

参赛队号：（由大赛组委会办公室填写）

2021 年（第七届）全国大学生统计建模大赛

参赛学校：

北京师范大学

论文题目： 数据新动能评价体系构建及分区实证研究

参赛队员：

徐墨姝 王艺杰 刘晓轩

指导老师：

李高荣 张欣

摘要

数据要素作为新型的生产要素,已随着信息经济的发展融入经济价值创造系统中,广泛影响着生产力的发展。而数据新动能,即是将数据作为一种新型生产要素,对其它要素产生协同促进作用,成为推动经济高质量发展的新型动能。然而遗憾的是,数据新动能这个概念提出时间较晚,且以往的研究大多是描述性的,主要聚焦于数据要素的理论依据、关于数据要素的政策等,实证分析比较少见。为了对数据新动能这一概念进行量化研究,对数据要素与经济发展之间的关系进行实证研究,我们小组开展了此次调查。

我们结合国内实际发展情况构建了数据新动能评价体系,用以测算全国数据新动能指数,并基于此评价体系建立数据新动能指数、数字普惠金融指数与人均国内生产总值之间的模型,以进行实证分析。在数据新动能评估指标下设置 5 个指标,采用熵值法进行指标权重计算,测度 2011-2020 年全国数据新动能指标权重情况。得出的结论是,全国数据新动能的发展势头良好,指数从 3772 上升至 40181,涨幅显著。随后,按照省份地理位置、是否为一线城市、地区生产总值排名选出 8 个代表省份,考察数据要素与地区生产总值的关系。建立了面板数据模型,探究数据要素在不同省份的发展和影响的差异。此外,辅以问卷调查的方式探究大众对数据的使用、感受、重视程度。在调查分析结果的基础上,给出相应结论与建议,强调数据版权保护,呼吁加强数字信息基础设施的建设。

关键词: 数据新动能; 熵值法; 面板数据模型;

目 录

摘 要.....	2
表格清单.....	5
插图清单.....	6
第一部分 绪论	7
一、研究背景.....	7
二、研究目的及研究意义.....	7
三、文献综述.....	8
(一) 国内关于数据要素发展的相应文献	8
(二) 国内关于构建新动能统计测度的相关文献	8
(三) 国内关于数据新动能对经济发展作用的文献	8
第二部分 数据要素在全国的发展	10
一、调查方案设计.....	10
(一) 调查目的	10
(二) 调查问卷设计思路	10
(三) 调查方法	10
二、调查结果统计及分析.....	12
(一) 问卷发放及回收情况	12
(二) 调查对象基本情况	12
(三) 大众对数据的使用程度	12
(四) 大众对数据的感受程度	15
(五) 大众对数据的重视程度	16
(六) 结论与建议	21
第三部分 数据新动能评价体系	22
一、普惠金融指数.....	22
二、数据新动能指数体系的构建.....	24
(一) 数据来源	24
(二) 数据新动能评估指标体系的构建	25
(三) 综合评价方法	25

(四) 数据新动能评估指标综合评估结果	28
三、总结	29
第四部分 基于数据新动能评价体系的实证分析	30
一、模型设定	30
(一) 数据新动能与全国生产总值	30
(二) 数据新动能与地区发展	30
二、数据来源与预处理	31
三、数据新动能与人均国内生产总值的关系	32
四、数据新动能与地区发展的关系	33
五、结论与建议	36
(一) 结论	36
(二) 建议	36
参考文献	37
附录 1 调查问卷	38
附录 2 Stata 代码	39
致 谢	40

表格清单

表 1	数字普惠金融指标体系 ^[11]	22
表 2	指标数据来源.....	25
表 3	全国数据新动能指数指标.....	25
表 4	2011-2020 年中国数据新动能指数	28
表 5	相关变量的描述性统计.....	31
表 6	数据新动能的发展与人均 GDP	32
表 7	数据新动能与地区发展回归结果.....	34
表 8	数据新动能指标单独回归结果.....	34

插图清单

图 1	调查对象分布地区.....	12
图 2	数据在工作中的使用情况.....	13
图 3	数据的使用来源.....	13
图 4	数据使用频率情况.....	14
图 5	数据软件的掌握情况.....	14
图 6	大数据的重要性客观体会调查结果.....	15
图 7	大众身边的数据出现情况.....	16
图 8	大数据的重要性主观体会调查结果.....	17
图 9	大数据应用和发展的重视情况.....	18
图 10	数据发展关注的问题调查情况.....	18
图 11	数据重要性第一位情况.....	19
图 12	数据重要性第二位情况.....	19
图 13	数据重要性第三位情况.....	20
图 14	数据重要性第四位情况.....	20
图 15	数字惠普金融指数.....	24
图 16	熵值法计算思路.....	26
图 17	全国数据新动能指数.....	28
图 18	2011-2020 年各省人均 GDP.....	31
图 19	各省数字惠普金融指数发展趋势.....	32
图 20	数据新动能指标系数.....	35

第一部分 绪论

一、研究背景

近年来,随着社会信息化的发展,数据被当成一种新型生产要素,成为推动经济高质量发展的新动能。2020 年 4 月 9 日,国务院印发《关于构建更加完善的要素市场化配置体制机制的意见》,明确将数据作为一种新型生产要素纳入《意见》。

生产要素的种类和形态随着经济发展和时代变化不断变化,21 世纪是信息经济的世纪,以大数据为代表的信息资源逐渐演进成一种生产要素,在经济社会中越来越重要,并且对生产力发展有广泛影响。中国把推进社会、经济的数字化,以及加强利用数据作为创新发展的重要动能,并在 2015 年推动了新一轮科技革命的迅猛发展,如今在人工智能、大数据等方面在国际上占有领先优势。

数据新动能这个概念提出时间较晚,现有文献缺乏对其的实证研究。以往的研究大多是描述性的,主要聚焦于数据要素的理论依据、关于数据要素的政策等,实证分析比较少见。考虑到中国当前经济下行的压力,以及建设现代化强国的政策目标,对数据新动能进行量化研究尤为重要。

二、研究目的及研究意义

数据作为新要素在中国的发展尚未成熟,未来还面临许多机遇和挑战。本研究构造了用于衡量数据要素发展情况的新指标,并用于实证分析中,帮助加深理解数据要素促进宏观经济增长的作用,为进一步使用数据要素的工作提供参考。

三、文献综述

（一）国内关于数据要素发展的相应文献

何玉长和王伟^[1]对数据和数据市场化进行了释义，讨论了数据要素市场运行机制，并对数据要素市场规范化提出了几点建议。王建冬和童楠楠^[2]提出了数据要素与其他要素联动的三层次模型，即基础层数字产业化、支撑层产业数字化和整合层全要素数字化，构建了数据要素与其他生产要素协同联动的“五链协同”机制。许宪春等人^[3]首先探讨了大数据与绿色发展的概念与特征，其次以货车帮、滴滴出行等大数据平台对绿色发展的影响为例，建立模型分析了大数据对于绿色发展的作用，最后得出大数据对绿色发展有显著的正向影响的结论，并给出了几点政策建议。然而，数据要素属于新型要素，数据新动能的概念也属于新概念，关于这方面的文献大多着重于概念诠释和与原有要素的对比，提出政策建议，缺乏对数据新动能的定量描述。

（二）国内关于构建新动能统计测度的相关文献

关于新动能的统计测度研究，在文献中体现出来的大致分成两类方法：

第一种方法是用综合评价的方法，选取指标进行评价。如邵明振^[4]等人选取了2009-2018年河南省各地市的面板数据，构建了河南省经济新动能指标评价体系，并结合层次分析法和熵值法对经济新动能进行了测度和评价分析，考察了经济新动能、经济增长和产业结构的关系。郑世林和熊丽^[5]利用2015-2018年全国各省的面板数据，选择功效系数法和标杆法相结合的方法作为综合指数评估方法。对结果进行评估和实证分析，估算了中国整体、四大板块和各个省市培育经济发展新动能的成效。

第二种方法是基于全要素生产率（TFP）的计量测度方法。如郑江淮等人^[6]基于柯布-道格拉斯生产函数，分解全要素生产率的增长动力来源，通过各指标的动能值进行测算，合成出经济增长的动能指数。然而，这两种方法都存在着一一些问题。主要体现在：部分指标的选取都存在不太严谨的地方，基于 TFP 的测量方法考虑的角度过于片面。

（三）国内关于数据新动能对经济发展作用的文献

陈昌兵^[7]从我国三大产业的增加值和就业结构的变化分析，提出我国三次产业发展的动力分别是农业经济、工业经济和服务产业及数字经济，得出经济高质量发展的根本不在于提高劳动生产率和全要素生产率，而在于将原有的发展方式向人力资本和技术创新驱动增长方式转型。荆文君和孙宝文^[8]从微观和宏观两个层面，分别探讨了数字经济与经济增长的关系，及其促进经济高质量发展的内在机理。在微观层面，数字经济提高经济的均衡水平是通过形成具有长尾效应的规模经济。在宏观层面，数字经济促进经济增长可以通过新的投入要素、新的资源配置效率和新的全要素生产率这三个方面。张勋等人^[9]建立了数字金融与包容性增长的双向固定效应模型，他们的数据来源是家庭追踪调查（CFPS）数据和北京大学数字普惠金融指数，发现数字金融的发展仅对农村居民的家庭收入有显著的正向影响，从而促进了中国的包容性增长，同时，文章还建立了数字金融与家庭创业的二元选择模型，发现数字金融的发展仅对农村居民的创业行为有显著的正向影响，从而满足数字金融到普惠金融，再到家庭创业机会均等化，最后到包容性增长的传导机制。孙玉环等人^[10]首先借助普惠金融指数对中国普惠金融的发展进行描述性统计，其次使用大连市数据和直到三级指标的普惠金融指数建立了固定效应面板模型进行实证研究，对城乡、收入阶层、受教育程度进行了分样本回

归。总体而言，针对数据新动能对经济发展的作用的实证研究偏少。

第二部分 数据要素在全国的发展

一、调查方案设计

（一）调查目的

数据要素在数字经济时代的推动作用已经不可小觑，为了探究数据要素在大众中的发展情况，我们将通过问卷调查，在全国范围内，通过线上发送问卷，帮助我们对数据这一“新型生产要素”的使用情况进行直观、主观的了解。

（二）调查问卷设计思路

设计本文问卷前，我们查找并学习了前人做过的关于数据使用的相关研究问题。寻求同类研究问卷设计规律，根据规律反映出一般同类研究多有一下几个部分：使用深度、覆盖广度以及使用感受等。本文问卷遵循规律，在问卷正式发放前进行试验，根据问卷初稿的反馈效果修改问卷，最终确定问卷内容，可以说本问卷是科学合理的，可以保证回收的数据真实可靠。

此次调查问卷题目一共设置 9 个题目，分为三大类题目分别为 A 类、B 类、C 类，代表受访者对数据的使用程度、对数据的感受程度和对数据的重视程度。

（三）调查方法

1. 文献分析法

我们通过搜集、学习大量文献，了解到关于数据的发展对于经济发展的种种

影响。然而，数据新动能这一概念提出的时间较晚，与此关联紧密的文献较少，所以我们主要搜集数据发展、经济发展、数字金融相关的文献，进行综合的比较和学习，以一种非介入式的方法快速了解我们关注的这几个领域的发展和动态。在此次研究中，文献调查法是主要使用的方法之一，贯穿于研究的始终。在研究初级阶段，我们通过文献调查法了解调查课题，而在研究的中后期，我们也从文献中学习到指标的编写方法。

2. 问卷调查法

本研究的调查主要使用的是问卷调查法。问卷调查的结果可以帮助我们对数据这一“新型生产要素”的使用情况进行直观、主观的了解，和本研究后续的统计建模实证分析的研究结果相对独立。虽然问卷调查法的结果数据不进入我们建立的模型，但是仍对此次研究具有重要意义。我们通过互联网发放问卷，且尽量覆盖各年龄、职业层次的受访者，目的是使调查结果不出现系统性偏差。问卷调查帮助我们通过较为快速、简便的方法，对我们感兴趣的几大问题进行针对性的精准探究，比起其他的二手数据更能反映我们关注的对象的情况。

3. 头脑风暴法

头脑风暴法是一种讨论、座谈的方法，旨在放松气氛，使得人们在轻松的氛围中发挥出自己想象力和创造力，积极地思考、积极地参与。在此次研究中，三位小组成员使用头脑风暴法讨论了新指标的构建形式，每位成员都充分发挥了创造力，畅所欲言，提出了自己个人的想法并且进行充分讨论。小组成员形成初步想法后，与导师深入讨论，并且根据导师的指导进行更深层次的思考和探究，最终确定了本次研究的指标构建模型。

二、调查结果统计及分析

（一）问卷发放及回收情况

本文的问卷通过问卷星进行线上发放，经统计，我们一共收到 116 份有效问卷，问卷的发放回收率达到 90%以上，此数据满足了市场调查问卷回收率要求，是有效的。

（二）调查对象基本情况

如图 1 所示，是回收问卷的 IP 地址来源分布图，颜色代表省份。此次问卷调查覆盖 13 个省/市/自治区，样本在全国分布较为均匀，可以用来进行数据分析。



图 1 调查对象分布地区

（三）大众对数据的使用程度

在此次调查回收的问卷中，我们对大众使用数据的情况及形式进行了统计，

可视化结果如图 2 所示。

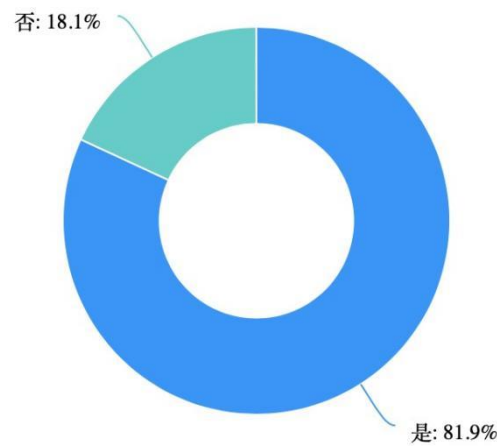


图 2 数据在工作中的使用情况

调查结果数据显示，18.1%的人主观上认为他们在工作中不需要使用数据，81.9%的人主观认为他们在工作中需要使用到各种形式的数 据。由此可见，绝大多数的人的日常工作涉及各类数据的使用。

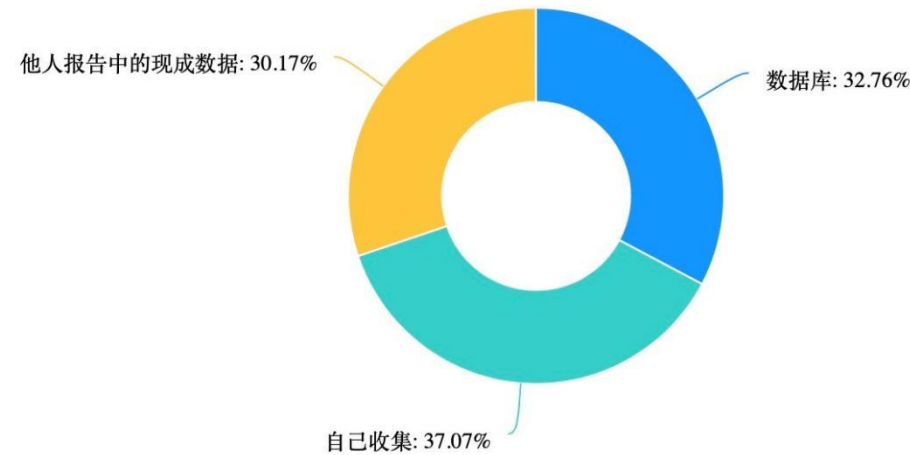


图 3 数据的使用来源

根据调查结果，如图 3 所示。受访者中，37.07%的人表示，其使用的数据源于自己收集的一首数据。其次为占比 32.76%的人群，他们表示自己使用的数 据来源于数据库。最少的是 30.17%的人群，他们使用的数据为他人报告中的现成数据。

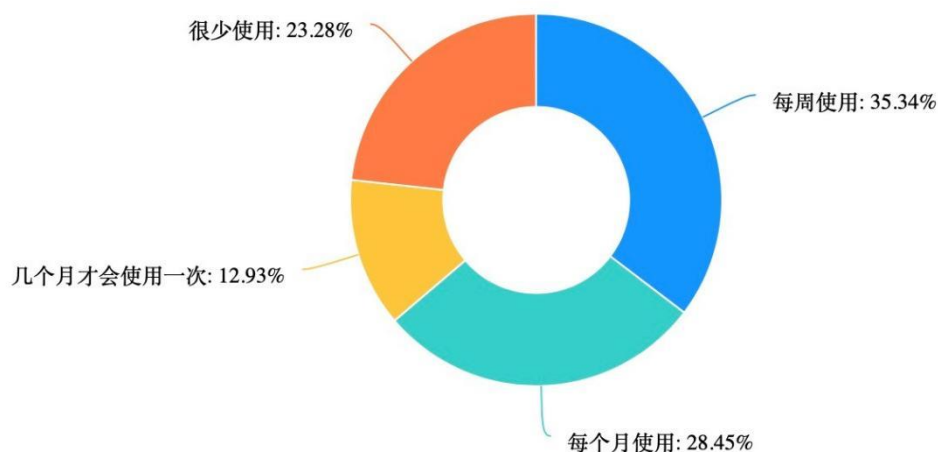


图 4 数据使用频率情况

关于使用数据的频率的调查结果如图 4 所示，35.34%的受访者表示每周都会使用数据。其次是 28.45%的受访者，她们每个月使用数据。接着 23.28%的受访者表示很少使用数据。最少的是 12.93%的受访者，他们表示几个月才会使用一次或基本不使用数据。

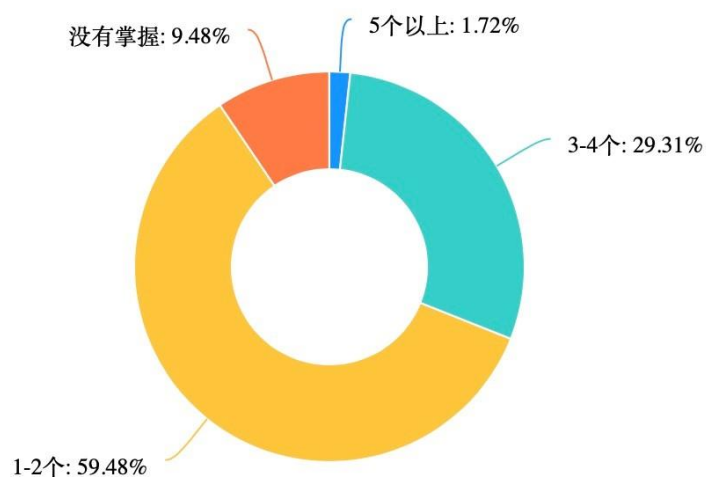


图 5 数据软件的掌握情况

根据图 5 所示，受访者中，59.48%的人掌握了 1-2 个数据处理工具(如 Excel)，人数超过一半。29.31%的人掌握了 3-4 个数据处理工具，1.72%的人掌握了 5 个以上的数据处理工具。没有掌握任何一个数据处理工具的人数占此次调研的

9.48%。

（四）大众对数据的感受程度

大众对数据使用情况的感受不尽相同，为了了解大众的感受程度，我们在问卷中设计了相应问题，调查结果如下：

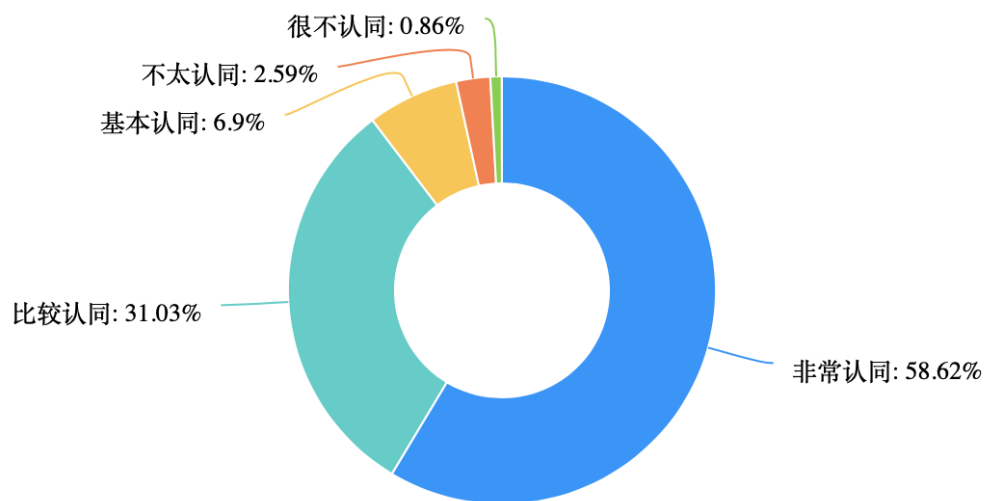


图 6 大数据的重要性客观体会调查结果

图 6 显示，对与“感受到大数据正在变得越来越重要”这一命题，58.62%的受访者表示非常认同，也即对于大数据日趋重要的感受非常强烈，人数超过半数。31.03%的受访者表示比较认同，也就是同样有此类感觉，但不是非常强烈。6.9%的受访者表示基本认同，也就是感受程度中等。还有 2.59%的受访者不太认同、0.86%的受访者很不认同，表示对这一现象并无明显感受。

大众中大多数感到数据要素的重要性正在提升，说明客观上数据要素存在着发展趋势，也是促使我们在后续的研究中寻找数据要素发展情况的指代变量进行定量的研究。

大众对数据的感受不仅有数据应用的发展情况，还有生活中接触的数据。这类数据不仅限于工作中需要处理的数据，还有嵌入到衣、食、住、行各方面的数

据应用，就如手机购物软件的智能推送，能够根据客户历史购物及浏览数据智能推送其最有可能心仪的商品。大数据应用极广，只要稍一留心就能发现它出现在我们身边。

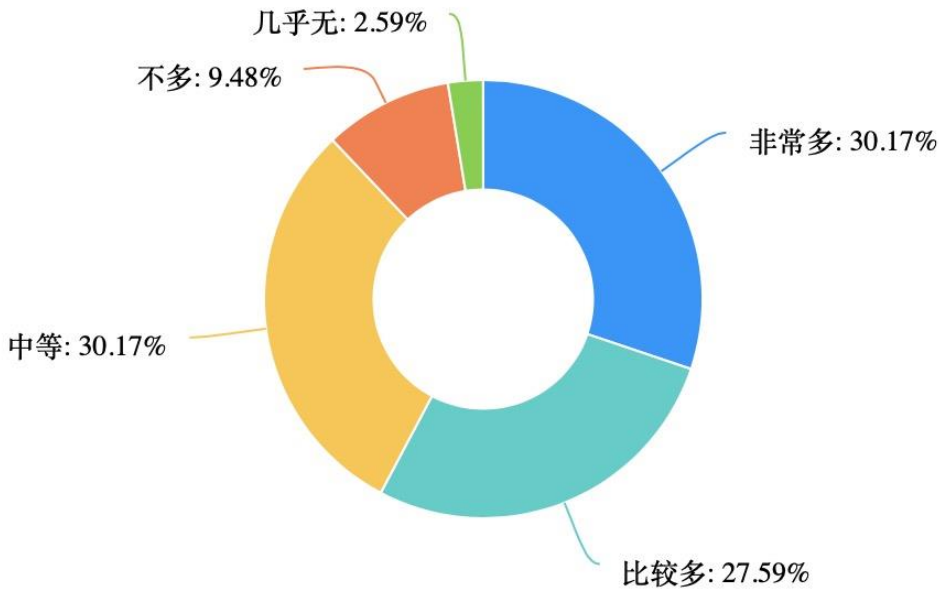


图 7 大众身边的数据出现情况

如图 7 所示，在受访人群中，30.17%的人感受到生活中接触的数据非常多，同样有 30.17%的人感到接触的数据应用中等。人数排在其次的是感到比较多的，占受访者比重的 27.59 个百分点。少数（9.48%）的受访者对于大数据的感受微弱，感到生活中接触的数据不多，这不仅与其生活中接触的数据多寡有关，同时也与其对于数据的感知力有关。极少（2.59%）的受访者表示，他们几乎无法感受到身边的数据。

（五）大众对数据的重视程度

数据的发展情况与大众对数据的重视程度不同。重视程度刻画了大众主观上对于数据发展前景的预测或偏好。虽然小样本的问卷调查不一定能够很好地代替总体情况，但是仍可以给我们一些重要信息。在关于受访者个人对“大数据正在

变得越来越重要”这一命题的看法调查中，我们得到的数据如图 8 所示：

可以看出，55.17%的受访者个人对于这一命题非常认同，也就是不论大数据目前发展情况如何，他们主观上的判断是，大数据将会变得越来越重要，持有这

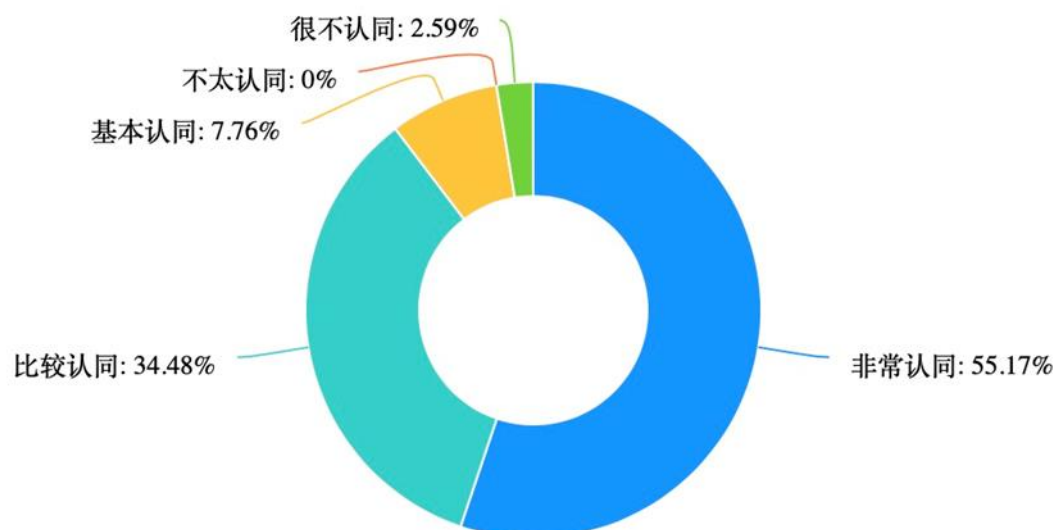


图 8 大数据的重要性主观体会调查结果

一看法的人数超过了半数，但也没有占到受访者的绝大多数。其次，34.48%的受访者比较认同这一命题，也就是他们主观上比较认同大数据将会变得越来越重要。接下来有 7.76%的人基本认同，持有较为中立的态度。有意思的是，0 人选择不认同，而剩下的 2.59%的受访者选择很不认同。

由此可得，对于大数据的发展，在本次问卷调查的受访者中，无人持有中等消极态度，而有非常积极、较为积极、中立和非常消极的态度分化。

当被问及平时是否会关注大数据的应用和发展，结果如图 9 所示。47.41%的受访者表示他们偶尔留心，占到了本题的最多数。有 28.45%的受访者表示积极关注，可以看出，一共 75.86%的受访者平时对于大数据的进展有着较为积极主动的关注，这也体现了大众对于大数据的重视程度。持有中立态度，也就是平时会接收相关信息、但也不会主动了解的人数占有受访者的 15.52%。还有 6.9%的

人表示不太关注，以及 1.72%的人一般会忽略相关新闻。

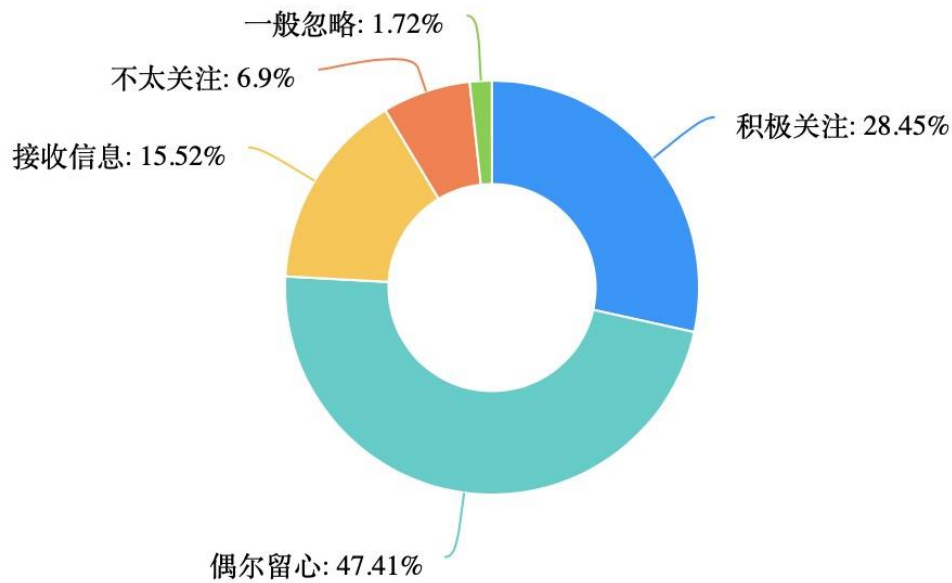


图 9 大数据应用和发展的重视情况

大数据发展的过程中也存在一些问题，如数据的真实性能否得到保障、数据的隐私保护问题、数据存在不当使用的情况、数据的版权问题等等。针对这些主要问题，我们邀请受访者进行了主观重要度排序，得到了受访者眼中数据发展需要关注的问题排序，结果如图 10 所示。

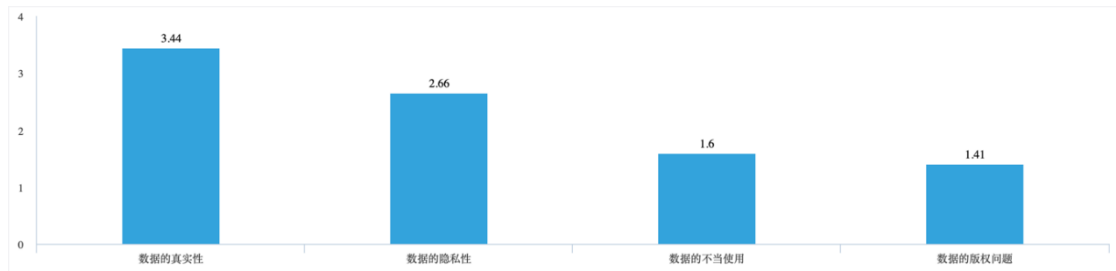


图 10 数据发展关注的问题调查情况

可以看出，大众最关心的是数据的真实性问题，数据的真实性平均综合得分为 3.44。说明，大众希望数据提供准确的信息，以供了解、分析乃至预测。人们不仅希望能够从数据中看出方向、趋势，还希望精确的数字能够代表可以量化的抽象问题。所以，保证数据的真实性是我们在今后的数据分析过程中需要注

意的问题。其次，是数据的隐私问题，平均综合得分为 2.66，仅次于数据的真实性。大众对于数据隐私的要求较高，因为近年来数据泄露、侵犯隐私的事偶有发生，会给大众带来困扰甚至是损失。这也提醒了各个数据部门做好数据的保密工作。再其次，是数据的不当使用问题，平均综合得分为 1.6，得分和排名第二的数据的隐私性问题有着一定的差距。调查中排在四个要素最后的是数据的版权问题，平均综合得分为 1.41，但和数据的不当使用问题得分差异较小，近乎并列关系。

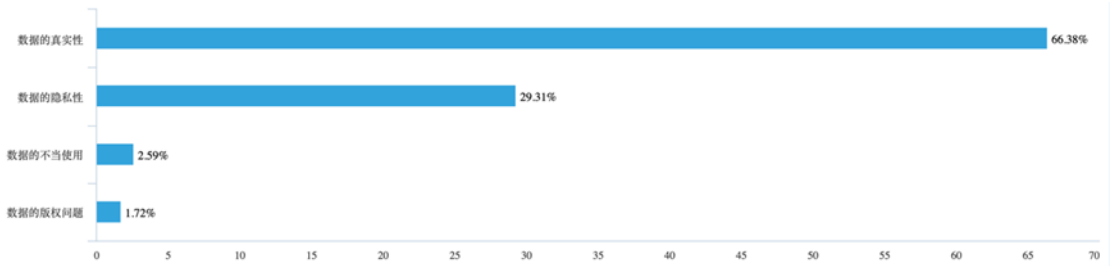


图 11 数据重要性第一位情况

如图 11 所示，66.46%的受访者将数据的真实性放在第一位，充分说明了真实数据的重要性。余下的人中，29.31%将数据的隐私性排在第一位，说明数据的隐私性在大众眼中也十分重要。在这两项之外，提名数据的不当使用、数据的版权问题的人数呈现断崖式的下跌，说明数据的真实性、隐私性这两项要素在大众心中的重要程度显著高于后两项。

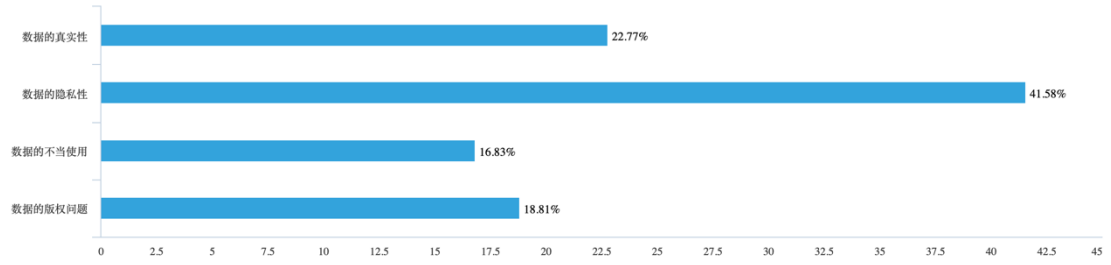


图 12 数据重要性第二位情况

如图 12 所示，在重要性的第二位上，要素的提名出现了散布情况。41.58%的受访者认为数据的隐私性应当排名第二，余下的人中，选择数据的真实性的人

数占比 22.77%，选择数据的不当使用的人数占比 16.83%，选择数据的版权问题的人数占比 18.81%。

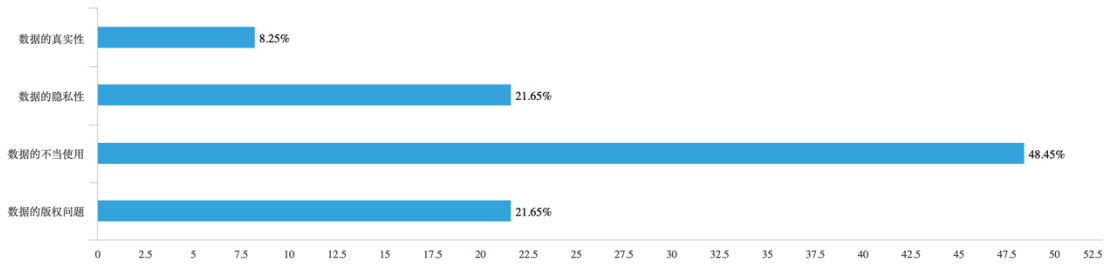


图 13 数据重要性第三位情况

如图 13 所示，在重要性的第三位上，数据的不当使用被提及得最多，占比 48.45%。这主要是因为数据的真实性、隐私性已经被大众广泛关注，而数据的不当使用在此前的排名中并没有被重视，所以在第三名的位置上得到了较多提名。余下的三项中，数据的隐私性、数据的版权问题被提及得一样多，均占比 21.65%，数据的真实性被提及较少，占比 8.25%，主要是因为大众对于数据真实性的重视程度较高。

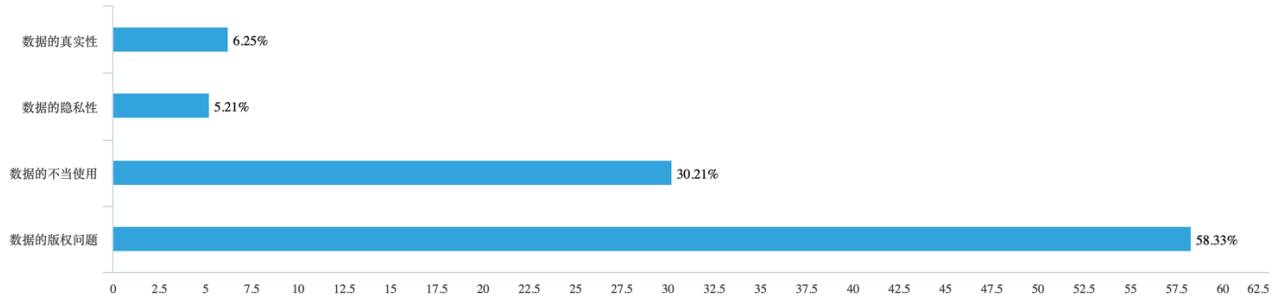


图 14 数据重要性第四位情况

如图 14 所示，在第四位上，数据的真实性占比 6.25%，数据的隐私性占比 5.21%，均较低，说明这两项得到了大众的充分关注。而数据的不当使用占比稍多，30.21%的受访者认为数据的不当使用是这四个要素中相对最不值得关注的，尽管这四项因素均是重要问题，但相比来说其他三项更加重要。在第四位上被提名最多的就是数据的版权问题，占比高达 58.33%，这充分说明了大众在这几项因素中，相对最不关心的就是数据的版权问题。

其实，数据本身不在著作权法的保护之下，但是如果数据经过了人们的处理、汇编，就应当收到著作权法的保护，人们在使用时应当注意版权问题。有些数据属于公共资源的，例如国家统计局公布的宏观经济数据等，这类公共资源可以自由转载。

然而，随着大数据的发展，数据挖掘等技术也在飞速提升之中，海量数据的准确性和全面性比起数据的编排有着更为宝贵的价值，反不正当竞争法就很好地保护了数据的内容价值不被侵犯。这一问题的回答让我们感受到，当今数据的版权问题尚没有引起大众足够的重视，是数据版权的一个隐患，也是我们今后可以努力提高的方向。

（六）结论与建议

1. 调查结论

①根据问卷调查，大众普遍对数据的使用、感受、重视程度较高。使用程度高体现在数据使用频率高、普遍掌握数据工具多。感受程度高体现在大众对于数据发展态势、生活中大数据应用的感知。重视程度高体现在大众主观上对于大数据发展趋势的判断高。

②根据问卷调查，大众关注数据的真实准确性和隐私性方面存在的问题。但是数据版权意识稍显薄弱。

2. 调查建议

①向社会中各类工作人群提供简单易学的数据分析处理软件教程。

②提升大众数据版权意识，做好数据版权保护。

第三部分 数据新动能评价体系

在数据新动能评价体系构建的部分，我们从数据要素的使用情况和数据的发展情况两个角度构造指标，在数据要素的使用方面选择了“普惠金融指数”，用以测度数据要素在经济市场的使用情况，同时，综合全国新基建 7 个领域中的人工智能、物联网、工业互联网、大数据以及公有云行业的发展情况，利用综合评价方法，建立了与工业结合更紧密的评价指标，更完善地测度了数据新动能在我国的发展情况。

一、普惠金融指数

数字金融是数据要素在经济市场中的一个重要的使用形式，然而大部分现有的统计指标不能很好的描述数字金融在金融体系中的情况。为了补充这方面的资料，北京大学数字金融研究中心和蚂蚁金服集团成立了联合研究组，编制了一套“北京大学数字普惠金融指数”。具体体系见表 1。

表 1 数字普惠金融指标体系^[11]

一级维度	二级维度	具体指标
覆盖广度	账户覆盖率	每万人拥有支付宝账号数量
		支付宝绑卡用户比例
		平均每个支付宝账号绑定银行卡数
使用深度	支付业务	人均支付笔数
		人均支付金额
		高频度（年活跃 50 次及以上）活跃用户数占 年活跃 1 次及以上比
	货币基金业务	人均购买余额宝笔数
		人均购买余额宝金额
		每万支付宝用户购买余额宝的人数
	信贷业务	每万支付宝成年用户中有互联网消费贷的用户 数
		人均贷款笔数
		人均贷款金额
		每万支付宝成年用户中有互联网小微经营贷的

		用户数
		小微经营者户均贷款笔数
		小微经营者平均贷款金额
	保险业务	每万支付宝用户中被保险用户数
		人均保险笔数
		人均保险金额
	投资业务	每万人支付宝用户中参与互联网投资理财人数
		人均投资笔数
		人均投资金额
	信用业务	自然人信用人均调用次数
		每万支付宝用户中使用基于信用的服务用户数 (包括金融、住宿、出行、社交等)
数字化程度	移动化	移动支付笔数占比
		移动支付金额占比
	实惠化	小微经营者平均贷款利率
		个人平均贷款利率
	信用化	花呗支付笔数占比
		花呗支付金额占比
		芝麻信用免押笔数占比(较全部需要押金情形)
		芝麻信用免押金额占比(较全部需要押金情形)
	便利化	用户二维码支付的笔数占比
		用户二维码支付的金额占比

数字普惠金融指数从三个方面描述了数字金融，分别是覆盖广度、使用深度和数字支持服务程度，采用了蚂蚁金服的交易账户大数据，从很大程度上代表了普惠金融数字化的程度。图 15 报告了 2011-2020 数字普惠金融指数的变化，全国的数字普惠金融指数由省级指数平均求得，由 2011 年的 40.00 上升到 2020 年的 341.22，表明数字金融在中国经历了快速显著的发展，佐证了数据新动能在金融方面的发展和作用。

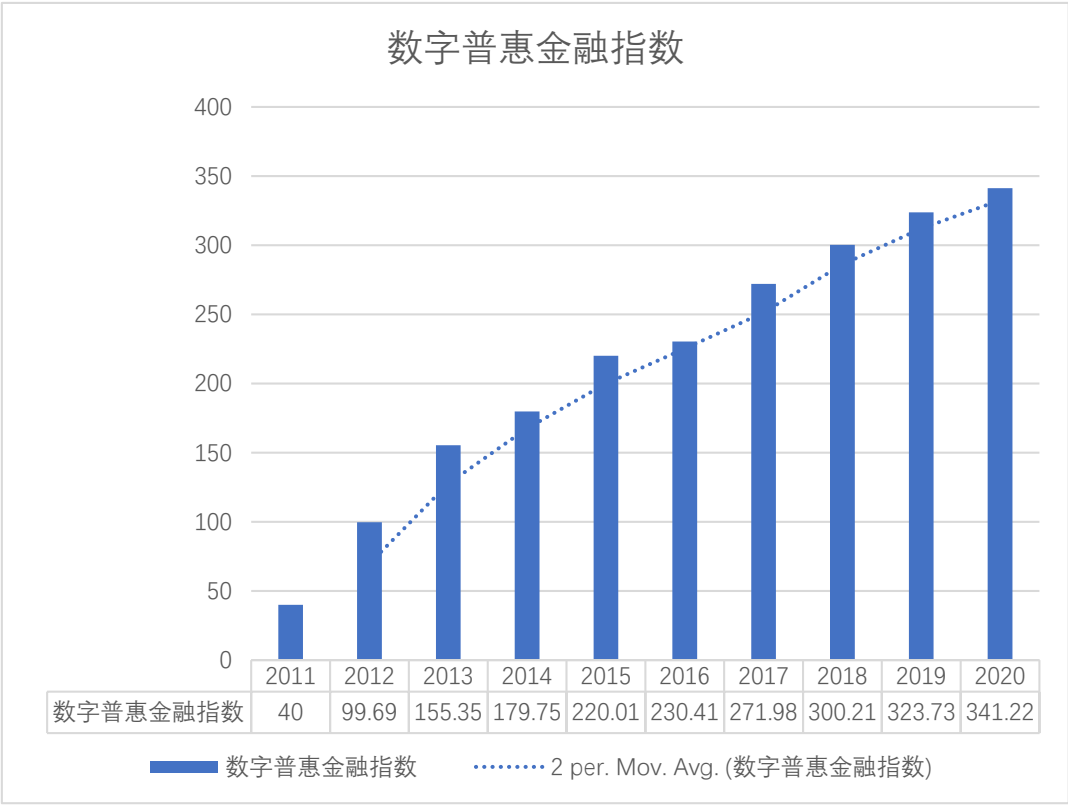


图 15 数字普惠金融指数

二、数据新动能指数体系的构建

（一）数据来源

本文选取 2011-2020 年全国的数据，对全国总体的数据新动能进行了评估。数据来源如表 2 所示。

表 2 指标数据来源

指标	数据来源
物联网行业发展	中国信息通信研究院
工业互联网行业发展	工业互联网产业联盟
公有云行业发展	IDG
人工智能行业发展	前瞻产业研究院
大数据行业发展	《大数据白皮书》

（二）数据新动能评估指标体系的构建

国务院于 2020 年在印发的《关于构建更加完善的要素市场化配置体制机制的意见》中提出将数据要素作为一种新型生产要素，考察它对其他要素效率的作用。本文提出一种指标体系，结合国内实际发展情况，测算全国数据新动能指数并进行分析。

在数字经济时代，数据成为了第一生产要素^[12]，这种“新型”生产要素，推动了我国物联网、工业互联网、公有云、人工智能等新型产业的发展。基于数据新动能的内涵界定，在数据新动能评估指标下设置 5 个指标，具体包括：物联网发展、工业互联网发展、公有云发展、人工智能发展、大数据发展。根据以上分析研究，选取的指标体系如表 3 所示。

表 3 全国数据新动能指数指标

总指数	指标分类	属性
全国数据新动能指数	物联网行业发展	正向
	工业互联网行业发展	正向
	公有云行业发展	正向
	人工智能行业发展	正向
	大数据行业发展	正向

（三）综合评价方法

1. 熵值法概述

熵值法是一种客观评价方法，它不同于层次分析法这种主观的综合评价的方法，它是以各个指标的信息熵为媒介，计算各个指标的相对变化程度，然后衡量它对指标体系在整体的影响，从而来决定对应的权重。

在熵值法中，信息熵的定义是离散随机事件的出现概率。应用在熵值法中，可以解释为：熵值越大说明系统越混乱，表示该指标携带的信息越少；熵值越小说明系统越有序，表示该指标携带的信息越多。

熵值法的主要计算思路如图所示：

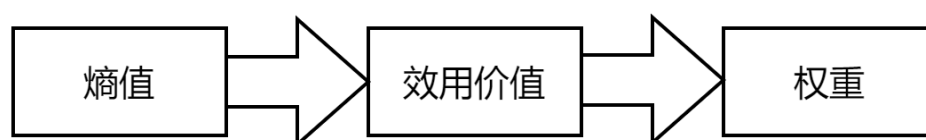


图 16 熵值法计算思路

熵值一般定义为：如果系统处于 n 种状态，每种状态出现的概率为 P_i 。则系统的熵为如下表达式（1）。

$$E = - \sum_{i=1}^n P_i \ln P_i \quad (1)$$

2. 熵值法构建指数的过程

接下来把选取的 5 个数据新动能指标的权重进行计算，本文采用熵值法对各个指标进行权重计算。因为熵值法是客观赋值法，各项指标的权重是通过熵这个概念进行计算的，它表示的是指标间相对的权重，可以克服主观因素造成的测量误差。因此，本文采用该方法测度全国数据新动能指标权重情况。

在正式计算熵值前，首先要对这些指标进行无量纲化处理，即运用归一法对各个数值进行标准化的处理。这样的做法目的是解决指标见量纲不同或者类型不同带来的问题。

x_{ij} 为第 i 年的第 j 个指标的数值。（ $i = 2011, \dots, 2020$ ； $j = 1, \dots, 5$ ）

归一法处理具体方法如式(2)、(3)所示:

对于正向指标

$$X_{ij} = \frac{x_{ij} - \text{Min}\{x_{ij}\}}{\text{Max}\{x_{ij}\} - \text{Min}\{x_{ij}\}} \quad (2)$$

对于负向指标

$$X_{ij} = \frac{\text{Max}\{x_{ij}\} - x_{ij}}{\text{Max}\{x_{ij}\} - \text{Min}\{x_{ij}\}} \quad (3)$$

综合评价的具体步骤如下:

第一步, 计算第*i*年占第*j*项指标的比重, 计算式如式(4)所示。

$$p_{ij} = \frac{X_{ij}}{\sum_{i=1}^n X_{ij}}, \quad j = 1, \dots, 5 \quad (4)$$

第二步, 计算第*j*项指标的熵值, 计算式如式(5)所示。

$$e_j = -K^* \sum_{i=1}^n p_{ij} \ln(p_{ij}) \quad (5)$$

其中 $K^* = \frac{1}{\ln n}$, *n*为方案的的数量, 在本文中为 10。

第三步, 计算第*j*项的差异系数。对第 *j* 项数据新动能指标, 它的差异系数 g_j 越大, 代表着该指标值的差异越大, 对方案评价的左右就越大, 则指标越重要; 反之, 指标越不重要。

定义差异系数为式(6)。

$$g_j = 1 - e_j \quad (6)$$

第四步, 求权值, 定义权重如式(7)所示。

$$W_j = \frac{g_j}{\sum_{j=1}^n g_j}, \quad j = 1, \dots, 5 \quad (7)$$

经计算, 各指标所得权重为(0.194, 0.189, 0.213, 0.202, 0.202)。

第五步, 计算综合得分, 定义综合得分如式(8)所示。

$$S_i = \sum_{j=1}^n W_j P_{ij}, \quad i = 2011, \dots, 2020 \quad (8)$$

经计算，全国 2011-2020 年数据新动能评估指标的综合得分如表 4 所示。

表 4 2011-2020 年中国数据新动能指数

年份	<i>SNE</i>
2011 年	3772.2237
2012 年	5166.7834
2013 年	6676.8355
2014 年	8276.9664
2015 年	11793.7459
2016 年	16743.3743
2017 年	22526.6665
2018 年	30019.9148
2019 年	35504.0954
2020 年	40181.2757

（四）数据新动能评估指标综合评估结果

运用上述方法计算出 2011-2020 年全国数据新动能评估指标的综合得分，如图 17 所示。

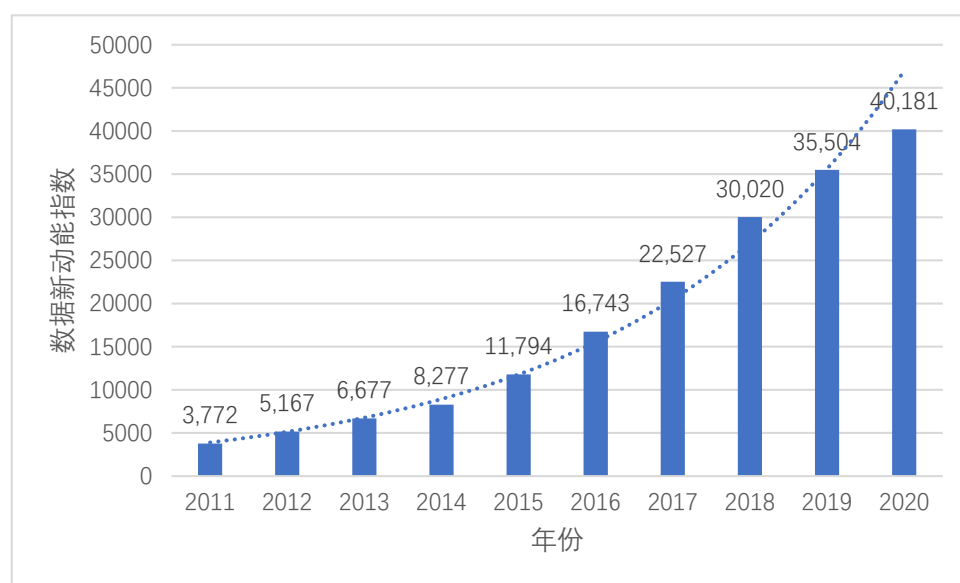


图 17 全国数据新动能指数

由图 17 可知，全国数据新动能指数呈逐步上升的趋势，显示出数据新动能的发展持续攀升。

从 2011 年到 2020 年，该指数从 3772 上升至 40181，涨幅非常显著。从图中可以看出，数据新动能的发展可以分为两个阶段：在 2011-2014 年，该指数增

长较慢，可能的原因是，数据行业处于发展的阶段，很多技术处于探索阶段；在2015年之后，该指数出现了大幅的上升，表明在这段时间，我国物联网行业、人工智能行业、大数据等行业迅速发展。

三、总结

我们将结合普惠金融指数和数据新动能指数，形成计量数据新动能的统计测度，接下来，我们将对这个体系进行实证分析，考察数据新动能对经济增长的影响。

第四部分 基于数据新动能评价体系的实证分析

一、模型设定

(一) 数据新动能与全国生产总值

首先建立数据新动能指数、数字普惠金融指数与人均国内生产总值之间的模型。 SNE_t 表示第 t 年的数据新动能指数， IF_t 表示第 t 年的数字普惠金融指数，第 t 年人均国内生产总值取对数，记为 $\ln(pgdp)_t$ 。实证模型如下：

$$\ln(pgdp)_t = b_0 + b_1 SNE_t + b_2 IF_{t-1} + b'_3 X_t + \varepsilon_t \quad (1)$$

在①式中， X_t 表示为了缓解遗漏变量偏误而设置的其他变量，包括宏观经济层面的居民消费水平和其他方面的总人口数， ε_t 为随机扰动项。为了减弱反向因果，将数字普惠金融指数滞后一期。根据模型①， b_1 衡量数据新动能的发展对宏观经济的总体影响。

但以上模型并不能体现出数据新动能在不同地区的影响的差异。本文接下来建立模型二，考察数据要素在各地的情况。

模型①将在 Stata 软件中进行估计。

(二) 数据新动能与地区发展

接下来建立数据新动能、数字普惠金融指数与不同地区生产总值之间的面板数据模型，按照中国的 4 个经济区域，把 31 省份划分到中部地区、东部地区、西部地区和东北地区，分组研究。本模型旨在探究数据要素在不同省份的发展和影响的差异。 $SNE_{i,t}$ 表示第 i 个省份第 t 年的数据新动能指数， $IF_{i,t}$ 表示第 i 个省份第 t 年的数字普惠金融指数， μ_i 表示地区固定效应，第 i 个省份第 t 年地区生产总值

取对数，记为 $\ln(pgdp)_{i,t}$ 。实证模型如下：

$$\ln(pgdp)_{i,t} = b_1 SNE_{i,t} + b_2 IF_{i,t-1} + \mu_i + \varepsilon_{it} \quad (2)$$

模型②将在 Stata 软件中用 LSDV 法进行估计，并使用聚类稳健标准误。

二、数据来源与预处理

为了估计模型①和②，本文使用以下三方面的数据：

第一部分的数据为本文使用综合评价法构造的数据新动能指数。

第二部分的数据来自国家统计局（stats.gov.cn），选取的样本年份区间均为 2011–2020 年，表 5 和图 18 是相关变量的描述性统计。

表 5 相关变量的描述性统计

变量	样本	均值	标准差
人均 GDP (以 2011 年为基期)	10	4.60	0.024
居民消费水平 (以 2011 年为基期)	10	12172.76	1051.56
人口总数 (万人)	10	137845.5	2005.63

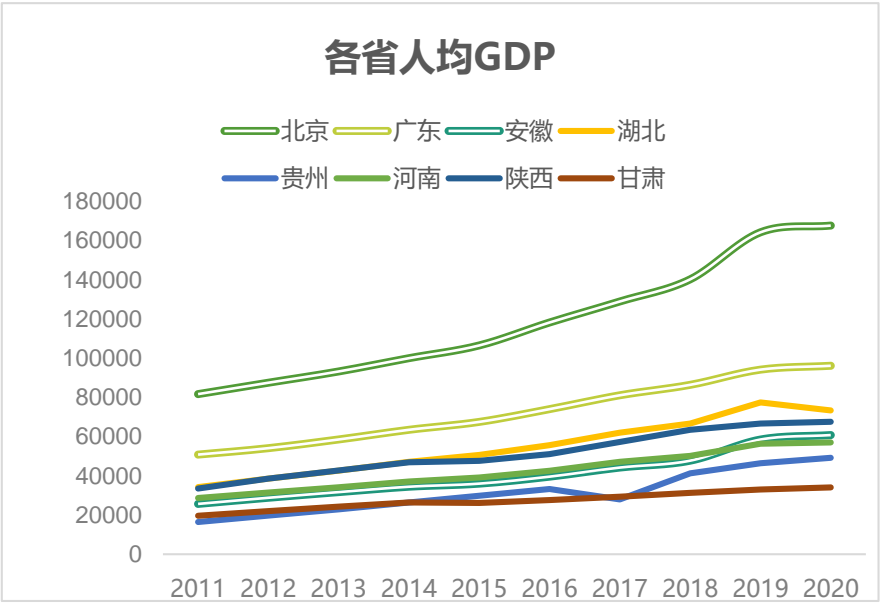


图 18 2011–2020 年各省人均 GDP

第三部分数据来自中国数字普惠金融指数，图 19 报告了八省份各自数字普惠金融指数的趋势。

三、数据新动能与人均国内生产总值的关系

我们对模型①在 Eviews 中使用加权最小二乘法进行估计，消除了异方差，表 3 报告了回归结果。

表 6 数据新动能的发展与人均 GDP

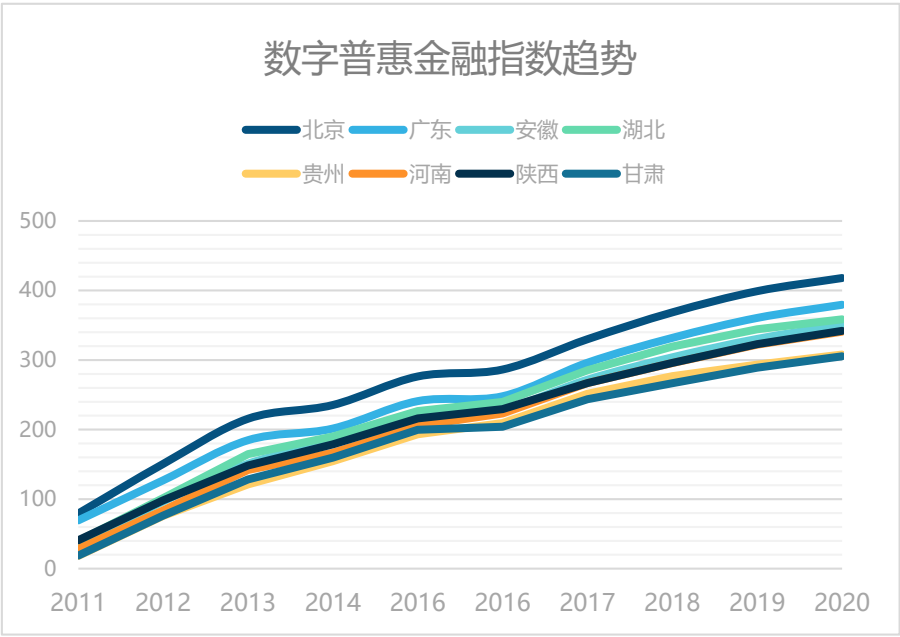


图 19 各省数字普惠金融指数发展趋势

因变量：人均 GDP	系数
数据新动能指数	0.177***
数字普惠金融指数	21.682***
居民消费水平	0.050
总人口数	-1.026**

注：***、**和*分别表示在 1%、5%和 10%的水平显著，下同。

回归方程的结果如下所示：

$$\ln(pgdp)_t = 172431.08 + 0.177SNE_t + 21.682IF_{t-1} + (0.050, -1.026)X_t$$

调整后可决系数为 98.1%，表明模型拟合效果优。数据新动能指数与数字普惠金融指数均显著为正，说明数据要素在技术层面的发展和金融层面的发展都对宏观经济有显著的正向影响。系数显示为，当数据新动能指数提升一个单位，人均 GDP 大约提升 0.177 个百分点，数字普惠金融指数提升一单位，人均 GDP 大约提升 21.682 个百分点。

四、数据新动能与地区发展的关系

对模型②使用 LSDV 法进行估计：

$$\ln(pgdp)_{i,t} = b_1SNE_{i,t} + b_2IF_{i,t-1} + \mu_i + \varepsilon_{it} (i = 1, \dots, 31; t = 1, \dots, 10)$$

用截距项 α_i 替换个体异质性 u_i ，将模型改写成：

$$\ln(pgdp)_{i,t} = b_1SNE_{i,t} + b_2IF_{i,t-1} + \alpha_i + \varepsilon_{it} (i = 1, \dots, 31; t = 1, \dots, 10)$$

接下来合并解释变量，用向量表示个体异质性，

$$\ln(pgdp)_i = X_i^* \beta^* + i\alpha_i + \varepsilon_t$$

其中：

$$i\alpha_i = \begin{bmatrix} i & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & i \end{bmatrix}_{310 \times 310} \begin{bmatrix} \alpha_1 \\ \dots \\ \alpha_{31} \end{bmatrix}_{310 \times 1}$$

用矩阵简化表述为

$$\ln(pgdp) = [X^* \text{ province}_1 \text{ province}_2 \dots \text{ province}_{31}] \begin{bmatrix} \beta^* \\ \alpha \end{bmatrix} + \varepsilon$$

转换后的公式满足外生性条件，可以采用 OLS 混合回归估计。表 7 报告了回归结果。

表 7 数据新动能与地区发展回归结果

解释变量		中部地区	东部地区	西部地区	东北地区
SNE	0.19 [*] (0.09)	0.55 ^{***} (0.08)	0.32 (0.30)	0.25 ^{***} (0.10)	-0.39 ^{***} (0.13)
IF	87.85 ^{***} (10.13)	32.33 ^{***} (7.56)	119.30 ^{***} (18.45)	58.86 ^{***} (7.88)	77.56 ^{***} (18.40)

从回归结果中可以看出，普惠金融指数和数据新动能指数在 4 个地区大小不同，体现了地区发展差异。其中，普惠金融指数主要体现了金融行业数字化的程度，数据新动能指数主要体现工业和高新科技行业数字化的程度，二者在不同地区的发展有差异数据，新动能指数的系数从大到小依次是中部、东部、西部、东北部。基本在发达程度越高的地区，对经济的促进作用越强。而金融行业数字化时间更早，相对工业和高新科技行业也更容易转型，所以在西部和东北部也能很快地发展。

为了进一步探究数据新动能在不同地区的作用，接下来我们选出 8 个代表省份北京市、广东省、湖北省、陕西省、安徽省、河南省、贵州省、甘肃省，对数据新动能指标进行单独回归，结果如下表 8 所示。

表 8 数据新动能指标单独回归结果

省份	数据新动能指数
北京市	2.298 ^{***}
广东省	1.211 ^{***}
安徽省	0.889 ^{***}
湖北省	1.079 ^{***}
贵州省	0.801 ^{***}
河南省	0.744 ^{***}
陕西省	0.868 ^{***}
甘肃省	0.335 ^{***}

将回归结果按照回归系数的大小进行排序，结果如图 15 所示。

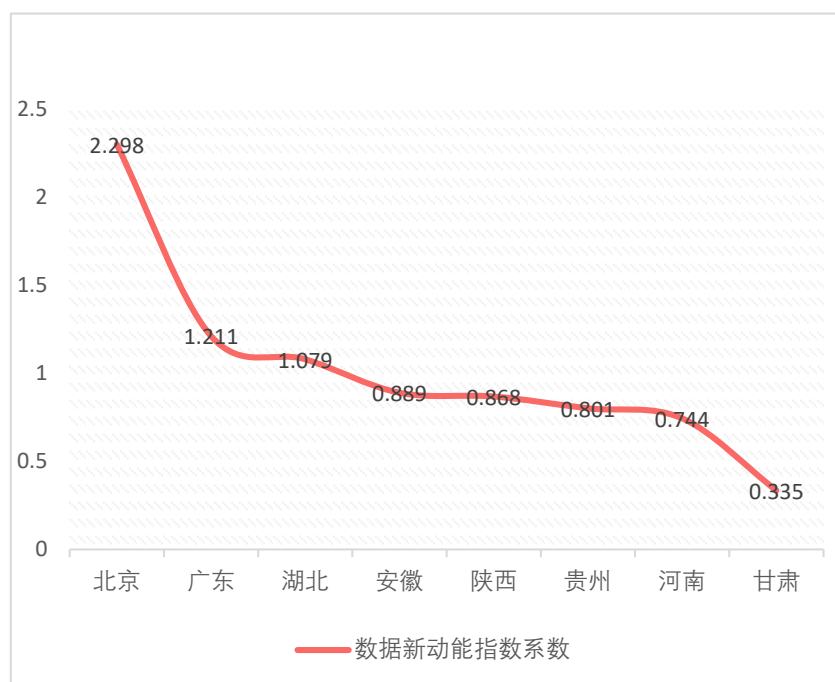


图 20 数据新动能指标系数

首先，数据新动能指数的系数在八个省份全部显著为正，说明其对不同省份的经济都有显著的正向影响。

其次，不同省份，数据新动能指数的系数大小不同，在发达程度越高的省份，数据新动能对宏观经济的促进作用越明显，其大小趋势基本符合地区生产总值的排名。推测原因有两个：第一，发达省份的部分或大部分经济活动都完成了数字化转型，信息基础设施也比较发达，数据要素与人民生活息息相关，所以数据新动能影响作用随之增大。第二，相对于成熟且广泛的传统业务，数据要素落地往往对硬件设备改造升级、软件集成等有更高的要求。从数据新动能指数构建的五个方面来讲，物联网的关键技术就是传感器、射频识别、全球定位系统等，其实现依托于这些设备的互相作用；人工智能的完整发展需要处理数据、设计网络、算法模型训练多机并行计算、应用端性能优化等。而部分省份数字化程度还不高，不够支撑数据新动能发挥全部作用。

最后，从地区总体来看的数据新动能指数系数排名和省域排名略有不同，对

于中部地区总体的数据新动能指数系数大于东部地区的情况,可能是由于中部地区发展空间比东部地区大,而东部处于边际效应递减的过程。

总体看来,发展数字化基础设施,才能从根本上将数据变成生产要素,提升数据新动能的水平。

五、结论与建议

(一) 结论

1. 数据要素在宏观经济中起到了正向作用。数据新动能指数、数字普惠金融指数的系数在模型①和模型②中均显著为正。

2. 代表金融行业数字化程度的普惠金融指数和代表工业和高新科技行业数字化程度的数据新动能指数均体现了省级差异,在不同省份的影响不同。越发达的城市正向影响越大,其原因可能是数据新动能发挥作用依托于完备的信息基础设施。其次,两个指数之间的影响也有差异,其原因可能是金融行业数字化时间更早,相对工业和高新科技行业也更容易转型,所以在西部和东部也能很快地发展。

3. 数据新动能指数系数在中国东部和中部发生反转,即促进效果与地区人均GDP出现反转,其原因可能是由于中部地区发展空间比东部地区大,而东部处于边际效应递减的过程。

(二) 建议

1. 坚持推进数据与各产业的融合的大方向。
2. 重视信息基础设施的建设,在配套设施完善的情况下善用数据。

3. 关注数据新动能在各地区之间的发展差异,警惕其是否促进了地区间发展不平衡。

参考文献

- [1]何玉长,王伟.数据要素市场化的理论阐释[J].当代经济研究,2021(04):33-44.
- [2]王建冬,童楠楠.数字经济背景下数据与其他生产要素的协同联动机制研究[J].电子政务,2020(03):22-31.
- [3]许宪春,任雪,常子豪.大数据与绿色发展[J].中国工业经济,2019(04):5-22.
- [4]邵明振,马舒瑞,屈小芳,张瑜.河南省经济新动能统计测度、经济效应及发展路径研究[J].统计理论与实践,2021(03):15-22.
- [5]郑世林,熊丽.中国培育经济发展新动能的成效研究[J].技术经济,2021,40(01):1-11.
- [6]郑江淮,宋建,张玉昌,郑玉,姜青克.中国经济增长新旧动能转换的进展评估[J].中国工业经济,2018(06):24-42.
- [7]陈昌兵.新时代我国经济高质量发展动力转换研究[J].上海经济研究,2018(05):16-24+41.
- [8]荆文君,孙宝文.数字经济促进经济高质量发展:一个理论分析框架[J].经济学家,2019(02):66-73.
- [9]张勋,万广华,张佳佳,何宗樾.数字经济、普惠金融与包容性增长[J].经济研究,2019,54(08):71-86.
- [10]孙玉环,张汀昱,王雪妮,李丹阳.中国数字普惠金融发展的现状、问题及前景[J].数量经济技术经济研究,2021,38(02):43-59.
- [11]郭峰,王靖一,王芳,孔涛,张勋,程志云.《测度中国数字普惠金融发展:指数编制

与空间特征》,2020,19(04):1401-1418.

[12]林志杰,孟政炫.数据生产要素的结合机制——互补性资产视角[J].北京交通大学学报(社会科学版),2021,20(02):28-38.

附录 1 调查问卷

您好:

我们是某高校的大学生,正在做一个关于大数据的调查。为了了解对数据的使用、感受、重视程度,我们需要您的看法。此问卷采取的是匿名调查,我们将不会记录您的个人姓名,并对您的个人答案信息进行严格保密。填写以下问卷大致会花费您一分钟的时间,在填完问卷之后您将会收到红包以表对您参与的感谢。我们衷心感谢您的支持!

1. 在您的工作中,是否需要使用各种形式的数据?

A 是 B 否

2. 您使用的数据主要是以下哪种形式?

A 数据库 B 自己收集 C 他人报告中的

3. 您工作中使用数据的频率?

A 每周使用 B 每个月使用 C 几个月才会使用一次 D 几乎很少使用

4. 您掌握的数据处理分析工具(例如 Excel)的数量?

A 5 个以上 B 三到四个 C 一到两个 D 没有掌握

5. 您感觉大数据正在变得越来越重要?

A 非常认同 B 比较认同 C 基本认同 D 不太认同 E 很不认同

6. 您生活中接触的数据多吗?

A 非常多 B 比较多 C 中等 D 不多 E 很少

7. 您个人认为大数据正在变得越来越重要?

A 非常认同 B 比较认同 C 基本认同 D 不太认同 E 很不认同

8. 您会关注大数据的应用和发展吗?

A 积极关注 B 留心关注 C 接收信息 D 不太关注 E 一般忽略

9. 您给大数据的发展过程中需要注意的问题, 重要程度排序:

数据的真实性 数据的隐私性 数据的不当使用 数据的版权问题

附录 2 Stata 代码

```
xtset province year
```

```
xtreg gdp sne puhui i.province,r
```

致 谢

此次统计建模大赛对于我们三位成员来说，意义重大，十分难忘。我们得到了来自学院、老师、同学、好友等各方面的支持和帮助，在此，我们想正式向他们表达我们衷心的感谢。

感谢学院对我们的栽培和指导，孙玉泉老师、吕绍川老师、金蛟老师、张欣老师，还有许许多多给过我们不论是知识还是方法上指导的亲爱的老师们，感谢你们对我们的教育和指导！你们带着我们学习了专业知识，修炼了专业技能，也教会了我们做研究的方法。尤其是感谢我们的指导老师李高荣、张欣老师，他们陪伴着我们比赛的始终，一直耐心细心地给予我们帮助。

感谢队友，是你们每一个人的参与使得我们可以将研究做出预期中的效果。这次比赛从选题到实施，少不了每一位队友的努力，三位队友每天进行会议讨论、头脑风暴，查阅无数文献，对研究指标进行数次修改，得出最终的方案。

感谢学校的图书馆，良好的自习氛围和设施无形之中也给了我们一些动力。

最后我们还想感谢这次比赛，这次统计建模大赛也促使我们去用自己的视角观察数据新动能，也拓宽了我们的视野，加深了我们对数据的理解。在这次大赛中，我们尽力以所学去做所想，用课堂上、课堂外学到的知识分析实际的问题，收获颇丰！