

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：上海对外经贸大学

论文题目：基于数据资产规模指数的经济增长新动能统计测度

参赛队员：欧晓轩、曹奕琳、毛泓元

指导老师：李佩瑾

目录

一、引言与研究现状.....	6
(一) 研究背景	6
(二) 研究内容	7
二、数据资产驱动增长理论基础	7
(一) 经济增长传统动能测度	7
(二) 经济增长新动能测度	8
(三) 数据资产规模测度	9
(四) 网络搜索数据应用研究	10
三、数据资产规模测度	10
(一) 指数因子的确定	11
(二) 样本选取及关键词的确定	12
1. 样本选取	12
2. 关键词的确定	12
(三) 数据的搜集	13
(四) 综合指数的确定	14
四、数据资产规模测度	16
(一) 模型的建立	16
(二) 控制变量	17
(三) 估计回归参数	19
(四) 模型的检验	19
1. 模型假设	19
2. 变量标准化	20
3. 变量的筛选	21
(五) 回归诊断	21
1. 残差分析	21
2. 共线性诊断	22
3. 误差项独立性检验	23
(六) 模型的分析	23
六、结论与展望	24

（一）结论	24
（二）创新点	25
（三）研究展望	26
参考文献	27
致谢	32

表格清单

表 1	指数因子对应含义	14
表 2	各主成分贡献率	15
表 3	主成分载荷	16
表 4	控制变量对应指标	17
表 5	T 检验与 P 值的关系.....	20
表 6	方差比例	22
表 7	汇总表	23

插图清单

图 1	关键词词云图	13
图 2	x_1 相关系数矩阵 A	15
图 3	标准化 A 矩阵的特征根 λ	15
图 4	残差正态概率图	22

摘要:

本文基于文本分析方法从热度、商业、学术三个方面,构建一个反映数据资产规模的综合指数。首先通过网络爬取方式对 2011-2020 十年间光明日报网、人民日报网和学术期刊等相关平台选取数据主题有关文章与新闻,采用 TF-IDF 算法对相关文章中词频进行统计研究,综合选取与数据关系较为密切的关键词。在热度方面选取十年间“数据”相关关键词每季度的平均百度搜索指数,在商业方面选取三家龙头数据企业 BAT 的季度资产总额,学术方面选取知网每季度“数据”相关关键字的文章相对篇数,基于主成分分析法建立数据相关综合指数,并测度十年间数据综合指数的增长率,用该指数增长率来反映数据资产规模的增长率。

传统以来国内 GDP 的增长主要受劳动、资本、土地和技术等基本要素的影响,随着数据要素的逐步发展,本文尝试用数据资产规模的增长率来反映数据要素增长对 GDP 的推动作用,客观查找出其他对 GDP 增长率有影响的变量共同建立多元线性回归方程,基于此来验证数据要素对国内 GDP 增长的推动作用,并对我国数据新动能的发展提出可行性建议。

关键词: 新动能; 数据资产规模; 多元线性回归; 主成分分析; TF-IDF 算法

Abstract:

Based on the content analysis method, this paper constructed a comprehensive index to reflect the scale of data assets from three aspects, which are popularity, business and academic. Firstly, through the network crawling method, this paper selected the literature and news which were related to data from Guangming Daily and People's Daily from 2011 to 2020. And then, TF-IDF algorithm was used to conduct statistical research on the word frequency of related articles, and to comprehensively select keywords relevant to the topic of data. In terms of popularity, the average Baidu Index of data-related keywords in each quarter in the past ten years was considered. In terms of business, the total quarterly assets of three leading data enterprises BAT were collected. In terms of academic research, the number of articles about data-related keywords in each quarter of CNKI was gathered. In the construction of the data-related index, the weight was determined based on principal component analysis. The growth of domestic GDP in the past was mainly affected by labor, capital, technology and other factors. With the development of the element of data, this paper attempted to use the growth rate of data asset scale to show the driving effect of the data element growth on GDP. And after objectively finding other variables that had an impact on GDP growth, the paper established a multivariable linear regression equation, thus verifying the driving effect of the data element on domestic GDP growth, that is, the new energy of data.

Key Words: the new energy of economic growth; data asset scale; multivariable linear regression; principal component analysis; TF-IDF algorithm

一、引言与研究现状

（一）研究背景

在 2020 年公布的《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》中，首次将数据与传统生产要素劳动力、资本、技术以及土地相提并论，将其纳入生产要素范围，成为第五大要素，并提出应加快培育数据要素市场，充分发挥数据要素对其他生产要素的倍增作用。近年来数据要素已逐渐融入我国经济价值创造体系中，成为数字经济时代不可或缺的战略资源性和重要生产力，数据为推动经济高质量发展提供了新的动力与活力。

2020 年，新冠肺炎疫情来势汹汹，波及全球，世界经济形势复杂严峻，受到极大挑战。在此背景下，“十四五”规划以及 2035 远景目标纲要指出，要推动数字经济与实体经济的深度融合发展，加快构建国内国际双循环相互促进的新发展格局。在 2021 年的《中国数字经济发展白皮书》中显示，在过去的 2020 年我国数字经济规模已达到 39.2 万亿元，占 GDP 比重为 38.6%，同比名义增长 9.7%，数字经济在逆势中不断迅猛发展，有效支撑着疫情防控和经济社会恢复发展。在 2020 年电力大数据协同创新高峰论坛上，国家信息中心相关发言人则表示数字经济将会继续保持持续高速增长，数字化转型将成为企业发展的主流。数据是驱动经济发展的新引擎，随着我国数字化发展的持续增强，提倡更高标准、更高质量地推进数字中国的建设发展，为奋进新征程提供新动能。

数字经济日渐发展，需对新型、复杂的数据资产规模需进行相关测度与计量，但目前其测度方法尚不完善，难以真实准确地测度数据资产规模。2021 年 3 月，中国软件网、海比研究院以及中国软件行业协会推出《2021 中国数据资产化工具市场研究报告》，根据报告显示，2020 年我国数据资产化市场规模为 280 亿元，未来五年的复合增长率将达到 40% 以上，但相关研究人员也表示数据资产规模的研究仍是当今市场亟需解决的问题。

本文通过建立数据相关综合指数在一定程度上代替数据资产规模，并用数据资产规模的增长率反应数据要素对 GDP 增长的驱动作用，旨在为数据资产评估、完善数据市场以及数据要素市场对国内新经济的推动，即数据新动能等方面提供思考方向。

（二）研究内容

本文首先基于文本分析的方法，通过网络爬虫技术获取 2011-2020 十年间光明日报、人民日报新闻以及学术期刊，并对此文本数据进行数据清洗、数据处理，随后采用 TF-IDF 算法对文本进行词频统计，选取“数据”相关的关键词。接着主要从热度、商业和学术三个方面构建数据相关综合指数。在热度方面，基于百度指数得到相关关键词的近十年季度平均搜索指数；在商业方面，选取三家龙头数据企业 BAT 的季度资产总额；在学术方面，选取知网“数据”相关关键字的近十年季度论文相对发表篇数，对其进行主成分分析建立综合指数，并用此综合指数反映数据资产规模。

综合选取传统上对 GDP 增长有影响的其他变量，在此基础上加入利用本文分析方法构建的反映数据资产规模的综合指数作为另一自变量，共同建立多元线性回归分析，并对该模型进行检验与分析，验证数据资产规模的增长对 GDP 增长的驱动作用，即数据要素的新动能。

二、数据资产驱动增长理论基础

（一）经济增长传统动能测度

以往的关于新动能的统计测度的研究一般分为构建指标体系进行综合评价和将全要素生产率视为新动能进行计量测度两类。

柴士改和李金昌（2021）基于对新动能本质特征的认识，对投入、产出两个方面的要素构成和表现指标进行梳理，设置了 2 个一级指标和 10 个二级指标，并进一步细分成 23 个三级指标和 82 个四级指标，分别赋以权重，构成经济增长新动能指标体系，并通过灰色关联度方法、计算相关系数、测算指标体系的熵等方法进行完整性检验。何强（2019）也类似地设置了 4 个一级分类指数和 9 个二级指标，用德尔菲法分别赋以权重，并使用因子分析法、熵值法和 TOPSIS 法进行稳健性分析，建立经济发展新动能指数；并用似不相关回归模型和单变量方程

计量模型来甄别经济发展新动能的影响因素。类似的，刘姝雯等（2021）选取了 1 个一级指标、3 个二级指标、8 个三级指标和 30 个四级指标，分别赋以权重，进行经济发展新旧动能转换的指数测算。此外，为了描述地区经济特征，王珍（2019）建立区域一体化和新旧动能转换关系的理论框架，并基于该理论模型选取了我国部分地区作为样本进行变量测度，做实证检验。

张豪等（2017）进行经济核算并用随机前沿超越对数生产函数对 TFP 增长率进行分解，以进一步挖掘 TFP 的内在增长要素。郑江淮等（2018）在柯布-道格拉斯生产函数上进行变换，刻画出 TFP 及其增长的构成项，用份额偏离分解等方法，最终得到了一个经济增长动能指数，并使用这个指数对中国经济增长进行解读。

此外，周荣荣和李佳（2019）在柯布-道格拉斯模型的基础上引入人力资本的扩展增长模型，并继续引入其他变量的扩展模型，采用一段时间内全国的时间序列数据进行测算，对中国经济增长动力因素的增长贡献进行实证检验。行帆等（2019）只通过对一些指标和数据进行了定性的描述，而对指标的选取建立在党的十九大报告的基础上。其中，方志耕（2006）将经济增长动能与物理学概念中的动能联系起来，可谓一种创新，但此外没有更多研究表明该种方法的可行性。

（二）经济增长新动能测度

“十四五”规划纲要指出，要打造数字经济新优势，提升企业技术创新能力，形成以企业为主体、市场为导向、产学研用深度融合的技术创新体系。而人工智能、区块链、大数据等数字技术打破了组织原有边界，使得企业内部的组织形态、运营模式、产品设计和研发流程等变得更加开放和动态，为数字经济下企业技术创新能力的提升提供了战略机遇。

王一鸣（2020）、穆荣平（2019）、余江等（2017）和柳卸林等（2020）均提出数据要素作为数字时代经济社会发展的基础性生产要素，已成为加快企业转型、加速技术经济与发展范式跃迁、全面提升国民经济循环效率的新动力。

自 2015 年起，多次在贵阳召开的数博会让贵州的大数据中心吸引了外界诸多关注。发展数字经济，让数字经济成为高质量发展的强劲引擎已经成为不可阻挡的时代潮流（贵州日报评论员，2021）；数字经济也将引领新型工业化高质量发展（刘小红，张人龙，2021）。

焦豪（2021）提出，要加快数字经济下企业技术创新能力提升，推动数据和业务协同化共融发展。尹西明等（2021）则对数据要素价值化动态过程机制进行研究，提出要培育发展数据要素市场，加速释放数据要素市场红利。

因此，我们认为数据资产的扩大能够一定程度上反映经济的增长，希望基于网络数据来对数据资产经济增长新动能进行统计测度。

（三）数据资产规模测度

数据很久以前就一直存在着，而近些年，云计算、区块链等技术产业应运而生，这使得数据的总数量有了飞速的上涨，其中的价值也逐渐被大众所认可。1974年，Richard Peters 率先提出“数据资产”一词，个人所持的各类债券等资产均可算作数据资产。众多现有文献表示，数据资产是无形资产。在目前的国民经济核算体系 SNA2008 中，生产产出与非生产资产中均不包含数据，因此其现有的核算方法无法记录实际的数据价值，需构建系统性数据核算方法以便更全面地体现数据价值。由于数据资产的复制成本几乎为零、异质性和规模性等特殊性质，生产数据的成本远远小于其内在的价值，且数据的交易市场有限，未来收益不易评估，因此无法使用无形资产的价值评估方法来进行数据资产价值的核算（李静萍，2020）。Ahmad 和 Ven（2018）针对数据资产评估提出了三种方法：结合数据市场的价格估计价值、根据数据的广告收入估计价值以及根据消费者对免费服务的支付意愿估计价值，但这三种方法距离广泛地运用于宏观的数据资产估值仍有很大的差距。在财务会的方面，李泽红（2018）等对大数据资产的确认、评估价值、计量和报告深入研究，为企业将数据资产引进财务报表提供了参考。

正值大数据时代，我们认为数据资产可以算作是推动经济发展不可缺少的动能之一，而对于数据资产规模的研究相对较少，以及对数据资产核算的体系尚不完善。本文将通过 TF-IDF 算法与主成分分析，构建数据资产规模相关的综合指数，在一定程度上反映数据资产的规模，并在前人研究的基础上，选用影响 GDP 增长的其他传统要素，从中引入数据要素，进而建立 GDP 增长模型，测度数据资产规模的的增长对经济推动的贡献程度。

（四）网络搜索数据应用研究

关于网络搜索数据，不同学者均提出了自己的观点。例如崔东佳（2014）使用百度搜索引擎得到相应的网络搜索指数，并依次建立回归预测模型，进行协整分析和 Granger 因果检验，最终对不同品牌汽车销量进行了预测。其他学者如唐一丁（2016）、张涛和刘宽斌（2019）、王勇和董恒新（2017）以及张瑞等（2020）也运用网络搜索数据在其他方面进行了应用。

因此，本文使用网络搜索数据来作为解释经济增长的一部分影响因素。

三、数据资产规模测度

数据资产是无形的，且数据多样化的形式与计量单位使其难以被完整准确地统计出来。当下是数据爆炸的时代，数据增长的速度远超人们想象，但目前针对数据资产规模方面的研究较少。本文将通过构建 D 指数，基于 TF-IDF 算法以及主成分分析，从热度、商业、学术三个方面测度数据资产的规模。

D 指数的构建主要可分为三个步骤：

1. 将数据资产分为热度、商业、学术三个主要方面进行测度，并各选取 3 个指标，作为指数因子 $x_i (i = 1, 2, \dots, 9)$ ；
2. 基于 TF-IDF 算法，通过 Python 中 jieba 模块，在与数据高度相关的样本文章中实现选取 3 个代表性词汇，作为热度与学术方面指标的关键词；
3. 从百度指数、公司季度报表、Wind、中国知网，收集 9 个指标因子在 2011 年至 2020 年间各季度的数据；
4. 根据 2011 年至 2020 年间 10 年 9 个指数因子的各季度数据，基于主成分分析构建反映数据资产规模的综合 D 指数：

$$D_t = \alpha_1 x_{1t} + \alpha_2 x_{2t} + \dots + \alpha_9 x_{9t} \quad i = 1, 2, \dots, 9, t = 1, 2, \dots, n \quad (1)$$

D_t 为第 t 个季度的 D 指数（以 2011 年第二季度作为第 1 个季度）， x_{it} 为第 i 个指数因子第 t 个季度的值，其中 α_i 为赋予第 i 个指数因子的权重， $\sum_{i=1}^n \alpha_i = 1$ 。 D_t 指数为第 t 个季度在热度、商业、学术三个方面的 9 个代表数据资产规模的指标的加权求和值，以此反映第 t 个季度的数据资产规模。

（一）指数因子的确定

用户浏览的数据、搜索的数据，都将会对平台给用户推送的内容产生巨大影响，并带来更多的流量。李秉祥、任晗晓（2021）认为这种通过在公域的平台搭建“私域流量”的营商方式，其为企业带来的价值可直接与用户建立联系，以便企业二次营销，从而降低营销成本，增强用户忠诚度。这些流量能对平台企业创造价值，是其未来利润的主要增长动力，而平台用户数据的价值则正是反映了这些平台企业的数据资产价值。本文将平台用户所带来的流量、用户数量等方面归统一纳为热度方面，作为测度数据资产规模的一大方面。

如今是互联网经济时代，不少学者认为数据行业的公司共享数据、发展数据衍生业务、促成数据交易，这已成为市场所需，需对该类型的数据资产价值或其规模构建评估模型。王静、王娟（2019）认为随着互联网金融企业数量的增长，其企业竞争力的加强以及相关产业制度的完备，互联网金融企业已成为我国经济发展的中流砥柱。其合理利用互联网大数据技术，逐步形成数据共享、利润共享，促进整个行业的发展。司雨鑫（2019）认为数据企业通过其线上便捷的网络渠道以及巨大的用户数量，已成为数据资产布局早且融合深的行业。当前数据企业依托其庞大的数据资产，实现精准营销，降低运营成本，提升其品牌影响力，数据资产所带来的价值日渐明显。所以，商业也是数据资产规模的一大方面。

2007年，计算机图灵奖得主吉姆·格雷（Jim Grey）在美国国家研究理事会计算机科学和远程通讯委员会（NRC-CSTB）的一个演讲报告中提出了科学研究的“第四范式”，即以数据为中心，以协同化、网络化与数据驱动为主要特征的数据密集型科学研究，强调了数据对现今科研工作的重大影响。而某一领域最前沿的研究成果通常被学术期刊所收录，当这些研究成果被运用于实际中，将会更好地为实体经济服务，这种数据资产的价值不言而喻，且不占少数。学术研究是数据资产规模的又一方面。

因此，本文将从热度、商业、学术三个方面构建D指数，并分别选取3个指标进行测度。

考虑到互联网为数据资产的存储、传播、规模扩大等提供了有利的条件，数据的网络热度在一定程度上也反映了数据资产的规模大小。在热度方面，本文以

百度指数中 3 个选定的“数据”相关关键词的第 t 个季度平均搜索指数作为指数因子 $x_{it}(i = 1, 2, 3, t = 1, 2, \dots, n)$ 。

数据行业的公司，以大量的数据作为原材料，进行加工处理后，又以数据作为最终产品为国家、社会、个人提供服务，其公司规模，在很大程度上同样也能作为反映数据资产规模的指标。在商业方面，选取数据行业的 3 家著名公司 BAT 的季度资产总额，分别作为指数因子 $x_{it}(i = 4, 5, 6, t = 1, 2, \dots, n)$ ，以亿元作为单位。

在学术方面，鉴于中国知网关于文献信息发布的权威性，本文在中国知网，以“数据”作为核心关键词，并联合 3 个选定的“数据”相关关键词作为筛选条件，将各季度的文献相对发表篇数作为指数因子 $x_{it}(i = 7, 8, 9, t = 1, 2, \dots, n)$ 。

（二）样本选取及关键词的确定

1. 样本选取

鉴于学术期刊通常代表某一领域最前沿研究成果，而报刊作为传播较为广泛的媒介之一也具有较强的传播性和权威性，本文基于 2011-2020 年间的中国学术期刊，获取代表性学术文献形成样本。针对学术文献，以“数据”为关键词选取了 100 篇文章，以引用量作为排序标准，确保文章来源于核心期刊；针对报刊杂志，采用网络爬取技术筛选出了人民日报和光明日报中与“数据”有关文章。

2. 关键词的确定

基于 TF-IDF 算法思想，通过 Python 的 jieba 模块，从样本文章中选取“数据”相关的代表性词汇。

（1）TF-IDF 值实际上是 $TF * IDF$ 。TF 即关键词词频，IDF 即关键词的逆向文件频率。对于某一词汇，当其在特定文件中的词频较高，且其在所有文件中的 IDF 较低时，其 TF-IDF 值所对应的权重便会相对较高。所以，此算法易将常见的词汇筛去，从而保留重要词语。

(2) 从挑选出的所有文章中随机抽取 20 篇，并利用 TF-IDF 算法进行关键词筛选，并从剩下的文章中再次随机抽取 20 篇，通过 TF-IDF 算法进行关键词筛选，以此类推直到不重复地抽取完所有样本。从所有出现的关键词中挑选出频次最高的 5 个，依次是“数据”“信息”“大数据”“个人信息”“技术”。考虑到“数据”一词是样本筛选的首要条件，虽作为出现频次最高的关键词，但其不具有代表性，所以本文在关键词的选取中剔除“数据”一词。“个人信息”一词可算在“信息”的范围内，为避免重复，因此剔除“个人信息”一词。最终将“信息”“大数据”“技术”3 个词汇作为本文代表数据的关键词，即 D 指数构建中热度与学术方面所采用的关键词。



图 1 关键词词云图

(三) 数据的搜集

本文搜集的样本为 2011 年至 2020 年各季度的相关数据，共 40 组样本：

在数据资产的热度方面，本文使用百度搜索引擎的指数工具，测算 3 个本文选定关键词“信息”“大数据”“技术”各季度的平均用户搜索指数，并分别将其作为指数因子 $x_{1t}, x_{2t}, x_{3t} (t = 1, 2, \dots, n)$ 的观察值。

在商业方面，以数据行业的三大龙头企业阿里巴巴、腾讯、百度的季度资产总额（单位：亿元）代表整个数据行业的发展规模，以此反映数据资产的规模。通过查阅三家企业官网以及 Wind 数据库，搜集阿里巴巴、腾讯、百度的各季度报表，并将其季度资产总额作为指数因子 $x_{4t}, x_{5t}, x_{6t} (t = 1, 2, \dots, n)$ 的观察值。

在数据资产的学术方面，本文在中国知网，获取各季度文献发表总篇数，将其记作 $a_t(t = 1, 2, \dots, n)$ ，同时分别以“数据”和“信息”、“数据”和“大数据”、“数据”和“技术”作为条件进行三次筛选，将筛选所得各季度的文献发表篇数记作 $k_{7t}, k_{8t}, k_{9t}(t = 1, 2, \dots, n)$ ，并剔除各季度文献发表篇数不一致的影响因素，将各季度的文献相对发表篇数作为指数因子 $x_{7t}, x_{8t}, x_{9t}(t = 1, 2, \dots, n)$ 的观察值，其中 $x_{it} = k_{it} \frac{a_t}{a_{t-1}}(i = 7, 8, 9, t = 1, 2, \dots, n)$ ，将 k_{i1} 记为 $x_{i1}(i = 7, 8, 9)$ 。

表 1 指数因子对应含义

指数因子	含义
x_1	关键词“信息”的季度平均百度指数
x_2	关键词“大数据”的季度平均百度指数
x_3	关键词“技术”的季度平均百度指数
x_4	阿里巴巴季度资产规模
x_5	腾讯季度资产规模
x_6	百度季度资产规模
x_7	知网“数据”“信息”相关文献季度相对发表篇数
x_8	知网“数据”“大数据”相关文献季度相对发表篇数
x_9	知网“数据”“技术”相关文献季度相对发表篇数

(四) 综合指数的确定

本文根据 2011 年至 2020 年间 10 年 9 个指数因子的各季度数据，通过 MATLAB 软件，利用主成分分析依据主成分的累计贡献率选取出第一、第二和第三主成分，构建一个反映数据资产规模的数据综合指数——D 指数。

以下是具体计算过程与结果：

1.0000	0.1742	0.3915	0.1380	0.1389	0.0750	0.1204	0.1557	0.1951
0.1742	1.0000	0.4886	0.4299	0.4333	0.6446	0.4279	0.6896	0.4240
0.3915	0.4886	1.0000	0.5402	0.5532	0.6080	0.4596	0.5931	0.5026
0.1380	0.4299	0.5402	1.0000	0.9954	0.9437	0.3390	0.8921	0.4166
0.1389	0.4333	0.5532	0.9954	1.0000	0.9423	0.3682	0.8925	0.4548
0.0750	0.6446	0.6080	0.9437	0.9423	1.0000	0.4142	0.9728	0.4633
0.1204	0.4279	0.4596	0.3390	0.3682	0.4142	1.0000	0.4977	0.9085
0.1557	0.6896	0.5931	0.8921	0.8925	0.9728	0.4977	1.0000	0.5076
0.1951	0.4240	0.5026	0.4166	0.4548	0.4633	0.9085	0.5076	1.0000

图 2 x_1 相关系数矩阵 A

0.0030	0	0	0	0	0	0	0	0
0	0.0058	0	0	0	0	0	0	0
0	0	0.0448	0	0	0	0	0	0
0	0	0	0.1031	0	0	0	0	0
0	0	0	0	0.4112	0	0	0	0
0	0	0	0	0	0.6630	0	0	0
0	0	0	0	0	0	1.0355	0	0
0	0	0	0	0	0	0	1.3853	0
0	0	0	0	0	0	0	0	5.3484

图 3 标准化 A 矩阵的特征根 λ

图 2 是每个指数因子 x_1 的相关系数矩阵 A, 图 3 为标准化后的矩阵 A 的特征根 λ , 而 $\frac{\lambda_i}{\sum_{i=1}^p \lambda_i}$ 称为第 i 个主成分 f_i 的贡献率, $\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i}$ 称为前 k 个主成分 $f_i (i = 1, 2, \dots, k)$ 的累计的贡献率, 显然, 贡献率越大, 表明该成分代表原始信息就越多。根据图 3 中的特征根 λ , 按从左上到右下的顺序依次计算出 f_9 至 f_1 的主成分贡献率。

表 2 各主成分贡献率

主成分	方差贡献率 (%)	累计贡献率 (%)
f_1	59.43	59.43
f_2	15.39	74.82
f_3	11.51	86.33
f_4	7.37	93.7
f_5	4.57	98.27
f_6	1.15	99.42
f_7	0.5	99.92
f_8	0.06	99.98
f_9	0.03	100

其中前 3 个主成分 f_1 、 f_2 、 f_3 累计的贡献率为 86.33%，是主要成分，说明 f_1 、 f_2 、 f_3 已涵盖了所有测量指标 $x_i(i = 1, 2, \dots, 9)$ 所包含的信息，在降维的同时，也对研究与解决实际问题更具帮助，所以本文只使用 f_1 、 f_2 、 f_3 个主成分解决数据资产规模测算问题。经计算，表 3 为 f_1 、 f_2 、 f_3 三个主成分的载荷：

表 3 主成分载荷

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
f_1	0.1079	0.2985	0.3171	0.3842	0.3887	0.408	0.2767	0.4107	0.2973
f_2	0.3081	0.0872	0.1786	-0.3262	-0.298	-0.2671	0.5468	-0.174	0.5193
f_3	-0.8391	-0.0176	-0.3189	-0.0201	-0.0053	0.0499	0.3458	0.0282	0.2658

这三个主成分的表达式分别为：

$$f_1 = 0.1078 * x_1 + 0.2985 * x_2 + 0.3171 * x_3 + 0.3842 * x_4 + 0.3887 * x_5 \\ + 0.408 * x_6 + 0.2767 * x_7 + 0.4107 * x_8 + 0.2973 * x_9$$

$$f_2 = 0.3081 * x_1 + 0.0872 * x_2 + 0.1786 * x_3 - 0.3262 * x_4 - 0.298 * x_5 \\ - 0.2671 * x_6 + 0.5468 * x_7 - 0.174 * x_8 + 0.5193 * x_9$$

$$f_3 = -0.8391 * x_1 - 0.0176 * x_2 - 0.3189 * x_3 - 0.0201 * x_4 - 0.0053 * x_5 \\ + 0.0499 * x_6 + 0.3458 * x_7 + 0.0282 * x_8 + 0.2658 * x_9$$

计算出三个主成分的系数之后，我们计算主成分得分并将其作为 D 指数。 f_1 、 f_2 、 f_3 对应的特征值分别为 $\lambda_1=5.3484$ 、 $\lambda_2=1.3853$ 、 $\lambda_3=1.0355$ ，则主成分得分 f 即 D 指数的计算方法为：

$$f = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} f_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} f_2 + \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} f_3 \\ = 0.6884 f_1 + 0.1783 f_2 + 0.1333 f_3 \quad (2)$$

四、数据资产规模测度

（一）模型的建立

基于数据资产规模指数 D 的建立，测度数据资产规模的季度增长率以及其他对 GDP 增长有贡献的变量的季度增长率对因变量 GDP 季度增长率的影响，共同

构建多元线性回归模型，得出数据资产规模增长与 GDP 增长的关系，其中 X_1 表示数据资产规模指数 D 的季度增长率， β_1 系数表示数据资产驱动经济发展的新动能。

模型基本形式为：

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \quad (3)$$

其中：

β_i 为回归系数， ε 为随机误差项， α 为回归方程截距，且满足以下假设：

- (1) 解释变量 $X_i(i=1,2,3,4,5,6)$ 是可控制的，为非随机变量；
- (2) 假设随机误差项 ε 独立为零均值，同方差的正态分布；
- (3) 因变量 Y 是连续性随机变量，且符合正态分布。

（二）控制变量

为了去除量纲的影响，对各指标进行标准化处理，并计算其对应的季度增长率。变量 X_1 为基于网络数据构建的数据资产规模指数 D 的季度增长率， X_2 为固定资产投资季度增长率， X_3 为国家财政支出季度增长率， X_4 为货物运输量季度增长率， X_5 为进出口总额季度增长率， X_6 为社会消费品零售总额季度增长率。

表 4 控制变量对应指标

变量	指标
Y	GDP 季度增长率
X_1	数据资产规模指数 D 的季度增长率
X_2	固定资产投资季度增长率
X_3	国家财政支出季度增长率
X_4	货物运输量季度增长率
X_5	进出口总额季度增长率
X_6	社会消费品零售总额季度增长率

数据资产规模指数 D：本文从热度、商业、学术三个方面，构建 D 指数，以此测度数据资产的规模。热度方面选取“数据”相关关键词每季度的平均百度搜索指数；商业方面选取三家龙头数据企业 BAT 的季度资产总额；学术方面选取

知网每季度有关“数据”的文章篇数。使用主成分分析进行降维构建一个反映数据资产规模的综合指数 D 指数,并得出 2011-2020 年相对的数据资产规模季度增长率。

固定资产投资（不含农户）：是指建造和购置固定资产的经济活动。固定资产投资额是以货币表现的建造和购置固定资产活动的工作量,它是反映固定资产投资规模、速度、比例关系和使用方向的综合性指标。固定资产投资增加,当期和远期的 GDP 增量都会增加。刘金全和于惠春（2002）通过向量自回归模型和 Granger 因果检验得出固定资产投资能显著影响实际产出；黄忠华等（2008）基于面板数据模型也得到了类似的结论,本文选取了固定资产投资这项指标。

财政支出：是指政府为履行职能可以支配的货币额,是政府分配活动的重要方面。财政对社会经济的影响作用主要是通过财政支出来实现的,因此财政支出的规模、结构往往反映政府所进行的活动范围与政策选择的倾向性。庄子银和邹薇（2003）运用时间序列分析与省际横截面分析,研究了公共支出对经济增长的促进作用,基于此本文选取了国家财政支出这项指标。

货物运输量：在一定时期内,各种运输工具实际运送的货物重量。该指标是反映运输业为国民经济和人民生活服务的数量指标,也是制定和检查运输生产计划、研究运输发展规模和速度的重要指标。周君（2006）通过货运量与 GDP 相关性模型的建立分析,发现物流业对区域经济增长有很高的贡献度；徐涵（2020）通过铁路货运量和社会经济的相关性分析也得到了类似的结论。本文选取货运运输量这项指标。

货物进出口总额：实际进出我国国境的货物总金额,包括对外贸易实际进出口货物,来料加工装配进出口货物,国家间、联合国及国际组织无偿援助物资和赠送品等。该指标能够用以反映一个国家在对外贸易方面的总规模。净出口是拉动经济增长的三驾马车之一,进出口与经济增长之间也有着密切的相关性,如熊启泉和杨十二（2005）以及高峰等（2005）指出。本文选取货物进出口总额这一指标。

社会消费品零售总额：企业通过交易直接售给个人、社会集团非生产、非经营用的实物商品金额,以及提供餐饮服务所取得的收入金额。作为拉动经济增长三驾马车之一的消费,其与经济增长之间的相关性已被多次验证,如林伯强（2003）、赵进文和范继涛（2007）、韩智勇等（2004）。本文选取社会消费品零售总额这一指标。

除此以外还有劳动参与率、人均储蓄率、人力资本投资、环境污染指标等。这也是之后的研究可以进一步完善的地方。本文选用以上六个更为传统的指标作为控制变量。

(三) 估计回归参数

估计回归最常用的方法是最小二乘法，此方法的基本思想是使所有观测点到直线的距离最小，以确定 α 和 β_i 。假定回归直线为：

$$Y = \alpha^* + \beta^* X$$

残差为观测点 (x_t, y_t) 同估计直线垂直方向的间隔：

$$e_t = y_t - (\alpha^* + \beta^* X_t)$$

最小二乘法参数估计即为求解使以下评价函数取得最小值的 α^* 和 β^* ：

$$V = \sum_{t=1}^N (y_t - \alpha^* - \beta^* X_t)^2$$

$$\text{对 } \alpha^* \text{ 和 } \beta^* \text{ 求偏导令其为零联立方程得：} \begin{cases} \beta^* = \frac{\sum_{t=1}^N (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^N (x_t - \bar{x})^2} \\ \alpha^* = \bar{y} - \beta^* \bar{x} \end{cases}$$

其中， $\bar{x} = \frac{1}{N} \sum_{t=1}^N X_t$, $\bar{y} = \frac{1}{N} \sum_{t=1}^N Y_t$ ，能够证明 α^* 和 β^* 分别是 α 和 β 的一致无偏估计量。

(四) 模型的检验

1. 模型假设

F 检验，即方差分析：是对模型总显著性的检验

原假设： $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6$ 说明模型拟合不好

备择假设： $H_1: \beta_i \neq 0 (i = 1, 2, \dots, 6)$ 说明模型拟合得较好

F 检验：因变量 y_1, y_2, \dots, y_N 的差异用总离差平方和：

$$TSS = \sum_{i=1}^N (y_i - \bar{y})^2, df_T = N - 1$$

造成这一差异主要有两方面原因：

一是由于 $\beta = 0$ 不真导致的回归平方和 $RSS = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2, df_R = 1$;
 二是其他一切随机因素导致的残差平方和 $ESS = \sum_{i=1}^N (y_i - \hat{y}_i)^2, df_E = N - 2$ 。
 在 $\beta = 0$ 为真时:

$$F = \frac{RSS/\sigma^2/1}{ESS/\sigma^2/(N-2)} = \frac{RSS}{ESS/(N-2)} \sim F(1, N-2)$$

设检验水平为 α , 则检验规则是:

若用样本计算的 $F < F_{\alpha}(1, N-2)$, 则接受 H_0 ;

若用样本计算的 $F > F_{\alpha}(1, N-2)$, 则拒绝 H_0 。

t检验: 是对模型单个参数的显著性检验

原假设: $H_0: \beta_i=0 (i = 1, 2, \dots, 6)$ 说明自变量对因变量的解释效果为 0;

备择假设: $H_1: \beta_i \neq 0 (i = 1, 2, \dots, 6)$ 说明自变量解释效果是显著的。

在 $\alpha = 0$ 为真时:

$$t = \frac{\frac{\hat{\alpha}}{\sigma \sqrt{1/N + \bar{x}^2 / \sum (x_i - \bar{x})^2}}}{\sqrt{\frac{ESS}{\sigma^2} / (N-2)}} \sim t(N-2)$$

表 5 T 检验与 P 值的关系

t	P值	差异显著程度
$t \geq t_{(df)} 0.01$	$P \leq 0.01$	差异非常显著
$t \geq t_{(df)} 0.05$	$P \leq 0.05$	差异显著
$t < t_{(df)} 0.05$	$P > 0.05$	差异不显著

2. 变量标准化

本文首先对因变量以及自变量进行标准化处理, 消除不同量纲的影响, 使数据更具可比性与客观性。考虑到国家统计局公开 GDP 数据为名义 GDP, 即按当年的价格计算, 本文统一使用国家统计局公布季度数据, 保持价格影响的统一性。

对序列 X_1, X_2, \dots, X_n 使用 z-score 标准化方法进行变换:

$$y_i = \frac{x_i - \bar{x}}{s} \quad (i=1, 2, \dots, n)$$

$$\text{其中 } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

则新序列 y_1, y_2, \dots, y_n 的均值为 0, 方差为 1, 且无量纲。

3. 变量的筛选

本文采用逐步筛选法对变量进行筛选, 逐步筛选法在按照向前选择法选入变量后, 可剔除使得 F 统计量的显著性水平低于某一定义的水平变量, 它是向前选择法与向后选择法的结合。

本文基于除去 2011 年第一季度外的近十年共 39 个季度数据对六个变量进行筛选, 首次逐步筛选法保留变量仅为 X_6 , 本文为进一步进行共线性诊断与 COOKS'D 强影响点诊断, 发现有较为明显的异常值, 最终剔除 2017 年第一季度与 2020 年第二季度两个样本, 重新拟合模型。经过变量相关分析等, 通过逐步筛选法最终保留 X_1, X_3, X_4, X_6 四个变量, 这其中包括数据资产规模指数 D 的增长率 X_1 变量, 最终选定建立以 X_1, X_3, X_4, X_6 为自变量, Y 为因变量的多元线性回归模型。

(五) 回归诊断

1. 残差分析

残差图通常是以残差为纵坐标, 某一合适的自变量为横坐标的散点图。若观察到数据满足模型假设中误差项是均值为零、方差相等、独立的正态分布随机变量, 则残差作为误差的无偏估计反映误差的基本假设特征。由图 4 正态概率分布图可看出, 样本点分布与 45 度线重合度较高, 大致符合正态分布的假设。

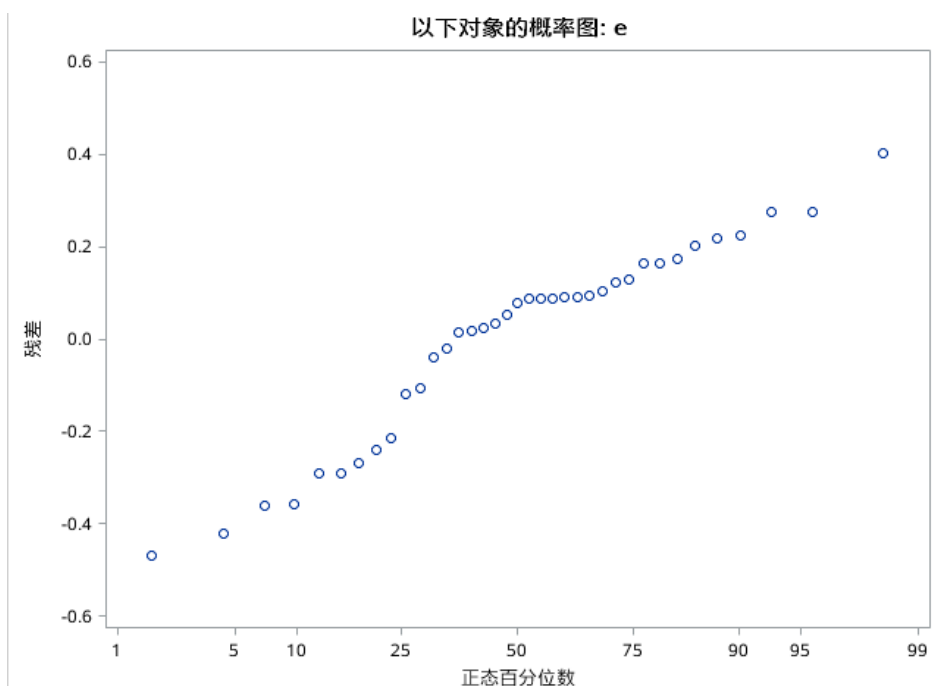


图 4 残差正态概率图

2. 共线性诊断

若回归分析中两个以及两个以上的自变量高度相关, 会引起最小二乘结果不精确。本文采用条件指数 **COLLIN** 和方差膨胀因子 **VIF** 统计量进行共线性诊断。

基于共线性判断, 若方差膨胀因子大于 10, 就可能具有共线性。若有一对变量偏差比例的值是对称的并且非常接近于 1, 那么这一对量就可能是具有共线性。

表 5 表示选取变量的方差比例, 也可看出各变量相互间不存在共线性。

表 6 方差比例

变量 维度	方差比例				
	常量 α	X_1	X_3	X_4	X_6
1	0.02	0.07	0.07	0.03	0.07
2	0.36	0.05	0.04	0.05	0.10
3	0.12	0.00	0.00	0.83	0.02
4	0.17	0.07	0.23	0.07	0.51
5	0.33	0.81	0.66	0.02	0.30

3. 误差项独立性检验

若误差项不独立，则参数的假设检验、无偏估计等都不能成立。本文采用残差的一阶相关性进行 Durbin-Watson 检验。原假设为随机误差项不存在一阶序列相关，备择假设为随机误差项存在一阶序列相关。

（六）模型的分析

表 7 汇总表

变量	回归系数	T 值	P 值	VIF	调整 R^2	F 值	P 值	DW 检验
X_1	0.14609	3.52	0.0013	1.066	0.9419	150.92	<0.0001	2.406
X_3	-0.4333	-3.40	0.0018	1.706				
X_4	0.13566	2.65	0.0124	1.536				
X_6	0.96705	16.86	0.000	2.276				

由表 7 分析可得：

对模型进行拟合优度检验，调整 R^2 值都达到 0.9419，接近 1，说明该模型拟合较有效；

对总回归模型进行 F 检验，对应的 P 值远远小于设定的显著性水平 0.05，则拒绝原假设 H_0 ，说明回归系数不全为零，回归方程建立是有效的；

对各变量对因变量的显著性进行 t 检验，各个变量所对应 t 检验的 P 值都小于显著性水平 0.05，说明四个控制变量都与因变量显著；

进行方差膨胀因子（VIF）检验自变量间的相关性，各变量 VIF 值都远小于 10，不存在共线性问题；

对误差项独立性（DW）进行检验自相关检验，其 DW 值为 2.406，当残差与自变量相互独立时，DW 值应该在 2 附近，可近似符合正态分布。

以上分析可确定 D 指数影响 GDP 增长的多元回归分析模型为：

$$Y = 0.14609x_1 - 0.4333x_3 + 0.13566x_4 + 0.96705x_6 \quad (4)$$

由上述回归方程可知， x_1 变量与 GDP 增长率 Y 呈正相关，其系数为 0.146，表示 D 指数每增长一个单位，都会使得 GDP 增长率平均提高 0.146 个单位，表明，D 指数所反映的数据资产规模对 GDP 增长的确有促进作用。且由上述回归方程可看出，社会消费品零售总额在此项模型中对 GDP 增长影响最大，系数高达 0.932，更能表示作为“三驾马车”之一的消费水平对 GDP 的推动作用，但由于 GDP 这个宏观变量本身受多种其他变量影响，本文建立的模型对国家财政支出和货物运输量总额的解释程度较小。

六、结论与展望

（一）结论

正值大数据时代，数据的总数量及增幅，因云计算等行业的崛起，有了飞速的上涨，然而在国民经济核算体系 SNA2008 中，数据资产并未被纳入 GDP 的核算之中。查阅相关文献资料，发现针对数据资产的核算方面的研究，有诸多学者对此提供了建议与解决方案，且认为数据可作为一种要素引进 GDP 核算之中，但目前仍没有统一广泛认可并被纳入实际应用之中的测度方法。其中针对数据资产的规模测度方面研究较少。

我们认为，数据资产是一类对于经济增长有着推动作用的、未被完全开发的、且被低估其实际价值的新动能。因此本文从热度、商业、学术三个方面选取 9 个指标，基于主成分分析构建 D 指数以表示数据资产规模。其中，热度方面，我们利用 TF-IDF 算法从各类新闻期刊和学术文章中提取出能够代表数据的三个关键词，之后在百度指数上搜索这三个关键词的季度指数，作为热度方面的指标；在商业方面，我们选取在数据、云计算方面具有代表性的‘BAT’互联网巨头，得到它们各季度的资产总额，作为商业方面的指标；在学术方面，根据之前利用 TF-IDF 算法得到的关键词，作为搜索指标，在知网上分别搜索出三个关键词所对应的文章数量，并提取季度数据，作为学术方面的指标，以此 D 指数在一定程度上反映数据资产规模的大小。随后，本文在 D 指数的基础上，通过多元线性回归的方法，构建 GDP 增长指数模型，将数据要素引入其中，测度数据要素

对经济增长的推动程度 $\beta=0.146$ ，旨在为数据资产评估、完善数据市场以及数据要素市场对国内新经济、新动能的推动等方面提供思考方向。

热度方面，应积极打造数据热潮，让数据引发新一轮搜索革命和思想革命，启发新的商业思维，让数据进一步走近人民，成为新兴技术与国民生活的粘合剂。

商业方面，以“BAT”为代表的互联网企业应重视其企业内部存储的数据资产，最大化地利用互联网平台的共享与联通红利，发挥数据资产的价值。让数据流动起来，用科技为数据赋能，让数据的力量推动企业加速发展，攀登新的高峰。

学术方面，应积极营造数据讨论氛围，促进数据资产核算相关研究及文献蓬勃发展，为之后数据资产价值的相关研究奠定坚实的基础。

数据的能量加速席卷着这个时代，数据及数据资产规模对于经济发展的推动作用已是板上钉钉的事实，而对于数据资产及其规模的测度与核算的论证及相关的实证研究却没有跟上数据发展的步伐。为实现数据资产与经济发展的相互推动与促进，形成良性循环，相关研究应该加快脚步，在确认数据可作为 GDP 核算的其中一种要素的基础上，将其纳入实际应用之中进行测度与计量，丰富数据资产价值维度，建立数据资产价值评估模型，推动数据资产价值评价指标体系的构建与完善。具体的，应指出哪些方面对数据资产规模的度量有代表性作用，在这些方面中又应从何种角度对其进行测算，如此才能对症下药，做到数据对经济发展推动的“精准打击”。本文大胆地从热度、商业与学术三个方面入手，选取了部分代表性指标，构建新的 GDP 增长指数模型，测度数据要素对经济增长的推动程度，试图迈出数据资产对经济发展推动的一小步。也许数据及数据资产规模的准确度量方法离我们还很远，但我们终能揭开它的神秘面纱。

（二）创新点

本文的创新点在于：

- （1）构建 D 指数反映数据资产规模。数据资产规模研究成果较少，本文通过热度、商业、学术三方面构建 D 综合指数，以测度其规模大小；
- （2）使用文本挖掘确定关键词。在 D 指数的构建过程中，依据 TF-IDF 算法，爬取近十年“数据”相关样本报纸，提取“数据”相关关键词；

(3) 数据要素驱动 GDP 增长。在传统的生产要素（劳动、资本、技术、土地）以外，将数据作为新型要素，加入 GDP 增长的多元线性回归模型中，以此测度数据新动能。

（三）研究展望

本文需改进的地方有：

(1) 在构建 D 指数的过程中，由于总体报刊文章数量过于庞大，本文所选取到的样本文章数量与其相比不充分大，因此根据这些样本文章所选择的“数据”关键词不足以完全代替“数据”一词的贡献；

(2) 在 GDP 增长指数模型中，部分指标的统计数据存在个别值缺失的情况，需基于现有数据通过一定的计算处理方式获得，因此经处理后的个别数据仍没有真实数据的可靠度高；

(3) 在 GDP 增长指数模型中的指标选取方面，即使前人已有研究表明某些指标对 GDP 的增长有促进作用，但由于其统计数据不宜获得因而被排除在选取范围内，本文的 GDP 增长指数模型中因素指标的选取仍有待改进。

参考文献

- [1] 苏晓.2020 年我国数字经济规模达到 39.2 万亿元[N].人民邮电报,2021-05-01(03).
- [2] 邓聪.数据驱动数字经济加速发展[N].人民邮电报,2020-12-17(03).
- [3] 前沿播报.2021 中国数据资产化工具市场研究报告今日正式发布[N].
https://www.sohu.com/a/456388508_120766754,2021-03-19.
- [4] 柴士改,李金昌.中国经济增长新动能统计测度研究[J].统计与信息论坛,2021,36(01):47-58.
- [5] 何强.中国农村经济发展新动能统计测度及提升路径研究[J].调研世界,2019(01):4-10.
- [6] 刘姝雯,杨胜刚,阳旸.中国农村经济发展新旧动能转换测度与评价[J].统计与决策,2021,37(08):73-76.
- [7] 王珍.论区域一体化城市经济增长的新旧动能转换——来自泛珠三角区域的考察[J].中国房地产,2019(24):49-57.
- [8] 张豪,张建华,谭静.中国经济增长的源泉与动能转换:1952-2015[J].经济问题探索,2017(09):12-24.
- [9] 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.
- [10] 周荣荣,李佳.全要素生产率提升与中国经济增长新动能成长[J].现代经济探讨,2019(09):17-24.

- [11]行帆,居桦,居占杰.新时代中国经济增长动能转换研究[J].河南工业大学学报(社会科学版),2019,15(03):53-60.
- [12]方志耕.产业经济发展动能及其测度模型研究[A].中国高等科学技术中心.“资源环境与区域发展中的计算问题”研讨会论文集[C].中国高等科学技术中心:中国高等科学技术中心,2006:32.
- [13]王一鸣.百年大变局、高质量发展与构建新发展格局[J].管理世界,2020,36(12):1-13.
- [14]穆荣平.国家创新体系与能力建设的有关思考[J].中国科技产业,2019(07):20-21.
- [15]余江,孟庆时,张越,张兮,陈凤.数字创新:创新研究新视角的探索及启示[J].科学研究,2017,35(07):1103-1111.
- [16]柳卸林,董彩婷,丁雪辰.数字创新时代:中国的机遇与挑战[J].科学学与科学技术管理,2020,41(06):3-15.
- [17]本报评论员.让数字经济成为高质量发展的强劲引擎[N].贵州日报,2021-05-27(001).
- [18]刘小红,张人龙.数字经济引领新型工业化高质量发展[N].贵州日报,2021-05-26(007).
- [19]焦豪.加快数字经济下企业技术创新能力提升[N].中国经济时报,2021-05-27(004).
- [20]尹西明,林镇阳,陈劲,林拥军.数据要素价值化动态过程机制研究[J/OL].科学学研究:1-18[2021-05-27].<https://doi.org/10.16192/j.cnki.1003-2053.20210524.001>.

- [21] PETERSON R E. A cross section study of the demand for money: The United States, 1960-62[J]. Journal of Finance, 1974, 29(1): 73-88
- [22]李静萍.数据资产核算研究[J].统计研究,2020,37(11):3-14.
- [23]Ahmad N, Ven P. Recording and Measuring Data in the System of National Accounts[J/OL].https://unstats.un.org/unsd/nationalaccount/aeg/2018/M12_3c1_Data_SNA_asset_boundary.pdf, 2018—11—02.
- [24]李泽红,檀晓云.大数据资产会计确认、计量与报告[J].财会通讯,2018(10):58-59+129.
- [25]崔东佳.大数据时代背景下的品牌汽车销量预测的实证研究[D].河南大学,2014.
- [26]唐一丁.网络搜索数据在预测房地产价格指数中的应用研究[D].吉林大学,2016.
- [27]张涛,刘宽斌.农业劳动力转移对中国经济增长的贡献研究——基于网络搜索大数据的视角[J].劳动经济研究,2019,7(01):120-138.
- [28]王勇,董恒新.大数据背景下中国季度失业率的预测研究——基于网络搜索数据的分析[J].系统科学与数学,2017,37(02):460-472.
- [29]张瑞,刘立新,唐晓彬,张斌儒.大数据背景下基于网络搜索数据商品零售价格指数预测研究[J].统计与信息论坛,2020,35(11):49-56.
- [30]李秉祥,任晗晓.大数据资产的估值[J].会计之友,2021(21):127-133.
- [31]王静,王娟.互联网金融企业数据资产价值评估——基于 B—S 理论模型的研究[J].技术经济与管理研究,2019(07):73-78.

- [32]司雨鑫. 互联网企业中数据资产价值评估模型研究[D].首都经济贸易大学,2019.DOI:10.27338/d.cnki.gsjmu.2019.000946.
- [33]吴锋.“大数据时代”科技期刊的出版革命及面临挑战[J].出版发行研究,2013(08):66-70.DOI:10.19393/j.cnki.cn11-1537/g2.2013.08.023.
- [34]周晓英.情报学进展系列论文之七 数据密集型科学研究范式的兴起与情报学的应对[J].情报资料工作,2012(02):5-11.
- [35]刘金全,于惠春.我国固定资产投资和经济增长之间影响关系的实证分析[J].统计研究,2002(01):26-29.
- [36]黄忠华,吴次芳,杜雪君.房地产投资与经济增长——全国及区域层面的面板数据分析[J].财贸经济,2008(08):56-60+72.
- [37]庄子银,邹薇.公共支出能否促进经济增长:中国的经验分析[J].管理世界,2003(07):4-12+154.
- [38]周君.区域物流业对地区经济增长的影响分析[J].统计与决策,2006(04):109-112.
- [39]徐涵.铁路货运量与国民经济发展的内在关系研究[J].铁道经济研究,2020(04):38-42.
- [40]熊启泉,杨十二.重新审视进口在经济增长中的作用——基于中国的实证研究[J].国际贸易问题,2005(02):5-10.
- [41]高峰,范炳全,王金田.我国进出口贸易与经济增长的关系——基于误差修正模型的实证分析[J].国际贸易问题,2005(07):5-9.
- [42]林伯强.电力消费与中国经济增长:基于生产函数的研究[J].管理世界,2003(11):18-27.

[43]赵进文,范继涛.经济增长与能源消费内在依从关系的实证研究[J].经济研究,2007(08):31-42.

[44]韩智勇,魏一鸣,焦建玲,范英,张九天.中国能源消费与经济增长的协整性与因果关系分析[J].系统工程,2004(12):17-21.

致谢

作为上海对外经贸大学贸易统计创新实验班的三名学生，通过本次统计建模比赛，我们结合自身专业知识，研究了数据资产规模对 GDP 增长的驱动作用。我们希望这次的研究能够带来一定的理论与实际意义，在此尤为感谢我们此次比赛的指导老师李佩瑾老师！从选题，到数据搜集，再到模型的建立，以及最后得出结论的整个过程中，李老师付出了巨大的心血，在学术方面，为我们提出了很多指导意见，并且在遇到我们遇到困难时，老师不断地为我们加油鼓劲，才让我们能够顺利完成这次的研究。在此感谢李老师的辛勤付出！