

A0485

# 数字经济对中国经济增长的贡献统计 测度

论文题目:数字经济对中国经济增长的贡献统计测度

参赛学校:东北财经大学

参赛成员(作者):钟宇峰、戴添宇、张津玮

指导老师:王鹏飞

# 目录

- 一、 研究背景及文献综述 ..... 1
  - (一) 研究背景 ..... 1
  - (二) 研究目标 ..... 2
  - (三) 相关文献综述 ..... 2
- 二、 研究思路和模型介绍 ..... 3
  - (一) 主要研究思路 ..... 3
  - (二) 模型介绍 ..... 4
    - 1. 柯布-道格拉斯生产函数模型 ..... 4
    - 2. 回归模型 ..... 5
    - 3. ARIMA 模型 ..... 6
    - 4. 新动能评价指标体系<sup>[3]</sup> ..... 7
- 三、 数据生产要素以一种新动能的方式对经济增长产生了重要推动作用 .... 8
  - (一) 新动能、新动能指数的定义 ..... 8
  - (二) 数据演化成生产要素的理论依据 ..... 10
  - (三) 数据对经济增长的推动作用 ..... 11
- 四、 模型的构建 ..... 12
  - (一) 数据生产要素对经济增长的贡献率——生产函数模型 ..... 12
  - (二) 数据要素相对于其他要素的倍增作用——多元线性回归模型 ..... 14
  - (三) 数字经济对中国经济增长的推动作用——ARIMA 模型 ..... 16
    - 1. 原始数据的处理 ..... 16
    - 2. 序列平稳性检验 ..... 17
    - 3. 模型建立和参数的估计 ..... 18
    - 4. 模型检验 ..... 21
- 五、 模型的应用 ..... 22
  - (一) 数据的来源 ..... 22
  - (二) GDP 预测 ..... 23
- 六、 总结与建议 ..... 27
  - (一) 结论 ..... 27
  - (二) 数字经济以及数据要素发展相关建议 ..... 27

七、	创新与不足 .....	30
(一)	创新 .....	30
(二)	不足 .....	30
参考文献	.....	31

表格和插图清单

图 1 主要研究思路 .....4

图 2 ARIMA 模型基本程序.....7

图 3 新动能定义及构成要素 .....9

图 4 数据要素对经济增长贡献形式 ..... 12

图 5 数据在生产力中的作用机制 ..... 12

图 7 自相关和偏相关检验结果..... 19

图 8 残差的自相关性检验 ..... 21

图 9 残差的正态性检验（QQ 图） ..... 22

图 10 三大产业增加值和融合度随时间变化 ..... 26

表 1 各变量参数估计结果 ..... 13

表 2 线性回归模型拟合结果 ..... 15

表 3 数据序列进行 ADF 检验结果 ..... 17

表 4 数据一阶拆分结果 ..... 18

表 5 参数调整结果 ..... 19

表 6 2019 年和 2020 年的各产业的预测值和真实值 ..... 23

表 7 2014 年和 2015 年各产业的预测值和真实值 ..... 24

表 8 2014 年和 2015 年 GDP 和三大产业收益率..... 25

表 9 2016-2019 年数字经济在三大产业之间的融合度 ..... 25

# 数字经济对中国经济增长贡献统计测度

## 摘要

大数据时代和第四次科技革命背景下，数据成为一种新型生产要素，并与资本、劳动、土地、技术四大传统生产要素同时出现在 2020 年国家印发的相关文件中，成为第五大生产要素。数据以一种渗透的方式参与生产过程，并替代了一些落后的生产要素组合方式，为生产过程做出突出贡献，提高了经济运行效率，加大了经济发展的潜力，提升了经济社会的活力，为经济增长注入新动能。近几年，以数字技术为依托，大数据、云计算、互联网、人工智能等应用型技术为支撑的数字经济的爆炸性发展又为经济的增长产生了重要的影响。本文对数据要素以及数字经济对中国经济增长的贡献展开研究。

首先，选取相关经济运行指标，观察数据要素和数字经济的发展带来的经济指标的变化，直观地了解到它们对中国经济增长的推动作用并在此基础上提出问题的假设。

其次，根据分析的问题建立合适的模型，并对建立的 ARIMA 模型进行检验，验证模型的适用性，然后选取相关数据，带入模型进行测算，测度数据要素对经济增长的贡献率和它相对其他生产要素对经济增长的倍增效应，并通过联系三大产业的增长情况来论证数字经济发展对中国经济增长产生的巨大影响。

最后，根据模型分析的结果得出相关的结论：数据要素对经济的质量和数量的增长具有显著的贡献，数字经济的发展对经济发展具有重要影响。并基于此提出一些数字经济发展的意见建议。

**关键词** 数据 数字经济 新动能 ARIMA 模型

## 一、 研究背景及文献综述

---

### (一) 研究背景

2020 年 4 月 9 日,国务院印发《关于构建更加完善的要素市场配置体制机制的意见》,文件中明确将数据作为一种新型生产要素,这是首次在国家战略层面提出将数据纳入生产要素,推动大量数据信息市场化的配置。这充分体现了数据生产要素的重要作用以及对经济发展产生的重要影响。

自从党的十九大以来,中国在加快构筑"智慧社会"和"数字中国"的宏伟蓝图引领下,加强了对数字与产业融合的力度,使得我国基于数字经济产业化、新兴产业数字化进程不断推进,基于数字经济所衍生的海量数据即大数据的主导地位和作用也日益得到凸显。

在第四次工业革命的背景下,随着互联网信息技术的不断发展和新一代移动通信技术的覆盖范围不断扩大,以人工智能、区块链、云计算、大数据等底层数字技术驱动的数字经济正在蓬勃发展。数字经济是以数字技术为依托,以数据的生产、流通、使用、消费等活动为支柱的全新经济形态。以阿里巴巴、腾讯、美团为首的互联网企业收集了海量的消费数据,以带动旗下产业的发展,并促进数字经济的迅速发展。此外,根据中国互联网络信息中心(CNNIC)公布的第 47 次《中国互联网络发展状况统计报告》中提供的数据,截止 2020 年 12 月,我国网民规模为 9.89 亿,互联网普及率达 70.4%,庞大的网民给消费市场贡献了海量数据。消费端数字化要素向供给侧的渗透又进一步带来产业互联网的发展,传统行业在 5G、移动互联网、物联网、云计算、大数据等数字技术的赋能下,可以实现更大范围的数字化和互联互通,包括来自生产端的机器设备,传感器等实时产生的海量数据。数据的海量聚集和爆发式增长是数字经济发展的一个长期趋势和特征。

## (二) 研究目标

数字经济时代，数据成为提高生产力、发展生产力的关键要素，蕴含着巨大的使用价值和潜能。鉴于数据这种新型要素对经济高质量发展所起的推动作用愈加明显，如何培育数据要素市场，使数据成为推动经济高质量发展的新动能，如何利用数据形成经济发展的新动能，是目前自然科学工作者和社会科学工作者共同关注的焦点。本文旨在研究数据作为新型生产要素是如何推动经济高质量发展和数据相较其他生产要素效率的倍增作用；以及探究以数字技术为依托、以数据的各类活动为支柱的数字经济对我国经济增长的贡献度；测度数字经济对中国经济发展的影响；并计算出数据对经济发展提供的新动能指数；最后预测我国数据产业发展的态势，针对数据行业发展如何带动经济发展提供一些建议，具有一定的现实意义。

## (三) 相关文献综述

近年来，国内外不少学者针对数据要素对经济增长贡献以及数字经济如何推动经济高质量发展展开了研究。本文主要参考、借鉴了以下学者的学术研究成果和相关期刊的研究报告：王胜利等（2020）从价值创造过程出发，强调数据生产要素可以缩短生产时间和流通时间，降低生产成本和流通成本，从而论证数据要素对经济增长的贡献；冯云廷（2020）认为数据规则和市场规则存在相互作用，这种相互作用使得数据要素“贡献度”的权衡和分布，成了一个很重要的问题，并提出做好数据作为生产要素的分配或分类，才能分辨出不同数据的贡献度；在此基础上，崔俊富等（2021）利用间接测算的方法，将数据这个生产要素带入柯布道格拉斯生产函数中测算出数据生产要素对中国经济增长的贡献率；《光明日报》（2020）中提到，与土地、资本传统生产要素相比，数据要素对推动经济增长具有倍增效应，数据可以提高经济运行效率，推动产

业转型升级，并提升政府治理效能；Honohan（2004）发现数字经济的繁荣能够减少社会贫困率，促进普惠金融与共享金融发展，优化分配结构与产业结构，从而能够促进国家经济呈指数式增长；Campbell（2013）则认为数字经济与实体经济增长规模之间并非是纯粹的线性关系，在分析两者联系时应充分考虑不同经济体内部的社会环境、文化背景、产业结构与体制机制等因素；在数字经济对经济发展的作用上，郭家堂和骆品亮（2016）认为以互联网为核心内容的数字经济可以通过思维、技术、平台与网络四种途径作用于我国全要素生产率，通过实证模型检验发现互联网经济可以显著促进我国全要素生产率的提升，但这种影响是非线性的，而是具有一定的门阀值，从而进一步的分析，互联网经济通过提升技术进步效率从而提高我国全要素生产率；张腾等（2021）采用空间计量模型测算了数字经济对于不同维度经济增长质量的影响，李天明（2020）等通过 ARIMA 模型测度了重大突发事件对贵州工业发展的影响。<sup>[6]</sup>

## 二、 研究思路和模型介绍

---

### （一） 主要研究思路

本文的主要研究思路是，首先通过选取相关指标运行态势的变化，分析数据要素以及数字经济的发展对最近 30 年中国经济增长的影响，体现在总产出（GDP）的变化。其次，借助柯布-道格拉斯生产函数模型，带入采集的国家统计局的统计数据，测度数据生产要素对浙江省经济增长的贡献率，再类比推断它对中国经济增长的贡献率；同时建立多元线性回归模型，对不同要素组合的投入和总产出进行回归，体现数据要素较其他生产要素效率的倍增效应；并通过近 30 年的 GDP 数据和三大产业数据建立 ARIMA 模型，对 2015 年以及 2014 年的三大产业增加值进行模拟计算，最后加总得到 GDP 预测值，并与实际的 GDP 数值进行比较，测度数字经济的发展对经济增长的影响。最后，结合中国



经济的新动能投入产出数据，根据指标体系与测算方法对 2015-2019 年的经济增长新动能指数进行测算，并针对未来数据要素市场发展以及数字经济的发展方向展开分析，提出一些相关的结论和建议。

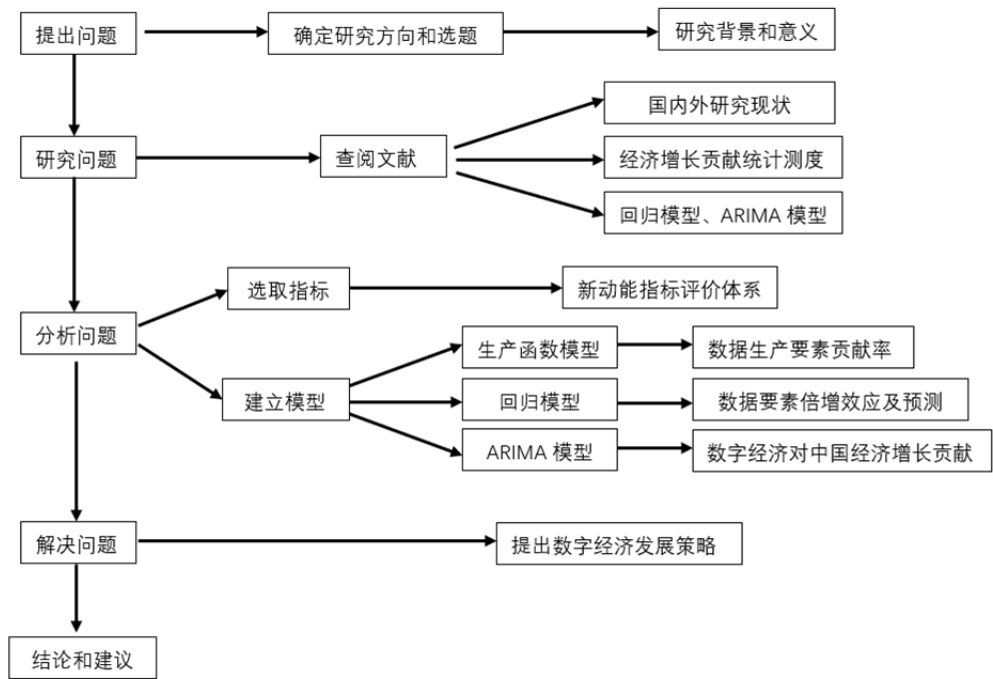


图 1 主要研究思路

(二) 模型介绍

1. 柯布-道格拉斯生产函数模型

宏观经济学中研究一个经济体的经济增长时采用了一种函数模型，用来测度生产要素投入与总产出的函数关系。该模型就是柯布 - 道格拉斯 (Cobb-Douglas) 生产函数模型，简称生产函数。该模型是美国数学家柯布和经济学家道格拉斯共同提出的。该模型引用古典经济学中的相关理论，起初假设生产过程中资本和劳动两个要素参与生产过程，且认为他们对生产过程的贡献程度有差异。随着科学技术的发展，柯布和道格拉斯发现技术资源在生产过程中发挥的作用和影响比较显著，随后便在此基础上进行改进，引入技术这一

生产要素,并且认为作用方式不同于资本和劳动。这一模型得到了广泛的应用,不仅可以用来预测国家或地区的劳动生产与产出增长的分析发展,而且在要素投入对经济增长的贡献的分析中也非常合适,在经济计量学和数理经济学中的研究与应用中都具有重要的地位和作用。

①生产函数的一般形式用公式表示为:

$$Y = AK^{\alpha}L^{\beta}, 0 < \alpha, \beta < 1 \quad \text{公式(1)}$$

其中Y表示总产出;A表示技术水平;K表示投入的资本量;L表示投入的劳动量; $\alpha, \beta$ 表示资本和劳动的产出弹性,且 $\alpha + \beta = 1$ 。

②在宏观经济学中,著名的柯布-道格拉斯(Cobb-Douglas)生产函数为:

$$Q(K, L) = aK^{\alpha}L^{\beta}, 0 < \alpha, \beta < 1 \quad \text{公式(2)}$$

式中Q, K, L分别代表产值、资金和劳动力; $a, \alpha, \beta$ 的值是由经济统计的数据确定的。

③本文在柯布-道格拉斯生产函数的基础上进行拓展,假设生产函数为:

$$Y = F(x_1, x_2, \dots, x_n) = x_1^{z_1} x_2^{z_2} \dots x_n^{z_n}, 0 < z_1, z_2, \dots, z_n < 1 \quad \text{公式(3)}$$

其中Y为总产出(以GDP衡量), $x_1, x_2, \dots, x_n$ 表示各种生产要素, $z_1, z_2, \dots, z_n$ 表示各种生产要素的贡献度或产出弹性。

## 2. 回归模型

回归模型(regression model)是一种对变量之间关系进行定量描述的数学模型,也是一种预测性的建模技术,它研究的是因变量(计量经济学中也称响应变量)和自变量(也称解释变量)之间的关系。这种分析模型广泛应用与分析预测、时间序列预测以及变量之间因果关系的探究。

回归分析是回归模型的基础也是其采用的方法,它是数据分析和数据建模的重要工具。它不仅能够揭示因变量和一个自变量之间的相关关系,而且还可

以揭示多个自变量对一个因变量的影响程度大小。

回归模型主要包括线性回归、广义线性回归、分位数回归和套索 (LASSO) 回归七种类型。本文采用最常见的线性回归模型进行生产要素投入与产出的分析。由于我们选择的自变量均为数值变量,且自变量个数大于 1。因此,本文选取的回归模型为多元线性回归模型。其一般形式用公式表示为:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \quad \text{公式(4)}$$

其中  $Y_i$  表示因变量第  $i$  个观测值,  $X_{ki}$  表示第  $k$  个自变量的第  $i$  个观测值,  $\beta_k$  表示第  $k$  个自变量的系数,  $u_i$  表示误差项。其中误差项满足同方差假定和正态性假定。

### 3. ARIMA 模型

ARIMA 模型 (Autoregressive Integrated Moving Average model), 即差分整合移动回归模型, 它是由 ARMA 模型扩展而来。主要应用于根据历史时间序列数据判断研究因变量的未来走势, 进行变量的预测, 帮助决策者做出决策。它是现代统计分析方法的应用, 是一种应用广泛的计量经济模型, 可以很好的对企业的未来进行预测。

ARIMA( $p, d, q$ ) 包含自回归和与移动平均。其中  $p, d, q$  三个参数分别表示自回归项数, 滑动平均数, 为平稳时间序列所进行的差分次数。

参数的确认。对原始数据进行差分 (若原始数据为平稳序列, 则参数  $d$  为 0), 通过 ADF 检验直到时间序列为平稳序列, 则差分次数就是参数  $d$  的值; 通过自相关函数和偏相关函数来确定参数  $p$  和  $q$  的值。

ARIMA( $p, d, q$ ) 模型是 ARMA( $p, q$ ) 模型的扩展。ARIMA( $p, d, q$ ) 模型可以表示为:

$$(1 - \sum_{i=1}^p \varphi_i L^i)(1 - L)^d X_i = (1 + \sum_{i=1}^q \theta_i L^i) \varepsilon_t \quad \text{公式(5)}$$

其中,  $L$  是滞后算子,  $d \in \mathbb{Z}, d > 0$ 。

ARIMA 模型进行预测主要有六个基本程序, 建立步骤如图 2 所示。

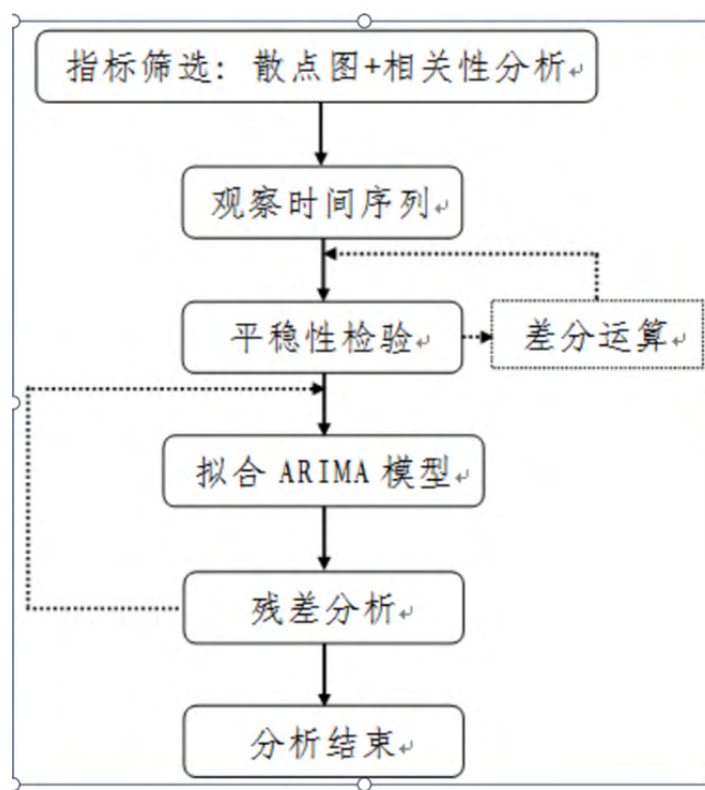


图 2 ARIMA 模型基本程序

#### 4. 新动能评价指标体系<sup>[2]</sup>

本文选取的经济增长新动能的指标是根据新动能的构成要素来划分的。新动能包括投入要素和产出要素。根据古典经济学和新古典经济学的相关理论, 新动能的投入要素应该包括新资本、新技术、新人才、和新资源( 数据资源等): 新资本主要是指金融体系对实体经济的支持力度、民间投资和创业风险投资; 新技术主要来源于自主创新, 在移动互联网、大数据, 人工智能、企业信息管理等方面的技术创新; 新人才主要是指各层次劳动力要素投入的新比例, 高精尖人才占比逐步提高; 新资源主要是践行绿色环保发展理念并代表动能发展方向的新型清洁动力能源以及数据这种刚被国家相关文件列入生产要素的要素资

源。

新动能的产出要素主要包括新创业、新企业、新产品、新产业、新市场以及新模式。新创业是在国家政策的支持下,进行高新技术中小企业的创业孵化,以减小创业风险,提高创业成功率和科研成果转化率;新企业就是指新诞生、新培育的民营企业,它们可以为经济社会的转型升级提供新动能;新产品是指一些高技术产品;新产业主要包括一些战略新兴产业和高技术产业;新市场是新动能的主要来源之一,主要是指高新技术市场;新模式是指一些创新的生产模式、商业模式、消费模式。

根据新动能的这些构成要素,本文在参考有关文献的基础上构建了一系列评价指标,根据指标的变化和对指标的系统性描述,判断新动能组成,测度数据要素发挥的新动能大小。

### 三、 数据生产要素以一种新动能的方式对经济增长产生了重要推动作用

---

#### (一) 新动能、新动能指数的定义

关于新动能的定义,国家统计局认为以新产业、新业态、新模式为代表的新经济就是新动能。国内研究学者认为知识、技术、信息、数据是新动能的支撑要素,新动能代表的是能够推动经济开启新一轮增长进程的新要素以及新要素之间的组合方式,并判断出未来经济增长新动能的来源,它将更多的来源于新一代数字或信息技术对经济社会发展的全面覆盖、渗透与改造。根据国家发改委在 2018 年发表的《中国产业发展报告 2018》相关文件,新动能主要来自科技创新、需求变化、制度变革三个方面,它是以新的生产要素作为支撑。

关于新动能的统计测度,已有文献主要分为两大类:一是构建指标体系进行综合评价。国家统计局于 2017 年 9 月首次对外公布经济发展的新动能指数,

它是由不同指数构成的，其中贡献构成从小到大排序主要包括知识活力指数、转型升级指数、创新驱动指数、经济活力指数、网络经济指数。二是将全要素生产率视为新动能进行计量测度。以 TFP 刻画的所有要素投入组合的产出效率可以较为客观准确的衡量经济增长的质量，一般被视作经济增长的新动能之一，因此，从经济增长的动力来源出发，将资本、劳动视为传统动能，全要素生产率视为新动能，进行测度。<sup>[2]</sup>两种方法都具有重要的参考意义，但同时具有一定的缺陷。

关于新动能的本质和特征，柴士进等在中国经济增长新动能的统计测度研究中提到新动能的前提在于“新”，新就是指创新；核心在于“动”，动就是指动力；目的在于“能”，能就是能量。此外文章中还指出新动能的构成要素以及指标评价体系和新动能指数测算方法。<sup>[2]</sup>

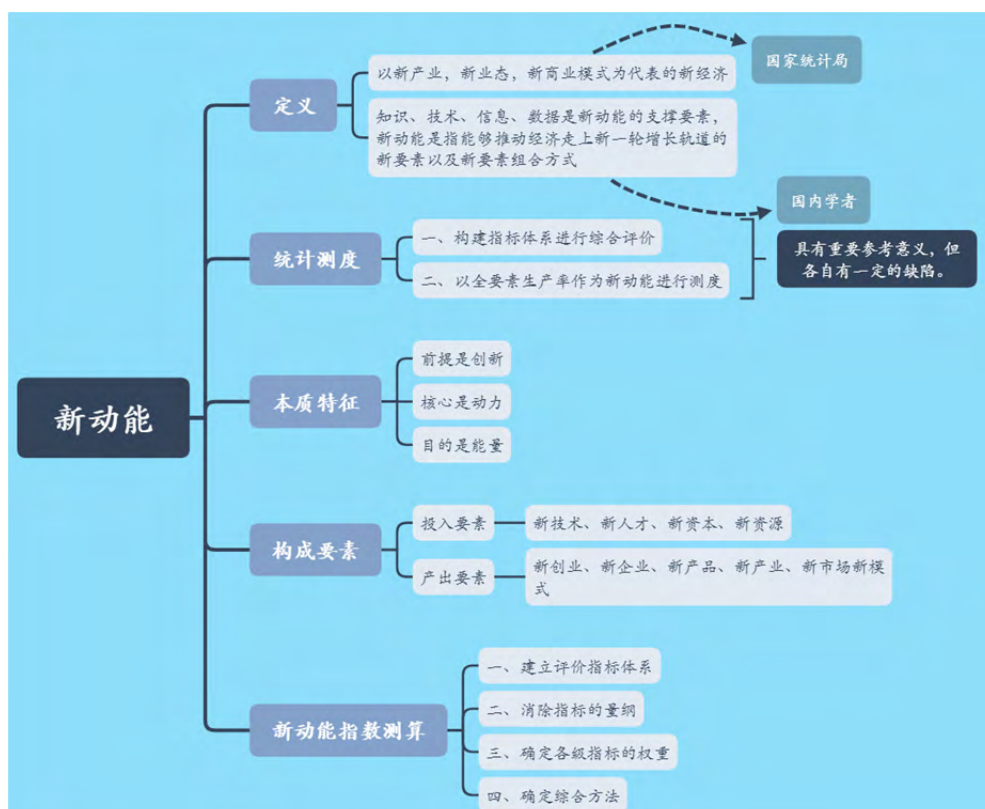


图 3 新动能定义及构成要素

## (二) 数据演化成生产要素的理论依据

数据演化为关键生产要素符合马克思对生产力内涵的阐述。在“农业经济—工业经济—数字经济”的发展过程中，面临着持续增长且不断高级化的需求发展，除劳动力以外的关键生产要素大体经历“土地—资本—企业家才能—数据”的演化。在农业经济时代，土地生产过程中具备极强的竞用性。进入工业经济时代，能够部分打破自然资源约束、有效扩大再生产的资本（生产性机器、设备）应运而生。资本能够有效地增强与劳动力的结合水平。进入工业化后期，企业家才能成为关键生产要素。企业家才能进一步降低了生产要素的竞用性，同时也提升了与其他生产要素的结合程度。

进入数字化经济时代，数据产生和流通的速度前所未有，整个经济系统呈现出极大的复杂性，供应链体系逐渐趋于数字化、智能化。企业家才能须通过数据，处理得到真实信息和新的知识，才能做出精准的生产决策来应对经济的不确定性。数据因其非竞用性和低损失率极大提高了知识生产效率，带来经济数量和质量上的增长。数据流通带来真实信息的互联又可以降低信息的不对称进而缓解市场失灵问题。<sup>[1]</sup>

同时，数据的生产要素属性是建立在数字技术进步的客观事实上的。生产部分只有基于算法、算力等数字技术普惠的客观条件下，才具备经济动机利用数据生产价值的能力。其次，生产要素的演进是消费需求引致的，而需求根植于人类价值观的发展水平。人类价值观的动态变化是促成数据发展成为生产要素的关键原因。原始生产要素不能有效满足消费者需求的动态变化，而数据可以做到对消费者价值观的精准映射。与工业时代仅满足人类的物质消费需求不同，数字经济时代更多地追求物质和精神财富的高水平发展。

### (三) 数据对经济增长的推动作用

数据对经济增长的贡献体现在直接生产过程和流通过程上。<sup>[3]</sup>在直接生产过程中，数据要素可以缩短生产时间。第一，数据生产要素可以有效缩短劳动生产时间，它的使用可以实现在相同劳动与生产资料结合时的时间段内获得更高的产量。可实现在较少的劳动者和生产资料结合的劳动时间内生产相同产量。第二，数据生产要素可以缩短生产资料准备时间，通过使用数据信息，生产者可以有效地记录一系列的时间信息，缩短生产资料准备时间，从而提高生产过程中的运行效率。第三，数据生产要素可以缩短劳动过程中中断的时间。此外，数据要素可以降低生产成本，实现相同成本创造更多价值。第一，它可以减少与劳动力相结合的生产资料的损耗。第二，可以降低原材料库存率以降低企业的管理成本。第三，可以减少因设备故障而使劳动过程中断造成的损失。

在流通过程中，数据要素提高了整个流通过程的运行效率，并降低了流通成本。数据生产要素的使用可以缩短购买和销售过程中消耗的时间，减少流通时间，从而加快资本的循环速度，进而创造出更多的价值。其次，在购买和售卖阶段中，数据生产要素可以有效地流通成本。购买过程中，数据可以降低搜索和匹配成本；售卖阶段中可以降低生产者的匹配成本。数据要素对经济增长过程中的贡献形式如图 4 所示。



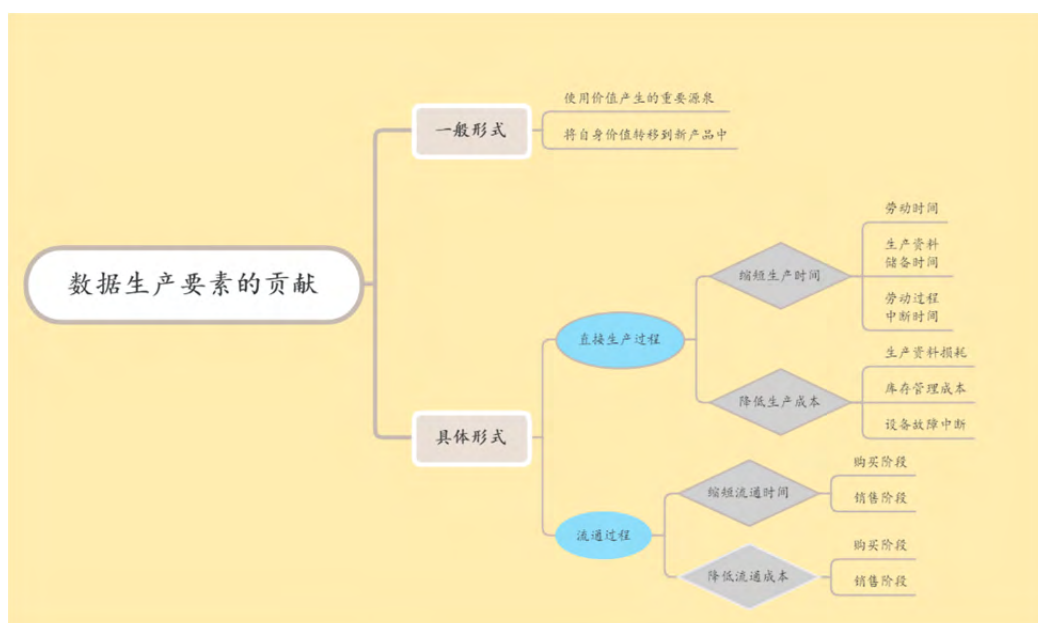


图 4 数据要素对经济增长贡献形式

最后，数据在生产力的作用机制中也发挥了推动作用。<sup>[1]</sup>数据不能直接形成生产力，而需要通过“数据—信息—知识”链转换之后才能被我们利用，从而形成价值，满足人类的需求。数据形成精神生产力，推动物质生产力向高级化发展。<sup>[1]</sup>作用机制如图 5 所示。

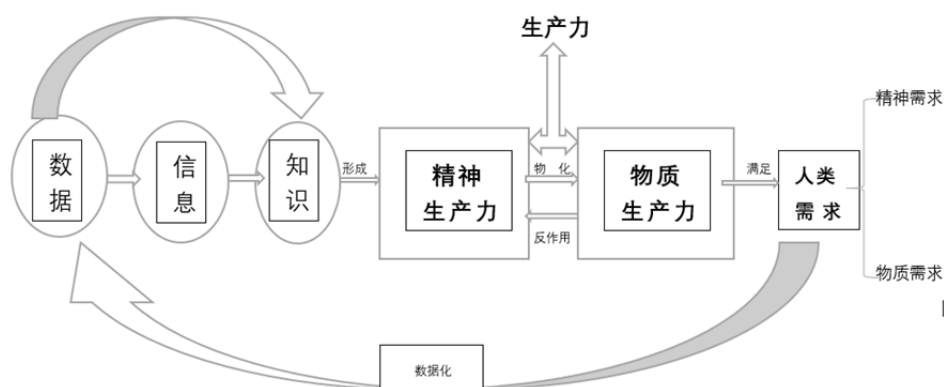


图 5 数据在生产力中的作用机制

## 四、模型的构建

### （一）数据生产要素对经济增长的贡献率——生产函数模型

关于数据要素，国内学者对其存在诸多争议，这让数据生产要素的价值难

以准确测度，所以我们无法对数据进行直接测度，采取间接测度的方式。

假设经济的增长满足道格拉斯生产函数形式

$$Y = k^{\alpha} l^{\beta} s^{\gamma} d^{\mu} \quad \text{公式(6)}$$

其中Y为总产出，这里用 GDP 衡量； $k, l, s, d$ 分别表示资本、劳动力、科技、数据生产要素。 $k, l, s$ 分别用固定资产投资额，劳动工资总额，科技专利授权数来衡量。对（6）式两边取对数得

$$\ln y = \alpha \ln k + \beta \ln l + \gamma \ln s + \mu \ln d \quad \text{公式(7)}$$

对两边同时关于时间求导，进行相关处理后，可得

$$1 = \frac{\dot{k}/k}{\dot{y}/y} + \frac{\dot{l}/l}{\dot{y}/y} + \frac{\dot{s}/s}{\dot{y}/y} + \mu \frac{\dot{d}/d}{\dot{y}/y} \quad \text{公式(8)}$$

对于贡献率  $\mu$  的测算可以通过  $1 - \quad - \quad -$  来间接测算。

这里本文将数字经济较为发达的浙江省作为研究对象，然后以此类推分析全国的数字经济对经济增长的贡献。借助 Eviews 经济数据分析软件进行回归分析，得到各个参数近年来的平均数值如表 1 所示。数据来源于《2020 年浙江统计年鉴》2016-2019 年各个指标的统计数据。计算得到数据生产要素对经济增长的平均贡献率  $\mu$  约为 19.2%，可以看出数据要素对经济增长的贡献较大，并根据前文的推测，这一贡献在未来可能将逐渐加大。

表 1 各变量参数估计结果

Dependent Variable:	Y
Method:	Least Squares
Date:	05/18/21
Time:	15:03
Sample:	2016 2019
Included observations:	4

Variable	Coefficient	Std.Error	t-Statistic	Prob.
X1	0.632689	0.341260	1.853979	0.3149
X2	0.170523	0.567325	0.300574	0.8141
X3	0.388403	0.228463	1.700070	0.3385
R-squared	0.997773	Mean dependent var	231658.0	
Adjusted R-squared	0.993319	S.D. dependent var	74359.76	
S.E. of regression	0.009880	Akaike info criterion	22.14836	
Sum squared resid	9.76E-05	Schwarz criterion	22.18808	
Log likelihood	15.56600	Hannan-Quinn criter.	21.88046	
Durbin-Watson stat	2.291280			

## (二) 数据要素相对其他要素的倍增作用——多元线性回归模型

本文选取总产出（GDP）作为因变量，并选取高新技术产业企业数(个)、工业企业数(个)、互联网企业数(个)三个变量作为解释变量，运用最小二乘的方法，带入 2008-2015 年的年度数据，在 Eviews 经济数据分析软件上进行模型拟合，得到结果如表 2 所示，

表 2 线性回归模型拟合结果

Dependent Variable:		Y		
Method:		Least Squares		
Date:		05/18/21		
Time:		18:51		
Sample:		1 8		
Included observations:		8		
Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	-37208.45	88257.47	-0.421590	0.6950
X1	2.636920	13.15838	0.200398	0.8509
X2	-0.050223	0.816360	-0.061520	0.9539
X3	8.141070	3.612032	2.253875	0.0873
R-squared	0.981488	Mean dependent var		231658.0
Adjusted R-squared	0.967604	S.D. dependent var		74359.76
S.E. of regression	13384.00	Akaike info criterion		22.14836
Sum squared resid	7.17E+08	Schwarz criterion		22.18808
Log likelihood	-84.59344	Hannan-Quinn criter.		21.88046
F-statistic	70.69125	Durbin-Watson stat		1.947519
Prob(F-statistic)	0.000639			

得到的模型如下：

$$Y=-37208.45+2.636920X_1-0.050223X_2+8.141070X_3$$

$$F=70.69125$$

$$R^2=0.981488$$

其中 $R^2$ 约为 0.98，表明 $Y$ 与 $X_1, X_2, X_3$ 具有高度的相关性，同时说明该模型的拟合度较高，模型较为理想，具有较强的适用性。

根据上述模型，变量 $X_2$ 的系数为负数，表明总产出与工业企业的数量之间呈现负相关关系，即生产中增加工业企业的数量反而会使得 GDP 下降，这可以充分解释为什么建造工业企业带来的 GDP 增长比建造其他任何数据型企业更低，同时也表明了数据生产要素具有的增长效应比传统的劳动力资本两要素相结合要素组合方式所具有的增长效应高。比较 $X_1$ 和 $X_3$ 两个变量的系数，以数据生产要素为支撑的互联网企业数量的增长比高新技术企业数量的增长能贡献更多的 GDP 增长，这又一次凸显出数据这种新型生产要素对于经济增长的推动作用。

### (三) 数字经济对中国经济增长的推动作用——ARIMA 模型

#### 1. 原始数据的处理

本文选取 1990-2019 年的中国 GDP 数据作为研究对象，由于这段时间，中国的 GDP 变化呈现指数增长态势，为了稳定序列数据，对所选取的 GDP 数据作取对数处理。得到结果如图 6 所示。

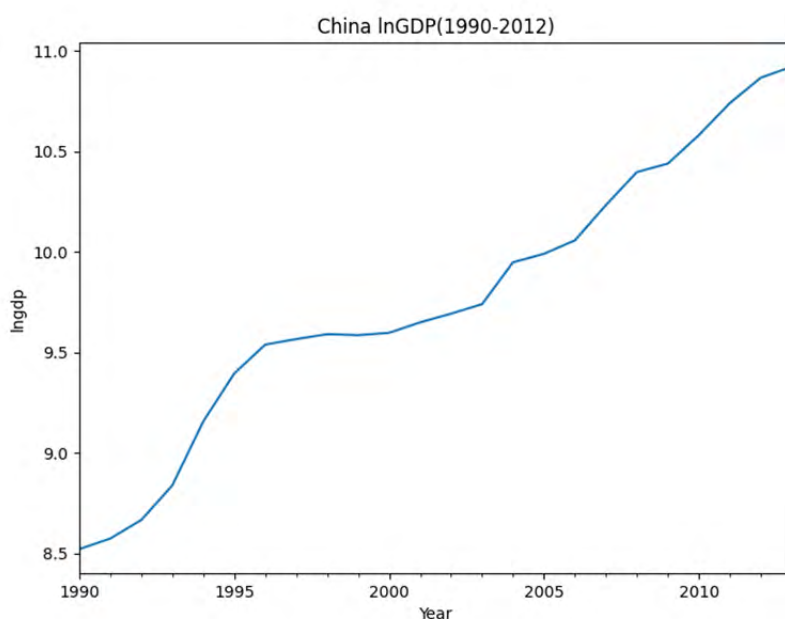


图 6 处理后的 GDP 时间序列数据

## 2. 序列平稳性检验

由上图可知，取对数后的 GDP 时间序列数据具有线性趋势。对该序列进行 ADF 检验，得到检验结果如下。

表 3 数据序列进行 ADF 检验结果

	value
Test Statistic Value	-0.63378
p-value	0.863165
Lags Used	9

Number of Observations Used	13
Critical Value(1%)	-4.06885
Critical Value(5%)	-3.12715
Critical Value(10%)	-2.70173

从表中可以看出，检验值为 3.5877，大于 5%和 10%的临界值，接近于 1%的临界值，可以认为该序列存在单位根，因而该序列是不稳定的。为了消除这一增长趋势，本文对这组数据进行一阶差分，拆分结果如表 4 所示。

表 4 数据一阶拆分结果

差分后的检验值为-0.63378，小于各个水平的临界值，由此可以判断差分

	value
Test Statistic Value	3.5877
p-value	0.998261
Lags Used	9
Number of Observations Used	14
Critical Value(1%)	-4.01203
Critical Value(5%)	-3.10418
Critical Value(10%)	-2.69099

后的序列为稳定的序列。

### 3. 模型建立和参数的估计

参数的确定。根据时间序列的识别规则，采用自相关分析图（ACF 图）、

偏相关分析图（PAC 图），AIC 准则（赤道信息量准则）和 BIC 准则（贝叶斯准则）相结合的方式来确定 ARMA 模型的阶数，根据模型的构建原则，本文选取使得 AIC 和 BIC 值达到最小的那一组作为模型的理想阶数。自相关和偏相关性检验如图 7 所示。

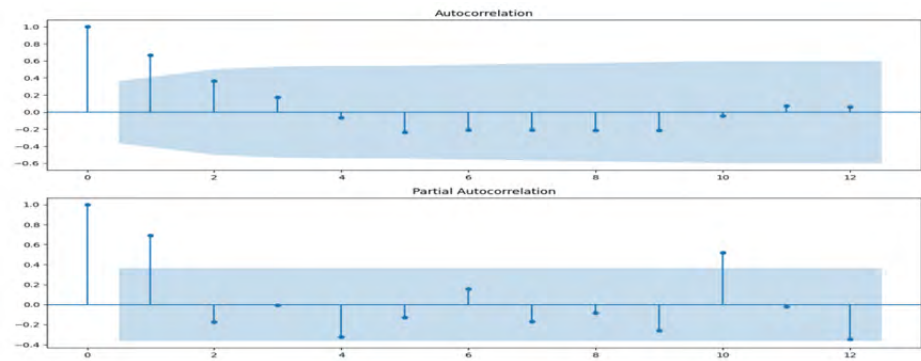


图 6 自相关和偏相关检验结果

根据分析可知，所得的自相关图与偏自相关图均呈现拖尾的分布状态，因而采用 ARMA 模型进行分析预测是合理的。选取的原始 GDP 数据经过一重差分得到平稳序列，所以参数 d 的取值为 1；通过模型的验证调整参数 p 和 q 的值使得 AIC 和 BIC 取值最小，参数调整结果如表 5 所示。

表 5 参数调整结果

Results: ARMA			
Model:	ARMA	BIC:	-44.9756
Dependent Variable:	Added value of primary industry	Log-Likelihood:	30.327
Date:	2021-05-29 15:33	Scale:	1.0000
No. Observations:	23	Method:	mle
Df Model:	4	Sample:	12-31-1991
Df Residuals:	19		12-31-2013



Converged:	1.0000			S.D. of innovations:	0.061	
No. Iterations:	54.0000			HQIC:	-49.225	
AIC:	-50.6530					
	Coef.	Std.Err.	t	P> t	[0.025	0.975
					]	
const	0.0950	0.0066	14.3774	0.0000	0.0821	0.1080
ar.L1.Added value of primary industry	1.3382	0.1568	8.5330	0.0000	1.0308	1.6455
ar.L2.Added value of primary industry	-0.6083	0.1658	-3.6702	0.0002	-0.9332	-0.2835
ma.L1.Added value of primary industry	-0.9999	0.1206	-8.2927	0.0000	-1.2362	-0.7636
	Real	Imaginary	Modulus	Frequency		
AR.1	1.0998	-0.6589	1.2821	-0.0859		
AR.2	1.0998	0.6589	1.2821	0.0859		
MA.1	1.0001	0.0000	1.0001	0.0000		

通过参数的确定，可以得到合适的  $p, d, q$  参数为 2, 1, 1。因此可以构建的模型为 ARIMA (2, 1, 1)。

#### 4. 模型检验

模型的检验主要分为两部分：一是残差的正态性检验，二是残差的自相关检验。残差需满足正态性，主要是为了使残差集中于某一个数值，如果该值与 0 很接近，则它实际服从均值为 0 的正态分布，即它是一个白噪声。白噪声在物理学上的定义是指功率谱密度在整个频域内均匀分布的噪声。白噪声或白杂讯，是一种功率频谱密度为常数的随机信号或随机过程。换言之，此信号在各个频段上的功率是一样的，此信号由于具有平坦功率谱的性质，因而被称作“白色的”，这是因为白光是由各种频率的单色光混合而成，这种信号也因此被称为白噪声。当残差为白噪声序列，则说明时间序列中有用的信息已经被提取，剩下都是随机扰动，无法被预测和使用。此外，残差还需要满足非自相关性，这主要是为了确保残差中不再包括 AR（自回归）或者 MA（移动回归）过程产生的序列。

自相关检验。对模型的残差进行自相关检验，得到检验结果如图 8 所示。分析可知，该序列具有无自相关的性质。

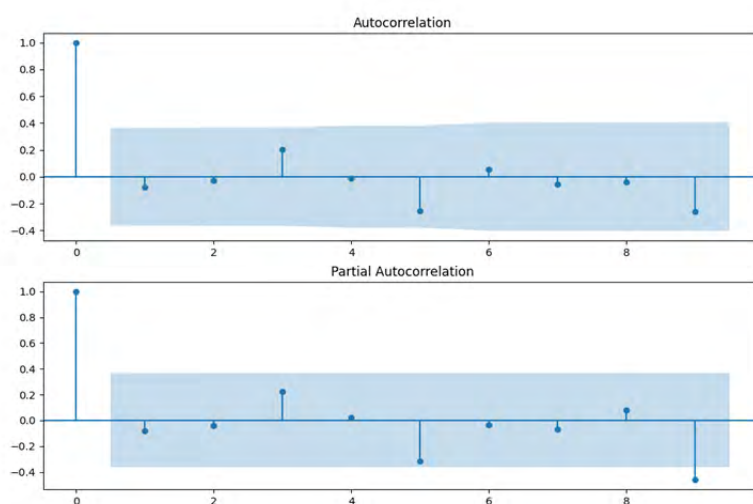


图 7 残差的自相关性检验

正态性检验。检验模型的残差是否服从正态分布，适合使用 QQ 图分析工

具。QQ 图通常用于检验某组数据是否来自某个分布总体，被广泛应用于正态分布的检验。对模型的数据分布进行 QQ 图检验，得到结果如图 9 所示。从图中可以看出，检验的数据在 QQ 图参考线两侧随机分布，只有少部分数据偏离直线较远，因此可以大致认为，残差服从正态分布。

经过一系列检验，可以判断，残差为白噪声序列，表明本文所建立的模型比较理想。

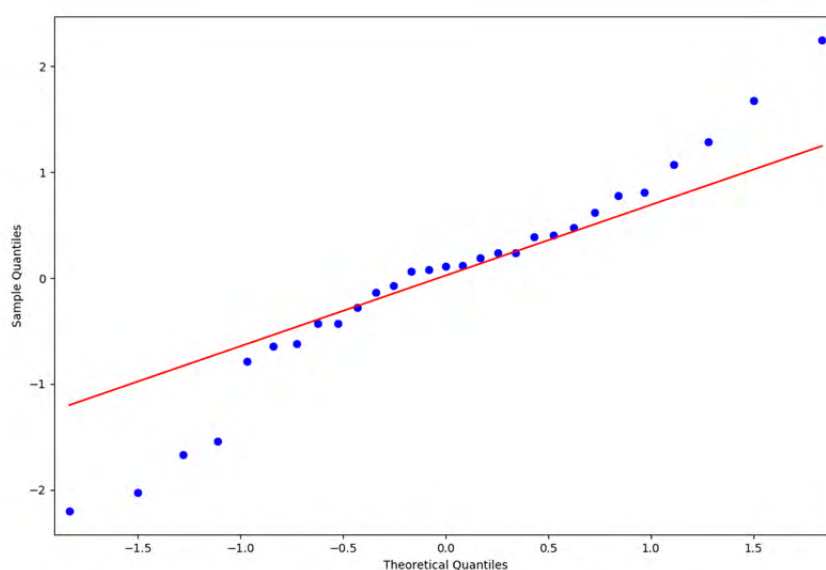


图 8 残差的正态性检验 (QQ 图)

## 五、 模型的应用

### (一) 数据的来源

本文所采用的数据均来自《中国统计年鉴》<sup>[7]</sup>、中国信息通信研究院，通过查找并获取 1990-2020 年 GDP 数据、一二三产业增加值来进行模型分析以及相关预测。

## (二) GDP 预测

根据国家有关政策和报道，数字经济的发展大致可以分为三个阶段：第一阶段：萌芽期（1994-2002 年），第二阶段：高速发展期（2003-2012 年），第三阶段：成熟期（2013 年-至今）。本文根据不同时期的 GDP 数据来预测各时期未来两年三次产业的 GDP 增长情况，并与真实值进行比较，来直观反映不同时期数字经济对于 GDP 的推动作用以及数字经济对三次产业发展的影响差异。

假设数字经济一直处于萌芽期，那么它对于 GDP 的推动作用并不明显。采用 1990-2018 的年度 GDP 数据预测 2019 和 2020 的 GDP 数据，分析数字经济的推动作用，分析结果如表 6 所示。

表 6 2019 年和 2020 年的各产业的预测值和真实值

时间	预测值（亿元）			真实值（亿元）		
	第一产业	第二产业	第三产业	第一产业	第二产业	第三产业
2020	83698.0	437176.0	584957.0	77754.0	384255.0	553977.0
2019	73083.3	400112.5	539074.6	70467.0	386165.3	534233.1

从表中数据可以看出，该模型对于 2019 年的各产业预测值都接近于真实值，但是 2020 年的预测值与真实值相差较大，究其原因，可能是 2020 年初突发新冠疫情导致经济大幅度衰退。同时可以发现，该模型较好的预测了各个产业的生产总值。

假设数字经济一直处于高速发展阶段，发展速度飞快，因此预期对经济增长的推动作用也十分明显。采用 1990-2012 年的数据作为训练数据，模拟中国数字经济一直处于高速发展阶段的 GDP 发展，并与实际值进行比较从而说明数字经济的推动作用，并验证数据生产要素是经济增长的新动能。所得结果如

表 7 所示。

表 7 2014 年和 2015 年各产业的预测值和真实值

时间 (年)	预测值 (亿元)			真实值 (亿元)		
	第一产业	第二产业	第三产业	第一产业	第二产业	第三产业
2015	53558.0	349010.7	404094.0	60863.0	282040.3	346149.7
2014	50249.0	299310.4	341603.0	58332.0	277571.8	308058.6

根据表中结果,2014 年、2015 年第一产业的预测均低于实际值,而第二三产业的预测值都要高于实际值,这表明数字经济的飞速发展会带来以数据要素为支撑的第二三产业的飞速发展,而相应的以传统的劳动、资本要素结合的第一产业会随着数字经济的飞速发展而衰退。

各个产业收益率分析。设置指标收益率为实际第三产业 GDP 与预测第三产业 GDP 的差值/实际第三产业 GDP ;总收益率为预测累加值与实际累加值的差值/实际累加值。

$$\text{收益率} = \frac{\text{预测 GDP} - \text{实际 GDP}}{\text{预测 GDP}},$$

$$\text{总收益率} = \frac{\text{预测累加值} - \text{实际累加值}}{\text{实际累加值}}。$$

所求的收益率如表 8 所示。结果表明,2014 年以及 2015 年数据要素投入带来第一产业的产出或收益率均为负;而给第二三产业带来的收益率均为正,且 2015 年的收益率都要显著高于 2014 年。此时,虽然各大产业的收益率存在巨大差异,但是仍然无法确实这是数据要素以新动能的方式发挥的推动作用。

表 8 2014 年和 2015 年 GDP 和三大产业收益率

时间（年）	GDP	第一产业	第二产业	第三产业
2014	1.0	-16.0	7.2	9.8
2015	19.8	-13.6	19.1	14.3
总收益率	20.8	-29.6	26.3	24.1

数字经济在三大产业之间的融合度分析。为验证各个产业的增长或下降与数字经济发展的联系，以更加充分的证明数字经济对经济增长的推动作用。本文在已有分析基础上，进行了三大产业与数字经济的融合度分析，分析结果如表 9 所示。

表 9 2016-2019 年数字经济在三大产业之间的融合度

年份	2016 年	2017 年	2018 年	2019 年
第一产业	6.2	6.5	7.3	8.2
第二产业	16.8	17.2	18.3	19.5
第三产业	29.6	32.6	35.9	37.8

将各产业增加值和各产业与数字经济融合度的关系用散点图表示，如图 10 所示。

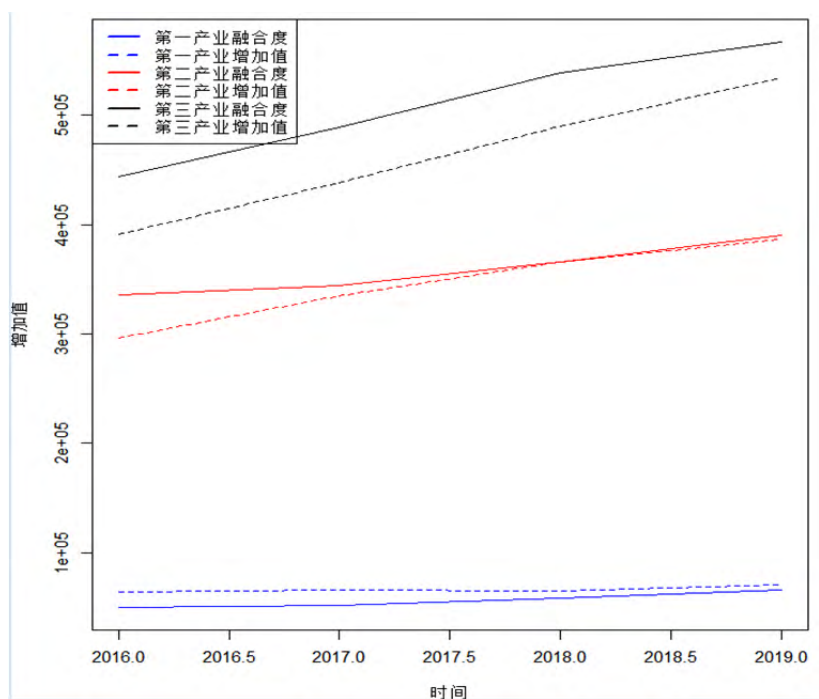


图 9 三大产业增加值和融合度随时间变化

从上表和图中可以看出，产业增加值与数字经济和产业融合度密切相关，呈正相关关系（融合度经过处理使得数量级上贴近产业增加值，方便比较两者趋势）。其中，产业融合度最高的第三产业的增加值最高，所以提高一二三产业与数字经济融合度有利于 GDP 的增长和经济的发展。

由于第一产业和数字经济融合度较低，可以判断：当数字经济发展增速减缓时，对于第一产业的产值增加并无显著影响，因此第一产业预测值低于真实值；然而由于第二产业和第三产业与数字经济融合度远高于第一产业，因此，当数字经济发展增速减缓，对于二三产业 GDP 增加值有负面影响，正因如此才说明迫切需要数据的支持、技术进步促使数字经济得到突破性发展，从而提高经济的活力，拉动 GDP 的增长。

## 六、 总结与建议

---

### (一) 结论

数据生产要素对经济增长具有突出贡献。数据作为一种新型生产要素，在推动经济增长发挥的作用要比传统的资本、劳动、土地等生产要素更加显著，并在未来的经济增长贡献中更具优势，继续发挥出巨大的力量，呈现出一种快速而稳定的增长态势。

数据对经济增长的倍增效应显著。本文在测度数据要素的倍增效应时，采用多元回归模型，对传统生产要素投入产出与数据生产要素的投入产出进行回归，发现数据要素的产出投入比要远高于其他传统生产要素，充分体现了数据要素相较于其他要素对经济增长的显著倍增效应。

数字经济的发展对经济增长的影响显著。数字经济以数字技术和海量数据为依托决定了数字经济的生产要素属性。在数据的赋能下，数字经济的发展有效地提高了经济的活力，释放了更广阔的发展空间，并加速了经济运行的效率，从而推动了我国经济增长，影响经济在质量和数量上的发展。

三大产业和数字经济的融合度决定了数字经济发展对各大产业的推动作用大小。本文从三大产业随着数字经济发展的增长情况出发，研究了数字经济对各大产业的推动作用差异，进而佐证数据生产要素对经济增长的强大推动作用以及要素投入产出效应显著高于其他传统要素。结果表明，第二三产业与数字经济的融合度要显著高于第一产业，而第二三产业随着数字经济的发展出现显著的增长。这充分验证了我们的结论，与数字经济融合度高的产业在如今数字经济高速发展的背景下，会有更大的发展潜力和更乐观的发展前景。

### (二) 数字经济以及数据要素发展相关建议

在大数据时代背景下，数据生产要素所发挥的作用正在逐渐变大，对经济



增长的影响也越来越强。我们需采取一些必要的措施，以便充分利用数据生产要素对我国经济发展的推动作用。

一是需要深刻认识到数据要素的重要作用。在数字时代大背景以及国家政策导向下，数据生产作用所发挥的作用愈加凸显，在各行业、各产业都应该及时贯彻落实国家在相关领域的政策部署，有效并充分的挖掘数据的价值，发挥数据的巨大潜力。

二是扎实推进数据与其他要素相结合的进程，充分发挥出数据的倍增效应。各种生产要素并不是独立发挥作用并创造价值的，数据也不例外，只有充分将数据与其他生产要素相结合，才能产生出“ $1+1>2$ ”的效应，从而更加高效地推动中国经济的发展。

三是进行数据生产要素的深度共享与融合。通过打通数据之间的相互联系，形成数据共享机制；推动社会、政府、不同企业、不同行业或不同领域之间的数据共享交换，促进数据有效地、充分地流动。另外，必须推动产业与数据的深度融合，促进数据产业化、产业数据化，通过数据来加速传统企业的转型升级和新产业的萌芽。

四是加紧完善数据生产要素市场分配机制。数据作为一种新型生产要素参与了生产过程，为生产做出了贡献，因此应该确保要素报酬分配的公平公正。这需要保证数据按照市场机制公平定制，从而保障数据的合法自由流通。

五是应加强数据生产要素领域的安全监管。数据的各种违规滥用和不法泄漏，容易对经济社会健康发展产生不良影响。必须通过完善数据的治理体系，建立健全相关的法律法规，对经济社会发展各领域的数据存储、流通和使用进行全面规范和有效监管。<sup>[5]</sup>

数字经济的发展也和经济增长息息相关，关于数字经济能否成为促进我国经济发展的新动能，本文给出如下参考建议。

首先，应该加大数字基础设施供给，提高数字技术的研发强度。数字经济以数字技术为基础，增加数字设施的供给、创新数字技术有利于数字经济的横向和纵向发展。一方面，通过构建大数据服务中心、数字通讯基站、国家重点实验室、建立数据创新创业中心等手段加强技术设施建设，为数字经济的快速发展提供必要的硬件支撑；另一方面，需加强云计算、物联网、大数据等基础性与公共适用性技术的研发强度，提高通用性数字技术的研发成功率和成果转化率，并推进量子通信等核心领域的深入研究，逐步提高我国数字经济的软实力与综合竞争力，从而培育数字经济的新动能。

其次，调整各地区数字经济发展步伐，提高数字经济发展协调性与均衡性。一方面，中央政府通过制定相关的政策、创新相关的制度，充分发挥资源调度和人才分配的作用，构建数字经济发展城市群，依靠发达地区带动相对落后地区的数字经济发展，以稳步实现数字经济发展的均衡性。另一方面，地方政府应加强区域的整体联动性，促进各区域之间的技术交流，同时贯彻并落实落实国家的有关政策，技术发达地区之间应倡导合作竞争，带动数字经济领域的创新，成为数字经济发展的领导者。落后地区应积极打破沟通壁垒，加强与发达地区之间的技术创新交流，学习相关知、制度和管理技术。

最后，应提升数字经济与教育科研领域的融合力度，促进数字经济与人力资本要素的高度结合，发挥出更大的发展效应。第一，可以开设数字经济领域有关课程和讲座，为在校学生提供关于数字领域的前沿知识和信息，提高学生群体的数字认识和应用能力。第二，数字经济企业应加强和高校科研团队合作，借助企业、高校的人力资本，充分加快数字经济领域和数字技术的创新进程。第三，完善数字经济领域的创新创业鼓励制度，加大对数字经济创新创业者的财政支持力度，从而拓宽数字经济在我国国民的覆盖范围，并促进数字主体多元化。<sup>[9]</sup>

## 七、 创新与不足

---

### (一) 创新

本文为测度数据生产要素以及数字经济的发展对经济增长的贡献并论证它们对经济增长的巨大推动作用，分别采用了经济学领域中的生产函数模型、计量分析领域中的多元线性回归模型以及统计分析领域中的 AIRMA 三种分析模型，每个模型之间相互联系，所得出的结论相互印证。分析方法更加多元，更加具体，更加全面，研究所得出的结论也更具说服力，并给出了合理的建议。此外，本文从三大产业的增长情况出发，寻找它们与数字经济发展之间的联系，无论是从正面，还是侧面，都很好的证明了我们研究前的假设和分析所得的结论。

### (二) 不足

本文对经济增长新动能的测度上没有建构出一套完整的指标体系，因此没有很好地完成数据要素对经济增长产生的新动能的测算。同时，数据的选择也缺乏一定的严谨性，预测出来的数据失去了部分精确程度。对问题的研究缺乏一定的深度和广度，对数据生产要素属性的分析并不全面。

## 参考文献

- [1] 戚聿东,刘欢欢.数字经济下数据的生产要素属性及其市场化配置机制研究[J].经济纵横,2020(11):63-76+2.
- [2] 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.
- [3] 王胜利,樊悦.论数据生产要素对经济增长的贡献[J].上海经济研究,2020(07):32-39+117.
- [4] 熙浅.时间序列模型(ARIMA)<https://www.jianshu.com/p/4130bac8ebec>.
- [5] 崔俊富,陈金伟.数据生产要素对中国经济增长的贡献研究[J].管理现代化,2021,41(02):32-34.
- [6] 李天明,田晓艳,占东明.财务视角下新冠肺炎疫情对贵州工业发展影响的统计测度[A].中国统计教育学会.2020年(第七届)全国大学生统计建模大赛优秀论文集[C].中国统计教育学会:中国统计教育学会,2020:25.
- [7] 国家统计局.中国统计年鉴[M]北京:中国统计出版社,2020.
- [8] 叶一标,陈海丹,王洵.重大突发事件对福建经济发展的影响统计测度——以新冠肺炎疫情为例[A].中国统计教育学会.2020年(第七届)全国大学生统计建模大赛优秀论文集[C].中国统计教育学会:中国统计教育学会,2020:19.
- [9] 张腾,蒋伏心,韦朕韬.数字经济能否成为促进我国经济高质量发展的新动能?[J].经济问题探索,2021(01):25-39.
- [10] 郭家堂,骆品亮.互联网对中国全要素生产率有促进作用吗?[J].管理世界,2016(10):34-49.
- [11] 冯云廷.如何认识数据作为生产要素的经济价值[J].国家治理,2020(38):40-43.
- [12] <https://baike.baidu.com/item/%E5%9B%9E%E5%BD%92%E6%A8%A1%E5%9E%8B>
- [13] Honohan, P. Financial Development, Growth and Poverty: How Close are the Links? [M]. Palgrave Macmillan, London, 2004: 1-37.
- [14] Campbell, F. Will Europe regulate over the top services on the mobile internet? [N]. United Liberty, 2013.
- [15] Yousaf Zahid, Radulescu Magdalena, Sinisi Crenguta Ileana, Serbanescu Luminita, Păunescu Loredana Maria. Towards Sustainable Digital Innovation of SMEs from the Developing Countries in the Context of the Digital Economy and Frugal Environment[J]. Sustainability, 2021, 13(10).

## 附录

---

本文代码：

### (1) R 代码

```
data1<-read.csv("C:/Users/tianyu/Desktop/统计建模/产业增加值与产业融合度.csv")

par(mai=c(0.7,0.7,0.1,0.1),cex=0.9)

gdp1=ts(data1$第一产业融合度,start = 2016,frequency = 1)
gdps1=ts(data1$第一产业增加值,start = 2016,frequency = 1)
gdp2=ts(data1$第二产业融合度,start = 2016,frequency = 1)
gdps2=ts(data1$第二产业增加值,start = 2016,frequency = 1)
gdp3=ts(data1$第三产业融合度,start = 2016,frequency = 1)
gdps3=ts(data1$第三产业增加值,start = 2016,frequency = 1)

ts.plot(gdp1*0.8e+04,gdp2*2e+04,gdp3*1.5e+04,gdps1,gdps2,gdps3,
col=c("blue","red","black","blue","red","black"),lty=c(1,1,1,2,2,2),xlab="
时间",ylab="增加值")

legend(x="topleft",legend=c("第一产业融合度","第一产业增加值","第二
产业融合度","第二产业增加值","第三产业融合度","第三产业增加值
"),col=c("blue","blue","red","red","black","black"),lty=c(1,2,1,2,1,2),lwd
=2)
```

### (2) Python 代码

```
import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt
```

```

from scipy import stats

import matplotlib.pyplot as plt

import statsmodels.api as sm

from statsmodels.graphics.tsaplots import acf, pacf, plot_acf, plot_pacf

from statsmodels.tsa.arima_model import ARMA

GDP = pd.read_csv('C:\\Users\\tianyu\\Desktop\\统计建模\\other
gdp.csv', index_col='date')

GDP.index =
pd.Index(sm.tsa.dates_from_range('1990','2020'))

n_sample = GDP.shape[0]

n=7

n_train = n_sample-n

ts_train = GDP.iloc[:n_train]['Added value of primary industry']

ts_train.plot(figsize = (12,8))

if n_train<10:

    plt.title("China GDP(1990-20"+"0"+str(n_train-12)+")")

else:

    plt.title("China GDP(1990-20"+str(n_train-12)+")")

plt.legend(bbox_to_anchor = (1.25,0.5))

plt.xlabel("year")

plt.ylabel("gdp")

sns.despine()

plt.show()

ts_train = np.log(ts_train)

```

```

logtrain=np.log(GDP.iloc[:,1]['Added value of primary industry'])

ts_train.plot(figsize=(8,6))

if n_train<10:

    plt.title("China lnGDP(1990-20"+"0"+str(n_train-12)+"")

else:

    plt.title("China lnGDP(1990-20"+str(n_train-12)+"")

plt.xlabel("Year")

plt.ylabel("lngdp")

plt.show()

t=sm.tsa.stattools.adfuller(ts_train, )

output=pd.DataFrame(index=['Test Statistic Value', 'p-value', 'Lags
Used', 'Number of Observations Used',"Critical Value(1%)","Critical
Value(5%)","Critical Value(10%)"],columns=['value'])

output['value']['Test Statistic Value'] = t[0]

output['value']['p-value'] = t[1]

output['value']['Lags Used'] = t[2]

output['value']['Number of Observations Used'] = t[3]

output['value']['Critical Value(1%)'] = t[4]['1%']

output['value']['Critical Value(5%)'] = t[4]['5%']

output['value']['Critical Value(10%)'] = t[4]['10%']

print(output)

ts_train=ts_train.diff(1)

ts_train = ts_train.dropna(how=any)

ts_train.plot(figsize=(8,6))

```

```

plt.show()

t=sm.tsa.stattools.adfuller(ts_train)

output=pd.DataFrame(index=['Test Statistic Value', 'p-value', 'Lags
Used', 'Number of Observations Used',"Critical Value(1%)","Critical
Value(5%)","Critical Value(10%)"],columns=['value'])

output['value']['Test Statistic Value'] = t[0]

output['value']['p-value'] = t[1]

output['value']['Lags Used'] = t[2]

output['value']['Number of Observations Used'] = t[3]

output['value']['Critical Value(1%)'] = t[4]['1%']

output['value']['Critical Value(5%)'] = t[4]['5%']

output['value']['Critical Value(10%)'] = t[4]['10%']

print(output)

fig = plt.figure(figsize=(12,8))

ax1 = fig.add_subplot(211)

fig = sm.graphics.tsa.plot_acf(ts_train.values.squeeze(), lags=9,
ax=ax1)

ax2 = fig.add_subplot(212)

fig = sm.graphics.tsa.plot_pacf(ts_train, lags=9, ax=ax2)

ax2.xaxis.set_ticks_position('bottom')

fig.tight_layout()

plt.show()

r,rac,Q = sm.tsa.acf(ts_train, qstat=True)

prac = sm.tsa.acf(ts_train)

```



```

table_data = np.c_[range(1,len(r)), r[1:],rac,prac[1:len(rac)+1],Q]

table = pd.DataFrame(table_data, columns=['lag', "AC","Q", "PAC",
"Prob(>Q)"])

print(table)

(p,
q)

=(sm.tsa.arma_order_select_ic(ts_train,max_ar=3,max_ma=3,ic='aic')[
'aic_min_order'])

d=1

arma_mod = ARMA(ts_train,(p,d,q)).fit(dispatch=-1,method='mle')

summary = (arma_mod.summary2(alpha=.05, float_format="%.8f"))

print(summary)

resid = arma_mod.resid

fig = plt.figure(figsize=(12,8))

ax1 = fig.add_subplot(211)

fig = sm.graphics.tsa.plot_acf(resid.values.squeeze(), lags=9, ax=ax1)

ax2 = fig.add_subplot(212)

fig = sm.graphics.tsa.plot_pacf(resid, lags=9, ax=ax2)

plt.show()

fig = plt.figure(figsize=(12,8))

ax = fig.add_subplot(111)

fig = sm.qqplot(resid, line='q', ax=ax, fit=True)

plt.show()

predict1=arma_mod.predict(0,30)

for i in range(30):

```

```
logtrain[i+1]=logtrain[i]+predict1[i]  
predict3=np.exp(logtrain)  
plt.plot(predict3,label="forecast GDP")  
plt.plot(GDP,color="red",label="real GDP")  
plt.legend()  
plt.show()
```

统计软件：Python3.9.0

Eviews10

R-3.0.1

数据源：见电子版论文压缩包

## 致谢

团队经过一个月左右的不断努力，终于将这篇统计建模大赛论文完成。从最初的参赛报名、到建模培训、确定选题、讨论研究思路、学习并选择合适的模型、收集数据、到最后的论文修改、排版、定稿，都少不了团队共同进行讨论的身影，感谢团队每个成员的认真负责，感谢参赛期间团队的不懈努力、相互鼓励。

此外，感谢我们的指导老师，XXX 老师，是他及时的、认真负责地指导，帮我们指点迷津，给模型分析和数据处理进行帮助和指导，才能有一篇合格论文的呈现。感谢老师的付出，感谢老师的悉心教导。

最后，感谢大赛组委会，感谢你们给我们这样一次锻炼自己的机会，让我们利用这次机会巩固学习统计相关的知识，并将课上所学的知识应用到实际问题的解决研究中来，这是一次宝贵的机会，我们学会了很多，不仅学会了很多数学模型、统计模型，以及统计建模思想，更重要的是，学会了团队合作，并培养了团队的综合素质。