及爬虫在某基金经责审计项目中的应用

贲文硕

一、实例概述

（一）项目背景介绍

XX 资本成立于 201X 年，实际控制人为深圳市国资委，以受托管理 XX 基金为核心任务，以母基金专业管理机构为核心定位。基金定位高、资金量大，在市属国资基金群中起到引领作用。通过分析财务数据发现，2017 年末 XX 基金合并口径银行存款余额为 19.34 亿元、银行理财产品余额为 134.62 亿元，合计 153.96 亿元，在总资产中占比 63.12%；2018 年末 XX 基金合并口径银行存款余额为 99.21 亿元、银行理财产品余额为 64.14 亿元，合计 163.35 亿元，在总资产中占比 52.31%，理财及银行存款余额连续两年在总资产中占比超过 50%。此外，2017 年末 XX 基金存款利息收入及理财收益共 3.94 亿元，在营业利润中占比 103.86%；2018 年末 XX 基金存款利息收入及理财收益共 4.65 亿元，在营业利润中占比 96.88%。从上述数据可以看出，XX 基金 2017、2018 年利润几乎全部来自于银行存款利息及理财收益。结合 XX 基金设立的有关材料，该基金及其管理公司成立时间较短，大部分投资仍处于投资期，尚未进入收益阶段。由于基金成立初期市委市政府即向其注入了大量资金，因此存款利息及理财收益成为了近

两年 XX 基金利润的主要来源。经过宏观的数据分析，项目组将审计重点放在了大额资金的计划管理、理财收益、子基金及所投底层项目公司的真实性和运营情况。

为核实被审计单位提供数据的真实性，项目组采用了内外部数据结合比对的方法，要求被审计单位在提供银行存款明细账的同时需同步提供有银行用印的对账单佐证。针对子基金及底层项目公司信息，通过网络第三方平台检索的方式核查其真实性及运作状态，内外部数据结合，充分利用网络数据库，有力推动了审计全覆盖。但是，面对海量的数据，在有限的时间内依靠有限的人员无法做到全面核查，传统的抽查审计方式势必会影响审计质量。因此本项目在审计实施过程中充分利用了 Python 编程，将大数据审计方法应用到了具体项目上，既做到了全覆盖又兼顾了效率，有效提升了审计质量。主要包括以下三个方面：

1. 是否存在大额资金闲置。XX 基金在成立之初，即承接了大量政府资金，由 XX 资本负责经营管理。在基金的运营过程中，资金的投出需经过尽调、论证、谈判、合同签署等诸多环节，在资金拨付前，针对账上的大额资金沉淀，XX 资本应做好计划，在保证资金安全的前提下，通过购买银行理财产品等方式提高国有资金收益。项目组要求被审计单位提供 XX 基金及其相关主体名下所有账户的银行存款明细账，将连续 15 天以上银行账户余额大于 2 亿元视为一次资金闲置。利用 Python 编写程序语言读

取上述数据，统计出审计期间内的所有资金闲置情况，并根据被审计单位过往银行理财收益率平均值测算出闲置资金的利息损失金额。

2. 是否存在银行理财产品收益率明显偏低。XX 基金及其相关主体日常经营中会频繁在各类商业银行开展理财业务，由于手中资金量较大，理财收益率的微小差别也会对国有资金最终收益造成较大影响。例如，10 亿元购买 6 个月期限的理财产品，如理财收益率相差 2%，最终的资金收益差额将达到 1000 万元。项目组要求被审计单位提供审计期间内所有银行理财业务明细数据，包含起止日期、托管银行、收益率、购买赎回金额等要素，按照不同计息天数统计出各时间段的平均收益率，例如 30-60 天区间的理财收益率平均值为 3.56%。利用 Python 编程，筛选出明显偏低的异常数据，同时根据平均收益率测算利息损失金额，重点筛查异常理财业务的决策程序是否合规，是否存在利益输送等违法行为。

3. 所投子基金及底层项目公司是否存在异常。XX 资本以母基金专业管理机构为核心定位，XX 基金的资金投向以子基金为主，直接投资企业数量较少。为核实基金所投资产质量，项目组利用中国证券投资基金业协会的私募基金信息公示功能，检索已投子基金信息，查看是否存在经营异常，造成国有资产损失。针对子基金已投底层项目公司，利用互联网“企查查”企业信息查

询网站，根据企业名称检索相关信息，查看企业是否存在破产清算、失联、重大诉讼等经营异常情况。以上信息检索过程通过 Python 爬虫功能实现，大幅度提升工作效率。

（二）项目审查数据范围

本次审计数据包括从被审计单位账务系统金蝶 K/3Cloud 获取的财务数据，后台数据库为 SQLServer2012R2，针对大额资金闲置问题，调取六个相关主体名下全部账户银行存款明细账，共 77 个银行账户。针对银行理财产品收益率偏低问题，调取六个相关主体全部理财产品收益明细数据，共 489 笔。从业务系统获取所投子基金及底层项目公司相关信息，后台数据库为 SQLServer2008R2。同时从中国证券投资基金业协会网站获取子基金运营状况信息，从“企查查”网站获取底层项目公司公开信息，包括成立时间、法人、经营状况等。

（三）案例特点

1. 在审计过程中，充分运用了 Python 编程语言，面对海量的项目信息和银行存款明细账数据，通过编写算法，精准定位到异常数据，为审计发现问题提供了有力的数据支撑。

2. 结合被审计单位业务特点，选择了大额资金闲置和理财收益率偏低作为突破口，此类问题在以往审计中涉及较少，容易被忽视。本项目在审计关注点上打破常规，提供了新的思路。

3. 充分利用网络公开信息，通过 Python 爬虫技术抓取关键

信息进行比对。在互联网普及、信息阳光化的今天，互联网已成为审计人员获取信息的重要渠道，将编程语言与互联网海量数据结合，大幅提升了审计效率及覆盖面。

二、项目具体实施过程

（一）成果一

将单个银行账户连续 15 天以上资金余额大于 2 亿元，视为一次资金闲置。在被审计期间内，XX 基金及其相关主体共发生资金闲置情况 12 次。例如 XX 基金一般户 2016 年 8 月 29 日至 2017 年 1 月 11 日期间，10 亿元资金以活期存款形式闲置在银行账户上，XX 资本未能根据业务需求及时制定资金计划，造成国有资金闲置浪费。

根据 XX 基金及其相关主体已购买银行理财产品明细数据，计算出不同期限理财产品的平均收益率，以此为依据测算闲置资金的利息损失（见表 1）。

时间段	平均收益率
0<计息天数≤15	2.87%
15<计息天数≤30	3.35%
30<计息天数≤60	3.56%
60<计息天数≤90	3.68%
90<计息天数≤120	4.01%
120 天以上（计算平均收益时，剔除 3%以下不合理数据）	4.12%

表 1 平均理财收益率

经估算，12 次资金闲置共造成利息损失金额为

192,275,866.65 元。

1. 审计思路

XX 基金在设立初期，市政府即向其注入了大量资金，但 XX 资本经营班子成员尚未全部到位，基金整体投资进度较慢。因此，项目组关注到了大额国有资金闲置的问题，是否有按照《关于加强市属国有企业银行账户管理及资金风险防范综合监管的指导意见》的有关规定，“控制闲置资金，提高资金效益。”由于 XX 资本控制的主体较多，且不同主体均开设了多个银行账户，为了全面准确的挖掘出被审计期间内的资金闲置问题，项目组想到使用 Python 编程软件，利用算法精准的找出问题所在。在编程前，首先要确定标准，什么是资金闲置，项目组根据过往审计经验，集体讨论决定将单个银行账户连续 15 天以上账户余额大于 2 亿元，视为一次资金闲置。在确定标准后，要求 XX 资本提供其控制所有主体的银行存款明细账，利用 Python 进行数据分析，筛选出全部资金闲置问题，并在表格上做出标注。

第二步，根据资金闲置情况，测算资金闲置造成的利息损失。测算损失的利率是根据 XX 资本控制主体过往所有理财业务收益率按不同计息天数计算出的平均值，例如 0 到 15 天的理财收益率平均值为 2.87%，15 到 30 天的理财收益率平均值为 3.35%，依此类推，计算过程同样通过 Python 编程实现。

2. 审计流程图（见图 1）

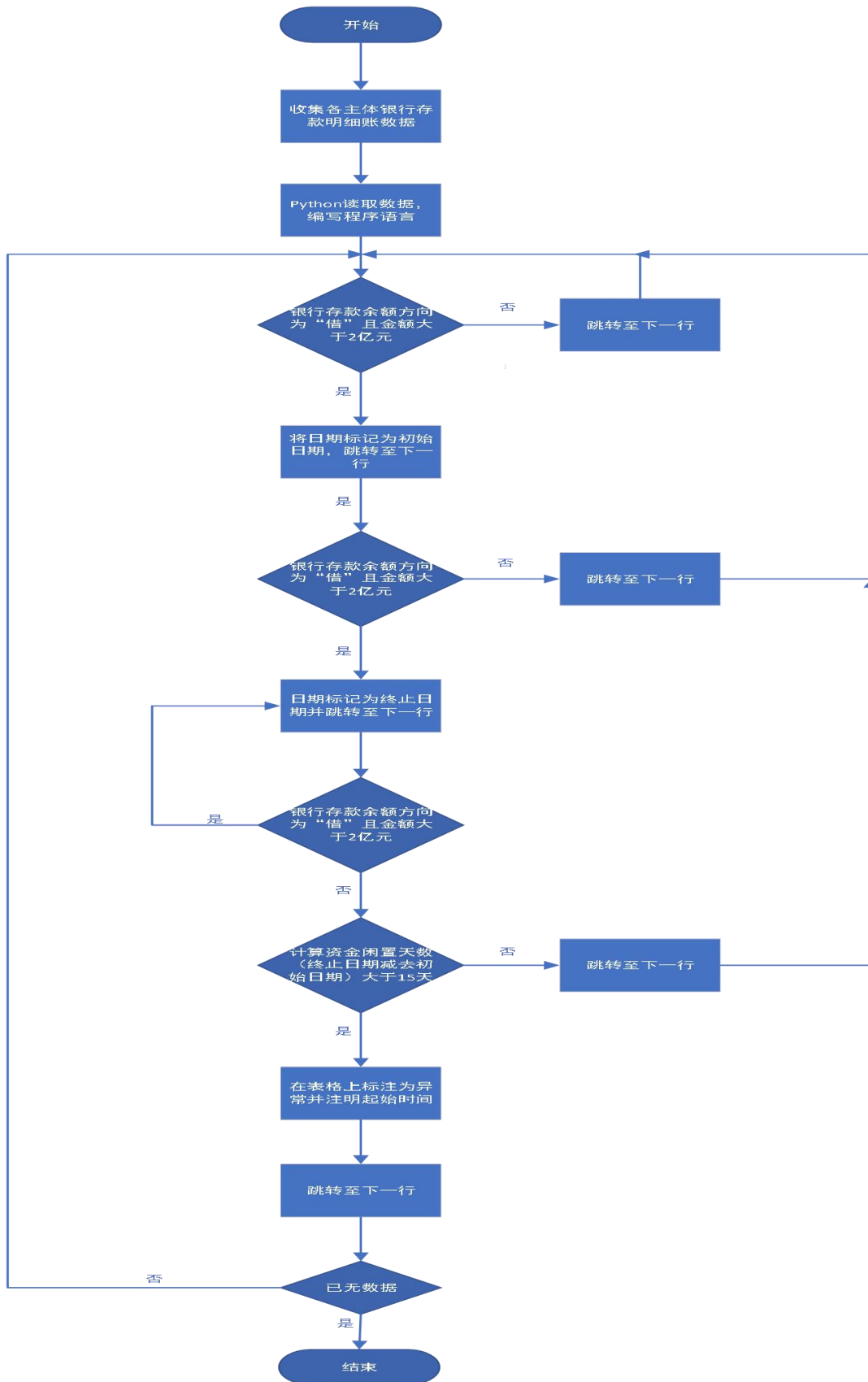


图 1 资金闲置状况统计流程

3. 审计方法和步骤

步骤一：收集 XX 资本控制的主体名下所有银行存款明细账资料；

步骤二：用 Python 读取数据，编写程序语言；

步骤三：按照时间顺序逐行读取数据，设置判断条件为“银行存款余额方向为借且金额 2 亿元”，判断为是则将日期标记为初始日期并跳转至下一行进行步骤四，判断为否则跳转至下一行重复步骤三（见图 2）；

科目编码	科目名称	凭证号	凭证摘要	对方科目编码	对方科目名称	借方金额	贷方金额	方向	余额
1002	银行存款		期初余额					平	
		记 4	8.29 收市引导基金注资款	4001	实收资本	1,000,000,000.00		借	1,000,000,000.00
			本日合计			1,000,000,000.00		借	1,000,000,000.00
		记 5	8.30 建行开户手续费	6603.04	财务费用_手续费		100.00	借	999,999,900.00
			本日合计				100.00	借	999,999,900.00
	银行存款		本期合计			1,000,000,000.00	100.00	借	999,999,900.00
	银行存款		本年累计			1,000,000,000.00	100.00	借	999,999,900.00
1002	银行存款		期初余额					借	999,999,900.00
		记 6	9.2 建行退开户手续费	6603.04	财务费用_手续	100		借	1,000,000,000.00
			本日合计			100.00		借	1,000,000,000.00
		记 5	8.25-9.21 建行利息收入	6603.02	财务费用_利息	191,666.66		借	1,000,191,666.66

图 2 银行存款明细账

步骤四：设置判断条件为“银行存款余额方向为借且金额 2 亿元”，判断为是则将日期标记为终止日期并跳转至下一行，判断为否则进行步骤五；

步骤五：设置判断条件为“终止日期减去初日日期大于 15 天”，判断为否则跳转至下一行重复步骤三，判断为是则在表格上标注为异常并注明起始时间（见图 3）；

借	1,001,613,779.49								
00 借	1,001,613,779.49								
借	1,001,613,779.49	异常	2016/08/29 到 2017/01/01	闲置资金大于2亿					
00 借	1,613,779.49								
00 借	1,613,779.49								

图 3 资金闲置数据

步骤六：跳转至下一行，设置判断条件为“已无数据”，判断为否则重复步骤三，判断为是则结束此流程。

步骤七：根据过往所有理财业务数据，按不同计息天数计算出收益率平均值，以此为依据测算闲置资金的利息损失。平均收益率的计算过程详见“成果二”。

（二）成果二

根据被审计单位及其相关主体已购买银行理财产品明细数据，计算出不同时间段理财产品的平均收益率，数据见表 2：

计息天数	平均收益率
0<计息天数≤15	2.87%
15<计息天数≤30	3.35%
30<计息天数≤60	3.56%
60<计息天数≤90	3.68%
90<计息天数≤120	4.01%
120 天以上（计算平均收益时，剔除 3%以下不合理数据）	4.12%

表 2 平均理财收益率

将单笔银行理财收益率低于相应时间段平均收益率 0.015 以上，视为收益率明显偏低的异常数据。经统计，异常理财数据共 30 笔（赎回一次算一笔），按照平均收益率估算利息损失为 54,856,589.69 元。

1. 审计思路

XX 资本控制的各主体资金量均较大，日常经营中会频繁的购买商业银行理财产品已提高资金收益。项目组在查阅资料时注意到，部分理财产品期限长，但收益率较低，明显低于市场平均水平。XX 资本控制的六个主体，累计购买理财产品（含七天通知存款、结构性存款等）共 489 笔，且计息天数各不相同。为了能够全面的筛查出理财收益过低问题，并测算相应的利息损失，项目组讨论决定使用 Python 编程实现上述目的。首先，收集 XX 资本控制所有主体被审计期间内的理财业务数据，包含理财产品名称、购买日期、赎回日期、金额、收益率等要素。按照不同的计息天数，计算出不同时间段理财产品的平均收益率，具体划分为六个时间段，0 到 15 天、15 到 30 天、30 到 60 天、60 到 90 天、90 到 120 天以及 120 天以上。

在计算出平均收益率后，将各个时间段理财收益率低于平均值 0.015 以上的数据通过 Python 筛选出来，再根据平均收益率计算出期望收益，期望收益减去实际收益，得出结果即为理财收益损失。例如某理财业务本金 1 亿元，计息天数 100 天，收益率 2.01%，实际收益 201 万元，相应时间段的平均收益率为 4.01%，计算出期望收益为 401 万元，期望收益减去实际收益得出利息损失估算金额为 201 万元。上述过程全部通过 Python 编程实现。

2. 审计流程图（见图 4.5）

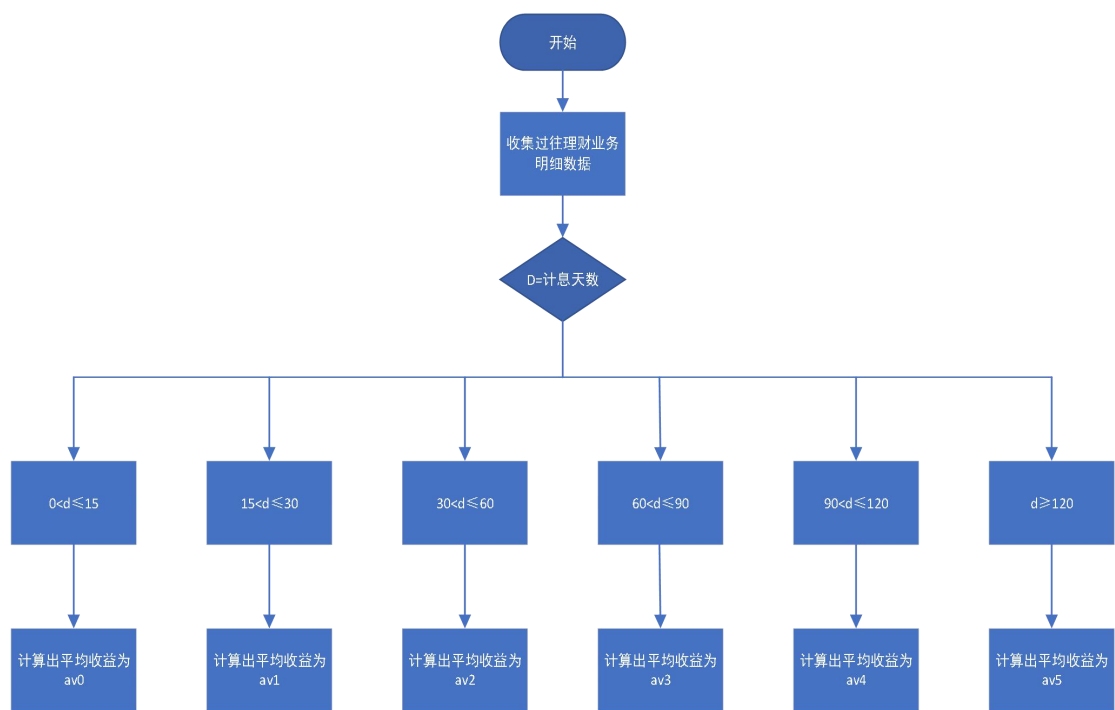


图 4 平均收益率计算流程

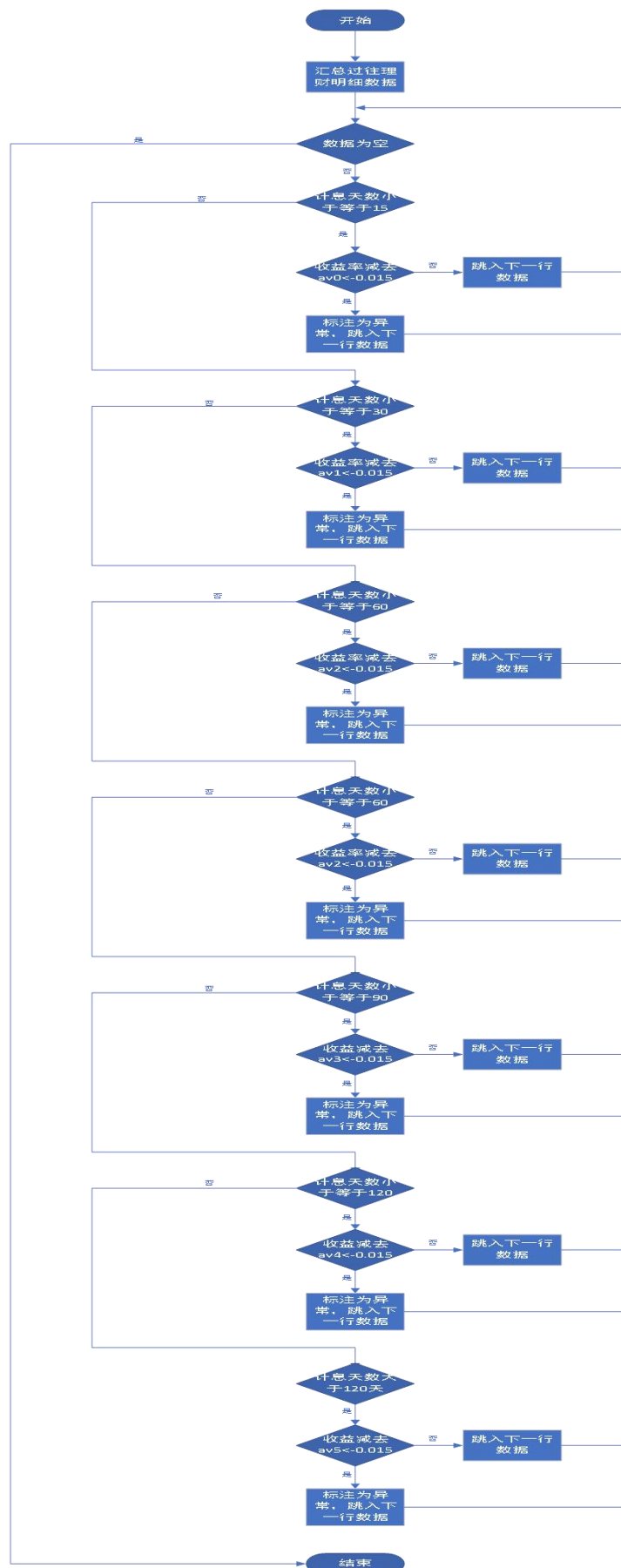


图 5 理财收益率异常判断流程

3. 审计方法和步骤

步骤一：收集整理被审计单位过往所有理财业务明细数据，包含产品名称、购买时间、赎回时间、计息天数、收益率、托管银行等要素；

步骤二：用 Python 读取数据，编写程序语言；

步骤三：按照计息天数将所有理财数据分为六个部分： $0 < \text{计息天数} \leq 15$ 、 $15 < \text{计息天数} \leq 30$ 、 $30 < \text{计息天数} \leq 60$ 、 $60 < \text{计息天数} \leq 90$ 、 $90 < \text{计息天数} \leq 120$ 、 $\text{计息天数} > 120$ ；

步骤四：计算六个部分理财业务的平均收益分别为 av_0 、 av_1 、 av_2 、 av_3 、 av_4 、 av_5 ；

步骤五：从第一行开始读取理财明细数据，设置判断条件为“数据为空”，判断为是，则流程结束，判断为否则进入步骤八；

步骤六：设置判断条件为“计息天数小于等于 15 天”，判断为是则进入步骤七，判断为否则进入步骤八；

步骤七：设置判断条件“收益率减去 av_0 小于 -0.015 ”，若判断为是则在表格中标注为异常，跳转至下一行并回到步骤五，若判断为否直接跳转至下一行并回到步骤五（见图 6）；

N	O	P	Q	R	S	T	U	V	W
是否异常	平均利率	备注							
异常	0.041243	第2行收益率异常, 收益率为0.0225,120+天平均收益率为0.041242806							
异常	0.035649	第3行收益率异常, 收益率为0.01755,30~60天平均收益率为0.03564902912621359							
异常	0.035649	第4行收益率异常, 收益率为0.01755,30~60天平均收益率为0.03564902912621359							
异常	0.040079	第5行收益率异常, 收益率为0.01755,90~120天平均收益率为0.04007863636363638							
异常	0.041243	第6行收益率异常, 收益率为0.01755,120+天平均收益率为0.041242806							
异常	0.035649	第7行收益率异常, 收益率为0.01755,30~60天平均收益率为0.03564902912621359							

图6 理财收益率异常数据

步骤八：设置判断条件为“计息天数小于等于 30 天”，判断为是则进入步骤九，判断为否则进入步骤十；

步骤九：设置判断条件“收益率减去 av1 小于-0.015”，若判断为是则在表格中标注为异常，跳转至一下行并回到步骤五，若判断为否直接跳转至下一行并回到步骤五；

步骤十：设置判断条件为“计息天数小于等于 60 天”，判断为是则进入步骤十一，判断为否则进入步骤十二；

步骤十一：设置判断条件“收益率减去 av2 小于-0.015”，若判断为是则在表格中标注为异常，跳转至一下行并回到步骤五，若判断为否直接跳转至下一行并回到步骤五；

步骤十二：设置判断条件为“计息天数小于等于 90 天”，判断为是则进入步骤十三，判断为否则进入步骤十四；

步骤十三：设置判断条件“收益率减去 av3 小于-0.015”，若判断为是则在表格中标注为异常，跳转至一下行并回到步骤五，若判断为否直接跳转至下一行并回到步骤五；

步骤十四：设置判断条件为“计息天数小于等于 120 天”，

判断为是则进入步骤十五，判断为否则进入步骤十六；

步骤十五：设置判断条件“收益率减去 av4 小于-0.015”，若判断为是则在表格中标注为异常，跳转至一下行并回到步骤五，若判断为否直接跳转至下一行并回到步骤五；

步骤十六：设置判断条件为“计息天数大于 120 天”，判断为是则进入步骤十七，此时已不存在判断为否的数据；

步骤十七：设置判断条件“收益率减去 av5 小于-0.015”，若判断为是则在表格中标注为异常，跳转至一下行并回到步骤五，若判断为否直接跳转至下一行并回到步骤五。

步骤十八：筛选出所有异常数据后，根据平均收益率计算出期望收益，期望收益减去实际收益，得出结果即为理财收益损失。

（三）成果三

根据 XX 基金已投子基金名称，通过 Python 爬虫在中国证券投资基金业协会检索相关数据。除政策性项目外，被审计单位已投 50 只子基金中 49 只运作状态正常，一只子基金受到假期及疫情影响尚未完成登记备案，已督促其尽快办理（见图 7）。

基金名称	运作状	基金协会特别提
合伙企业（有限合伙）	正在运作	
疗产业投资基金（有限合伙）	正在运作	
基金	正在运作	
兴产业基金	未找到公司信息	

图 7 基金运作状态数据

根据被审计单位提供的底层项目公司信息，通过 Python 爬

虫在“企查查”企业信息网站抓取项目公司工商信息、经营状况等数据。除政策性项目外，447 家底层项目公司中有 3 家经营状态异常，异常原因均为“工商行政管理部门在依法履职过程中通过登记的住所或经营场所无法取得联系”，经过逐一排查，被列入异常名单的 3 家公司都已移出，底层项目公司整体运营状况良好（见图 8）。

AJ	AK	AL
成立时间	异常原因	
2010-04-20	异常原因： 时间:2019-03-19 00:00:00~2019-04-12 16:14:54原因:通过登记的住所或者经营场所无法联系的	
2015-02-06	异常原因： 时间:2015-12-28 00:00:00~2016-01-15 00:00:00原因:通过登记的住所或者经营场所无法联系的	
2003-06-09	异常原因： 时间:2018-12-19 00:00:00 2019-02-22 00:00:00原因:通过登记的住所或者经营场所无法联系的时间:2018-12-19 00:00:00~2019-02-22 08:55:53原因:通过登记的住所或者经营场所无法联系的	

图 8 公司运作状态异常数据

1. 审计思路

XX 基金通过投资子基金间接投资到具体公司，通过这一方式起到引导放大社会资本的作用。截至被审计期末，XX 基金已投子基金 50 只，底层项目公司 447 家，数据量较大。项目组在获取了投资项目的明细数据后，首先考虑验证数据真实性，即提供的子基金和底层项目公司是否真实存在，另外要核查投资项目的运营状况，是否存在异常，如已破产清算、存在大额诉讼、重大违规事项等。

对于子基金的核查，根据《私募投资基金管理人登记和基金备案办法（试行）》规定，“私募基金管理人应当在私募基金募集完毕后 20 个工作日内，通过私募基金登记备案系统进行备案”，备案后可在中国证券投资基金业协会信息公示一栏查询到备案信息，包括基金名称、管理人、当前运作状态等数据。对于底层项目公司，通过企查查企业信息网，可查询到企业的工商信息，包括公司名称、成立日期、法人、经营范围、股东信息等数据，另外可看到企业的经营状态，是否存在破产清算、大额诉讼等异常状况。

由于上述信息都是通过互联网检索获取且数据量较大，项目组想到运用 Python 爬虫功能，以子基金和底层项目公司名称为关键信息，通过编写程序语言，利用 Python 自动在网络上抓取相关数据并输出到 excel 表格上，相对于人工逐一搜索，审计效率得到大幅度提升。

2. 审计流程图（见图 9）

（1）子基金信息核查

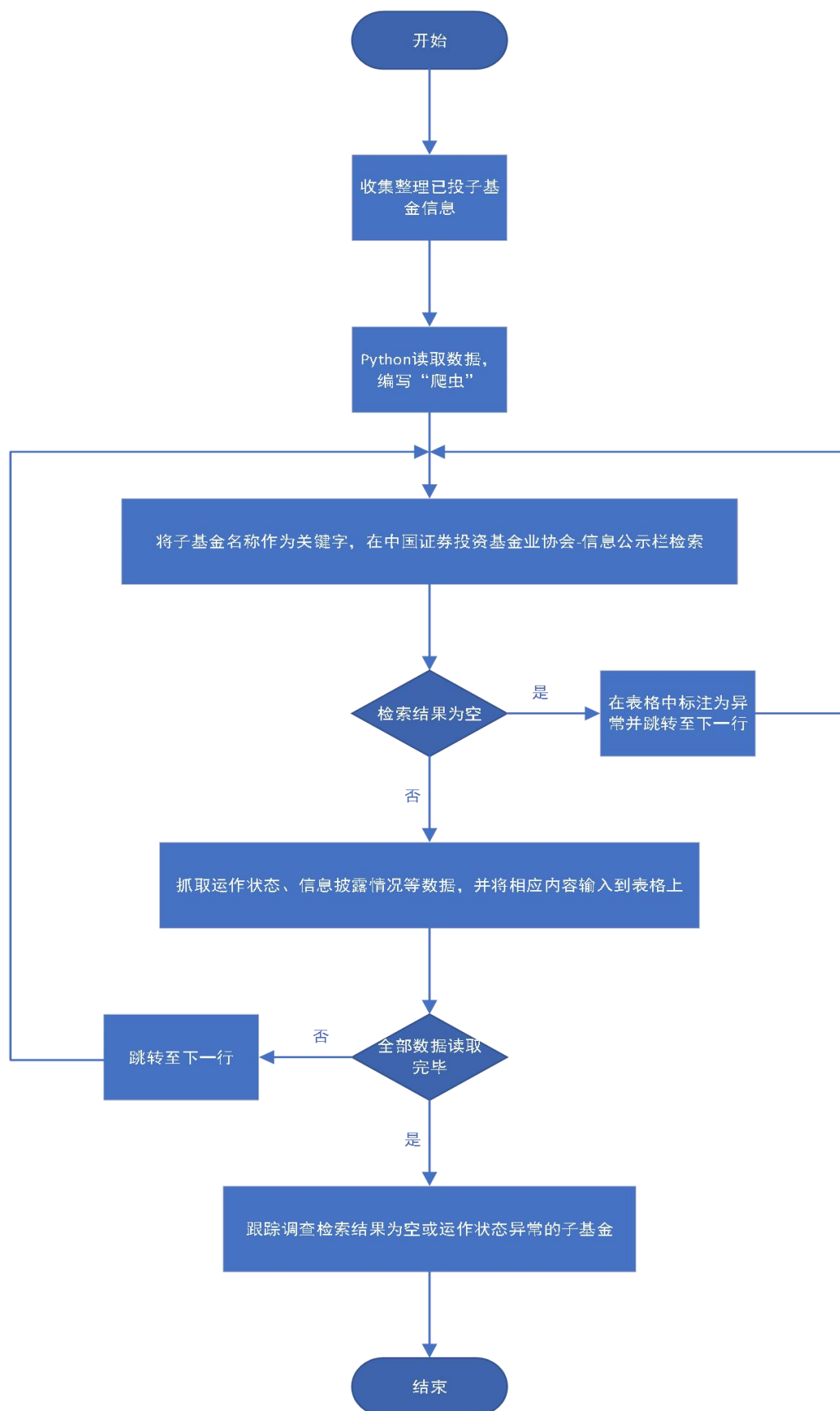


图 9 子基金信息核查流程

(2) 底层项目公司信息核查（见图 10）

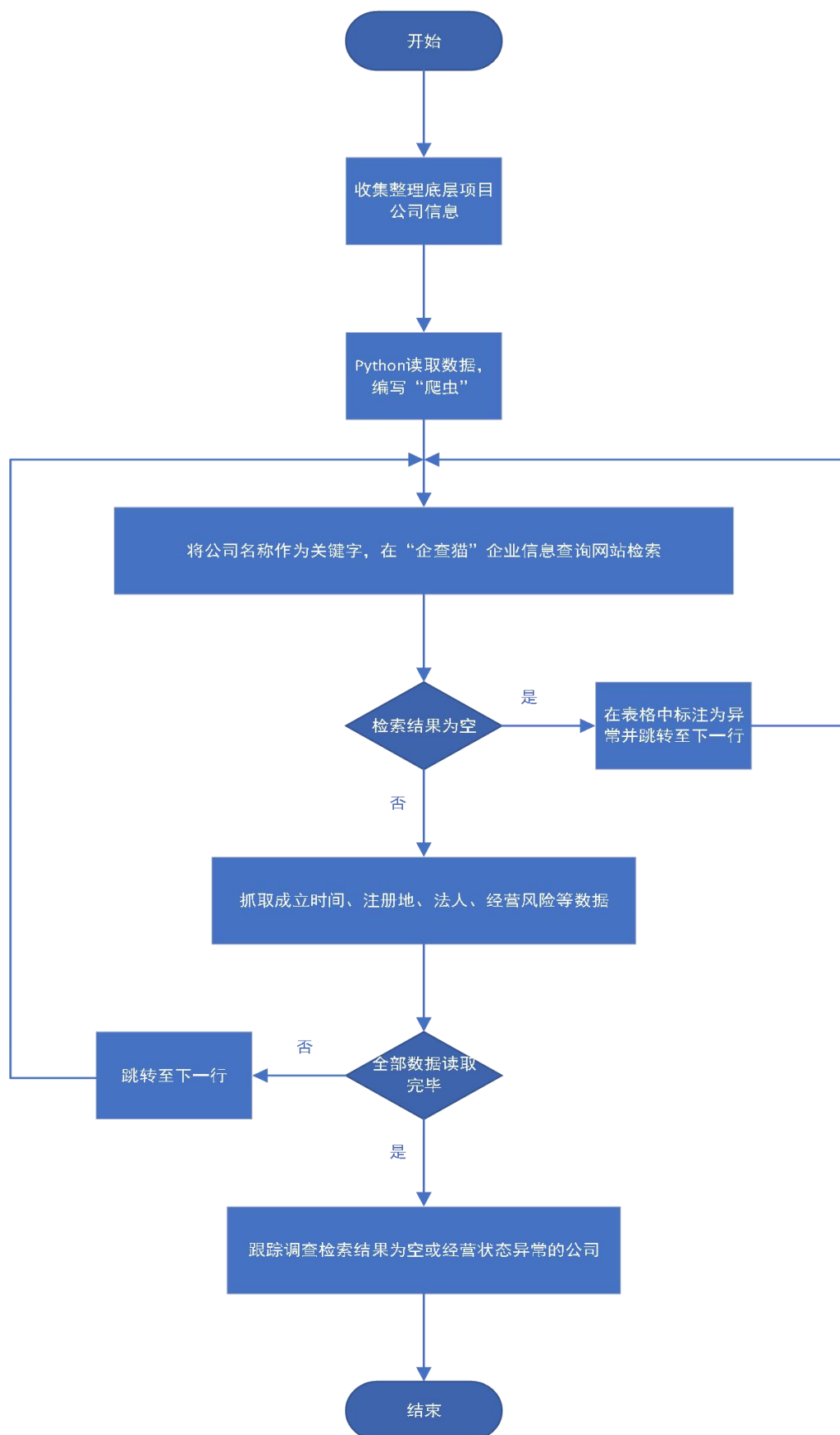


图 10 底层项目公司信息核查流程

3. 审计方法和步骤

(1) 子基金信息核查

步骤一：收集整理已投子基金信息；

步骤二：Python 读取数据，编写“爬虫”；

步骤三：将子基金名称作为关键字，在中国证券投资基金业协会上-信息公示栏检索；

步骤四：设置判断条件为“检索结果为空”，若判断为是，则在表格上标注为异常并跳转至下一行，回到步骤三，若判断为否，则进入步骤五；

步骤五：抓取基金运作状态、信息披露情况等数据，并将内容输出到表格上，跳转至步骤六（此处可尽可能多的获取信息，以便后期核查）；

步骤六：设置判断条件为“全部数据读取完毕”，若判断为否则跳转至下一行回到步骤三，若判断为是则进入步骤七；

步骤七：跟踪调查检索结果为空或运作状态异常的子基金，流程结束。

(2) 底层项目公司信息核查

步骤一：收集整理子基金已投底层项目公司信息；

步骤二：Python 读取数据，编写“爬虫”；

步骤三：将公司名称作为关键字，在“企查查”企业信息网站进行检索；

步骤四：设置判断条件为“检索结果为空”，若判断为是，则在表格上标注为异常并跳转至下一行，回到步骤三，若判断为否，则进入步骤五；

步骤五：抓取公司成立时间、注册地、法人、经营风险等数据，并将内容输出到表格上，跳转至步骤六（此处可尽可能多的获取信息，以便后期核查）；

步骤六：设置判断条件为“全部数据读取完毕”，若判断为否则跳转至下一行回到步骤三，若判断为是则进入步骤七；

步骤七：跟踪调查检索结果为空或经营状态异常的公司，流程结束。

三、案例总结

近年来，我市国资国企改革不断探索理念创新、模式创新，积极打造基金群体系，目前市属国资国企主导或参与设立的各类基金总规模已超过 3000 亿，预计在 2020 年将超过 5000 亿，进一步提升我市基金业中心城市地位。面对规模庞大的国资基金，审计机关如何做好国有资产的看门人和守护者，如何利用大数据全面、准确的揭示基金运行过程中存在的问题，希望本案例能为大家提供一些新的思路。

首先，针对我市国资基金资金量较大的特点，项目组注意到了大额资金沉淀及理财收益偏低的问题。选用 Python 作为数据分析工具，Python 被看作当前数据分析的首选语言，具有简单、免费开源、可移植性、可拓展性等特点，并且 Python 拥有丰富而强大的库，在数据分析领域可起到关键作用。通过 Python 编程，项目组可以全面、准确的了解被审计单位的资金闲置情况，同时准确筛选出理财收益偏低的异常数据。当被审计单位控制主体多、不同主体银行账户多，上述方法可大幅度提高审计效率及精确性，数据越庞杂、格式越混乱，越能体现出 Python 数据分析的优势。

此外，本案例应用到了 Python 爬虫功能，利用网络第三方的公开数据，核查基金投资项目的真实性及质量。第三方数据具有公开、独立的特点，以此信息作为参照，可增强审计结果的说服力。如没有应用现代化的计算机审计方法，当数据量较大时，只能通过抽查去推断整体状况，势必会影响审计成果的准确性，且人工逐一检索效率低、容易出错。大数据审计完美克服了上述缺点，既避免了抽查造成的结果失真，又提高了审计效率和质量。审计全覆盖是党中央、国务院对审计工作提出的明确要求，只有真正将大数据审计方法应用在具体的项目中才能有效做到全覆盖，充分发挥审计监督作用。

（作者单位：深圳市审计局）