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: 南京邮电大学

论文题目:

数据要素赋能城市群一体化发展

——基于三大国家级城市群的面板数据

参赛队员:

王欣、田阿松、苏宏韬

指导老师:

范兆媛

数据要素赋能城市群一体化发展

——基于三大国家级城市群的面板数据

目录

摘要	I
一、引言	2
(一)研究背景与意义	2
(二) 研究目标	3
(三) 研究的理论基础	3
1. 数据动能测度研究的现状	3
2. 城市群一体化测度研究的现状	5
3. 文献述评	6
(四)技术路线	6
(五)研究的创新点	7
二、指标构建与测算	7
(一) 数据新动能指标构建与测算	7
1. 指标构建	7
2. 测算过程	8
3. 测算结果	9
(二)城市群一体化测度的构建与测算1	2
1. 指标构建1	2
2. 测算过程	3
3. 测算结果	4
(三) 变量选择与描述1	5

1. 控制变量选择15
2. 控制变量处理15
(四)数据来源说明16
三、模型建立与实证分析16
(一)模型设定16
(二) 建模步骤流程图17
(三) 建模过程18
1. 控制变量筛选18
2. 描述性分析19
3. 古典假设21
4. 面板单位根检验22
5. chow 检验(F 检验)22
6. Hausman 检验22
7. 基准回归结果22
8. 稳健性检验23
(四)数据新动能对城市群一体化影响的实证结果分析23
(五)数据新动能对城市经济发展模型建立24
1. 描述性分析 24
2. Chow 检验(F 检验)26
3. Hausman 检验27
4. 基准同归结果

5. 稳健性检验
(六)数据新动能对城市经济发展影响的实证结果分析28
四、结论与对策建议29
(一)结论29
(二)对策建议29
参考文献3
致谢34
图表目录
图 1-1 技术路线图
图 3-1 建模步骤流程图18
图 3-2 各城市群一体化趋势图
图 3-3 各城市群数据新动能变化趋势图20
图 3-4 CITYBELT 与 SJXDN 散点图21
图 3-5 长三角各城市经济发展变化趋势图25
图 3-6 长三角各城市数据新动能变化趋势图25
图 3-7 被解释变量与核心解释变量的散点图26
表 2-1 数据新动能测度构成指标
表 2-2 京津冀城市群数据新动能指标测算结果10
表 2-3 长三角城市群数据新动能指标测算结果1
表 2-4 珠三角城市群数据新动能指标测算结果12
表 2-5 城市群一体化测度13

表 2-6 三大城市群一体化测算结果	15
表 3-1 基本面板回归	19
表 3-2 古典假设检验	21
表 3-3 Chow 检验 (F 检验)	22

摘要

随着信息时代来临,数字经济迅速发展,数据的价值和地位不断提升,日益 成为我国重要的战略资源和新型的生产要素。"十四五"规划纲要明确提出,以 促进城市群发展为抓手,全面形成"两横三纵"城镇化战略布局。国家一直主张 加快城市群一体化进程,鼓励发挥大城市的中心辐射作用,带动周围城市发展。 可见,探索数据动能与城市群一体化二者之间的关系对国家的协调可持续发展意 义重大。为探究数据动能对城市群一体化的影响情况,首先通过搜集 2000-2019 年《中国城市统计年鉴》、《中国科技统计年鉴》的数据,在查阅相关文献以及政 策后,构建了数据要素、城市群一体化程度的指标体系,并利用熵值法为城市群 的一体化程度和数据动能赋值。其次,根据 chow 检验以及 Hausman 检验的结果, 选择固定效应模型进行回归。再次,通过固定效应回归方法,测算出变量系数, 利用动态面板回归方法检验变量系数的稳健性。最后,研究发现城市群的数据动 能及一体化程度都呈现逐年增长的趋势,且数据要素对城市群一体化程度有显著 的促进作用。为进一步研究数据动能对城市经济发展的作用机制,以长三角 27 个城市为例,进行 Hausman 检验后,选择固定效应面板模型,探究数据动能对城 市经济发展的推动作用。基于上述结论,提出以下建议: 1.国家政策扶持产业发 展,引导数据科技新兴产业健康蓬勃发展;2.提高劳动者素质,引进专业型人才; 3. 城市积极搭建数字合作平台,开放数据交易中心: 4. 各城市发布人才优惠政 策,保留当地人才并吸引更多人才。

关键词:数据动能;城市群一体化;熵值法;面板数据模型

一、引言

(一) 研究背景与意义

信息时代来临,大数据技术蓬勃发展,不断推动着经济发展与社会进步,数据所释放的动能在社会服务、城市运作、政府运行、企业运营等各个方面发挥着越来越重要的作用,也逐渐受到政府以及全社会越来越多的重视。中共中央于2020年4月9日颁布《关于构建更加完善的要素市场化配置体制机制的意见》的文件,强调利用数据要素带动其它要素发展,重视数据要素的市场发展,使数据动能为经济发展贡献力量。"数字产业化"和"产业数字化"这两个概念也是"十四五"规划中提出来的,发展数字经济被多次强调。数据的价值在不断发展的经济社会中提升很快,我国已经将其列为重要的战略资源,是竞争一个方面。党的十八大以来,我国数字化发展能力持续增强,数字经济规模不断壮大,信息便民惠民加速普及,数字治理格局日益完善,数字中国建设拥有坚实基础,展现出更加广阔的前景。各地积极响应,打造数字驱动发展的高地。数据在社会生产的各个方面都发挥着其重要作用,而对于城市群一体化的进程,数据动能是否也发挥着一定的推动作用是本研究所探讨的。

"十四五"规划纲要还提出,在全国范围内形成"两横三纵"的战略布局以加快促进城市群一体化发展,促使城市群内基础设施协调布局、公共服务共享以及生态共建环境得以共治。由此可见,推进城市群一体化已然成为国家层面战略规划。但是,大而不强、内部协调性不足、体制机制障碍等现实性问题使得我国城市群内部发展较为困难,这使得与世界上一流城市群相比,我国的国家级城市群仍存在创新力不足、联动性不足的问题。所以,对城市群发展潜能的提升,城市群一体化进程的推动,对提高我国城市群发展水平有非常重要的意义。

(二) 研究目标

- 1. 构建数据动能与城市群一体化的指标体系。基于发展数字经济、挖掘数据动能以及城市群一体化发展的政策要求,本文通过构建数据动能与城市群一体化指标体系,进行测度,结合相关研究,从生产设施、产业运营以及市场创新的角度测算数据动能。此外,结合以往文献中对城市群一体化程度的测度研究,将其分解为市场一体化、政策一体化、基础设施一体化、生态环境一体化四个方面进行测算。
- 2. 本文着重分析数据动能对城市群一体化的动态影响,探索其影响机制,并 根据研究结果为数据动能对城市群一体化的正向作用最大化提供理论导向,探索 数据动能以及城市群一体化的发展规律,并探究两者之间的关系,找出主导因素, 以主导因素为目标导向,高效率提高城市群一体化水平,建设高质量的城市群, 促进国内空间协调发展,以期实现国内各地区经济协调全面发展。
- 3. 本研究将选取长三角地区进行案例研究,探究数据动能对城市经济发展的 影响作用机制,为数据动能通过推动城市发展进而促进城市群一体化发展提供理 论基础。

(三)研究的理论基础

1. 数据动能测度研究的现状

目前,将数据视为一种生产要素并对其进行测度研究的文献较少,但对数字 经济的测度研究已经比较成熟,并且已经形成较为完整的体系。数字经济体现一 个城市利用数据动能的能力,对数据动能的测算具有一定借鉴意义。

徐清源等(2018)测算了国内外数字经济能力,使用了很多市场经济类的三级指标,欧盟数字经济与社会指数(DESI)指标体系国际电信联盟 ICT 发展指数指标体系等构成了国外数字经济的指标。中国信息通信研究院数字经济指数指

标体系、上海社科院全球数字经济竞争力指数指标体系等代表性较高。这些指标 体系所使用的二级指标如软件和信息技术服务业规模、互联网普及率、ICT 设备 及应用、数据交易中心数量、高技术产业专利情况等也可应用于数据动能的测算。 刘军等(2020)以信息化发展指标、互联网发展指标、数字经济发展指标为一级 指标,以信息化基础、信息化影响、固定端互联网基础、移动端互联网基础、数 字交易基础、数字交易影响等为二级指标,选取相应的三级指标构建数字经济测 度体系。范合君等(2020)提出了使用生产数字化、消费数字化、流通数字化和 政府数字化四个维度构建数据体系指标,生产数字化使用企业 ERP 普及率指数、 企业 MES 普及率指数等指标来测度: 消费数字化使用移动电话普及率指数、互 联网普及率指数等指标来测度: 流通数字化可以使用运用电子商务规模指数、电 子商务成长指数等指标测度: 政府数字化主要有两项指标, 即服务方式完备度指 数、服务事项覆盖度指数。数字化经济、数字化程度的指标体系为本文数据动能 的测度研究提供了一定启发与借鉴意义。Beomsoo Kim 等(2002)使用实验数字经 济(EDE),提出一种新的研究方法,虚拟领域实验,使其可行、有效的测试研究 假设与所需的实验控制水平和探索成功的商业策略在一个真实的商业世界。该文 对数字经济研究有以下三大结论,数字市场的效率、数字市场(公开价格市场和 拍卖)的有效性,以及质量认证的影响。刘静等(2020)构建了大数据系统指标 体系,包括3个一级指标,18个二级指标,所选取的一级指标有数据生产设施、 数据产业运营、数据市场创新,选取的二级指标有集成电路产量、软件业务收入、 互联网宽带接入用户、数字产业总产值、数字贸易总量、数字专利申请数、高等 教育入学率等。不难发现,数据动能的发展需要基础设施的支撑。Susanto Basu 等(2007)指出,城市化可以带来积极的经济成果,但如果没有适当基础设施支 持如信息和通信技术(ICT)和交通基础设施的不受限制地迁移到城市地区,可能 会导致负面副作用,如拥堵、贫民窟的形成和其他规模经济失调等。本文所构建

的数据动能测度体系结合了数字经济相关研究以及现有的对数据赋能的测算研 究。

2. 城市群一体化测度研究的现状

纵观以往的文献,已有不少学者对城市群一体化程度进行测度研究。汤放华 等(2018)对长株潭城市群一体化进行测度研究,将布局一体化、要素一体化、 制度一体化作为一级指标,将交通设施、产业分工、生态空间、人口流动、货物 流动、资金流动、信息流动、技术流动、战略协议、政策认同、交通设施、产业 分工、生态空间、人口流动、货物流动、资金流动、信息流动、技术流动、战略 协议、政策认同作为二级指标,并选取相应内容作为三级指标,利用熵值法赋权, 并得出测算结果。刘修岩等(2021)对中国主要的城市群一体化水平进行综合评 价,二级指标共有五类,分别为功能一体化、制度一体化、经济一体化、文化一 体化、产业一体化作为二级指标,将通高速铁路城市比例、通普通铁路城市比例、 通高速公路城市比例、是否成立市长联席会议制度、成立市长联席会议制度的年 数、是否获批为国家级城市群、获批为国家级城市群的年数、跨省级行政区划数、 人均 GDP 标准差以及经济开放度标准差等作为三级指标。徐境(2011)对呼包鄂 区域一体化发展进行评价研究,从支持系统、实力水平、绩效水平、生态安全角 度构成区域一体化发展的合意性评价指标体系。陈航航等(2018)梳理了区域一 体化理论和发展,主要从尺度、联系和边界等角度分析区域一体化理论发展的历 史进程,阐述了区域一体化逐渐具有网络化特征和多维联系的特征,该文中提出 GPNs 关注价值是如何经过创造、提高和捕获的过程而出现在大众视野的,以及 行动者如何参与和融入,明确地体现了制度背景和政策行动的重要性:强调地区 之间要存在相互依赖性,方可共同发展,"全球一地方"联系以及地方与生产网 络的再融合和有自己独立运营的经济模式是地方与全球经济建立联系的重要机 制。由此可见,制度与政策对城市群一体化的重要性。城市群一体化程度的测度

研究已形成较为规范成熟的体系,为本研究提供了极大的便利。本文结合相关政策与现有研究,并考虑到相关数据的可获得性,构建了一套测算城市群一体化程度的指标体系。

3. 文献述评

纵观以往研究,对数据动能进行测度研究的文献较少,但可以参考数字经济 以及大数据测度研究构建数据动能指标体系。关于城市群一体化的测度研究目前 已经形成比较成熟的体系,本研究借鉴以往的研究构建了城市群一体化程度的指 标体系。对于数据新动能对城市群一体化的影响研究,目前尚未有研究涉及,属 于学术上的空缺部分。本文将借鉴数字经济、大数据系统以及城市群一体化相关 研究为数据动能以及城市群一体化程度构建指标体系,并探究数据动能对城市群 一体化程度的影响机制。

(四) 技术路线

本研究基于提出问题、分析问题、解决问题的思路,阅读参考文献是提出问题阶段,构建指标体系、构建模型并进行实证分析是提出问题阶段,根据实证结论提出建议是解决问题阶段。技术路线图如下:

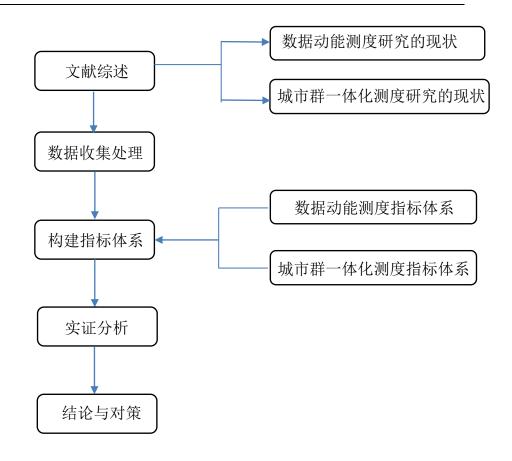


图 1-1 技术路线图

(五) 研究的创新点

研究角度新颖,以往文献中关于数据新动能对城市群一体化的影响相关研究 较少,本文构建面板回归模型,对二者之间的影响进行研究,属于较为新颖的研 究角度。

二、指标构建与测算

(一) 数据新动能指标构建与测算

1. 指标构建

本文借鉴了刘静、惠宁的论文、国家发布的相关政策和数字经济方面的论文作为数据新动能指标体系构建的理论基础,同时考虑数据的可获得性,从中选取

三级指标。首先,利用产业经营测度(A)、生产设施测度(S)和市场创新测度(D)三个二级指标构建一级指标,具体如下:

$$X = a * A + b * S + c * D$$

公式(1)

其中采取了熵值法来计算各三级指标的权重。

其次,通过求得的权重,计算出产业经营测度(A)。

产业经营测度(A)公式如下:

$$A = \sum p_i * O_i \qquad \qquad \triangle \vec{\mathbf{x}} \tag{2}$$

同理可测度生产设施测度(S)。

对每个二级指标,利用三级指标进行测度,具体如表 2-1 所示。

三级指标 一级指标 二级指标 物流产业收入 通信产业收入 产业经营测度 进出口贸易总额 具有数据交易平台以及个数 电信业务 数据新动能测度 科学技术支出 生产设施测度 互联网宽带用户接入数 信息传输业就业人数 在校大学生数(本专科) 专利申请数 市场创新测度 发表科技论文数 高校专利申请数

表 2-1 数据新动能测度构成指标

2. 测算过程

关于数据新动能三级指标的处理,首先,通过熵值法计算各三级指标占的权重。熵值法计算公式如下:

首先,确定指标数值,正项指标归一化公式如下:

负向指标归一化公式如下:

$$x_{it} = (\max(x_{it}) - x_{it}) / (\max(x_{it}) - \min(x_{it}))$$
 $\triangle \stackrel{\wedge}{\Rightarrow} (4)$

其次, 计算每一年的第 n 个指标的比重:

然后, 计算第 n 个指标的熵值:

其中, $k = \frac{1}{\ln n} > 0$,n=20,所以满足e ≥ 0

计算信息熵冗余度:

$$d_t = 1 - e_t (t=1, 2, 3, \dots, m)$$
 公式 (7)

计算 m 项指标中各项指标的权重:

得到每项二级指标下的三级指标对应的权重。

再次,通过得到的三级指标的权重来计算对应的二级指标的测度,公式如下:

$$Y = \sum p_i * Q_i \qquad \qquad \triangle \mathfrak{I}(9)$$

最后,根据下述公式,测度数据新动能。

$$X = a * A + b * S + c * D$$
 公式(10)

3. 测算结果

例如,在长三角城市群中,2016年电信业务为22268716万元,科学技术支出为10615155万元,互联网宽带用户接入数为51240000万人,信息传输业就业人数为761124万人。此时,计算出三级各指标2000到2019年的最大值和最小值,利用归一化公式,再取均值,计算该年该三级指标所对应的分数。然后对该分数进行熵值法求权重,近似权重为0.495、0.365、0.410、0.260,然后根据

对应的二级指标来算出二级指标的权重,根据所得二级指标的权重计算出该年一级指标生产设施测度所对应的值。

测得的三大城市群的数据新动能指标结果如下:

表 2-2 京津冀城市群数据新动能指标测算结果

时间	产业经营测度	生产设施测度	市场创新测度
2000	0	0	0.492
2001	11.342	8.940	1.097
2002	12.418	11.263	1.263
2003	13.040	13.683	2.427
2004	15.009	16.204	0.231
2005	15.880	18.973	0
2006	18.884	21.555	3.896
2007	22.419	27.799	4.204
2008	25.679	32.782	4.439
2009	26.262	37.359	62.960
2010	27.471	41.405	6.452
2011	32.297	29.015	54.888
2012	13.801	53.338	59.793
2013	42.630	30.958	72.136
2014	59.361	63.121	67.511
2015	64.026	74.805	73.163
2016	68.940	68.946	73.965
2017	72.353	79.679	77.651
2018	94.330	77.880	80.848
2019	100	80.518	87.580

表 2-3 长三角城市群数据新动能指标测算结果

		41 3X 1/11 1/1 1/1 1/1 1/1 1/1	
时间	产业经营测度	生产设施测度	市场创新测度
2000	0	0	0
2001	18.773	14.076	0.419
2002	21.028	16.065	1.033
2003	21.681	22.747	1.297
2004	28.121	32.405	2.420
2005	30.862	32.482	2.641
2006	33.991	35.173	2.693
2007	33.576	37.489	3.073
2008	46.071	49.061	3.408
2009	51.124	56.749	48.041
2010	56.183	71.307	53.023
2011	45.502	53.692	57.235
2012	43.533	56.698	65.658
2013	45.267	64.876	72.311
2014	46.842	68.243	72.554
2015	55.095	68.871	79.216
2016	68.500	74.924	79.576
2017	72.569	80.893	83.584
2018	89.300	87.900	87.616
2019	90.818	91.811	93.967

表 2-4 珠三角城市群数据新动能指标测算结果

时间	产业经营测度	生产设施测度	市场创新测度
2000	0	0	0.575
2001	16.097	10.467	0.743
2002	17.484	14.871	0
2003	18.533	16.894	0.019
2004	20.680	19.280	0.370
2005	22.043	20.356	0.843
2006	25.423	23.764	2.087
2007	29.286	27.267	2.241
2008	32.839	30.406	2.415
2009	33.793	32.820	5.834
2010	35.600	34.969	6.086
2011	40.554	42.337	40.449
2012	22.547	27.453	47.192
2013	52.203	53.299	59.289
2014	54.104	58.782	58.245
2015	59.565	64.395	61.827
2016	65.476	73.341	71.222
2017	69.952	77.911	80.683
2018	93.474	93.569	88.777
2019	100	97.215	99.501

(二) 城市群一体化测度的构建与测算

1. 指标构建

本文借鉴了国家发布的城市群相关文件、以往对城市群一体化的相关研究作为城市群一体化测度构建的理论基础。首先,利用市场一体化测度(Q)、政策一体化测度(W)、基础设施一体化测度(E)和生态环境一体化测度(R)四个二级指标测度城市群一体化,具体表达如下:

关于其中的二级指标权数的确定,本文采取熵值法来计算对应二级指标的权重。产业经营测度(A)公式如下:

 $Q = \sum p_i * O_i$ $\triangle \stackrel{<}{\operatorname{\operatorname{d}}} (12)$

同理,可以得到政策一体化测度(\mathbf{W})、基础设施一体化测度(\mathbf{E})和生态环境一体化测度(\mathbf{R})。

对应的三级指标的选取见表 2-5。

表 2-5 城市群一体化测度

一级指标	二级指标	三级指标
		人均 GDP
	市场一体化	FDI/GDP
		在岗工人平均工资
		所跨省级行政区域个数
	政策一体化	是否获批国家重点城市群
松字型 从44和南 —		财政支出/GDP
城市群一体化程度 —	基础设施一体化	中小学数量
		医院、卫生院个数
		客运总量
_		污染治理/GDP
	生态环境一体化	人均绿地面积
		工业废水排放达标率

2. 测算过程

关于城市群一体化三级指标的处理,首先,通过熵值法来计算各三级指标所占的权重。熵值法计算公式如下:

先确定指标数值,正项指标归一化公式如下:

$$x_{ij} = (x_{ij} - \min(x_{ij})) / (\max(x_{ij}) - \min(x_{ij}))$$
 $\triangle \rightrightarrows (13)$

负向指标归一化公式如下:

$$x_{ij} = (\max(x_{ij}) - x_{ij}))/(\max(x_{ij}) - \min(x_{ij}))$$
 公式(14)

然后计算每一年的第j个指标的比重:

其次, 计算第j个指标的熵值:

其中, $k = \frac{1}{\ln n} > 0$,n=20,所以满足 $e_j \ge 0$

计算信息熵冗余度:

计算 m 项指标中各项指标的权重:

因而近似得到每项二级指标下的四个三级指标对应的权重。

接着,通过得到的三级指标的权重来计算对应的二级指标的测度,公式如下:

$$Q = \sum p_i * O_i$$
 $\triangle \stackrel{\cdot}{\operatorname{I}} (19)$

其次,根据二级指标通过熵值法近似求权重,根据求出的二级指标权重进行 一级指标——城市群一体化的指标数值。

城市群一体化指标公式如下:

3. 测算结果

最终得到三大城市群各年一体化结果如表 2-6 所示。三大城市群的一体化值都保持在 0.2-0.4,整体呈现上升的趋势。2010年之后三大城市群的一体化程度较高,均维持在 0.3 以上。

表 2-6 三大城市群一体化测算结果

时间	长三角	珠三角	京津冀
2000	0.2879	0.2617	0.2746
2001	0.2827	0.2674	0.2721
2002	0.3071	0.3255	0.2634
2003	0.3032	0.3304	0.2781
2004	0.3190	0.3139	0.2848
2005	0.3238	0.3239	0.2904
2006	0.3096	0.3351	0.3067
2007	0.3108	0.3296	0.3055
2008	0.3136	0.3370	0.3048
2009	0.3199	0.3189	0.3043
2010	0.3085	0.3192	0.3068
2011	0.3257	0.3112	0.3019
2012	0.3114	0.3199	0.3064
2013	0.3165	0.3037	0.3095
2014	0.3170	0.3109	0.3063
2015	0.3529	0.3427	0.3581
2016	0.3507	0.3364	0.3623
2017	0.3577	0.3421	0.3646
2018	0.3559	0.3402	0.3619
2019	0.3598	0.3515	0.3548

(三) 变量选择与描述

1. 控制变量选择

对于控制变量,本研究选用了人口(百万人),经济(亿元),土地面积(平方公里),教育资源(所)四个指标作为该模型的控制变量。

对城市群一体化的影响因素除了本文所重点考察的数据动能外,人口多少、经济发展水平、土地面积、教育资源这四个变量对城市群一体化程度会存在一些影响。

2. 控制变量处理

对于控制变量的处理,本研究采用对数处理法。首先,对控制变量求平均值,

然后对平均值进行求对数处理,即该值为该年该控制变量的三级指标。

取平均值的目的是为了去除每个城市群的城市群城市数量大小不一样带来的误差。

处理后,对该数值取对数,即为该控制变量的指标数值。取对数的主要原因 是因为数值过大,直接计算十分不便,所以本研究采用对数处理法来处理控制变 量。

(四)数据来源说明

对于城市群一体化指标,市场一体化选取的外商直接投资额 (FDI) 占地区当年 GDP 的比值、人均 GDP、在岗工人平均工资作为三级指标,数据来源于2000 年至 2019 年的《中国城市统计年鉴》; 政策一体化的所跨行政区域个数、是否获批国家重点城市群的数据来源于网络搜索,财政支出与 GDP 之比来源于历年《中国城市统计年鉴》; 用于测量基础设施一体化与生态环境一体化的三级指标数据均来自历年《中国城市统计年鉴》。对于数据新动能测度指标,除了产业运营指标下的是否具有数据交易平台以及个数来源于网络搜索之外,其余三级指标数据均来自历年《中国城市统计年鉴》《中国科技统计年鉴》。需要说明的是,《中国科技统计年鉴》中并没有城市级的数据,因此,用上海市、江苏省、浙江省、安徽省的数据代替长三角整体的数据,珠三角城市群则是以广东省的数据进行代替,京津冀城市群以天津市、北京市、河北省的数据进行代替。

三、模型建立与实证分析

(一) 模型设定

每个城市群每年的情况有所差别,可能会存在个体效应和时间效应,因此选择变截距面板数据模型。在混合面板数据模型、固定效应面板数据模型以及随机

效应面板数据模型中进行检验挑选。初步设定的模型如下:

$$CITYBELT_{it} = \mu_i + \lambda_t + \beta_1 SJXDN_{it} + \beta_2 \ln RK_{it} + \beta_3 \ln XZMJ_{it} + \beta_4 \ln GXSL_{it} + \beta_5 \ln JJSP_{it} + \beta_0 + \xi_{it} \quad (i = 1, 2, 3; t = 1, 2, 3 \dots 20)$$

CITYBELT_{it} 表示第 i 各城市群第 t 年城市群一体化程度; μ_i 表示个体效应, λ_t 表示时间效应, μ_i 和 λ_t 里包含初始的截距项 β_0 , μ_i 和 λ_t 可能和某个解释变变量存在相关性; SJXDN_{it} 为核心解释变量,表示第 i 各城市群第 t 年数据新动能发展情况; $\ln RK_{it}$ 、 $\ln XZMJ_{it}$ 、 $\ln GXSL_{it}$ 、 $\ln JJSP_{it}$ 均为控制变量,表示第 i 个城市群第 t 年的人口、行政区域面积、高校数量取对数; ξ_{it} 表示随机扰动项,包含未考虑入模型的变量带来的扰动。

(二) 建模步骤流程图

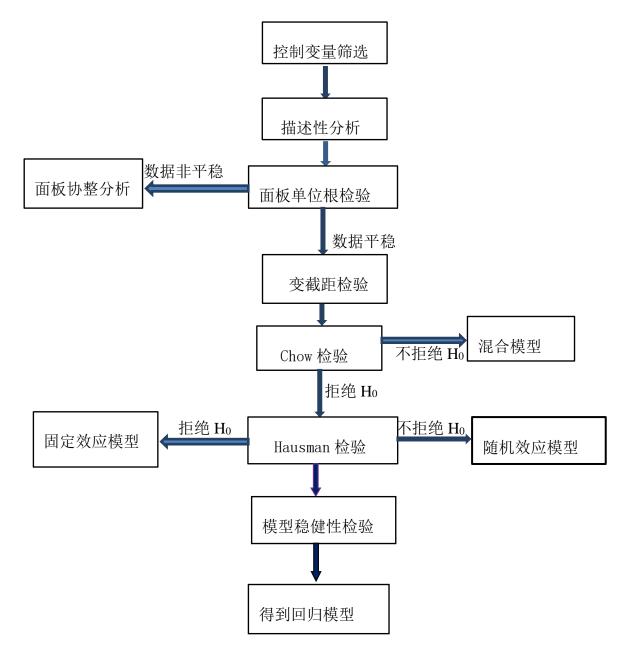


图 3-1 建模步骤流程图

(三) 建模过程①

1. 控制变量筛选

本文除了核心解释变量 SJXDN(数据新动能)外,还选取了 LnRK(人口状况)、lnJJSP(经济水平)、lnXZMJ(行政区域面积)、lnGXSL(高校数量)作为控

附注^① 本过程均在 STATA16 中完成

制变量。回归结果如表 3-1 所示。

表 3-7 基本面板回归

CITYBELT	Coef.	t 检验 P 值	F检验P值
SJXDN	0.036	0.114	
lnRK	-0.058	0.413	
lnJJSP	0.028	0.244	0.000
lnXZMJ	0.005	0.907	
lnGXSL	0.029	0.480	
_cons	0.323	0.005	

从表中可以看出,虽然模型通过了F检验,表明模型中至少存在一个重要的解释变量,但是模型的稳健性却很差,所有解释变量的系数均未通过t检验。对模型进行多重共线性检验,得到方差膨胀因子12.94>10,说明模型的各解释变量之间存在着严重的多重共线性。考虑到lnJJSP(经济水平)是以人均GDP进行数据处理的,而数据新动能的三级指标也选取了GDP,因此两者可能具有很强的相关性,求解出SJXDN与lnJJSP的相关系数为0.900。两解释变量之间的相关性很高,因此lnJJSP属于冗余的变量,为减轻多重共线性,应当剔除。所以,最终选择控制的变量为人口状况、行政区域面积、高校数量。

2. 描述性分析

录入数据后,对数据的基本特征做描述性分析。数据无缺失值,属于平衡面板数据,被解释变量以及核心解释变量的趋势图如 3-2、3-3 所示(y表示被解释变量,X1 为核心解释变量数据新动能,图 1、2、3 依次代表长三角城市群趋势图、珠三角城市群趋势图以及京津冀城市群趋势图)。

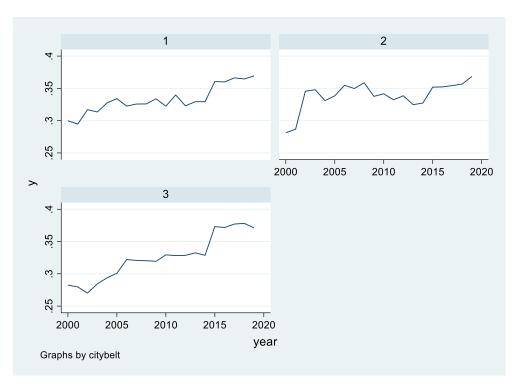


图 3-2 各城市群一体化趋势图

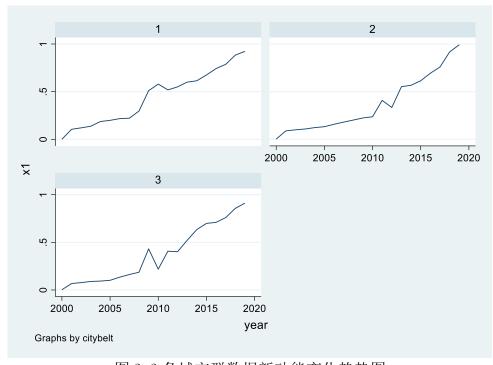


图 3-3 各城市群数据新动能变化趋势图

由上图可知,各城市群的一体化程度和数据新动能都呈上升趋势,长三角城 市群一体化与数据新动能发展波动性较小,比珠三角和京津冀更加稳定。为了解 数据新动能和城市群一体化的影响情况,在做混合回归之前先对城市群一体化和数据新动能做回归,得到的散点图 3-4。

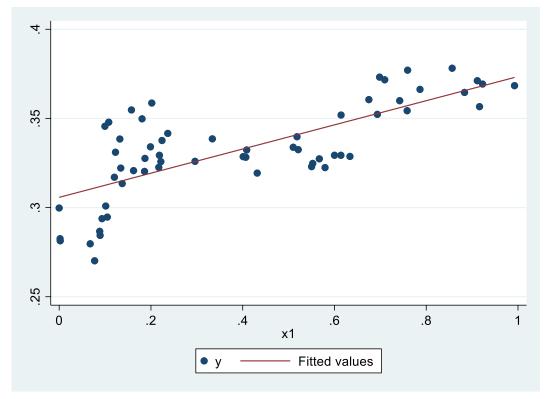


图 3-4 CITYBELT 与 SJXDN 散点图

由核心解释变量与被解释变量的偏回归图可知,数据新动能对城市群一体化 呈正向影响,这为后续的面板回归分析提供很大的帮助。

3. 古典假设

由于变截距面板模型后续的很多检验需要满足古典假设。对扰动项进行组间 异方差检验、组内自相关检验、组间同期相关检验。经检验,原数据存在异方差, 在对数据进行处理后,三种检验的结果如表 3-2 所示。

4. 面板单位根检验

面板数据处理好后,需进行面板单位根检验,检验数据的平稳性,防止出现 伪回归现象。这里对被解释变量 CITYBELT 进行面板单位根检验。在含个体固 定效应和线性时间趋势的数据检验中,拒绝原假设,数据平稳。

5. chow 检验 (F 检验)

为检验模型的截距是否发生变化,需进行 chow 检验,得到的结果如表 3-3 所示。

CITYBELT	Coef.	P-value	Chow 检验 (F 检验) P 值
SJXDN	0.064	0.001	
lnRK	-0.211	0.250	
lnXZMJ	0.011	0.967	0.000
lnGXSL	0.082	0.030	
_cons	0.739	0.541	

表 3-9 Chow 检验 (F 检验)

在 95%的显著性水平下, Chow 检验 (F 检验) 所得到的 P 值为 0.000<0.05, 拒绝原假设, 因此可以选择变截距模型。

6. Hausman 检验

为检验变截距模型的固定效应,需进行 Hausman 检验,由于前面已经完成异方差和自相关的处理,这里可以直接进行传统的 Hausman 检验。

在 95%显著性水平下, Hausman 检验 P 值为 0.000<0.05, 拒绝原假设, 因此, 选择固定效应模型。

7. 基准回归结果

通过上述的检验与分析,最终选择固定效应模型。以数据新动能为核心解释变量,来三大城市群一体化与数据新动能的关联性,同时,控制人口、行政区域面积、高校数量三个变量。所得基准回归结果如表 3-4 所示。

		·	
CITYBELT	Coef.	P	F检验P值
SJXDN	0.064	0.001	
lnRK	211	0.250	
lnXZMJ	0.011	0.967	0.0000
lnGXSL	0.082	0.030	
_cons	0.739	0.541	

表 3-4 基准回归结果

8. 稳健性检验

为检验回归结果的稳健性,将被解释变量 CITYBELT 滞后一期纳入模型进行 回归,核心解释变量 SJXDN 的系数为 0.025,并未发生符号上的改变,依然具有 显著的正向影响。因此回归结果具有稳健性。

(四) 数据新动能对城市群一体化影响的实证结果分析

根据上述探究分析,最终得到的回归结果为:

$$CITYBELT_{it} = 0.739 + \mu_i + \lambda_t + 0.064SJXDN_{it} - 0.211 \ln RK_{it} + 0.011 \ln XZMJ_{it} + 0.082 \ln GXSL_{it} + \xi_{it} (i=1,2,3;t=1,2,3.....20)$$

由混合回归模型结果可知,数据新动能对城市群一体化具有明显的正向影响, 在其他变量不变的条件下,数据新动能每增加一个单位,可以使城市群一体化平 均增加 0.064 个单位,说明在城市群一体化发展的过程中,数据动能发挥了显著 的推动作用。由数据产生的动能可以使城市群中的城市在市场、政策、基础设施、 生态环境等方面协同发展。数据新动能所包含的新的生产设施、产业运营、科技 创新等对城市群的发展起着重要作用。数据新动能可以更好、更高效地发挥城市 群里中心城市的作用(例如上海、北京、广州、深圳等),使中心城市快速辐射 周围城市,真正做到先富带动发展。

在控制变量方面,行政区域面积与高校数量对城市群一体化均发挥正向作用。各城市群行政区的面积越大,行政区域面积差异越小,对其城市群一体化程度的推进具有积极影响。由于行政区域面积基本变化不大,这方面,长三角、珠三角的表现要好于京津冀。高校是为城市群一体化提供人才的地方,新兴的数据产业需要高端人才,因此在高校数量分布密集的城市群,城市群一体化发展会较好。在高校数量方面,长三角、京津冀表现好于珠三角,这也是珠三角区域近年来发展高效教育的原因。人口与城市群一体化呈负向的影响关系,由于城市群中各城市的人口量分布很不均匀,导致周围城市的发展远落后于中心城市。人口是影响城市发展的一个重要因素,小城市要采取一定措施,减少人口迁向大城市,避免阻碍城市群一体化发展。

(五)数据新动能对城市经济发展模型建立②

城市群是由城市构成的,城市群一体化的发展离不开城市群内部每个城市的 经济发展。为探究数据新动能对城市经济发展影响情况,以城市历年 GDP 为被 解释变量,城市数据新动能为核心解释变量,控制各城市的人口、行政区域面积、 高校数量,选取长三角底下 27 个城市作为样例,构建变截距面板回归模型如下:

 $\ln CITY_{it} = \mu_i + \lambda_t + \beta_1 SJXDN_{it} + \beta_2 \ln RK_{it} + \beta_3 \ln RK_{it} + \beta_4 \ln GXSL_{it} + \xi_{it}$ (i=1,2,3······27;t=1,2,3······20)

1. 描述性分析

数据无缺失值,属于平衡面板数据,N=27,T=20,共计 540 个数据,属于长

^② 本过程均在 STATA16 中完成

面板数据。被解释变量 InCITY 与核心解释变量 SJXDN 的趋势如图 3-5、图 3-6 所示(1-27 分别代表长三角的 27 个城市)。



图 3-5 长三角各城市经济发展变化趋势图



图 3-6 长三角各城市数据新动能变化趋势图

由上图可知,长三角 27 个城市中,上海(图中的 1)的数据动能发展水平最高且发展速度最快,远远优于其它城市。在 2010年以前除上海外的其它 26 个城市数据动能发展很缓慢,2010年后,苏州(图中的 6)的数据动能开始迅速发展,随后杭州(图中的 11)、南京(图中的 2)、无锡(图中的 3)、宁波(图中的 12)等也开始发展。

被解释变量 CITY 和核心解释变量 SJXDN 做回归,得到的结果如图 3-7 所示。

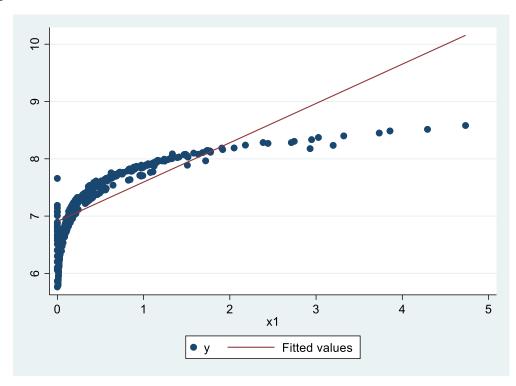


图 3-7 被解释变量与核心解释变量的散点图

由上图可知,数据新动能对城市经济发展呈正向影响,且一开始的趋势变化 较陡,后续区域平缓。这为后续的回归分析提供了很大的帮助。

由于本部分属于案例分析,因此一些检验将进行适当省略。

2. Chow 检验 (F 检验)

为检验数据新动能对城市经济发展的影响是否属于变截距面板模型,对其进行 chow 检验,得到的结果如表 3-5 所示。

表 3-5 Chow 检验 (F 检验)

lnCITY	Coef.	P-value	P-value
SJXDN	0.3702123	0.000	
lnRK	1.865021	0.000	
lnXZMJ	-0.119693	0.311	0.0000
lnGXSL	0.9336321	0.000	
_cons	1.892206	0.000	

在 95%显著性水平下, Chow 检验(F 检验)的 P 值为 0.0000<0.05,因此拒绝原假设,模型为变截距面板模型。

3. Hausman 检验

为检验模型的固定效应,对其进行 Hausman 检验。

在 95%显著性水平下, Hausman 检验 P 值为 0.0000<0.05, 拒绝原假设, 因此, 选择固定效应模型。

4. 基准回归结果

通过上述检验,最终应当选择固定效应模型,建立固定效应模型所得到的结果如表 3-6 所示。

表 3-6 基准回归结果

lnCITY	Coef.	P-value	F检验P值
SJXDN	0.370	0.000	
lnRK	1.865	0.000	
lnXZMJ	-0.120	0.311	0.000
nGXSL	0.934	0.000	
_cons	1.892	0.000	

5. 稳健性检验

为检验回归结果的稳健性,将被解释变量 CITYBELT 滞后一期纳入模型进行 回归,核心解释变量 SJXDN 的系数为 0.225,并未发生符号上的改变,依然具有 显著的正向影响。因此回归结果具有稳健性。

(六) 数据新动能对城市经济发展影响的实证结果分析

通过上述的检验分析,得到的固定效应模型为:

$$\ln CITY_{it} = 1.892 + \mu_i + \lambda_t + 0.370SJXDN_{it} + 1.865 \ln RK_{it} - 0.120 \ln XZMJ_{it} + 0.934 \ln GXSL_{it} + \xi_{it} \quad (i=1,2,3\cdots 27;t=1,2,3\cdots 20)$$

由回归模型可知,数据新动能对城市发展具有正向影响,在其它条件不变的情况下,数据新动能每增长一个单位,城市发展就平均增长 0.370%。相较于城市群,城市的体量较小,因此,城市层面的数据新动能对城市发展影响更为显著。位于长三角的城市大多较为发达,它们受数据新动能的带来的利益较大。也是对数据更具敏感性的城市,能够更早地利用数据来推动自己的经济发展。例如位于长三角的新一线城市杭州近年来依托数字新动能,发展城市数字经济,推动城市产能朝更加智能化的方向转型,城市经济发展迅速。

值得注意的是,对城市群一体化建立模型时,控制变量人口的系数为负,其对城市群一体化的影响是反方向的。由城市级的数据所进行的面板回归中,人口变量的系数为正,其对城市经济发展呈现正向的影响。在城市群一体化测度研究时,不仅有与经济相关的指标,还有与基础设施,政策等指标,因此城市群的人口太多,不便于城市群一体化管理,基础设施的分配。因此,人口对城市经济增长的贡献为正,而对城市群一体化程度是负向的影响。同样,行政区域面积这个控制变量在两个回归模型中的符号也是相反的。同样,在前面解释过,由于行政区域面积每年变化情况不是很大,因此只是用来比较各城市行政区域面积对城市经济发展影响的差异。在该案例中,行政区域面积较小的上海、南京等经济发展

较好,因此该控制变量的系数为负。无论是对城市群一体化发展,还是对城市经济发展,高校数量的影响均为正,这也凸显了教育资源和人才的重要性。

四、结论与对策建议

(一) 结论

- 1. 三大城市群的数据动能呈逐年增长的趋势,2019 年数据动能测算值从大到小依次为珠三角、长三角、京津冀城市群,分别为99.244、92.277、91.136,总体得分较高。三大城市群由于其核心城市的带动作用,聚集了较多的科技人才,具备良好的基础设施,为数据动能的发展奠定了一定的基础。随着数字经济、数据科技、互联网行业的迅猛发展,数据要素对经济发展的贡献度逐渐增大,释放了较大的乘数效应,促使经济大幅度增长。
- 2. 由面板回归结果可知,数据动能对城市群一体化发挥着明显的正向促进作用。核心城市利用数据要素释放了较多的能量,带动周边城市成长,同时其他城市利用内部拥有的数据动能,促进自身发展,城市间的发展差异明显减少,驱动城市群一体化进程。
- 3. 城市的人口、面积以及高校数量也影响了城市的发展,城市群内部协调发展须注意调节各城市的人口数量以及高校数量,这影响了当地的人才输送机制,科技为第一生产力,保留住人才是一个城市保持活力的根本措施。

(二) 对策建议

由面板回归模型结果可知,数据动能对促进城市群一体化发展,带动城市经济发展,发掘经济新增长点具有显著的促进作用。因此,采取有效的措施激发数

据动能,加快释放乘数效应是促进城市经济发展,加快城市群一体化进程,迎合国家协调发展布局的重要途径。

1. 国家政策扶持产业发展, 引导数据科技新兴产业健康蓬勃发展

基于数据动能对城市群一体化显著的正向影响,相关地区应当大力发展数据科技新兴行业,推动"互联网+"产业的发展。目前,产业互联网一片蓝海,"互联网金融"、电子商务等发展迅速。"互联网+"产业的发展依托于数据科技,因此,规范发展"互联网+"行业对数据动能的效益最大化有较大作用。2020年,《意见》首次正式将数据列为生产要素之一。数据要素的作用受到广泛的重视。然而,国家仍然需要发布关于深化数据要素作用的政策,使得全国各地的企业以及个人都意识到数据的重要作用,激发数据相关产业,如大数据、云计算、区块链等加速发展。通过数据要素的正向影响,带动城市群一体化进程,由此早日完成各地区协调发展的战略计划。

2. 提高劳动者素质, 引进专业型人才

毋庸置疑,数字产业为高新技术产业,数据科技的进步需要专业人才的支撑。 因此,政府应当加强教育资源的投入以及对科研机构的支持力度,加快培育一批 数据领域领军人才。高校也应当注重学生计算机能力、编程能力的培养,为日后 数据行业输送人才。相关企业提高岗位的适配度,引进专精人才,推进企业技术 升级,加快培育数据要素市场,实现产业链升级,带动城市群发展。

3. 城市积极搭建数字合作平台, 开放数据交易中心

我国目前的数据交易中心较少,开放更多的数据交易平台有利于数据动能的加速释放。第四届数字中国建设峰会的溢出效应助力中国数字的建设,国家数字经济创新发展试验区、厦门金砖创新基地等数字合作平台为数字经济高质量发展提供契机。峰会展示了目前的数字应用场景,对接了许多数字产业合作项目,为全国各地数字经济的产业合作搭建平台,中国新型经济发展格局逐渐形成。数字

合作平台的搭建,为中国经济发展带来了新的机遇,是一大重要的经济增长极。 在城市群内积极搭建合作平台,有助于促进城市间的数字经济合作共享,区域协 调发展,进而推动城市群整体的发展。

4. 各城市发布人才优惠政策,保留当地人才并吸引更多人才

各个城市须建立并完善对高新技术人才的优惠制度,保留住促进城市高质量、可持续发展的新鲜血液,为高新技术产业传输源源不断的动力,减少各个城市间技术上的差异,进而促进城市群协调发展。

参考文献

[1]徐清源,单志广,马潮江.国内外数字经济测度指标体系研究综述[J].调研世界.2017(11)

[2]刘军,杨渊鋆,张三峰.中国数字经济测度与驱动因素研究[J].上海经济研究. 2020(06)

[3]范合君,吴婷.中国数字化程度测度与指标体系构建[J].首都经济贸易大学学报.2020,22(04)

[4]KIM B, BARUA A, WHINSTON A B. Virtual field experiments for a digital economy: a new research methodology for exploring an information economy [J]. Decision support systems, 2002, 32(3): 215-231

[5] Pradhan Rudra P.; Arvin Mak B. Nair MahendhiranUrbanization, transportation infrastructure, ICT, and economic growth: A temporal causal analysis[J]. Cities Volume,2021

[6]刘静,惠宁,南士敬. 数据赋能驱动文化产业创新效率的非线性研究——基于 STR 模型的实证检验[J]. 经济与管理研究.2020,41(07)

[7]汤放华,吴平,周亮.长株潭城市群一体化程度测度与评价[J]. 经济地理.

2018,38(02)

[8]刘修岩,梁昌一.中国城市群一体化水平综合评价与时空演化特征分析——兼 论城市群规模的影响[J].兰州大学学报(社会科学版).2021,49(02)

[9]徐境.呼包鄂区域一体化发展评价研究[J].干旱区资源与环境. 2011,25(05)

[10]陈航航, 贺灿飞, 毛熙彦. 区域一体化研究综述: 尺度、联系与边界[J]. 热带地理, 2018, 38(1): 1-12

[11]廖信林,杨正源.数字经济赋能长三角地区制造业转型升级的效应测度与实现路径[J]. 华东经济管理. 2021,35(06)

[12]黄桂林,许如意,苏义坤. FDI、技术创新与城镇化互动关系研究--基于省际面板数据的 PVAR 实证分析[J]. 数学的实践与认识,2021

[13]杨传开,宁越敏. 中国省际人口迁移格局演变及其对城镇化发展的影响 [J]. 地理研究,2015,34(8): 1492-1506

[14]楚芳芳, 蒋涤非. 基于能值理论的长株潭城市群一体化研究 [J]. 科技进步与对策, 2012, 29(5): 45-50

[15]黄新飞,郑华懋. 区域一体化、地区专业化与趋同分析——基于珠江三角洲地区城市的实证分析[J]. 统计研究,2010,27(1):90-96

[16]娄文龙. 京津冀、长三角和珠三角区域经济一体化测量和比较[J]. 统计与决策, 2014(2): 90-92

[17]陈红霞,李国平. 1985-2007 年京津冀区域市场一体化水平测度与过程分析[J]. 地理研究, 2009, 28(6): 1 476 - 1483

[18]汤凯.中原经济区区域一体化建设研究[J]. 郑州航空工业管理学院学报,2013,31(2): 15-19

[19]Skeldon R. Trends in international migration in the Asian and Pacific Region [J]. International Social Science Journal, 2000, 52(3): 369 - 382.

[20]徐朝阳, 白艳, 王韡. 要素市场化改革与供需结构错配[J]. 经济研究, 2020(2): 20-35

[21]尼夫.数字经济 2.0: 引爆大数据生态红利[M]. 大数据文摘翻译组,译.北京:中国人民大学出版社,201

[22]程啸.论大数据时代的个人数据权利[J]. 中国社会科学, 2018(3): 102 - 122, 207-208

[23]SHAMIM S, ZENG J, SHARIQS Metal. Role of big data management in enhancing big data decision making capability and quality among Chinese firms: a dynamic capabilities view [J]. Information & Management, 2019, 56(6): 103—135

[24]马中东,宁朝山.数字经济、要素配置与制造业质量升级[J].经济体制改革,2020(3): 24-30

[25]黄群慧, 余泳泽, 张松林.互联网发展与制造业生产率提升: 内在机制与中国经验 [J]. 中国工业经济, 2019(8): 5-23

[26]陶长琪,周璇.产业融合下的产业转型优化升级效应分析——基于信息产业与制造业耦联的实证研究[J].产业经济研究,2015(3):21-3

[27]BLOOM N L,GARICANO R SADUN R, etal.The Distinct Effects of Information Technology and Communication Technology on Firm Organization [J]. Management Science, 2014, 60 (12): 2859-2885

[28]FORMAN C, ZEEBROECK N V. From Wires to Part- ners: How the Internet has Fostered R&D Collaborations within Firms[J]. Management Science, 2012, 58(8): 1549-1568

[29]李捷,余东华,张明志.信息技术、全要素生产率与制造 业转型升级的动力 机制——基于"两部门"论的研究 [J].中央财经大学学报,2017(9):67-78

[30]郭晗, 廉玉妍.数字经济与中国未来经济新动能培育[J]. 西北大学学报(哲学社会科学版), 2020,50(1): 65-72

[31]焦勇.数字经济赋能制造业转型:从价值重塑到价值创 造[J].经济学家,2020 (6): 87-94

[32]何帆,刘红霞.数字经济视角下实体企业数字化变革的 业绩提升效应评估[J]. 改革,2019(4):137-148.

[33]肖旭, 戚聿东.产业数字化转型的价值维度与理论逻辑 [J].改革,2019(8): 61-70

致谢

在此论文撰写完成之际,回首近期的共同努力,我要由衷地感谢我的队员。 从论文的选题、数据收集整理、模型构建、实证分析到最后的论文撰写与修改, 本研究探讨问题、交流想法、相互协作,在每个深夜里挑灯夜读、毫不懈怠。正 是有了互相的鼓励和支持,本研究的论文才得以在规定时间内顺利完成。此外, 本研究还要感谢指导老师,感谢她一直以来对本研究的关心和耐心指导。在论文 选题、数据整理、梳理框架、模型构建等环节,老师为本研究付出了很多时间和 心血,为本研究在编写论文的大方向上提供了宝贵的意见以及关键性的指导。当 本研究在研究中遇到棘手的问题时,询问老师,老师也都耐心指导,让本研究少 走了很多弯路。在此衷心感谢指导老师对本研究的无私帮助!