

投资者问卷调查实证研究

耿中华*

摘要:目前,随着资本市场的发展,及时、准确、有效了解投资者的关注点、态度、意见建议越来越重要。这就对实际工作提出了一定的要求:了解你的保护对象即投资者,特别是中小投资者变得越来越重要;真实地展现投资者真实的想法越来越重要;在有效性保证的前提下,以较低成本开展调查工作越来越重要。理论上,投资者调查依旧属于社会调查,依据的最主要工具是各类型问卷,通过对问卷所获的数据进行分析,进而获取相应的结论。但在对未知或模糊的信息具体化过程中,如何去转化需求、如何去分析获得的数据,一直是投资者调查的重难点。笔者认为,首先,本文介绍了社会调查问卷的主要分类——量表问卷和非量表问卷;其次,本文介绍了量表问卷和非量表问卷在获取数据后应该选择或者可以适用的数据处理方法、问卷分析步骤和分析模型;最后,本文以中证中小投资者服务中心每年都进行的投资者知情权、行权、维权现状调查为案例进行实证分析,探究投资者在受到因虚假陈述带来的经济损失后,不可变的自身因素对投资者的维权选择是否有统计意义上的实际影响。

该表涵盖四个方面,包括近百个问题,马克思以此来全面了解工人的劳动、生活和思想状况。20 世纪以来,结构式的调查问卷被越来越多地使用,特别是在满足定量需求方面。目前,调查问卷通常被定义成由一系列与研究目标有关的问题组成的问题集 (questionnaire)。这些问题对实现调查目的至关重要,以需要调查的内容为依据设计而成,形式上多采用各种类型的题目,如单选、多选、跳转、量表、表格等,以了解、探究关注的内容,如态度、行为、事实现状、问题原因等,从而知道被调查对象身上存在什么、发生了什么。将调查、调查问卷与统计学上的抽样相结合可以在保证调查有效性的基础上降低人力、物力,提高时效性,因而已成为当前最主要的调查方式。调查问卷常常有不同的分类,但通常根据调查目的和问卷属性将问卷分为两大类,即量表问卷和非量表问卷,而在二者中选择合适的问卷对于调查是否能达到目的至关重要。

(一) 量表问卷

量表问卷相对在学术性研究中出现得较多,其特点是问卷本身相对会更更多地包含可量化表示的题目和选项,既可以客观展现调查对象对于某事物的态度、看法等,又可以通过对各变量之间关系的研究,发现变量之间隐藏的逻辑关系。量表的问卷设计上一个特殊的、重要的点是把核心题设计为量表题,将关心或者需要研究的问题用量表展现出来,这样才可以将被调查者的回答转换成数字编码,再经过量化处理得到精确的分析结果,以反映出研究对象的倾向或某种结局。量表问卷从设计架构上来看,可以分为六个部分,依次为筛选问题、样本基本背景资料问题、数据特征资料问题、数据基本态度问题、核心研究变量问题以及其他题。量表问卷的六个部分在结构与结构之间是关联的,单一结构内部选项又是相互独立的,这样才能较为全面系统地展现关注的问题。

(二) 非量表问卷

非量表问卷是针对某一主题或目的设定的一系列相关问题集合,可使用多种问题类型,如单选、多选、排序、填空、跳转等。在问卷设计方面,非量表问卷除了要注意结构与结构之间的关联性,单一结构内部选项的相互独立性外,还要重点呈现出一份问卷从几个大的层级或维度出发,描述了几大主要的问题,并说明每个大的层级或维度又细分成几个小的层级或维度,以及相对应地需要被了解的问题,从而较为全面系统地展现关注的问题。

(三) 量表问卷和非量表问卷的区别

非量表问卷与量表问卷最大的差异体现在四个方面:一是应用范围。由于非量表

续表

项目	完全同意	同意	无所谓	不同意	完全不同意
很难看	1	2	3	4	5
采用好的材质	5	4	3	2	1
设计太差	1	2	3	4	5
颜色太少	5	4	3	2	1
款式太少	1	2	3	4	5

(二) 语义差异量表

语义差异量表 (semantic differential scale) 是一种态度量表, 结构上是由两部分构成。第一部分是由一组极端的、反义的形容词组成, 以圈定调查对象作出的评价上限和下限, 如“非常好”与“非常不好”、“非常满意”与“非常不满意”。第二部分是在两个极端的形容词之间设置若干等级, 被调查对象可以选出一个代表自己的观点或态度。语义差异量表可用来测量人们对某个事物、某一现象、某一观点的主观评价。(见表 2)

表 2 语义差异量表示例

请对您的任课老师做出评价 (请在下列空格处打“√”)								
8	7	6	5	4	3	2	1	0
有创见								无创见
知识广博								知识贫乏
拘谨								洒脱
精干								平庸
耐心								急躁
冷淡								热情
死板								灵活
认真								马虎
糊涂								聪明
羞怯								大胆
负责								不负责
有条理								没条理
细心								粗心
友好								不友好

注: 0~8 为每项评价的分数。

(三) 二分量表

二分量表 (dichotomous scale) 是对于极端表述的调查, 通常是对两个相互排斥的项目进行回答。二分量表一般以“是”和“否”, 或者“同意”和“不同意”形式出现, 可以理解成是语义差异量表的一种极端简化。(见表 3)

表 3 二分量表示例

事项	是	否
1. 今天你是否忘记服药?	是	否
2. 过去一周你是否忘记服药?	是	否
3. 治疗期间, 当你感觉身体不适时, 你是否咨询过医生?	是	否

(四) 等级顺序量表

等级顺序量表 (ordinal scale) 通常是用来对多个待了解者排序。等级顺序量表与其他量表最大的不同在于, 该量表是将全部待了解对象同时展示给被调查者, 并要求他们对这些对象排序或打分。例如, 表 4 所示就是请被调查者对 7 个汽车品牌进行排名, 每个品牌都会有相应的、唯一的序号, 1 代表“最不喜欢”, 7 表示“最喜欢”。

表 4 等级顺序量表示例

品牌	排序
大众	
丰田	
特斯拉	
比亚迪	
沃尔沃	
奥迪	
雷克萨斯	

(五) 加总量表

加总量表, 也称古特曼量表, 由路易斯·古特曼 (Louis Guttman) 于 1950 年提出。该量表的特点是每一个量表都由一一对应的一组特定回答组成, 从而避免出现最终的得分相同但态度组成不同的现象, 有利于提高区分精度, 因此也被称为最难设计的量表。该量表的设计逻辑是如果调查对象接受了某个较强的问题, 原则上就会接受相对弱的问题, 这样调查对象的答案就会出现较为明显的层级。

量表

后
数
分
程

样本自京办初



卷作答时长与投资者回答真实有效性之间的关系、作答激励对投资者问卷作答真实性的影响三方面。对于影响关系分析,由于因变量属于二元变量,可采用 Logistic 回归分析。为了尽量消除样本量不同对结果的影响,使对比效果更准确,本文通过选取 2019 年、2021 年的投资者知权、行权、维权现状调查数据做实证对比研究,探究影响关系,其中 2019 年有效样本 12,116,2021 年有效样本 12,464。

(四) 2019 年实证分析

本文涉及 10 个影响投资者准确做出维权意愿的自变量

表 5 2019 年变量两两之间的相关性

变量	答题时	地区	年龄	性别	教育背景	职业	金融从业人员	入市时间	入市资金	股票资金占比
皮尔逊相关性	-.029**	.000	-.181**	.078**	.059**	.063**	.228**	-.041**	-.007	-.012
Sig. (双尾)	.001	1.000	.000	.000	.000	.000	.000	.000	.453	.191
个案数	12, 116	12, 116	12, 116	12, 116	12, 116	12, 116	12, 116	12, 116	12, 116	12, 116

注: ** 在 .01 级别 (双尾), 相关性显著。

* 在 0.05 级别 (双尾), 相关性显著。

通过相关性分析发现, 维权意愿的真实表达与投资者的答题用时、年龄、性别、教育背景、职业、是否属于金融从业人员、入市时间、股票资金占家庭资产之比有相关性, 与投资者的地区、入市资金无相关性。

3. 多变量叠加分析
进行单变量分析后, 本文通过 SPSS 的 Logistic 回归模型验证多变量叠加后的综合作用, 探究多因素对投资者回答真实性的影响, 结果见表 6。

表 6 2019 年全变量 Logistic 回归检验

变量	B	标准误差	瓦尔德	自由度	显著性	Exp(B)
答题用时	.000	.000	12.257	1	.000	1.000
地区	-.001	.002	.355	1	.551	.999
年龄	-.409	.027	226.664	1	.000	.664
性别	.234	.050	22.111	1	.000	1.264
教育背景	.094	.027	11.932	1	.001	1.098
职业	.011	.010	1.226	1	.268	1.011
金融从业人员	1.124	.055	422.805	1	.000	3.078
入市时间	-.115	.019	35.130	1	.000	.891
入市资金	-.023	.022	1.083	1	.298	.977
股票资金占比	-.079	.026	9.561	1	.002	.924
常量	1.121	.223	25.310	1	.000	3.067

注: 输入的变量有答题用时、地区、年龄、性别、教育背景、职业、金融从业人员、入市时间、入市资金、股票资金占比。

通过对 10 个变量的数据回归分析, 可以发现地区、职业背景、入市资金这 3 个因素对投资者做出准确回答没有统计上的显著影响。虽然答题用时检测结果是显著的, 但回归系数为 0, 这说明实际上答题时间对投资者真实意愿反映没有实质影响。

但通过线下面访发现，投资者回答在“入市资金”这一类问题时，相对于单一地回答入市资金总额，往往不太愿意回答入市资金占家庭总资产的比例。其背后原因一可能是入市资金属于一个绝对数，在记忆和计算上更方便；另一方面可能是基于隐私的考虑，投资者希望保护自己的财富信息。为此，需要验证入市资金和股票资金占家庭总资产之比对最终回答结果的实际影响有多大区别，再结合投资者偏好做出保留谁的决断。

下面先进行保留股票资金占家庭总资产之比计算，而后进行保留入市资金计算，结果见表7。

表7 2019年 Logistic 回归结果(保留股票资金占比)

变量	B	标准误差	瓦尔德	自由度	显著性	Exp(B)
年龄	-.396	.026	234.055	1	.000	.673
性别	.230	.049	21.695	1	.000	1.259
教育背景	.091	.026	12.122	1	.000	1.095
金融从业人员	1.097	.054	411.190	1	.000	2.995
入市时间	-.116	.019	36.854	1	.000	.891
股票资金占比	-.108	.020	27.907	1	.000	.898
常量	1.266	.196	41.618	1	.000	3.547

注：输入的变量有年龄、性别、教育背景、金融从业人员、入市时间、股票资金占比。

经计算发现，如果只保留股票资金占家庭总资产之比这个变量，回归模型有效，纳入模型的各项变量全部显著，结果见表8。

表8 2019年 Logistic 回归结果(保留入市资金)

变量	B	标准误差	瓦尔德	自由度	显著性	Exp(B)
年龄	-.401	.026	239.212	1	.000	.670
性别	.226	.049	20.812	1	.000	1.253
教育背景	.098	.027	13.526	1	.000	1.103
金融从业人员	1.066	.053	400.211	1	.000	2.905
入市时间	-.125	.019	42.143	1	.000	.882
入市资金	-.071	.018	16.047	1	.000	.931
常量	1.286	.197	42.562	1	.000	3.620

注：输入的变量有年龄、性别、教育背景、金融从业人员、入市时间、入市资金。

经计算发现，如果只保留入市资金这个变量，回归模型依旧有效，各变量依旧全

部显著。

上述两个变量的计算结果显示,股票资金占家庭资产之比与入市资金占家庭资产之比这两个变量各自都对投资者准确回答有显著影响,当两者合并出现时入市资金占家庭资产之比会更为显著。因此,可以推断,当调查对象是广泛性的投资者或入市资金占家庭资产之比敏感度不高时,基于问卷应该有的无差别接受性前提要求,应该选择入市资金占家庭资产之比这个绝对指标。当调查对象是小部分的特殊群体或调查本身是对某一类群体的专享调查时,应该优先选择股票资金占家庭总资产之比这个相对指标。

4. 2019年三权调查数据实证结论

通过全变量回归检测,再结合相关性分析可以发现:
第一,投资者的年龄、性别、教育背景、是否金融从业、入市时间、入市资金、股票资金占家庭总资产之比会对投资者的维权意愿产生统计意义上的显著影响。
第二,投资者的地区分布不会对维权意愿产生统计意义上的显著影响,即不同地域的投资者受到损失后,维权意愿没有显著差异。
第三,投资者答题用时也不会对投资者真实意愿的表达产生实际影响,不会对投资者真实意愿的表达产生统计意义上的显著影响。

(五) 2021年实证分析

2021年投资者知权、行权、维权现状调查问卷相较于2019年做了一定的优化,一方面剔除了一些不显著的变量,另一方面对一些关心的变量进行了再次验证,从而将10个变量减少到6个变量,分别为答题用时、地区、年龄、教育背景、入市时间、入市资金。同时,增加了一个新的变量,风险承受等级(以下简称风险等级)。上述7个自变量,除了答题用时以外,剩余的6个都是投资者自身的基本属性,是不可改变的。

1. 数据处理

同2019年数据处理方法一样,本文通过赋值的形式对变量进行数据化。数据赋值原则是:两项选择变量赋值0或者1;多项选项中如果有时间顺序或者递增递减规律,则按时间顺序或增减规律赋值;投资者问卷作答时间以填写问卷实际耗时为依据;投资者遭受虚假陈述带来的损失后,愿意去维权赋值1,不愿意去维权赋值0。

2. 单变量分析

此处研究与2019年类似,在变量研究中,通常除了研究单一自变量对因变量的影响,还需考虑自变量之间的交互影响,虽然单个变量两两之间的相关性不一定会产生实际的影响。对于投资者知权、行权、维权现状调查,投资者最终作答的意愿是7

个自变量综合影响后的结果,因此,对于7个变量之间是否有相关性以及关联性的强弱,不是本次实证分析重点关注的內容。本次重点关注投资者在受到损失后,7个单因素对维权选择是否有影响,7个因素的叠加影响使投资者在问卷上反映出的信息是否真实有效,以及推断投资者对问卷的可接受度。

本文通过 SPSS 软件,运用双变量相关性检验判断初始影响因素,结果见表 9。

表 9 2021 年双变量之间的相关性

变量	所用时间	省市	年龄	教育背景	入市时间	入市资金	风险承受等级
皮尔逊相关性	-.128 **	.013	-.115 **	-.055 **	-.058 **	.073 **	-.071 **
Sig. (双尾)	.000	.133	.000	.000	.000	.000	.000
个案数	12,464	12,464	12,464	12,464	12,464	12,464	12,464

注: ** 在 0.01 级别(双尾),相关性显著。
* 在 0.05 级别(双尾),相关性显著。

通过对 7 个变量相关性分析发现:维权意愿的真实表达与投资者的答题所用时间、投资者的年龄、投资者的教育背景、投资者的入市时间、投资者的入市资金、投资者的风险承受等级有相关性,与投资者的地区,即所在省市无相关性。

3. 多变量叠加分析

进行单变量分析后,本文通过 SPSS 的 Logistic 回归模型验证多变量叠加后的综合作用,探究多因素对投资者回答真实性的影响,结果见表 10。

表 10 2021 年全变量 Logistic 回归模型

变量	B	标准误差	瓦尔德	自由度	显著性	Exp(B)
答题用时	-.001	.000	205.926	1	.000	.999
年龄	-.234	.021	126.180	1	.000	.791
教育背景	-.237	.022	114.607	1	.000	.789
入市资金	.234	.018	167.622	1	.000	1.263
风险承受等级	-.168	.018	88.791	1	.000	.845
地区	.001	.001	.880	1	.348	1.001
常量	.997	.125	63.533	1	.000	2.710

注:输入的变量有答题用时、年龄、教育背景、入市资金、风险承受等级、地区。

通过回归模型验证,可以再次确认地区不同不会对投资者维权作答真实性产生显著影响。

剔除地区因素以后,再进行 Logistic 回归分析,回归模型依旧有效,纳入模型的各

变量全部显著,结果见表 11。

表 11 2021 年变量 Logistic 回归模型(剔除地区因素)

变量	B	标准误差	瓦尔德	自由度	显著性	Exp(B)
答题用时	-.001	.000	206.303	1	.000	.999
年龄	-.234	.021	126.731	1	.000	.791
教育背景	-.238	.022	115.587	1	.000	.788
入市资金	.233	.018	166.808	1	.000	1.262
风险承受等级	-.169	.018	89.534	1	.000	.845
常量	1.055	.109	94.006	1	.000	2.871

注:输入的变量有答题用时、年龄、教育背景、入市资金、风险承受等级。

4. 2021 年投资者知权、行权、维权现状调查数据实证结论

通过全变量回归检测,再结合相关性分析发现:
 一是投资者的年龄、教育背景、入市资金、风险承受等级会对投资者的维权意愿产生统计意义上的显著影响,这与 2019 年的结论一致。
 二是投资者的地区分布不会对维权意愿产生统计意义上的显著差别,即不同地域的投资者受到损失后,维权意愿没有显著差异,这与 2019 年的结论一致。
 三是物质激励的消失的确使得答题时间对真实意愿的表达产生了一定影响。2021 年之前投资者作答有一定的抽红包激励,目的是提升投资者参与作答的积极性,保证回答的准确性。2021 年开始,投资者调查工作取消了抽红包激励。通过 2021 年数据分析可以发现,红包激励的消失的确对投资者的真实作答有一定的影响——答题用时在回归方程中有 -0.01 的影响系数,即表示作答时间越长,投资者真实意愿表达越容易失真。但考虑到其影响效果仅为 1%,因此可以说实际上的影响微乎其微。

五、结论及建议

(一) 实证结论

基于 2019 年和 2021 年的实证结果及对比,可以得出如下结论:
 一是投资者的年龄、性别、教育背景、是否金融从业、入市时间、入市资金或股票资金占家庭总资产之比等因素会对投资者的维权意愿产生统计意义上的显著影响,

职业因素却不会。

二是地域因素的确是众多调查中需要重要考虑的因素,但对于投资者受损后的维权意愿而言,投资者的地区分布不同不会对维权意愿产生统计意义上的显著影响。

标。当调查涉及的对象是小部分的特殊群体或调查本身是对某一类群体的专项调查时,应该优先选择股票资金占家庭总资产之比这个相对指标。这些措施可以在