

分类号 C8/309
UDC

密级
编号 10741

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

(专业学位)

论文题目 中国家庭数字鸿沟时空特征及影响因素分析

研究生姓名: 杨知雨

指导教师姓名、职称: 孙玉环教授

学科、专业名称: 统计学 应用统计硕士

研究方向: 大数据分析

提交日期: 2022年5月30日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 杨知雨 签字日期： 2022.5.30

导师签名： 孙玉环 签字日期： 2022.5.30

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意”/“不同意”）以下事项：

- 1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；
- 2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 杨知雨 签字日期： 2022.5.30

导师签名： 孙玉环 签字日期： 2022.5.30

Analysis on the Spatiotemporal Characteristics and Influencing Factors of the Digital Divide in Chinese Households

Candidate : Yang Zhiyu

Supervisor: Sun Yuhuan

摘要

随着互联网的逐步普及和数字化进程的不断推进,不同国家、地区、行业、家庭和个体之间在拥有和应用网络信息技术上的差异造成了数字鸿沟,引发信息落差及贫富进一步分化等问题。数字鸿沟反映了各群体在信息时代所面临的机遇差别,是网络经济社会中个人、群体获取知识、掌握创新思想的能力以及创造财富的能力差距的表现,任由信息化非均衡发展必将催生新的极化格局。比起宏观地区整体水平差异,微观层面的信息技术极化现象更能体现其本质。因此,系统分析数字鸿沟问题现状,量化家庭间数字鸿沟程度,考察其时空演变特征,明确各方面影响因素及其作用机制,在缩小和弥合家庭数字鸿沟、促进数字化和智能化均衡发展等方面具有重要意义。

本文在梳理现有文献的基础上,基于2014-2018年中国家庭追踪调查(CFPS)数据,从互联网使用层面选取指标,测算家庭信息化水平,计算地区数字鸿沟指数,从城乡、省际和八大综合经济区尺度分析二者的时间演化趋势和空间分布特征;从家庭、个人和地区三个方面选取解释变量,以家庭数字鸿沟指数为被解释变量,构建平衡面板数据,经检验选取固定效应模型进行拟合,分析各因素对家庭数字鸿沟的影响程度,进而为缓解数字鸿沟提出针对性建议。

主要结论包括:①我国互联网渗透率已超70%,信息接入层面差距逐渐消失,但互联网应用使用率极化程度仍然较高,信息使用差距较大;②数字鸿沟水平整体呈现缩小趋势,城镇数字鸿沟降幅明显,农村数字鸿沟下降缓慢,二者差距逐年扩大;③家庭信息使用水平持续升高,呈现“东高西低,南高北低”分布特征;地区数字鸿沟逐年减轻,呈现“东低西高,北重南轻”的分布特征;④各级家庭信息使用水平与地区数字鸿沟水平的地区分布相似,信息使用水平越高的地区内部极化程度越低,且发展越快的地区数字鸿沟弥合速度越快,但地区间差距有加重的态势;⑤微观因素和宏观因素共同作用于家庭数字鸿沟,家庭、个人和地区因素均对家庭数字鸿沟产生影响;⑥二级数字鸿沟影响因素与一级数字鸿沟存在差异,性别、城市化水平和信息基础设施发展水平对数字鸿沟的影响逐渐消失。

本文的主要工作及创新有:研究视角上,同时考虑宏观和微观层面,对比分析地区数字鸿沟和家庭信息使用水平的分布特征,并探究家庭数字鸿沟受宏观和微观因素的影响程度;指标选取上,依据二级数字鸿沟(技术使用差距)定义,

选取家庭网络使用行为相关变量，测算家庭信息使用水平，从家庭、个人和地区三个层面选取解释变量，以家庭数字鸿沟指数为被解释变量，构建平衡面板数据；研究方法上，结合计量经济学理论，通过检验判断数据是否存在由不可观测的个体差异或异质性造成的遗漏变量问题，以此选取最优面板数据回归模型。

关键词：数字鸿沟 家庭信息使用水平 影响因素 固定效应模型

Abstract

With the gradual popularization of the Internet and the continuous advancement of the digitalization process, the differences in the ownership and application of network information technology among different countries, regions, industries, families and individuals have created a digital divide, causing information gaps and further polarization between the rich and the poor. The digital divide reflects the differences in opportunities faced by various groups in the information age. It is a manifestation of the gap in the ability of individuals and groups to acquire knowledge and master innovative ideas in the network economy society to create wealth. If the unbalanced development of informatization is allowed, it will give birth to a new polarization pattern. Compared with the overall level difference of macro-regions, the phenomenon of micro-level information technology polarization can better reflect its essence. Therefore, systematically analyzing the status quo of the digital divide, quantifying the degree of the digital divide among households, examining its temporal and spatial evolution characteristics, and clarifying various influencing factors and their mechanisms, is of great significance to narrow and bridge the household digital divide and promote the balanced development of digitalization and intelligence.

On the basis of combing the existing literature, this paper based on the 2014-2018 China Family Tracking Survey (CFPS) data, select indicators

from the Internet use level, measure the level of household informatization, calculate the regional digital divide index, and analyze their temporal evolution trends and spatial distribution characteristics from the scales of urban and rural, inter-provincial, and eight comprehensive economic zones. Then, explanatory variables are selected from three aspects of households, individuals, and regions, and the household digital divide index is used as the explained variable to construct balanced panel data. After testing, a fixed-effect model was selected for fitting, and the degree of influence of each factor on the household digital divide was analyzed, and then put forward targeted suggestions for alleviating the digital divide.

The main conclusions include: ①my country's Internet penetration rate has exceeded 70%, and the gap in the information access level has gradually disappeared, but the degree of polarization in the utilization rate of Internet applications is still high, and the gap in the use of information is large. ②The overall level of the digital divide has shown a narrowing trend. The urban digital divide has dropped significantly, while the rural digital divide has declined slowly, and the gap between the two has widened year by year. ③The level of household information usage continues to rise, showing the distribution characteristics of "high in the east and low in the west, high in the south and low in the north". Meanwhile, the regional digital divide has been reduced year by year, showing the distribution characteristics of "low in the east and high in the west, heavy

in the north and light in the south". ④The level of household information use at all levels is similar to the distribution of the regional digital divide. The higher the level of information use, the lower the degree of internal polarization in the region, and the faster the development of the region, the faster the digital divide will be bridged, but the gaps between regions are increasing. ⑤ Micro-factors and macro-factors work together on the household digital divide. Family, personal and regional factors all have an impact on the household digital divide. ⑥ The factors affecting the second-level digital divide are different from the first-level digital divide. The influence of gender, urbanization level and information infrastructure development level on the digital divide gradually disappears.

The main work and innovations of this article are: In terms of research perspective, considering both the macro and micro levels, comparative analysis of the regional digital divide and the distribution characteristics of household information use levels, and exploring the degree to which the household digital divide is affected by macro and micro factors. In terms of index selection, according to the definition of secondary digital divide (technology use gap), variables related to family Internet use behavior were selected to measure the level of family information use. Explanatory variables were selected from three levels of family, individual and region, and the family digital divide index was taken as the explained variable to construct balanced panel data. In terms of research method, combined with

econometric theory, the data is tested to determine whether there is omitted variable bias caused by unobservable individual differences or heterogeneity, so as to select the optimal panel data regression model.

Keywords: Digital Divide; Household information use level; Influencing factors; Fixed Effects Model

目 录

1 引言	1
1.1 研究背景.....	1
1.2 研究内容.....	2
1.3 创新点与局限性.....	3
1.3.1 主要工作及创新.....	3
1.3.2 局限性.....	3
2 国内外研究综述	5
2.1 数字鸿沟的测度.....	5
2.1.1 数字鸿沟测度模型指标选取.....	5
2.1.2 数字鸿沟指数计算方法.....	6
2.2 数字鸿沟时空特征及驱动机制研究方法.....	7
2.3 数字鸿沟影响因素分析.....	8
2.4 简要述评.....	9
3 理论基础和主要方法	10
3.1 数字鸿沟定义.....	10
3.2 数字鸿沟测算.....	11
3.2.1 家庭信息使用水平.....	11
3.2.2 地区数字鸿沟指数.....	12
3.2.3 家庭数字鸿沟指数.....	12
3.3 面板数据回归模型.....	12
3.3.1 面板数据.....	12
3.3.2 面板数据模型类型.....	13
4 中国家庭数字鸿沟时空特征分析	15
4.1 全国互联网发展概况.....	15
4.1.1 互联网普及率.....	15
4.1.2 互联网应用使用率.....	16

4.2 地区数字鸿沟指数测算.....	17
4.2.1 数据来源.....	17
4.2.2 时空范围界定.....	18
4.2.3 指标选取和数据预处理.....	18
4.2.4 地区数字鸿沟指数计算.....	19
4.3 家庭信息使用水平分布特征.....	23
4.4 地区数字鸿沟测算结果分析.....	25
4.4.1 城乡地区数字鸿沟.....	26
4.4.2 省际地区数字鸿沟.....	27
5 中国家庭数字鸿沟影响因素分析.....	31
5.1 指标选取与数据预处理.....	31
5.1.1 指标选取.....	31
5.1.2 数据来源.....	34
5.1.3 数据预处理.....	35
5.2 模型检验与对比.....	36
5.3 家庭数字鸿沟影响因素.....	38
5.3.1 模型估计结果.....	38
5.3.2 家庭影响因素.....	39
5.3.3 个人影响因素.....	40
5.3.4 地区影响因素.....	41
6 结论与建议.....	43
6.1 主要结论.....	43
6.1.1 地区数字鸿沟时空特征总结.....	43
6.1.2 家庭数字鸿沟影响因素总结.....	44
6.2 参考建议.....	45
参考文献.....	47
后 记.....	53

图表目录

图 4.1	2010-2020 年全国城乡互联网普及率对比折线图	16
图 4.2	2020 年全国互联网应用使用率.....	17
图 4.3	2014、2016 和 2018 年全国家庭信息使用水平空间分布图.....	25
图 4.4	2014、2016 和 2018 年全国城乡地区数字鸿沟水平对比折线图.....	27
图 4.5	2014、2016 和 2018 年中国地区数字鸿沟空间分布图.....	28
表 4.1	家庭信息使用水平指标体系	19
表 4.2	<i>KMO</i> 和 <i>Bartlett's</i> 球形检验结果	20
表 4.3	方差解释率	21
表 4.4	旋转后因子载荷矩阵	22
表 4.5	因子得分	22
表 4.6	家庭信息使用水平分级标准	23
表 4.7	地区数字鸿沟水平分级标准	26
表 4.8	家庭信息使用水平和地区数字鸿沟水平分布对比	29
表 5.1	个人特征变量赋值方法	32
表 5.2	指标变量说明	34
表 5.3	解释变量描述统计结果	36
表 5.4	解释变量多重共线性检验结果	37
表 5.5	模型对比假设检验结果	38
表 5.6	固定效应模型估计结果	39

1 引言

随着信息技术的不断变革和数字技术的广泛应用,以互联网为主要载体的数字经济迅速发展,成为经济和社会发展的主要推手。但由于信息和电子技术在获取和使用上的差异,造成了不同国家、地区、行业、个体之间的不平等,即数字鸿沟(Digital Divide)。

1.1 研究背景

习近平总书记在向 2021 年世界互联网大会乌镇峰会致贺信中指出,数字技术正以新理念、新业态、新模式全面融入人类经济、政治、文化、社会、生态文明建设各领域和全过程,给人类生产生活带来广泛而深刻的影响。而信息化、数字化、网络化、智能化发展不均衡所衍生的数字鸿沟问题将导致经济分化加剧、技术技能使用不平等以及文化进步程度加速扩大等一系列问题,严重阻碍经济社会发展(薛伟贤和张飞燕, 2009)。

2020 年初以来,新冠肺炎疫情导致社交隔离,工作、学习、交易、娱乐等业务加速线上迁移,继而引发社会生活和工作学习方式的深刻变革。普及手机支付,有效减少直接接触;外卖快递增加无接触配送方式,降低病毒传播风险;线上直播教学实现“停课不停学”,远程办公保证企业正常运营;手机“健康码”、“行程卡”实时更新,成为通行凭证……疫情加速了数字化进程,但也放大了不同人群在互联网融入方面的差异。2020 年 8 月,一段“老人无健康码乘地铁受阻”的视频在网上热传,引发公众对于老年人遭遇数字鸿沟窘境的热议,社会各群体间在数字资源的使用上存在较大差距。

此外,微观家庭层面的数字鸿沟也不容忽视。家庭互联网连接性的技术差异以及教育状况、收入水平和对待互联网的态度等因素导致家庭使用互联网的状况存在明显差距(Mills 和 Whitacre, 2003),而这种差距将在多个层面对家庭产生影响,研究表明数字鸿沟的存在显著降低了家庭金融投资的概率(张正平和卢欢, 2021),对家庭收入也存在显著的负向作用(尹志超, 2021)。

数字鸿沟存在于社会各个层次,是经济、地理、社会、文化、政策等宏观因素和性别、年龄、技能、受教育程度、收入水平等微观因素共同作用的结果(薛伟贤和张飞燕, 2009; 张正平和卢欢, 2020),其影响因素有着多层次的复杂关

系结构（薛伟贤和刘骏，2008）。

因此，应理清当前我国数字鸿沟问题现状，测算我国家庭信息使用水平，解析我国地区数字鸿沟的时间演变趋势和空间分布特征，明确家庭数字鸿沟的主要影响因素，并采取针对性治理措施，进而为缩小和弥合家庭数字鸿沟、促进数字化和智能化均衡发展等方面提供理论参考和现实依据。

1.2 研究内容

本文基于 2014-2018 年中国家庭追踪调查（CFPS）面板数据，采用因子分析、固定效应模型等方法，测算我国家庭信息使用水平，从时间和空间两个维度，分析城乡和省际尺度下我国地区数字鸿沟的静态特征和动态路径，同时从宏观和微观角度探究家庭数字鸿沟的主要影响因素，继而针对性提出措施建议。文章的主要结构如下。

第一章为引言。介绍中国家庭数字鸿沟现状及影响因素研究的背景和意义，阐明研究思路和框架，指出主要创新点和不足。

第二章为国内外研究综述。结合研究主题，主要从数字鸿沟的测度、数字鸿沟时空特征及驱动机制研究方法和数字鸿沟影响因素分析三个方面进行文献梳理和述评，并总结已有经验和尚可改进之处。

第三章为理论基础和主要方法。主要从数字鸿沟的概念辨析、地区和家庭数字鸿沟指标体系构建、因子分析法和面板数据回归模型等本文涉及到的理论基础与方法框架。

第四章为中国家庭数字鸿沟测算及时空特征分析。简单阐述我国互联网发展现状，基于 2014-2018 年中国家庭追踪调查（CFPS）面板数据，从信息技术使用维度，选取 5 个指标，使用虚拟变量法和因子分析法测算我国家庭信息使用水平，并通过变异系数法计算各地区数字鸿沟指数，进而从时间和空间两个维度，分析城乡和省际尺度下我国家庭信息使用水平及地区数字鸿沟的静态特征和动态路径。

第五章为中国家庭数字鸿沟影响因素探究。本部分基于 2014-2018 年 CFPS 数据、Wind 数据库、《中国统计年鉴》、《中国住户调查主要数据》和各省市统计年鉴，从家庭、个人和地区三个层面选取解释变量，以家庭数字鸿沟为被解释变

量,比较三种面板数据回归模型,依此对比检验进而选取最优形式,从宏观和微观角度探究家庭数字鸿沟的主要影响因素。

第六章为主要结论与措施建议。根据实证分析中国地区数字鸿沟时空分布特征和家庭数字鸿沟影响因素,总结本文所得结论;基于不同区域的家庭数字鸿沟及城市发展状况,对应提出措施建议。

1.3 创新点与局限性

1.3.1 主要工作及创新

研究视角上,同时考虑宏观和微观层面,测算我国宏观地区数字鸿沟指数和微观家庭数字鸿沟指数;在时空特征部分,分别从城乡和省际尺度描述地区数字鸿沟静态分布和动态路径,并结合国务院发展研究中心提出的八大综合经济区划分标准,对比分析家庭信息使用水平和地区数字鸿沟分布特征,揭示其潜在规律;在影响因素部分,从宏观和微观方面选取解释变量,探究家庭数字鸿沟的驱动机制。

指标选取上,基于中国家庭追踪调查(CFPS)面板数据,依据二级数字鸿沟(技术使用差距)定义,选取家庭网络使用行为相关变量,测算家庭信息使用水平,计算地区数字鸿沟指数,将二者作为时空特征分析的研究主体;在影响因素部分,从家庭、个人和地区三个层面选取解释变量,以家庭数字鸿沟指数为被解释变量,构建平衡面板数据。

研究方法上,结合计量经济学理论,分别使用混合回归模型、固定效应模型和随机效应模型进行拟合,通过检验判断数据是否存在由于不可观测的个体差异或异质性造成的遗漏变量问题,以此选取最优面板数据回归模型。

1.3.2 局限性

首先,受限于中国家庭追踪调查(CFPS)的调查范围,本文研究空间范围未能覆盖我国全部地区,同时由于早期相关数据严重缺失,且2020年数据尚未发布,本文采用2014年、2016年和2018年的面板数据。其次,由于数据难以获得,导致无法测度家庭成员间的作用机制,可在未来研究中尝试探索能反映所有

家庭成员综合水平的测算方法。最后，在影响因素探究部分，虽然本文从研究视角上有所创新，但部分指标更多作用于一级数字鸿沟，对二级数字鸿沟的影响微弱，可在后续研究中进一步探索。

2 国内外研究综述

随着数字鸿沟问题的日益突出，数字鸿沟的测度方法、时空演化特征和形成原因等科学问题受到国内外组织机构和学者的广泛关注。结合本文研究主题，本部分将主要从数字鸿沟的测度、时空特征和鸿沟影响因素三个方面进行文献梳理和述评。

2.1 数字鸿沟的测度

数字鸿沟问题涉及经济、政治、社会、科技和文化等多个方面，是一种综合性差异（薛伟贤和刘骏，2010），考虑到其内涵的复杂性，对数字鸿沟的测度方法一直是研究的热点。从现有文献来看，数字鸿沟测度模型以指数法为主，各个机构和学者的测度思路大体相似，均选取能够代表数字鸿沟具体表现的指标计算“鸿沟大小”，其本质为量化差距，但由于研究视角与研究对象的不同，所用指标和计算方法也存在一定差异（张正平和卢欢，2020）。

2.1.1 数字鸿沟测度模型指标选取

目前，绝大多数研究均采用多个指标构建数字鸿沟指标体系，仅有少数文献由于特定的研究内容或受限于相关数据选取单一指标进行测度。大部分指标体系的核心都是计算一个类似信息化水平的指数值，在研究视角上大致可以分为宏观视角和微观视角两类。

从宏观视角出发，研究对象多为不同国家和地区，对区域间的数字鸿沟进行分析。例如，Sicherl（1999）采用互联网普及率分析西欧和北美地区达到同一发展水平的时间差距；刘骏和薛伟贤（2012）选取信息技术意识、信息技术接入、信息技术利用和信息技术环境作为一级指标，测算我国城乡数字鸿沟；罗延锦和茶洪旺（2018）通过人均拥有电脑数量、互联网宽带接入用户比率、信息消费量、城镇化率和信息技术交易量等 23 个指标，计算我国 31 省市的数字鸿沟相对指数。

从微观视角出发，研究对象多为不同群体或个体，更多地考虑到不同人口统计特征，有助于探究数字鸿沟的深层次成因，为数字鸿沟的宏观分析提供微观依

据。例如，美国国家远程通信和信息管理局（National Telecommunications and Information Administration, NTIA）使用电话、电脑、互联网普及率测算不同收入、教育程度、种族、年龄和家庭结构人群之间的数字鸿沟（NTIA, 1999）；祝建华（2002）将人群分为不同阶层，通过网民比例计算各阶层网民占比与总体人口网民占比之差的绝对值，对其加权求和后得到指数值；Lucendo-Monedero 等（2019）从家庭和个人对电子商务、电子银行和电子政务服务等方面选取指标，构建家庭和个人数字发展指数。

2.1.2 数字鸿沟指数计算方法

在现有研究中，数字鸿沟的计算方法主要有以下几种：①绝对差距法和相对差距法（NTIA, 1999；金兼斌, 2003；刘骏和薛伟贤, 2012），在分析趋势时常采用绝对差距法，目前该方法的应用最为广泛，相对分析法则更适用于分析数字鸿沟的现状和程度，这两种方法直观易懂，但也存在明显弊端，当用相对差距法得出数字鸿沟在缩小时，用绝对差距法却可能得到相反的结论（张新红, 2008）；②基尼系数法（相对集中度指数），该方法利用基尼系数测算多组对象间不同信息技术方面的不均衡状况，但不够直观明晰，且不适合直接分析两个对象间的差异，应用范围较窄（张新红, 2008）；③时间距离法（Sicherl, 1999），由斯洛文尼亚教授 Pavle Sicher 提出，基于时间序列数据，计算不同地区互联网普及率达到同一水平的的时间差距，以此测度数字鸿沟，该方法提供了一种动态比较数字鸿沟的新思路，但主要用于单一指标，不适合分析较复杂的多组研究对象；④相对差距综合指数法（张新红, 2008；罗延锦和茶洪旺, 2018），是相对差距法的一种扩展，可以静态或动态分析一国或地区多种数字鸿沟的总水平和时空演变特征，国家信息中心信息化研究部 2007 年发布的《中国数字鸿沟报告》中使用的便是此方法；⑤虚拟变量法（宋红岩, 2016；张丽等, 2018；张正平和卢欢, 2021），将反映 ICT 水平的指标用 0 或 1 表示，该方法主要用于衡量个体间数字鸿沟；⑥变异系数法（祝坤子等, 2015；宋周莺和刘卫东, 2013），对区域内各省信息化发展指数求变异系数，以此度量该区域内省际之间信息化发展的差异，数据值越离散，说明该区域数字鸿沟越大。

计算指数时，不同学者确定各指标权重的方法也不尽相同，最为常用的是算

术平均法（张新红，2008），该方法较为简单，但各指标权值相同无法体现其贡献大小；另有学者选择层次分析（张彬，2009）、因子分析（刘骏和薛伟贤，2012；张正平和卢欢，2021）等方法，相比前者更为科学。

2.2 数字鸿沟时空特征及驱动机制研究方法

时空特征和驱动机制是目前对数字鸿沟的主要研究方向之一，由于对数字鸿沟的界定、研究视角和指标选取各不相同，相关文献选用的研究方法和所得结论也不尽一致。

在时空分布的相关研究中，时间演变趋势与空间分布特征相结合，现有文献多基于年份尺度，通过分析对比各国家/地区网民数量（柯惠新和王锡苓，2005）、互联网渗透率（汪明峰，2005）和各类数字鸿沟指数（刘骏和薛伟贤，2012；李健和邬晓鸥，2018）等指标，描述归纳数字鸿沟演化特征，具体研究方法有 ArcGIS 空间统计分析（祝坤子等，2015）、Lorenz 曲线（邱娟和汪明峰，2010）、ICTs 扩散模型（薛伟贤和刘骏，2011）等。研究的尺度主要为国家（汪明峰，2005；祝坤子等，2015）、城乡（邱娟和汪明峰，2010；Philip 等（2017）；李健和邬晓鸥，2018）和省际（刘骏和薛伟贤，2012；祝坤子等，2015）。

在驱动机制的相关研究中，针对不同数据类型，分析方法也有所不同。①基于时间序列数据的研究较少，薛伟贤和王涛峰（2006）选取 1997-2004 年我国相关数据，分别对经济发展水平、信息资源、科教投入和人才结构四个影响因素进行回归分析，由此对比各因素影响程度，但未顾及地区异质性。②使用横截面数据时，多采用线性回归模型（胡鞍钢和周绍杰，2002；薛伟贤和王江泉，2011）判断各影响因素是否与数字鸿沟显著相关。③基于一段时间内多个区域的面板数据研究较为普遍，研究方法以简单多元回归模型（Guillen 等，2005；刘芸，2007；Cruz-Jesus 等，2018）为主，但此方法忽略了个体特征差异。④在分析通过问卷调查法获得的数据时，因变量可能是分类变量，这种情况下更适合使用 Logistics 模型（徐越等，2020）。

此外，有个别学者对数字鸿沟影响因素的关系结构进行研究，均通过构建解释结构模型（ISM）分析各因素间的互相影响，得到清晰的影响因素拓扑图（薛伟贤和刘骏，2008；张彬等，2009；黄金和赵冬梅，2011）。但在选取影响因素时

多通过文献计量法和专家经验判断，无法避免主观影响。

2.3 数字鸿沟影响因素分析

数字鸿沟的成因较为复杂，是多种因素共同作用的结果，既涵盖经济、地理、政治、社会等宏观因素，又包括性别、年龄、收入水平、受教育程度、个体技能等微观因素。

①经济因素对数字鸿沟的影响几乎在所有相关文献中都非常显著，是形成数字鸿沟的根本原因（张正平和卢欢，2020）。Chinn 和 Fairlie（2007）对 161 个国家进行调研，研究发现导致全球性数字鸿沟的最主要原因就是经济发展水平。在国家层面上，Billon 等（2010）分析 142 个国家的数据，发现 GDP 是唯一对所有国家数字化发展均有显著影响的因素；汪明峰（2005）基于空间层面探究影响我国互联网普及程度的影响因素，回归结果显示人均地区生产总值（GRP）的作用最为明显。

②技术层面差异也是数字鸿沟的主要驱动因素之一。Mumporeze 和 Prieler（2017）采用 2005-2013 年调查数据对巴西数字鸿沟成因进行研究，指出创新技能的差异会影响 ICT 产品的使用，与其他因素相互作用下最终导致了数字鸿沟的形成。信息基础设施是使用 ICT 的物质基础（刘芸，2007），也是推动互联网快速发展的关键要素，不过目前传统信息基础设施普及度已相对饱和，无法有效反映互联网发展的省际差异，只有较为先进的信息基础设施与互联网发展水平存在较为显著的联系（邱娟和汪明峰，2010）。

③区域开放程度在一定程度上体现了某一国家/地区获取和利用外部知识的程度，进而决定了该区域的信息技术连接能力（汪明峰，2005），且开放程度对数字鸿沟的影响存在递增效应，会随着时间的增长而累积，产生持久的长期影响（韦路和谢点，2015）。

④城市化率往往成为决定一个地区互联网使用状况的主要因素之一。汪明峰（2005）研究发现我国各省区的城市化水平与互联网渗透率之间存在非常显著的关联性，计量回归结果表明城市化率对地区互联网使用水平的解释力高达 77%。同时，非农化水平对于我国省际数字鸿沟的影响正在不断增加（邱娟和汪明峰，2010）。

⑤还有个人和家庭层面的影响因素。陆峰等（2007）指出同一社会群体中，数字鸿沟的形成原因有收入水平、文化程度、年龄、性别等方面；王云生（2007）则认为智力、技能和种族等因素也会产生影响；袁勤俭等（2005）对 1984-2000 年美国不同类型家庭的计算机普及率进行研究，发现已婚且孩子小于 18 岁双亲家庭始终最高，且各类家庭间的差距在逐步扩大；Mumporeze 和 Prieler（2017）指出，数字文盲是巴西数字鸿沟的主要决定因素之一，尤其是在老年人中；也有少量研究显示，性别不再是主要成因，男女之间的数字鸿沟正在消失（洪海娟和万跃，2014）。

2.4 简要述评

综上所述，国内外学者已针对数字鸿沟的测度、时空特征和影响因素等科学问题展开大量研究，在研究视角、数据来源、空间范围和研究方法等方面存在异同，同时可能具有一定局限性：①宏观视角下区域间信息化水平差异的分析，虽然考虑了区域和社会的异质性，但忽略了数字鸿沟微观层面的成因；②基于微观视角的研究多以个体为研究对象，如老年人、大学生等，鲜有学者以家庭、企业等单位探讨数字鸿沟；③微观层面研究大多使用单次问卷调查所得的截面数据，无法分析此类数字鸿沟的时间演化趋势；④对于各因素影响程度的研究大多使用简单多元线性回归模型，但此方法忽略了数据的个体效应。本文将基于 2014-2018 年中国家庭追踪调查（CFPS）面板数据，分别测算地区数字鸿沟指数和家庭数字鸿沟，并通过检验选取计量经济模型分析家庭数字鸿沟的影响因素。

3 理论基础和主要方法

本文将通过因子分析法测算家庭信息使用水平,计算地区数字鸿沟指数和家庭数字鸿沟指数,以考察我国数字鸿沟的时空演化特征,并基于面板数据回归模型探究家庭数字鸿沟的影响因素。据此,本部分主要介绍数字鸿沟概念辨析、数字鸿沟测算方法和面板数据回归模型。

3.1 数字鸿沟定义

“数字鸿沟”这一概念出现在 20 世纪 90 年代末期,其来源众说纷纭。起初,数字鸿沟被简单定义为能够获得和不能获得信息通信技术 (Information and Communications Technology, ICT) 的人之间的差距 (Riggins 和 Dewan, 2005; Dijk, 2006),这一定义仅体现了数字鸿沟在个人层面的单一现象 (Hwang 和 Joo-Seong, 2004),忽略了宏观层面以及 ICT 在使用上的差异 (Brandtzaeg 等, 2011)。此后,许多国际组织和机构都对数字鸿沟的内涵进行了扩展。经济合作与发展组织 (Organization for Economic Cooperation and Development, OECD) 提出了一个具有代表性的概念,即数字鸿沟是“不同社会经济水平的个人、家庭、商业部门和地理区域,在接入信息通信技术以及利用互联网从事各种活动的机会上存在的显著差距”(OECD, 2001)。国际电信联盟 (International Telecommunication Union, ITU) 认为,数字鸿沟是由于贫富、受教育水平、教育设施等因素而形成的,获取信息和通信新技术方面的不平等,发达国家与发展中国家之间、城市与农村之间、年轻一代与老一代之间均存在数字鸿沟。

随着研究的逐步深入,学者们根据不同表现形式将数字鸿沟分为三个层次。一级数字鸿沟是指由基础设施服务差异导致的技术接入差距,是数字鸿沟最基础的形式,但随着互联网渗透率的逐步提高,这一差距正在不断缩小 (Hwang 和 Joo-Seong, 2004),因此,一级数字鸿沟已经不是研究的主要方向;二级数字鸿沟指由经济发展水平差异导致的技术使用差距 (Riggins 和 Dewan, 2005; 许竹青等, 2013),尽管越来越多的地区和个体可以获得 ICT,但在应用层面仍然呈现两极分化态势 (张正平和卢欢, 2020),这也成为当前研究的热点问题;三级数字鸿沟则是指前两级接入和使用差异所造成的不平等后果及影响 (Robles 等, 2011; Wei 等, 2011; Raileanu, 2018),数字鸿沟不只是 ICT 普及的问题,它是

经济不平等、知识差距、阶层分化在信息时代的体现（薛伟贤和刘骏，2010），是包括由此产生的贫富分化、机会不均和社会不公等问题的复杂现象（张正平和卢欢，2020），三级数字鸿沟正在逐渐引起关注。

结合现有研究结果和我国互联网发展现状，本文基于二级数字鸿沟，即技术使用差距层面，测算和分析我国数字鸿沟时空演化特征及影响因素。

3.2 数字鸿沟测算

从现有文献来看，数字鸿沟测度模型以指数法为主，首先从信息技术接入、使用等层面选取多个指标计算信息化水平，然后根据数据类型和分析问题的需要采用绝对差距、相对差距和虚拟变量等方法得出数字鸿沟指数。

3.2.1 家庭信息使用水平

目前以微观视角测算信息化水平的研究较少，尚未有统一的计算标准，本文借鉴张正平和卢欢（2021）的方法，选取能表示受访者使用互联网进行各项活动频率的虚拟变量，进行因子分析。因子分析法是多元统计分析中的典型方法之一，通常用于降维和简化数据（朱建平，2016）。它通过研究众多原始变量之间的相关关系，探求观测数据中的基本结构，用“抽象”变量表示原始变量中的数据结构，这些抽象变量就是“因子”。原始变量是容易收集的可观测的显在变量，因子则是无法直接得到的不可观测的潜在变量，因子分析就是寻找潜在的起支配作用的因子模型的方法。设有原始变量 $x_1, x_2, x_3, \dots, x_n$ ，它们与潜在因子之间的关系可以表示为：

$$\begin{cases} x_1 = b_{11}z_1 + b_{12}z_2 + b_{13}z_3 \cdots + b_{1n}z_n + e_1 \\ x_2 = b_{21}z_1 + b_{22}z_2 + b_{23}z_3 \cdots + b_{2n}z_n + e_2 \\ x_3 = b_{31}z_1 + b_{32}z_2 + b_{33}z_3 \cdots + b_{3n}z_n + e_3 \\ \vdots \\ x_n = b_{n1}z_1 + b_{n2}z_2 + b_{n3}z_3 \cdots + b_{nn}z_n + e_n \end{cases} \quad (3.1)$$

式中， $z_1 \sim z_n$ 为 n 个潜在因子，是各原始变量都包含的因子，称共性因子； $e_1 \sim e_n$ 为 n 个只包含在某个原始变量之中的，只对一个原始变量起作用的个性因子，是各变量特有的特殊因子，共性因子与特殊因子相互独立。

3.2.2 地区数字鸿沟指数

为探究不同区域划分标准下我国家庭数字鸿沟的空间分布特征,更好地度量各地区间家庭信息使用水平发展的差异,同时消除数据尺度和量纲的影响,本文参考祝坤子等人(祝坤子等,2015;宋周莺和刘卫东,2013)的做法,选用变异系数法,对不同地区各个家庭的信息使用水平求变异系数,得到地区数字鸿沟指数,数据值越离散,说明该区域家庭数字鸿沟越大。

3.2.3 家庭数字鸿沟指数

为探究微观层面影响家庭数字鸿沟的因素,解释变量中含有个人特征变量和家庭特征变量,为统一口径,被解释变量不宜选取地区数字鸿沟指数,而应针对每个受访家庭计算家庭数字鸿沟水平。因此,本文沿用张正平和卢欢(2021)的计算方法,用信息技术使用水平最高的家庭,即家庭信息使用水平最大值,分别减去每个家庭信息使用水平的值,得到各个家庭的数字鸿沟指数。

3.3 面板数据回归模型

3.3.1 面板数据

面板数据(panel data)是指在一段时间内跟踪一组个体的数据,既有横截面维度(n 位个体),又有时间维度(T 个时期),当 n 较大, T 较小时,被称为“短面板”;反之,如果 n 较小, T 较大,被称为“长面板”。在面板数据中,如果每个时期在样本中的个体完全一样,则称为“平衡面板数据”,反之,则称为“非平衡面板数据”。在面板模型中,如果解释变量包含被解释变量的滞后值,被称为“动态面板”;反之,则称为“静态面板”。

相比于截面数据和时间序列数据,面板数据的样本容量更大,估计精度更高,可以解决由于不可观测的个体差异或异质性造成的遗漏变量问题,而且能提供更多个体动态行为信息,但收集成本较高,不易获得(陈强,2014)。

3.3.2 面板数据模型类型

为了避免极端策略导致忽略个体间的异质性或者共性,估计面板数据一般假定个体的回归方程拥有相同的斜率和不同的截距,这种模型被称为“个体效应模型”,基本形式可以表示为:

$$y_{it} = x_{it}'\beta + z_i' + u_i + \varepsilon_{it} \quad (i=1, \dots, n; t=1, \dots, T) \quad (3.2)$$

式中, z_i 为不随时间而变的个体特征, x_{it} 为随个体和时间而变的解释变量, $(u_i + \varepsilon_{it})$ 为复合扰动项, 其中, u_i 代表不可观测的个体异质性的截距项。如果 u_i 不存在, 则称为“混合回归”; 如果 u_i 与某个变量相关, 则称为“固定效应模型”

(FE 模型); 如果 u_i 与所有解释变量均不相关, 则称为“随机效应模型”(RE 模型)。在实证检验中, 一般采用豪斯曼检验来判断应该选取固定效应模型还是随机效应模型。

(1) 混合回归

混合回归将面板数据看成截面数据, 假设不存在个体效应, 有时甚至假定解释变量是严格外生的, 模型中所有个体拥有完全相同的回归方程, 则式 (3.2) 可写为:

$$y_{it} = \alpha + x_{it}'\beta + z_i' + \varepsilon_{it} \quad (3.3)$$

式中, x_{it} 不包括常数项的随个体和时间而变的解释变量。对于混合回归的基本假设, 必须进行统计检验, 以判断是否存在个体固定效应和随机效应。

(2) 固定效应模型

个体固定效应模型解决了不随时间而变但随个体而异的遗漏变量问题。给定个体 i , 将式 (3.2) 两边对时间取平均, 再用式 (3.2) 减去平均后的方程可得原模型的离差形式, 再定义 $y_{it} \equiv y_{it} - \bar{y}_i$, $x_{it} \equiv x_{it} - \bar{x}_i$, $\varepsilon_{it} \equiv \varepsilon_{it} - \bar{\varepsilon}_i$, 则可得到:

$$y_{it} = x_{it}'\beta + \varepsilon_{it} \quad (3.4)$$

若 ε_{it} 与 x_{it} 不相关, 则可以用 OLS 一致地估计 β , 称为“组内估计量”, 记为 β_{FE} , β_{FE} 使用了每位个体的组内离差信息, 即使个体特征与解释变量相关, 也

可以得到一致估计。

引入时间固定效应 $\lambda_t \equiv \gamma S_t$ (S_t 不可观测), 则可以解决不随个体而变但随时间而变的遗漏变量问题。模型设定为:

$$y_{it} = x_{it}'\beta + z_i'\delta + \lambda_t + u_i + \varepsilon_{it} \quad (3.5)$$

若使用 LSDV 法, 对每个时期定义一个虚拟变量, 然后把 $(T-1)$ 个时间虚拟变量保活在回归方程中, 则既考虑了个体固定效应, 又考虑了时间固定效应。

(3) 随机效应模型

对于式 (3.2), 随机效应模型假定 u_i 与解释变量 $\{x_{it}, z_i\}$ 均不相关, 但由于扰动项由 u_i 和 ε_{it} 组成, OLS 不是最有效率的。假设不同个体之间的扰动项互不相关, 由于 u_i 的存在, 同一个体不同时期的扰动项之间存在自相关:

$$Cov(u_i + \varepsilon_{it}, u_i + \varepsilon_{is}) = \begin{cases} \sigma_u^2, & \text{若 } t \neq s \\ \sigma_u^2 + \sigma_\varepsilon^2, & \text{若 } t = s \end{cases} \quad (3.6)$$

式中, σ_u^2 为 u_i 的方差 (不随 i 变化), σ_ε^2 为 ε_{it} 的方差 (不随 i 、 t 变化)。当 $t \neq s$ 时, 其自相关系数为:

$$\rho \equiv Corr(u_i + \varepsilon_{it}, u_i + \varepsilon_{is}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2} \quad (3.7)$$

可以看出, ρ 不随时间距离 ($t-s$) 而改变, ρ 越大, 复合扰动项 ($u_i + \varepsilon_{it}$) 中的个体效应部分 (u_i) 越重要。

4 中国家庭数字鸿沟时空特征分析

本部分将基于 2014-2018 年中国家庭追踪调查 (CFPS) 面板数据, 测算我国家庭信息使用水平, 计算地区数字鸿沟指数, 进而从时间和空间两个维度, 分析城乡和省际尺度下我国家庭信息使用水平及地区数字鸿沟的静态特征和动态路径。

4.1 全国互联网发展概况

在测度地区数字鸿沟水平前, 首先从普及度和使用率两个角度对我国互联网发展情况进行简要描述。

4.1.1 互联网普及率

2010 年 6 月 8 日, 国务院新闻办公室首次发表《中国互联网状况》白皮书, 说明了中国政府关于互联网的基本政策: “积极利用、科学发展、依法管理、确保安全”。此后十年间, 中国的互联网事业稳步提高、持续飞速发展, 是世界上信息化发展最快的国家之一。中国互联网络信息中心 (CNNIC) 统计数据显示, 截至 2010 年 12 月 31 日, 中国网民规模为 4.57 亿人, 其中, 手机网民 3.03 亿人, 约占 66.3%; 全国互联网普及率 (图 4.1) 首次超过 30%, 城镇互联网普及率达 50.2%, 但农村互联网普及率仍然较低, 仅为 16.1%。2015 年, 全国互联网普及率突破 50%, 农村互联网普及率也首次达到 30% 以上。“十三五”期间, 我国在以互联网为代表的信息通信技术各相关领域均取得前所未有的历史性成就, 互联网基础设施建设全面覆盖, 互联网普惠深入推进, 数字经济欣欣向荣, 高新技术加快探索, 互联网治理逐步完善。截至 2020 年 12 月, 我国网民规模达 9.89 亿, 约占全球网民的五分之一, 为全球最大的数字社会, 其中, 手机网民规模为 9.86 亿, 占比高达 99.7%; 互联网普及率达 70.4%, 同比增长 9.2 个百分点, 农村互联网普及工作力度加大, 已达 55.9%, 虽然与城镇互联网普及率之间的差异仍超过 20%, 但已出现缩减趋势。

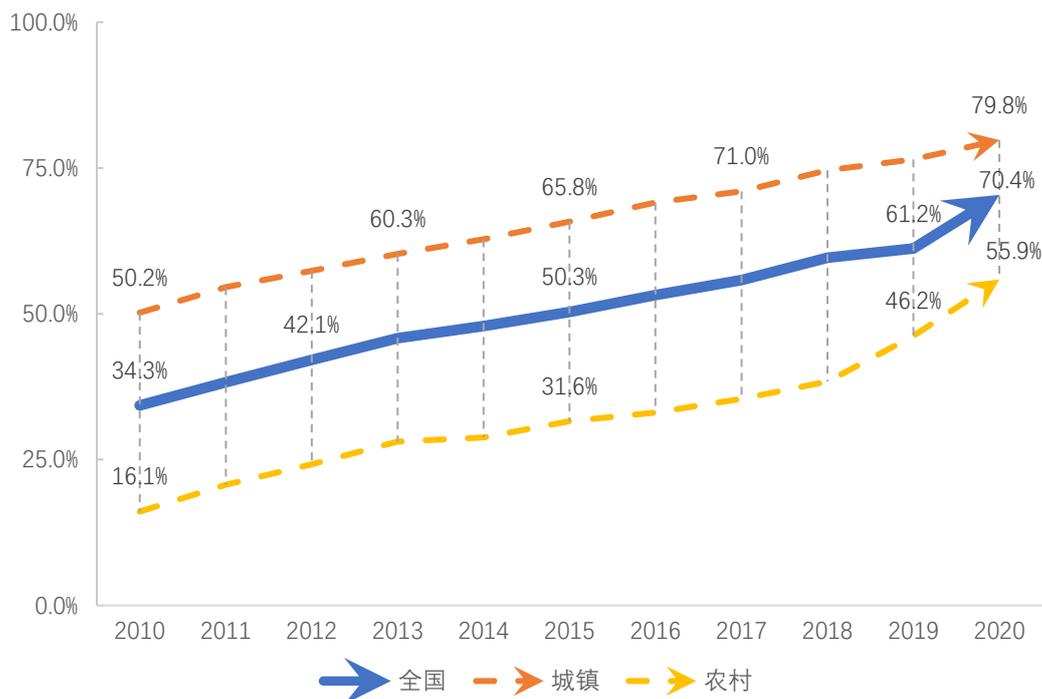


图 4.1 2010-2020 年全国城乡互联网普及率对比折线图

4.1.2 互联网应用使用率

尽管我国居民在互联网接入层面的差异日益缩小，接入维度数字鸿沟正在逐渐消失，但在使用维度仍存在较大不平衡。

CNNIC 发布的第 47 次《中国互联网络发展状况统计报告》公布了 2020 年我国网民各类互联网应用用户和使用率（图 4.2）情况，主要分为基础应用、商务交易、网络娱乐和公共服务四类。基础应用类中，即时通信的使用率高达 99.2%，为所有互联网应用渗透率第一，搜索引擎和网络新闻均超过 75%，但较 2019 年有所下降，远程办公用户在新冠肺炎疫情影响下出现大幅增长，但仍处在发展初期，使用率为 34.9%；商务交易类应用使用率差异相对较大，网络支付最高，达 86.4%，互联网理财仅 17.2%，为所有应用渗透率最低；网络娱乐类应用发展较快，网络视频用户占有所有互联网用户的 93.7%，仅次于即时通信，其中，短视频拥有绝对优势，88.3% 的互联网用户均会使用短视频应用，此外，网络音乐、网络直播和网络游戏使用率均超过 50%；公共服务类应用使用率普遍较低，且增速较低，仍有很大的发展空间。

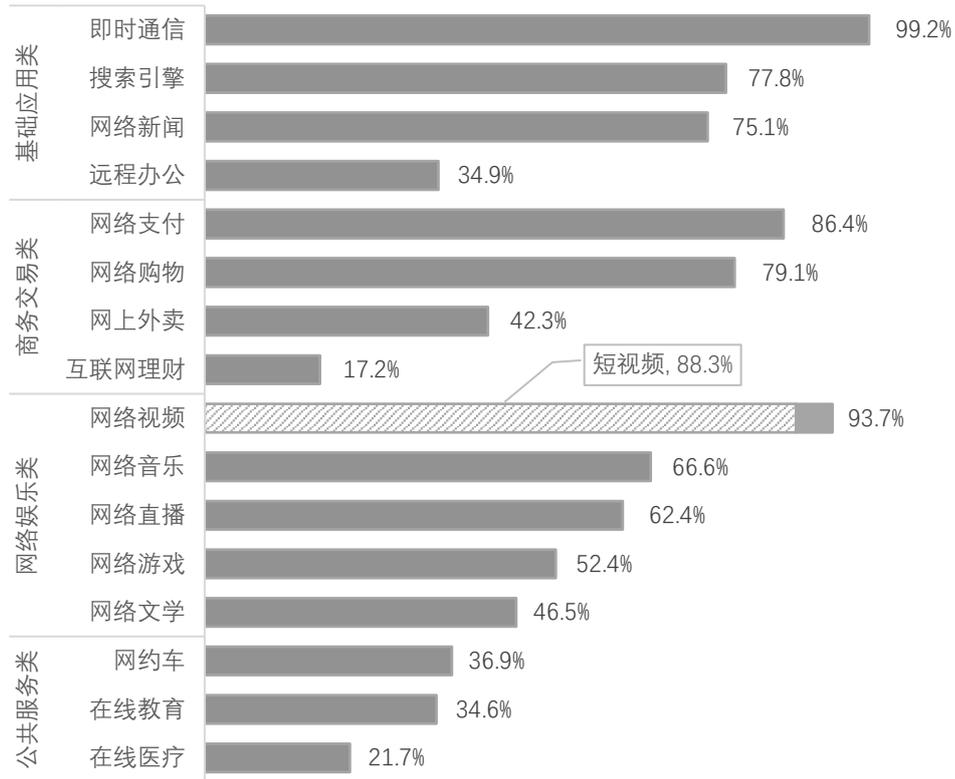


图 4.2 2020 年全国各类互联网应用使用率

4.2 地区数字鸿沟指数测算

基于中国家庭追踪调查（CFPS）数据，从家庭信息技术使用维度选取指标，对问卷数据进行清洗合并，采用因子分析法计算家庭信息使用水平，并通过变异系数法得出各地区数字鸿沟指数。

4.2.1 数据来源

本部分从中国家庭追踪调查（CFPS）获取数据，该调查是一项全国性、综合性的社会追踪调查项目，由北京大学研究团队设计、北京大学以及自然科学基金资助、北京大学中国社会科学调查中心（ISSS）负责实施。CFPS 以个人和家庭为研究主体，调查包括家庭关系、家庭经济、收入、教育和婚姻等主题，并特别调查了受访者的互联网使用情况。基线样本覆盖 25 个省/市/自治区、14960 户

家庭、42590 位个人，代表了中国 95%的人口，并对个人样本进行两年一期的追踪调查，以期反映中国社会、经济、教育和身心健康等方面的变迁(谢宇等,2014)，具有较高权威性和学术研究价值。CFPS 数据主要分为家庭关系、家庭经济、成人库、少儿库和村居库，本部分所需数据主要来自成人库。

4.2.2 时空范围界定

研究时间维度上，由于 2010 年和 2012 年家庭信息相关数据缺失 80%以上，不具有代表性，且 2020 年 CFPS 数据尚未发布，因此选取 2014 年、2016 年和 2018 年三期调查数据展开研究。

研究空间维度上，CFPS 的基线调查对象为中国（不含香港、澳门、台湾以及新疆维吾尔自治区、西藏自治区、青海省、内蒙古自治区、宁夏回族自治区、海南省）25 个省/市/自治区中的家庭户和样本家庭户^①中的所有家庭成员^②（Xie 和 Lu, 2015），本部分及后续研究均选取与之相同的空间范围，并在部分分析中参考国务院发展研究中心提出的八大综合经济区^③划分我国地理区域。

4.2.3 指标选取和数据预处理

当前，随着互联网渗透率的逐步提高，一级数字鸿沟（技术接入差距）正在不断缩小、逐步弥合（Hwang, 2004），但二级数字鸿沟（技术使用差距）仍处于较高水平，各个家庭在信息技能和应用方面呈现两极分化态势（张正平和卢欢, 2020），加剧了人们在社会生活中的不平等地位（Chadwick, 2010）。因此，本文基于二级数字鸿沟的定义，从信息技术使用维度选取指标，计算各个家庭的信息化水平。

家庭关系应该是由多个核心连接起来的树状的网络，因此，CFPS 克服其他

^① 居住在传统居民住宅内的、家中至少有一人拥有中国国籍的一个独立经济单元，便可视为一个满足项目访问条件的家庭户。

^② CFPS 定义的家庭成员指样本家庭中经济上联系在一起的直系亲属，或经济上联系在一起、与该家庭有血缘/婚姻/领养关系且连续居住时间满 3 个月的非直系亲属。

^③ 东北地区（辽、吉、黑）、北部沿海地区（京、津、冀、鲁）、东部沿海地区（沪、苏、浙）、南部沿海地区（闽、粤、琼）、黄河中游地区（陕、晋、豫、蒙）、长江中游地区（鄂、湘、赣、皖）、西南地区（滇、黔、川、渝、桂）以及大西北地区（甘、青、宁、藏、新）。

调查选取单一成员为“户主”的局限性，从经济角度出发，设计了主事者、决策者、财务管理者 and 房产所有者的概念。考虑到后续分析数字鸿沟影响因素的指标选取，本文使用财务管理者的网络使用行为代表该家庭户的信息使用水平，以此计算数字鸿沟指数。从学习、工作、社交、娱乐和商业活动五个使用维度选取变量构建指标体系，具体构成如表 4.1 所示。

表 4.1 家庭信息使用水平指标体系

变量	取值范围	取值标签
使用互联网学习的频率 (x_1)	0-6	0 代表“从不”
使用互联网工作的频率 (x_2)	0-6	1 代表“几个月一次” 2 代表“一月一次”
使用互联网社交的频率 (x_3)	0-6	3 代表“一月 2-3 次”
使用互联网娱乐的频率 (x_4)	0-6	4 代表“一周 1-2 次” 5 代表“一周 3-4 次”
使用互联网商业活动的频率 (x_5)	0-6	6 代表“几乎每天”

除所在地区、城乡分类和上述五个变量以外，还保留了家庭编码和个人编码变量，便于后续分析时与个人特征变量和家庭经济数据库进行匹配合并，确定指标后对原始数据进行清洗。首先，提取指标后删除回答为“无法判断”、“缺失”、“不适用”、“拒绝回答”和“不知道”的样本；其次，为进一步得到合理有效的因子分析结果，将使用维度五个变量的取值由 1-7（依次按频率降序赋值）改为 0-6（依次按频率升序赋值）；此外，人口流动导致部分受访家庭所在地区为新疆维吾尔自治区、西藏自治区、青海省、内蒙古自治区、宁夏回族自治区和海南省，不在 2010 年基线调查时限定的空间范围内，但仅占 6.2%，不足 50 户，缺乏对上述地区家庭的代表性，因此予以剔除。最终得到 2014 年 4057 户、2016 年 5070 户和 2018 年 5186 户家庭样本。

4.2.4 地区数字鸿沟指数计算

首先对变量 $x_1 - x_5$ 进行 *KMO* 和 *Bartlett's* 球形检验，结果如表 4.2 所示。2014

年、2016年和2018年的 *KMO* 值均大于 0.6, 说明变量间偏相关性较强; *Bartlett's* 球形检验显著性均小于 0.05, 可以拒绝原假设, 即相关阵不是单位阵, 说明变量之间存在相关关系。检验表明数据适合做因子分析, 且效果较好。

表 4.2 *KMO* 和 *Bartlett's* 球形检验结果

年份	<i>KMO</i>	<i>Bartlett's</i> 球形检验		
		近似卡方	<i>df</i>	<i>p</i>
2014	0.680	2179.087	10	0.000
2016	0.714	4031.301	10	0.000
2018	0.713	5220.326	10	0.000

由于变量 $x_1 - x_5$ 均表示受访者使用互联网进行各项活动的频率, 取值范围及标签一致, 不存在量纲差异, 因此可以直接使用原始变量进行因子分析, 不需要进行标准化处理。分析后得到因子旋转前和旋转后的特征根和方差解释率, 结果如表 4.3 所示。可以看到 2014 年、2016 年和 2018 年均保留两个因子, 其初始特征根均大于 1, 累积贡献率均大于 60%。

旋转后的因子载荷阵 (表 4.4) 显示了各个因子在每个变量上的载荷, 可以看出, 2014 年、2016 年和 2018 年的因子载荷非常相似, 均有一个因子在变量 x_1 和 x_2 上有较大载荷, 另一个因子则在变量 x_3 和 x_4 上有较大载荷; 但变量 x_5 (使用互联网商业活动的频率) 在 2014 年与变量 x_3 (使用互联网社交的频率) 和 x_4 (使用互联网娱乐的频率) 有较强相关性, 在 2016 年和 2018 年则与变量 x_1 (使用互联网学习的频率) 和 x_2 (使用互联网工作的频率) 更为相关。

表 4.3 方差解释率

年份	因子	初始特征值			旋转后方差解释率		
		特征根	方差解释率 (%)	累积 (%)	特征根	方差解释率 (%)	累积 (%)
2014	f_{11}	2.102	0.420	0.420	1.596	0.319	0.319
	f_{12}	1.083	0.217	0.637	1.589	0.318	0.637
	f_{13}	0.678	0.136	0.773			
	f_{14}	0.643	0.129	0.901			
	f_{15}	0.495	0.099	1.000			
2016	f_{21}	2.201	0.440	0.440	1.783	0.357	0.357
	f_{22}	1.018	0.204	0.644	1.437	0.287	0.644
	f_{23}	0.695	0.139	0.783			
	f_{24}	0.600	0.120	0.903			
	f_{25}	0.485	0.097	1.000			
2018	f_{31}	2.333	0.467	0.467	1.865	0.373	0.373
	f_{32}	1.017	0.204	0.670	1.485	0.297	0.670
	f_{33}	0.663	0.133	0.803			
	f_{34}	0.589	0.118	0.921			
	f_{35}	0.397	0.080	1.000			

表 4.4 旋转后因子载荷矩阵

变量	2014 年		2016 年		2018 年	
	f_{11}	f_{12}	f_{21}	f_{22}	f_{31}	f_{32}
x_1	-0.009	0.847	0.822	0.036	0.854	0.068
x_2	0.200	0.819	0.836	0.116	0.869	0.126
x_3	0.741	0.134	0.143	0.758	0.167	0.760
x_4	0.793	-0.010	0.073	0.809	0.076	0.820
x_5	0.615	0.427	0.617	0.441	0.590	0.463

由因子得分系数矩阵（表 4.5）可以得出各个因子的因子得分表达式：

$$f_{11} = -0.196x_1 - 0.044x_2 + 0.487x_3 + 0.556x_4 + 0.333x_5 \quad (4.1)$$

$$f_{12} = 0.596x_1 + 0.530x_2 - 0.072x_3 - 0.185x_4 + 0.162x_5 \quad (4.2)$$

$$f_{21} = 0.518x_1 + 0.507x_2 - 0.100x_3 - 0.157x_4 + 0.284x_5 \quad (4.3)$$

$$f_{22} = -0.179x_1 - 0.119x_2 + 0.566x_3 + 0.625x_4 + 0.195x_5 \quad (4.4)$$

$$f_{31} = 0.516x_1 + 0.510x_2 - 0.097x_3 - 0.170x_4 + 0.246x_5 \quad (4.5)$$

$$f_{32} = -0.173x_1 - 0.132x_2 + 0.553x_3 + 0.624x_4 + 0.207x_5 \quad (4.6)$$

表 4.5 因子得分

变量	2014 年		2016 年		2018 年	
	f_{11}	f_{12}	f_{21}	f_{22}	f_{31}	f_{32}
x_1	-0.196	0.596	0.518	-0.179	0.516	-0.173
x_2	-0.044	0.530	0.507	-0.119	0.510	-0.132
x_3	0.487	-0.072	-0.100	0.566	-0.097	0.553
x_4	0.556	-0.185	-0.157	0.625	-0.170	0.624
x_5	0.333	0.162	0.284	0.195	0.246	0.207

为进一步进行综合评价,还需要将因子以各自的方差贡献率占累计贡献率的比重作为权重来加权计算综合得分:

$$F_1 = \frac{0.319}{0.637} f_{11} + \frac{0.318}{0.637} f_{12} = 0.501f_{11} + 0.499f_{12} \quad (4.7)$$

$$F_2 = \frac{0.357}{0.644} f_{21} + \frac{0.287}{0.644} f_{22} = 0.554f_{21} + 0.446f_{22} \quad (4.8)$$

$$F_3 = \frac{0.373}{0.670} f_{31} + \frac{0.297}{0.670} f_{32} = 0.557f_{31} + 0.443f_{32} \quad (4.9)$$

式中, F_1 、 F_2 和 F_3 分别表示2014年、2016年和2018年的家庭信息使用水平。

按年份分别对各地区各样本家庭户的家庭信息使用水平求变异系数,以此反映各地区家庭之间信息使用水平的离散程度,即地区数字鸿沟水平,计算公式如下:

$$DDI_i = \frac{\sigma_i}{\mu_i} \quad (4.10)$$

式中, DDI_i 表示*i*地区数字鸿沟指数, σ_i 表示*i*地区家庭信息使用水平标准差, μ_i 表示*i*地区家庭信息使用水平均值。

4.3 家庭信息使用水平分布特征

分别求出25个省/市/自治区家庭信息使用水平的样本均值,对其进行描述性分析。本文基于Quartile分类方法,以25个省/市/自治区家庭信息使用水平的25%、50%和75%分位数为界,整体划分成I-IV级,信息使用水平逐级增高,具体分级标准如表4.6所示。

表 4.6 家庭信息使用水平分级标准

描述性统计量		地区家庭信息使用水平
最小值		2.5405
分位数	25%	2.9046
	50%	3.1741
	75%	3.3854
最大值		4.7247

根据各地区家庭信息使用水平级别，分别绘制 2014 年、2016 年和 2018 年的地图以展示其时空特征（图 4.3）。整体来看，时间维度上家庭信息使用水平持续升高，IV 级地区逐年递增，占比由 8% 增至 40%，I 级地区则从 11 个降为一个，发展速度较快；空间维度始终呈现“东高西低，南高北低”的梯形分布，与经济发展水平空间特征较为相似。其中，北京和上海始终处于最高水平；河北、山东、江西、河南三年连升，分布于北部沿海、长江中游和黄河中游区域；四川、重庆和广西先降级后升级，均位于西南地区，重庆变化尤为显著，从 III 级跌至 I 级后又跃升为 IV 级；仅地处西北的甘肃一直为最低水平；剩余地区均呈现持平或上升。

分年份来看。2014 年，IV-I 级地区数量比为 0.08 : 0.20 : 0.28 : 0.44，仅北京和上海处于 IV 级；江苏、浙江、重庆和广东为 III 级，均为经济发展水平较高、人口较为密集的地区，分布于东部沿海、南部沿海和西南综合经济区；II 级地区（黑龙江、辽宁、四川、贵州、广西、湖北和湖南）位于东北、西南和长江中游区域；其余 11 个地区均为 I 级水平，占研究区域的 44%。可以看出，2014 年我国家庭信息使用水平整体偏低，存在很大发展空间。

2016 年，IV-I 级地区数量比为 0.28 : 0.16 : 0.28 : 0.28。IV 级地区升至七个，提高 20 个百分点，分布于北部沿海、东部沿海、南部沿海、和长江中游，其中，湖北和湖南直升两级；III 级和 II 级地区数量与 2014 年基本持平，但变动较大，其中，安徽、福建和云南从 I 级跃升至 III 级，分别位于长江中游、南部沿海和西南地区，重庆则与之相反，由 III 级降至 I 级；I 级地区减少为七个，占 28%，与 2014 年比下降 16 个百分点，主要位于西北、西南、黄河中游和东北地区。

2018 年，IV-I 级地区数量比为 0.40 : 0.40 : 0.16 : 0.04。IV 级和 III 级地区均增为十个，覆盖 80% 的研究区域；仅甘肃省仍为 I 级，也是调查对象中唯一一个位于西北的地区；东部沿海地区全为最高水平，北部沿海、南部沿海、黄河中游和长江中游也全部升至 III 级及以上；四川和广西与 2014 年持平，仍为 II 级；与 2016 年相比，仅云南有所下降。可以看出，2018 年我国家庭信息使用水平整体较高，但发展是否均衡仍要作进一步分析。

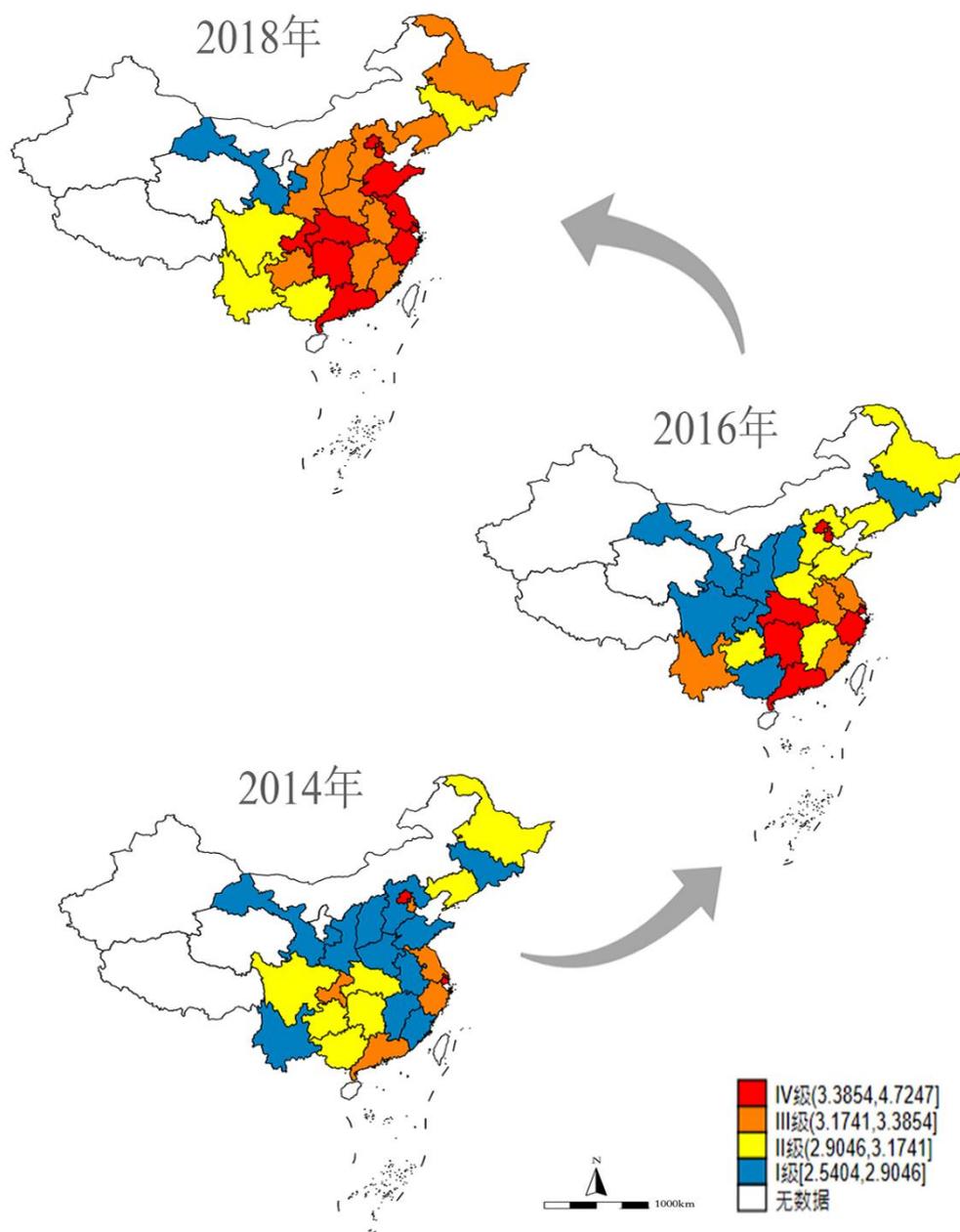


图 4.3 2014、2016 和 2018 年全国家庭信息使用水平空间分布图

4.4 地区数字鸿沟测算结果分析

根据地区数字鸿沟测算结果,分别从城乡和省际尺度分析我国地区数字鸿沟时间演变趋势和空间分布特征。

4.4.1 城乡地区数字鸿沟

依旧使用 Quartile 分类方法，将地区数字鸿沟指数划分成I-IV级，数字鸿沟水平逐级增加，具体分级标准如表 4.7 所示。

表 4.7 地区数字鸿沟水平分级标准

描述性统计量		地区数字鸿沟水平
最小值		0.3193
分位数	25%	0.4567
	50%	0.4938
	75%	0.5536
最大值		0.6451

由图 4.4 可知，2014-2018 年，我国数字鸿沟水平整体呈现缩小趋势，2014 年和 2016 年分别为 0.534 和 0.520，均处于 III 级（0.4938-0.5536），2018 年降至 II 级（0.4567-0.4938），为 0.475，与 2014 年相比下降 11%。我国城镇数字鸿沟始终小于全国水平，且降幅明显，2014-2018 年由 III 级降至 I 级（0.3193-0.4567），2018 年较 2014 年下降 19.2%。农村数字鸿沟则始终高于全国水平，且下降缓慢，2014 年和 2016 年均为最高水平，2018 年略低于 IV 级（0.5536-0.6451），较 2014 年仅下降 4.2%，且仍未达到四年前的全国水平。

尽管城镇与农村数字鸿沟均持续下降，但二者之间的差距逐渐扩大，2018 年城乡数字鸿沟之差（0.142）是 2014 年（0.069）的两倍以上。城镇家庭在使用层面的信息化水平趋于均衡，极化程度较低，农村家庭信息使用水平则极化明显，发展极不均衡，这一趋势在一定程度上阻碍了弥合全国数字鸿沟的进程，并将进一步加深城乡之间的数字鸿沟。

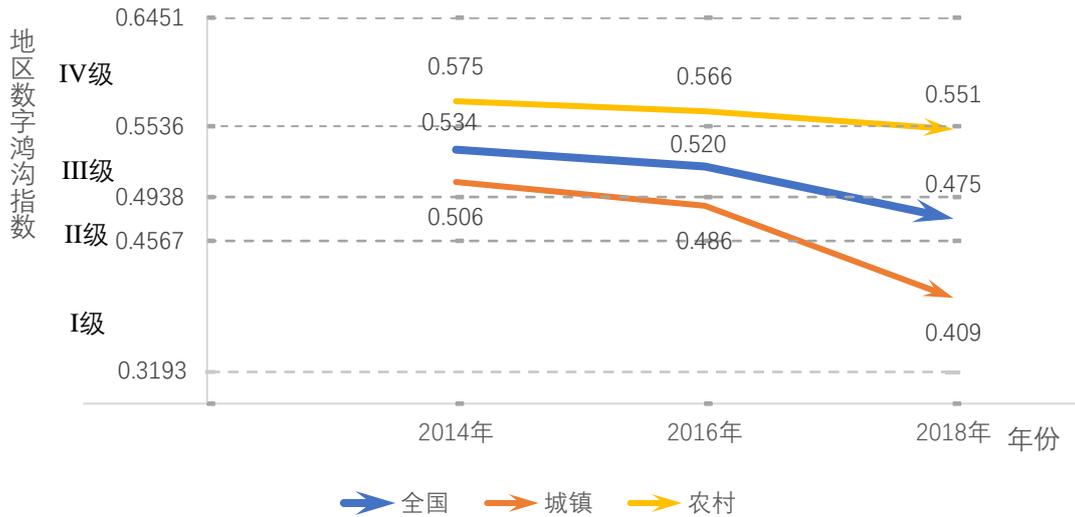


图 4.4 2014、2016 和 2018 年全国城乡地区数字鸿沟水平对比折线图

4.4.2 省际地区数字鸿沟

根据各地区数字鸿沟指数级别，分别绘制 2014 年、2016 年和 2018 年地理图，以展示其时空演变特征，如图 4.5 所示。

(1) 省际数字鸿沟情况整体分布

整体来看，时间维度上，我国数字鸿沟问题逐年减轻，有明显缓解，IV 级地区占比由 11 个降至 2 个，III 级以下地区由 28% 升至 80%；空间维度上呈现“东低西高，北重南轻”的分布特征。

细化至各个地区，天津、重庆、云南、辽宁、吉林三年均为同一水平，其中，天津、重庆始终为最低水平（I 级），吉林则始终为 IV 级数字鸿沟；河北、山东、山西、河南、福建、江西三年连降，其中，冀鲁晋豫四省毗邻，地处北部沿海和黄河中游，数字鸿沟均由 IV 级降至 II 级；其余省/市/自治区整体下降或持平，没有数字鸿沟持续加深的地区。

(2) 各级数字鸿沟地区变动趋势

2014 年，I-IV 级地区数量比为 0.16 : 0.12 : 0.28 : 0.44。I 级地区（天津、江苏、湖北和重庆）数字鸿沟最小，家庭信息使用水平极化程度较低，分别定位在北部沿海、东部沿海、长江中游和西南经济区；II 级地区为北京、广东和云南，位于北部沿海、南部沿海和西南地区；III 级地区（7 个）和 IV 级地区（11 个）

共占 72%。可以看出，2014 年我国整体数字鸿沟水平较高，家庭信息使用水平差异性较大，呈极化态势。

2016 年，I-IV 级地区数量比为 0.20 : 0.20 : 0.36 : 0.24。随着家庭信息使用水平整体提升，各地区数字鸿沟级别变动较大，整体呈减轻趋势，但也有部分地区不降反升。湖南由最高水平连降三级，福建连降两级，均达到 I 级水平；黑龙江、江苏、陕西、湖北和贵州数字鸿沟有所加剧，分别定位在东北、东部沿海、黄河中游、长江中游和西南地区，其中，江苏连升两级，由最低水平变为 III 级数字鸿沟。

2018 年，I-IV 级地区数量比为 0.40 : 0.40 : 0.12 : 0.08。浅色部分明显增多，I 级和 II 级地区数量翻番，均为十个，I 级主要分布在北部沿海、东部沿海、南部沿海、长江中游和西南地区，经济发展水平较高，且人口较为密集；2014 年和 2016 年均为 IV 级的四川和广西降至 III 级；IV 级仅余吉林和甘肃，分别位于东北和西北。与 2016 年相比，没有出现数字鸿沟反向加深的地区，信息使用极化现象得到明显改善。

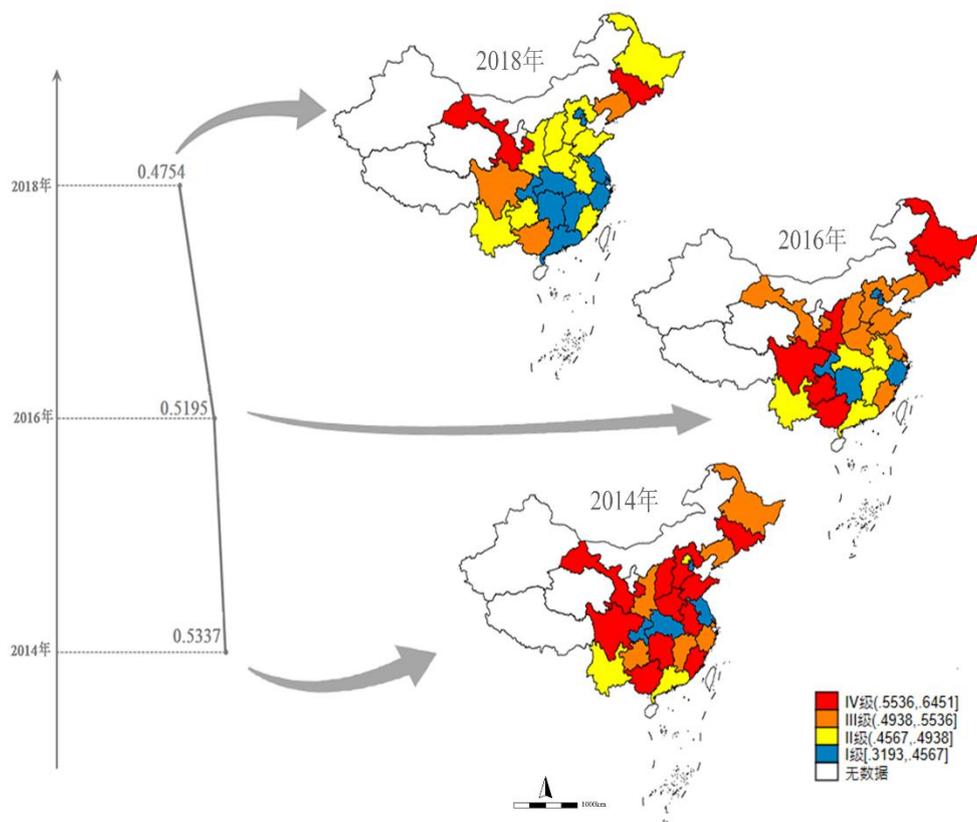


图 4.5 2014、2016 和 2018 年中国地区家庭数字鸿沟空间分布图

(3) 地区家庭信息使用水平与数字鸿沟水平分布对比

进一步对比分析地区信息使用水平和地区数字鸿沟水平分布情况, 将 2014-2018 年各级别地区所在综合经济区进行汇总, 信息使用水平由高到低降序排列, 数字鸿沟水平由轻到重升序排列, 整理结果如表 4.8。

表 4.8 家庭信息使用水平和地区数字鸿沟水平分布对比

年份	地区信息化水平		地区数字鸿沟	
	级别	所处经济区	所处经济区	级别
2014	IV	北部沿海、东部沿海	北部沿海、东部沿海、 长江中游、西南	I
	III	北部沿海、东部沿海、 南部沿海、西南	北部沿海、南部沿海、西南	II
	II	东北、长江中游、西南	东北、东部沿海、黄河中游、 长江中游、西南	II
	I	东北、北部沿海、南部沿海、黄 河中游、长江中游、西南、西北	东北、北部沿海、南部沿海、黄 河中游、长江中游、西南、西北	IV
2016	IV	北部沿海、东部沿海、 南部沿海、长江中游	北部沿海、东部沿海、 长江中游、西南	I
	III	东部沿海、南部沿海、 长江中游、西南	南部沿海、长江中游、西南	II
	II	东北、北部沿海、黄河中游、 长江中游、西南、西北	东北、北部沿海、东部沿海、 南部沿海、黄河中游、西北	III
	I	东北、黄河中游、西南、西北	东北、黄河中游、西南	IV
2018	IV	北部沿海、东部沿海、南部沿 海、长江中游、西南	北部沿海、东部沿海、南部沿 海、长江中游、西南	I
	III	东北、北部沿海、南部沿海、 黄河中游、长江中游、西南	东北、北部沿海、南部沿海、 黄河中游、长江中游、西南	II
	II	东北、西南	东北、西南	III
	I	西北	东北、西北	IV

由表 4.8 可知,我国家庭信息使用水平发展程度分布与极化程度分布存在一定相似性,信息使用水平越高的地区极化程度越低,且发展越快的地区数字鸿沟弥合速度越快,2018 年,IV-II 级信息使用水平与 I-III 级数字鸿沟的综合经济区分布已经完全一致。但相比于家庭信息使用水平的稳步增长,家庭数字鸿沟的发展存在一定波动性,部分地区在提高使用水平的过程中难以保证所有家庭同步升级,导致家庭间差异增大,继而加重数字鸿沟问题。

分解地区来看,沿海地区(北部沿海、东部沿海、南部沿海)信息使用水平整体领先于其他地区,且数字鸿沟水平整体较低,虽有过加剧时期,但能很快缓和恢复,整体发展较为均衡;长江中游和黄河中游在 2014 年信息使用水平较低,数字鸿沟程度较深,但发展非常迅速,2018 年已全部处于前两级,直追沿海地区;西南地区各省/市/自治区经济水平差异较大,信息使用和数字鸿沟级别也较为离散,整体发展较为缓慢和不均衡;东北和西北始终处于较低使用水平和较高极化程度,尤其是西北地区,2014-2018 年排名均靠后,且没有明显改善。从全国角度来看,我国各地区信息使用水平仍存在极化现象,各地区发展程度和发展速度差异较大,将在一定程度上加重地区间的数字鸿沟。

5 中国家庭数字鸿沟影响因素分析

本部分将基于 2014-2018 年 CFPS 数据、Wind 数据库、《中国统计年鉴》、《中国住户调查主要数据》和各省市区统计年鉴，从家庭、个人和地区三个层面选取解释变量，以家庭数字鸿沟为被解释变量，对比检验三种面板数据回归模型，依此选取最优形式，从宏观和微观角度探究家庭数字鸿沟的主要影响因素。

5.1 指标选取与数据预处理

本小节对面板数据模型的被解释变量及解释变量进行介绍，并简单阐述数据来源及数据预处理。

5.1.1 指标选取

参考已有文献，本部分以家庭数字鸿沟指数为被解释变量，从多角度选取 14 个指标作为解释变量，以探究微观和宏观层面我国家庭数字鸿沟的影响因素，具体介绍如下。

(1) 被解释变量

家庭数字鸿沟指数：为从微观视角分析数字鸿沟的驱动机制，本文依据第五部分测算结果，用信息技术使用水平最高的家庭，即家庭信息使用水平最大值，分别减去每个家庭信息使用水平的值，得到各个样本的家庭数字鸿沟指数，以此作为被解释变量，记为 DDI。

(2) 解释变量

根据理论框架，结合其他学者数字鸿沟影响因素研究的相关成果，从家庭特征、个人特征和地区特征三个角度选择解释变量，涵盖家庭规模、家庭经济、性别、年龄、学历、婚姻状况、经济发展水平、城市化水平、信息基础设施发展水平、经济开放程度和科技发展水平等多个方面选取解释变量。

① 家庭特征

家庭规模：家庭人口规模在一定程度上决定了家庭的资源分配方式，不同规模的家庭消费状况也有所不同。同等收入家庭间，规模小的家庭消费水平更高，会对网络信息产品的支出占比产生正向影响；规模大的家庭在生活、教育和医疗等方面花费较多，进而压缩网络信息产品的消费需求，消费结构差异会加深家庭

间数字鸿沟。本文选取家庭人口数量反映家庭规模，记为 *familysize*。

家庭经济：无论是从数字鸿沟的各类定义还是现有文献的研究结果来看，经济因素都是形成数字鸿沟的根本原因之一，在微观层面，家庭经济水平差异会导致家庭信息化水平极化程度增高。家庭纯收入和家庭净资产分别代表了家庭经济的增量和存量，反映了该家庭的消费能力和购买力水平。本文以人均家庭纯收入和家庭净资产作为衡量家庭经济水平的指标，分别记为 *afincome* 和 *nettasset*。

② 个人特征

中国人常以家庭为单位从事经济活动和社会交往，家庭是代际关系的主要机制，也是个人在物质与精神情感方面极其重要的支持来源，家庭成员的观念和行为会相互影响，其中，家庭主要负责人对其余成员的影响尤为显著。因此，在研究家庭时，有必要将户主的个人特征纳入分析范畴。从已有文献来看，不同特征群体的信息接入和使用水平存在一定差距，性别、年龄、文化程度、婚姻情况和主观态度等方面都有可能是决定数字鸿沟的因素 (Mumporeze 和 Prieler, 2017; 陆峰等, 2007)。也有研究表明，随着经济的高速发展和思想观念的不断进步，部分群体间的数字鸿沟正在逐步弥合 (洪海娟和万跃, 2014)。本文选取家庭财务管理者的性别 (*gender*)、年龄 (*age*)、最高学历 (*edu*)、婚姻状况 (*mar*) 和互联网作为信息渠道的重要性 (*impo*) 五个变量代表家庭负责人的个人特征，其赋值方法见表 5.1。

表 5.1 个人特征变量赋值方法

变量名称	赋值方法
性别 (<i>gender</i>)	0 代表“女性”，1 代表“男性”
年龄 (<i>age</i>)	受访者实际年龄 (周岁)
最高学历 (<i>edu</i>)	0 代表“文盲/半文盲”，1 代表“小学”，2 代表“初中”，3 代表“高中/中专/技校/职高”，4 代表“大专”，5 代表“大学本科”，6 代表“硕士”，7 代表“博士”
婚姻状况 (<i>mar</i>)	0 代表“未婚”，1 代表“在婚 (有配偶)”，2 代表“同居”，3 代表“离婚”，4 代表“丧偶”
互联网作为信息渠道的重要性 (<i>impo</i>)	按照重要程度从 1-5 赋值 (1 表示非常不重要，5 表示非常重要)

③ 地区特征

经济发展水平：中国信息化发展是经济发展的延续，经济发达地区良好的区位、发达的交通、众多的人才、活跃的市场，使信息基础设施的布局、信息的消费与生产等都优先得到发展（祝坤子，2015）；经济发达的国家或地区，其经济活动对于区域互联网发展的需求相对较大（邱娟和汪明峰，2010）；此外，第五部分测算的我国地区家庭信息使用水平和地区数字鸿沟也与区域经济发展水平具有相似分布特征。考虑到我国各地区发展状况差异较大，为保证可比性，本文选取人均 GDP 衡量各地区经济发展水平，记为 $aGDP$ 。

城市化水平：城市化不断发展可能导致资源要素过度集中，截至 2020 年末，我国城市互联网渗透率仍比农村地区高 20 个百分点，第五部分测算的我国城市和农村家庭信息使用水平与极化程度也有较大差距，可以看出无论是接入层面还是使用层面，城镇化水平都对区域互联网的发展产生较大影响，本文将城市化率作为测度城市化进程的指标，记为 $urRate$ 。

信息基础设施发展水平：信息基础设施是互联网发展的重要支撑，其发展水平代表了该地区居民的互联网接入水平，也是判断区域间数字鸿沟程度最重要的依据之一。但近年来，随着传统信息基础设施的飞速普及，不同家庭在互联网接入上的差异越来越小，该因素对数字鸿沟的影响力也有所下降。本文选取万人长途光缆线路长度来反映信息基础设施发展水平，记为 $aOptical$ 。

经济开放程度：互联网作为一种全球性的信息传输载体，是与开放世界连接的手段，区域经济开放会正向促进互联网的发展，互联网使用水平的整体提升可以减轻区域内家庭数字鸿沟，但同时也将加重区域间的差异程度。本文用实际利用外商直接投资占 GDP 比重代表经济开放程度，记为 $FDIrate$ 。

科技发展水平：科技文化水平与互联网发展水平息息相关，一方面互联网的使用需要一定的计算机硬件和软件知识；另一方面互联网所带来的信息与知识需要具备一定的文化水平才能深刻理解。另外，科教投入对“数字鸿沟”的影响非常直接，投入高的地区往往信息化水平也更高。本文选取 3 种专利万人申请受理量和研究与试验发展经费投入强度作为衡量科技发展水平的变量，分别记为 $aPatent$ 和 $RDRate$ 。

综上，本部分选取的具体指标变量及其表示符号如表 5.2 所示。

表 5.2 指标变量说明

指标属性	指标类别	指标名称	表示符号
被解释变量	-	家庭数字鸿沟指数	DDI
解释变量	家庭特征变量	家庭人口数量（人）	familysize
		人均家庭纯收入（元/人）	afincome
		家庭净资产（元）	netasset
	个人特征变量	性别	gender
		年龄（周岁）	age
		最高学历	edu
		婚姻状况	mar
		互联网作为信息渠道的重要性	impo
	地区特征变量	人均 GDP（元/人）	aGDP
		城市化率（%）	urRate
		万人长途光缆线路长度（公里）	aOptical
		实际利用外商直接投资占 GDP 比重（%）	FDIrate
		3 种专利万人申请受理量（件）	aPatent
研究与试验发展经费投入强度（%）		RDrate	

5.1.2 数据来源

家庭特征变量和个人特征变量分别来自 2014-2018CFPS 家庭经济库和成人库，所选样本与第五部分清洗合并后的数据一一对应，其中，个人特征变量的主体为各家庭户财务管理者；地区特征变量中，城市化率来自 Wind 数据库和历年《中国住户调查主要数据》，实际利用外商直接投资来自 Wind 数据库和各地区统计年鉴，3 种专利申请受理量来源于国家知识产权局官网，其余指标均来自历年《中国统计年鉴》和各地区统计年鉴。

5.1.3 数据预处理

为保证数据质量和样本代表性，对所选变量进行缺失值插补、汇率转换、对数处理，并构建平衡面板数据。

(1) 缺失值

最高学历这一变量的数据缺失值较多，经过比对，个人文化程度三年全缺失的样本极少，且在调查间隔（两年）内几乎没有变化，因此本文使用往期调查最高学历或后期调查最低学历对缺失值进行插补；吉林省经济技术合作局自 2017 年起不再向省统计局提供实际利用外商直接投资数据^①，因此用 2016 年同比增长率推算 2018 年缺失数据值。

(2) 汇率转换和对数处理

根据历年《中国统计年鉴》中人民币对美元汇率将实际利用外商直接投资额转换为人民币，并与地区 GDP 统一单位（亿元）。为缓和部分变量条件分布的异方差和偏态性问题，本文参考 J.M.伍德里奇在《计量经济学导论：现代观点》中提出的选取标准，对人均家庭纯收入、家庭净资产、人均 GDP、城市化率、万人长途光缆线路长度和 3 种专利万人申请受理量取其自然对数，在原有指标表示符号前加“L”，生成新变量。

(3) 构建平衡面板

首先，生成年份变量，对所选样本按 2014 年家庭和个人编码进行匹配，跨年、跨表合并家庭、个人和地区特征变量；其次，小部分家庭在 2016-2018 年间分家，这些样本可能出现系统性变化，也有家庭由于各种原因在部分年份没有受访，尽管以上两种情况的样本量非常小，但都会导致非平衡面板，因此在不影响样本地区分布的前提下予以剔除，得到平衡面板数据。

预处理后得到 2014-2018 年 25 个省区市 3057 个家庭样本的平衡面板数据，14 个解释变量均有 9171 条观测，其描述性统计结果见表 5.3。

^① 信息来源：http://tjj.jl.gov.cn/hdjl/hdfk/hdfk/201812/t20181228_5453235.html

表 5.3 解释变量描述统计结果

变量	均值	标准差	最小值	最大值
familysize	3.79	1.86	1.00	21.00
Lafincome	9.37	1.14	-0.69	15.24
Lnetasset	12.31	1.47	3.40	18.20
gender	0.52	0.50	0.00	1.00
age	51.48	14.19	14.00	95.00
edu	2.70	1.61	0.00	10.00
mar	2.24	0.88	1.00	5.00
impo	2.20	1.54	1.00	5.00
LaGDP	10.79	0.43	10.13	11.94
LurRate	4.04	0.20	3.69	4.50
LaOptical	1.73	0.52	0.50	2.62
FDIrate	1.86	1.17	0.04	10.89
LaPatent	2.63	0.81	0.63	4.59
RDRate	1.75	0.88	0.60	6.17

5.2 模型检验与对比

为选择最优模型，消除数据的自相关性和异方差性，本文基于处理后的平衡面板数据，首先进行多重共线性检验，然后选取稳健标准误，分别使用混合回归模型、固定效应模型和随机效应模型进行拟合，最后通过 LM 检验及豪斯曼检验确定最终模型。

(1) 多重共线性检验

首先对 14 个解释变量进行多重共线性检验，结果如表 5.4。可以看到所有解释变量的方差膨胀因子 (*VIF*) 均小于 10，且 *VIF* 均值也小于 10，故本文所选取的 14 个解释变量之间相关性不强，不存在多重共线性问题，可以构建模型。

表 5.4 解释变量多重共线性检验结果

解释变量	VIF	1/(VIF)
familysize	1.31	0.763035
Lafincome	1.58	0.634909
Lnetasset	1.56	0.641868
gender	1.02	0.981782
age	1.22	0.822541
edu	1.26	0.792813
mar	1.08	0.927388
impo	1.13	0.882745
LaGDP	8.89	0.112542
LurRate	5.30	0.188743
LaOptical	3.25	0.307285
FDIrate	1.85	0.541257
LaPatent	3.52	0.283914
RDrate	5.29	0.189124
Mean VIF	2.73	-

(2) 混合回归模型与固定效应模型

虽然通常假设扰动项为独立同分布，采用普通标准误进行计算，但同一家庭在不同时期的扰动项之间一般存在自相关，与这一假定不符，故本文所有模型均使用聚类稳健标准误（cluster-robust standard error）。

使用固定效应模型（组内估计量）进行拟合，输出结果中 ρ 为 0.58，说明复合扰动项（ $u_i + \varepsilon_{it}$ ）的方差更多来自个体效应 u_i 的变动；同时，对于原假设“ $H_0: all u_i = 0$ ”（不存在个体异质性截距项）， F 检验的 p 值为 0.0000，故强烈拒绝原假设，固定效应模型优于混合回归模型，允许每位个体拥有自己的截距项。

(3) 混合回归模型与随机效应模型

使用可行广义最小二乘法（FGLS）对随机效应模型进行估计，对其进行 LM

检验， p 值为 0.0000，强烈拒绝原假设“ $H_0: \sigma_u = 0$ ”（不存在个体随机效应），即认为存在个体随机效应。作为对照，对随机效应模型进行最大似然估计（MLE），其系数估计值与随机效应 FGLS 有所不同，但性质类似，且同样强烈拒绝原假设（ p 值为 0.0000）。因此，随机效应模型优于混合回归模型。

（4）固定效应模型与随机效应模型

在处理面板数据时，通常使用豪斯曼检验判断应该选择固定效应模型还是随机效应模型。豪斯曼检验的原假设是“ $H_0: u_i$ 与 x_{it}, z_i 不相关”（即随机效应模型更有效），输出结果中 p 值为 0.0000，故强烈拒绝原假设，应该使用固定效应模型。

综上，本文最终选择固定效应模型进行拟合，上述各假设检验结果汇总如表 5.5。

表 5.5 模型对比假设检验结果

原假设	原假设含义	p 值
$all u_i = 0$	不存在个体异质性截距项	0.0000
$\sigma_u = 0$	不存在个体随机效应	0.0000
u_i 与 x_{it}, z_i 不相关	随机效应模型比固定效应模型更有效	0.0000

5.3 家庭数字鸿沟影响因素

根据固定效应模型，对预处理后的平衡面板数据进行回归分析，并使用聚类稳健标准误，以消除个体不同时期扰动项之间的自相关性，回归结果表明各个解释变量对家庭数字鸿沟的影响程度及作用方向。

5.3.1 模型估计结果

由输出结果可知（表 5.6），共有七个解释变量通过显著性检验，其中，人均家庭纯收入(Lafincome)、最高学历(edu)、互联网作为信息渠道的重要性(impo)、

人均 GDP (LaGDP)、3 种专利万人申请受理量 (LaPatent) 和研究与试验发展经费投入强度 (RDrate) 对家庭数字鸿沟 (DDI) 产生负向显著影响, 年龄 (age) 则正向影响显著; 家庭规模 (familysize)、家庭净资产 (Lnetasset)、性别 (gender)、婚姻状况 (mar)、城市化率 (LurRate)、万人长途光缆线路长度 (LaOptical) 和实际利用外商直接投资 (FDIrate) 占 GDP 比重未通过显著性检验。

表 5.6 固定效应模型估计结果

解释变量	系数	聚类稳健标准误	<i>t</i> 值	<i>p</i> 值
familysize	0.0397	0.0414	0.96	0.338
Lafincome	-0.0719*	0.0449	-1.81	0.071
Lnetasset	-0.0368	0.0310	-1.19	0.236
gender	-0.0217	0.0898	-0.24	0.809
age	0.0366***	0.0100	3.66	0.000
edu	-0.1158***	0.0479	-2.61	0.009
mar	0.1519	0.0975	1.56	0.119
impo	-0.2424***	0.0281	-8.64	0.000
LaGDP	-1.6051***	0.4261	-3.77	0.000
LurRate	-1.0201	0.9026	-1.13	0.259
LaOptical	0.4537	0.3435	1.32	0.187
FDIrate	0.0575	0.0336	1.62	0.104
LaPatent	-0.6342***	0.1550	-4.09	0.000
RDrate	-0.8084***	0.2045	-3.95	0.000

注: ***, **、*分别表示在 0.01、0.05、0.1 显著性水平下显著。

5.3.2 家庭影响因素

整体来看, 家庭因素对数字鸿沟的影响较小, 且主要为家庭经济。三个家庭特征变量中仅人均家庭纯收入 (Lafincome) 通过显著性检验; 家庭规模 (familysize) 和家庭净资产 (Lnetasset) 对家庭数字鸿沟没有明显影响。

(1) 人均家庭纯收入

人均家庭纯收入在 0.1 显著性水平下显著，系数估计值为负，说明在其他变量不变的情况下，人均家庭纯收入增加，该家庭的数字鸿沟程度将降低。收入对技术应用具有非常关键的影响，人均收入直接影响家庭信息技术设备的购买和使用，也影响居民对信息服务的需求，对于贫困弱势群体，即使在经济发达地区也不能很好地融入数字化经济中，导致信息使用水平明显低于高收入群体，收入水平的极化程度直接导致了信息使用水平的差异化。

(2) 家庭规模和家庭净资产

自 1982 年开始推行计划生育起，我国提倡少生优生，以控制人口数量、提高人口素质，促使各家庭人口数差距逐渐缩小、家庭规模日益均衡，这一趋势导致家庭规模在居民消费结构和消费需求差异上的影响力持续降低，对家庭数字鸿沟水平无明显正向或负向刺激作用；家庭净资产也未通过显著性检验，说明家庭经济在影响家庭数字鸿沟时，更多是依靠增量（人均家庭纯收入），而非存量（家庭净资产），资产变动对居民购买和使用信息技术设备决策的作用力微乎其微。

5.3.3 个人影响因素

在微观层面，相比家庭因素，数字鸿沟受个人因素影响更大。个人特征变量中年龄（age）、最高学历（edu）和互联网作为信息渠道的重要性（impo）均通过显著性检验；性别（gender）和婚姻状况（mar）对家庭数字鸿沟没有明显影响。

(1) 年龄

年龄变量在 0.01 显著性水平下拒绝原假设，系数估计值为正，说明在其他变量不变的情况下，家庭财务管理人的年龄越大，该家庭的数字鸿沟程度越深。信息技术作为一种发展极其迅速的新兴技术，其普及程度和普及速度虽然已达到较高水平，但在不同年龄阶段的人群中发展并不均衡，年龄越大的群体信息使用能力越弱，且有恐惧、认为成本较高且难以学会等抵触情绪，导致老年人无法跨越数字鸿沟，越来越难以跟上社会智能化发展进程。

(2) 最高学历

最高学历变量可以通过 0.01 显著性检验，系数估计值为负，说明在其他变量不变的情况下，家庭负责人学历越高，该家庭的数字鸿沟程度越低。不同学历

水平的群体间信息技能的差距明显,并将在获取知识和信息并加以应用方面产生差异,这直接造成了个体间的数字鸿沟。受教育程度越高,对信息技术的学习能力、接受程度和掌握程度越高,同时,高学历人才在工作中对互联网应用有更多的需求和更深的依赖。

(3) 主观意识

家庭负责人主观意识(互联网作为信息渠道的重要性)在 0.01 显著性水平下显著,系数估计值为负,说明在同一地区、信息技能掌握能力相同的人群中,主观意识差距仍会导致数字鸿沟,认为互联网作为信息渠道重要性越高的群体,越愿意通过互联网进行各项活动,对互联网应用有更高接受度。

(4) 性别

尽管有部分早期研究表明男性和女性之间存在数字鸿沟,但大多停留在信息接入层面,随着互联网渗透率的逐步提高和人们观念意识的升级,性别已经不再是影响数字鸿沟的因素之一,性别数字鸿沟已经基本弥合。

5.3.4 地区影响因素

除了微观因素,宏观地区因素对微观家庭数字鸿沟水平也存在一定效应。地区特征变量中人均 GDP (LaGDP)、3 种专利万人申请受理量 (LaPatent) 和研究与试验发展经费投入强度 (RDrate) 通过显著性检验;城市化率 (LurRate)、万人长途光缆线路长度(LaOptical)和实际利用外商直接投资占 GDP 比重(FDIrate)则对家庭数字鸿沟无明显影响。

(1) 人均 GDP

人均 GDP 变量在 0.01 显著性水平下显著,系数估计值为负,且明显高于其他影响因素,说明在其他变量不变的情况下,区域人均 GDP 增加,该地区家庭的数字鸿沟程度将显著降低。经济发展水平是区域互联网发展的基础,也是形成数字鸿沟根本原因。互联网基础设施建设、用户上网设备、用户上网费用的支付都是由经济水平决定的,经济因素对区域互联网发展的影响是多方面的。地区经济水平的提升会促使消费结构升级、数字经济利益分配更加均衡,继而提高互联网的使用密度。

(2) 3 种专利万人申请受理量和研究与试验发展经费投入强度

3 种专利万人申请受理量和研究与试验发展经费投入强度变量在 0.01 显著性水平下显著，系数估计值均为负，说明在其他变量不变的情况下，区域科技发展水平越高，该地区家庭数字鸿沟越小。科技发展水平差异是导致数字鸿沟产生的重要原因之一，区域科教投入在一定程度上决定了国民接受和使用信息的能力。

（4）万人长途光缆线路长度

近年来，国家加大互联网接入的建设力度，全国互联网普及度已超过 70%，各地区信息基础设施发展水平差距逐渐减小，信息接入对互联网使用的影响力正在逐渐消失，对接入层面数字鸿沟没有明显作用。

6 结论与建议

本文在现有文献的基础上,测算我国家庭信息化水平,分别从城乡和省际两种尺度分析各地区数字鸿沟时空演变特征,并使用固定效应模型,从家庭、个人和地区三个方面,探究家庭数字鸿沟的宏观和微观影响因素,由此得出系列结论并提出针对性建议。

6.1 主要结论

本部分将分别从 2014-2018 年我国地区数字鸿沟时空特征和家庭数字鸿沟影响因素两方面入手,对本文的研究结果进行归纳总结,具体阐述如下。

6.1.1 地区数字鸿沟时空特征总结

分析我国互联网发展现状,基于 CFPS 数据测算我国家庭信息使用水平和地区数字鸿沟情况,从城乡、省际和综合经济区三种尺度描述其时空特征,主要得出以下结论。

(1) 我国互联网渗透率已超 70%,信息接入层面差距逐渐消失,但互联网应用使用率极化程度仍然较高,信息使用差距较大。

互联网普及程度上,截至 2020 年 12 月,我国网民规模约占全球网民的五分之一,为全球最大的数字社会,手机网民占比高达 99.7%,互联网普及率达 70.4%,城乡互联网普及率之间的差异仍超过 20%,但已出现缩减趋势。互联网应用使用程度上,即时通信的使用率高达 99.2%,为所有互联网应用渗透率第一,短视频用户次之,占网络视频用户的 94.2%,互联网理财仅 17.2%,为所有应用渗透率最低,此外,公共服务类应用使用率普遍较低,且增速较低,仍有很大的发展空间。

(2) 数字鸿沟水平整体呈现缩小趋势,城镇数字鸿沟降幅明显,农村数字鸿沟下降缓慢,二者差距逐年扩大。

2014-2018 年,我国城镇数字鸿沟始终小于全国水平,共下降 19.2%,由 III 级降至 I 级,城镇家庭在使用层面的信息化水平趋于均衡,极化程度较低;农村数字鸿沟则始终高于全国水平,仅下降 4.2%,由 IV 级降至 III 级,至 2018 年仍

未达到四年前的全国水平，农村家庭信息使用水平则极化明显，发展极不均衡；城乡数字鸿沟差距持续增大，2018年城乡数字鸿沟之差已达2014年的两倍以上。

(3) 家庭信息使用水平持续升高，呈现“东高西低，南高北低”分布特征；地区数字鸿沟逐年减轻，呈现“东低西高，北重南轻”的分布特征。

家庭信息使用水平持续升高，发展迅速，始终呈现“东高西低，南高北低”的梯形分布，与经济发展水平空间特征较为相似。IV级地区逐年递增，占比由8%增至40%，I级地区则从11个降为一个，发展速度较快；北京和上海始终处于最高水平，仅地处西北的甘肃一直为最低水平，剩余地区均呈现持平或上升。

数字鸿沟问题逐年减轻，有明显缓解，空间维度上呈现“东低西高，北重南轻”的分布特征。IV级地区占比由11个降至2个，III级以下地区由28%升至80%；天津、重庆始终为最低水平，吉林则始终为最高水平，其余省/市/自治区整体下降或持平，没有数字鸿沟持续加深的地区。

(4) 各级家庭信息使用水平与地区数字鸿沟水平的地区分布相似，信息使用水平越高的地区内部极化程度越低，且发展越快的地区数字鸿沟弥合速度越快，但地区间差距有加重态势。

2018年，IV-II级信息使用水平与I-III级数字鸿沟的综合经济区分布已经完全一致。沿海地区（北部沿海、东部沿海、南部沿海）信息使用水平整体领先于其他地区，数字鸿沟水平整体较低；长江中游和黄河中游数字鸿沟弥合速度较快；西南地区信息使用和数字鸿沟级别较为离散，整体发展缓慢且不均衡；东北和西北始终处于较低使用水平和较高极化程度，尤其是西北地区，2014-2018年排名均靠后，且没有明显改善。相比于家庭信息使用水平的稳步增长，家庭数字鸿沟的发展存在一定波动性，部分地区在提高使用水平的过程中难以保证所有家庭同步升级，导致家庭间差异增大，继而加重数字鸿沟问题。

6.1.2 家庭数字鸿沟影响因素总结

从微观和宏观视角选取家庭特征、个人特征和地区特征解释变量，使用固定效应模型测度其对家庭数字鸿沟的影响作用，主要得出以下结论。

(1) 存在个体效应，固定效应模型优于混合回归模型及随机效应模型。

预处理后数据不存在多重共线性问题，经检验，数据存在随个体而异的遗漏

变量问题和一定随机效应，不应使用混合回归模型；豪斯曼检验结果显示固定效应模型比随机效应模型更有效。

(2) 微观因素和宏观因素共同作用于家庭数字鸿沟，家庭、个人和地区因素均对家庭数字鸿沟产生影响。

经过一系列检验，最终选择固定效应模型。七个变量通过显著性检验，其中，有四个微观影响因素，分别为人均家庭纯收入、年龄、最高学历、主观意识，三个宏观影响因素，分别为人均 GDP、3 种专利万人申请受理量和研究与试验发展经费投入强度；家庭负责人年龄越大，该家庭数字鸿沟程度越深，其余变量对数字鸿沟均有显著的负向缓解作用。

(3) 二级数字鸿沟影响因素与一级数字鸿沟存在差异，性别、城市化水平和信息基础设施发展水平对数字鸿沟的影响逐渐消失。

固定效应模型估计结果中，性别、城市化率和万人长途光缆线路长度未通过显著性检验。性别、城市化水平和信息基础设施发展水平对个体和地区互联网接入有决定性作用，是一级数字鸿沟（技术接入差距）的主要影响因素，但随着我国互联网普及程度稳步提高，一级数字鸿沟已逐步弥合，二级数字鸿沟（技术使用差距）则日渐加深，上述三类因素对我国数字鸿沟的影响不再显著。

6.2 参考建议

针对实证分析发现的问题，分别从信息技术使用成本、区域协同措施、多元应用普及、互联网教育投入和信息技术适老化等几个角度提出建议。

(1) 提高农村及经济落后地区互联网应用程度，降低信息技术使用成本。

通过探究家庭数字鸿沟影响因素可知，数字鸿沟与地区经济发展水平和家庭收入水平密切相关，时空特征测算结果也显示城镇与农村、沿海地区与西部落后地区数字鸿沟水平差距明显，且弥合速度较慢，甚至有不断扩大的趋势。因此，为缓解由经济差异造成的信息使用差异，一方面，应着重发展农村及经济落后地区的信息化水平，推进农业数字化转型，确保互联网服务的公平提供，在继续提高互联网普及程度的同时，依据低收入群体的需求，研发和推广相应智能终端应用。另一方面，应降低信息技术使用成本，优化电信资费方案，免费开放各项基本服务功能，以鼓励低收入群体使用更多网络资源，培养其用户习惯。

(2) 加强区域信息化协同发展，促进互联网技术及应用共享。

从分布特征来看，各级家庭信息使用水平和各级数字鸿沟水平在空间分布上具有聚集性，东部地区发展水平更高且发展速度更快，西部和东北地区则始终处于落后地位且发展缓慢，导致整体极化问题难以解决。因此，应加强信息化建设的区域协调作用，由东部和中部地区带动西北和东北地区共同发展，建立互联网合作和互助机制；经济发达地区往往拥有更高的科研创新水平，在加速突破人工智能、区块链和量子信息等新一代信息技术的同时，也应与经济落后地区共享核心技术成果及实际应用，避免因科技水平差异增大造成更深的“鸿沟”。

(3) 优化线上公共服务平台，推进互联网应用多元化。

从具体应用类型来看，各类互联网应用使用程度并不均衡，基础通信、线上支付和网络娱乐较为普及，但公共服务类使用率整体较低，用户结构较为单一。因此，应推进互联网多元化应用，提高医疗、教育等公共服务应用便捷程度，加速技术融合，拓宽应用场景，针对不同用户的差异化需求优化升级服务平台使用体验，以扩展用户群体，提高使用率。

(4) 提高信息技术教育投入，引导积极网络观念。

个人受教育程度和主观意识对家庭数字鸿沟存在显著影响，即使提供了低成本服务和新型互联网应用，并不意味着会被居民广泛接受和使用，对信息技术的了解程度和观念意识有时候会阻碍网络资源的均衡分配。因此，应提高相关教育投入，加强网络技术和网络安全等方面知识普及和信息化人才培养，并从低学历和中老年群体入手，引导积极正确的网络观念，帮助互联网应用推广普及。

(5) 推进互联网应用适老化改造，同时保留传统服务方式。

随着我国人口老龄化程度进一步加深，老年人技术使用困难问题日益严重，在新冠肺炎疫情背景下尤为明显，飞速发展的互联网技术对“数字弱势群体”并不友好，带来难以消弭的个体间数字鸿沟。对此，一方面，应推进互联网应用适老化改造，疫情期间各地推出“他人健康码代查”、“老幼健康码便捷自查”等便捷功能，效果显著，可在其他领域应用推广；另一方面，数字化可以支撑服务便利化，但不等同于便利化，数字弱势群体往往对数字化设备存在恐惧和抵触情绪，难以快速掌握信息技术使用能力，过度普及互联网应用将适得其反，应保留部分传统服务方式和咨询服务，避免个体数字鸿沟带来更多不平等后果及影响。

参考文献

- [1] Aparicio M, Bacao F, Oliveira T. Grit in the Path to e-Learning Success[J]. *Computers in Human Behavior*,2017,66: 388-399.
- [2] Billon M, Lera-Lopez F, Marco R. Differences in digitalization levels: a multivariate analysis studying the global digital divide[J]. *Review of World Economics*, 2010, 146(1): 39-73.
- [3] Brandtzæg P B, Heim J, Karahasanovi A. Understanding the new digital divide—A typology of Internet users in Europe[J]. *International Journal of Human - Computer Studies*, 2011, 69(3): 123-138.
- [4] Chen C. Searching for intellectual turning points: Progressive knowledge domain visualization[J]. *Proc Natl Acad Sci U S A*, 2004, 101(suppl): 5303-5310.
- [5] Chinn M D, Fairlie R W. The determinants of the global digital divide: A cross-country analysis of computer and Internet penetration[J]. *Oxford Economic Papers*, 2007, 59(1): 16-44.
- [6] Cruz-Jesus F, Oliveira T, Bacao F. The Global Digital Divide: Evidence and Drivers[J]. *Journal of Global Information Management*, 2018, 26(2): 1-26.
- [7] Dijk J. Digital divide research, achievements and shortcomings[J]. *Poetics*, 2006, 34(4): 221-235.
- [8] Goncalves G, Oliveira T, Cruz-Jesus F. Understanding individual-level digital divide: Evidence of an African country[J]. *Computers in Human Behavior*, 2018, 87(OCT.): 276-291.
- [9] Guillen, Mauro F, Sandra L. Suarez. Explaining the Global Digital Divide: Economic, Political and Sociological Drivers of Cross-National Internet Use[J]. *Social Forces*, 2005, 84(2): 681-708.
- [10]Hwang, Joo-Seong. Digital Divide in Internet Use Within the Urban Hierarchy: The Case of South Korea[J]. *Urban Geography*, 2004, 25(4): 372-389.
- [11]Lucendo-Monedero A L, Ruiz-Rodriguez F, Gonzalez-Relano R. Measuring the Digital Divide at Regional Level. A Spatial Analysis of the Inequalities in Digital Development of Households and Individuals in Europe[J]. *Telematics and*

- Informatics, 2019, 41(AUG.): 197-217.
- [12] Marislei, Nishijima, Terry, Macedo, Ivanauskas, Flavia, Mori, Sarti. Evolution and determinants of digital divide in Brazil (2005–2013) [J]. Telecommunications Policy, 2017, 41(1): 12-24.
- [13] Mills B F, Whitacre B E. Understanding the Non-Metropolitan-Metropolitan Digital Divide[J]. Growth & Change, 2003, 34(2): 219–243.
- [14] Mumporeze N, Prieler M. Gender digital divide in Rwanda: A qualitative analysis of socioeconomic factors[J]. Telematics and Informatics, 2017, 34(7): 1285-1293.
- [15] NTIA. Falling Through the Net: Defining the Digital Divide[R]. Government Printing Office USA, 1999.
- [16] OCED. Understanding the Digital Divide[R/OL]. <http://www.oecd.org>, 2001.
- [17] Philip L, Cottrill C, Farrington J, Williams F, Ashmore F. The digital divide: Patterns, policy and scenarios for connecting the ‘final few’ in rural communities across Great Britain[J]. Journal of Rural Studies, 2017, 54: 386-398.
- [18] Puspitasari L, Ishii K. Digital divides and mobile Internet in Indonesia: Impact of smartphones[J]. Telematics and Informatics, 2016, 33(2): 472-483.
- [19] Raileanu S M. New insights from a multilevel approach to the regional digital divide in the European Union[J]. Telecommunications Policy, 2018, 42(6): 452-463.
- [20] Riggins F J, Dewan S. The Digital Divide: Current and Future Research Directions[J]. Journal of the Association for Information Systems, 2005, 6(12): 298-337.
- [21] Robles J M, Torres-Albero C, Marco S D. Spanish E-government and the Third Digital Divide: A Sociological View[J]. Journal of US-China Public Administration, 2011, 8(4): 12.
- [22] Sicherl P.A New Perspective in Comparative Analysis of Information Society Indicators [J]. Informatic, 1999, 23: 455-460.
- [23] Wei K K, Teo H H, Chan H C, Tan B C Y. Conceptualizing and Testing a Social Cognitive Model of the Digital Divide[J]. Information Systems Research, 2011, 22(1): 170--187.

- [24] Xie Y, Lu P. The sampling design of the China Family Panel Studies (CFPS)[J]. Chinese Journal of Sociology, 2015, 1(4): 471-484.
- [25] Yucer Y, Rizvanoglu K. Battling gender stereotypes: A user study of a code-learning game, "Code Combat," with middle school children[J]. Computers in Human Behavior, 2019, 99(OCT): 352-365.
- [26] 陈强.高级计量经济学及 Stata 应用[M].北京:高等教育出版社:2014.
- [27] 达摩达尔.N.古扎拉蒂,唐.C.波特.计量经济学基础[M].北京:中国人民大学出版社:2011.
- [28] 董君,洪兴建.数字鸿沟的内涵、影响因素与测度[J].中国统计,2019(12):71-73.
- [29] 何铨,张湘笛.老年人数字鸿沟的影响因素及社会融合策略[J].浙江工业大学学报(社会科学版),2017,16(04):437-441.
- [30] 洪海娟,万跃华.数字鸿沟研究演进路径与前沿热点的知识图谱分析[J].情报科学, 2014,32(04):54-58.
- [31] 侯艳辉,郝敏.基于 DEMATEL 的数字鸿沟影响因素辨识[J].统计与决策, 2013(14):15-17.
- [32] 胡鞍钢,周绍杰.中国如何应对日益扩大的“数字鸿沟”[J].中国工业经济, 2002(03):5-12.
- [33] 黄金,赵冬梅.基于 ISM 模型的数字鸿沟影响因素分析[J].统计与决策, 2011(08):41-43.
- [34] J.M.伍德里奇.计量经济学导论[M].北京:中国人民大学出版社:2003.
- [35] 金春枝,李伦.我国互联网数字鸿沟空间分异格局研究[J].经济地理, 2016,36(08):106-112.
- [36] 柯惠新,王锡苓.亚太五国/地区数字鸿沟及其影响因素分析[J].现代传播, 2005(04): 88-94.
- [37] 李健,邬晓鸥.统筹城乡发展背景下重庆市城乡数字鸿沟及其影响因素的实证研究[J].现代情报,2018,38(03):87-91+132.
- [38] 李平.互联网应用适老化改造加速助力老年人跨越“数字鸿沟”[J].网信军民融合, 2021(04):28.
- [39] 刘骏,薛伟贤.城乡数字鸿沟测度指标体系及其实证研究[J].预测,2012,31(05): 68-73.

- [40] 刘晓倩,韩青.农村居民互联网使用对收入的影响及其机理——基于中国家庭追踪调查(CFPS)数据[J].农业技术经济,2018(09):123-134.
- [41] 刘艳.基于知识图谱的国内数字鸿沟研究现状与演进趋势[J].国家图书馆学刊,2020,29(01):99-113.
- [42] 刘怡,张旭阳,高璐.知识生产视角下大学生群体数字鸿沟的影响因素研究——基于全国 29 个省市大学生的实证调查[J].中国电化教育,2019(03):127-133.
- [43] 刘芸.关于国际数字鸿沟影响因素的实证分析[J].统计与决策,2007(09):87-89.
- [44] 鲁元平,王军鹏.数字鸿沟还是信息福利——互联网使用对居民主观福利的影响[J].经济学动态,2020(02):59-73.
- [45] 陆峰,李新,周汝瑞.我国数字鸿沟的成因和影响因素及消除策略研究[J].科技情报开发与经济,2007(32):93-96.
- [46] 罗廷锦,茶洪旺.“数字鸿沟”与反贫困研究——基于全国 31 个省市面板数据的实证分析[J].经济问题探索,2018(02):11-18+74.
- [47] 彭继增,陶旭辉,徐丽.我国数字化贫困地理集聚特征及时空演化机制[J].经济地理,2019,39(02):169-179.
- [48] 戚聿东,刘翠花.数字经济背景下互联网使用是否缩小了性别工资差异——基于中国综合社会调查的经验分析[J].经济理论与经济管理,2020(09):70-87.
- [49] 邱娟,汪明峰.进入 21 世纪以来中国互联网发展的时空差异及其影响因素分析[J].地域研究与开发,2010,29(05):28-32+38.
- [50] 宋红岩.“数字鸿沟”抑或“信息赋权”?——基于长三角农民工手机使用的调研研究[J].现代传播(中国传媒大学学报),2016,38(06):132-137.
- [51] 宋周莺,刘卫东.中国信息化发展进程及其时空格局分析[J].地理科学,2013,33(03):257-265.
- [52] 汪明峰.互联网使用与中国城市化——“数字鸿沟”的空间层面[J].社会学研究,2005(06):112-135+244.
- [53] 王云生.数字鸿沟的内涵、成因及其对策探讨[J].河南图书馆学刊,2007(04):16-18.
- [54] 韦路,谢点.全球数字鸿沟变迁及其影响因素研究——基于 1990-2010 世界宏观数据的实证分析[J].新闻与传播研究,2015,22(09):36-54+126-127.

- [55] 吴鹏,马述忠.包容性发展与全球数字鸿沟[J].上海商学院学报,2021,22(05):14-26.
- [56] 谢宇,胡婧炜,张春泥.中国家庭追踪调查:理念与实践[J].社会,2014,34(02):1-32.
- [57] 谢宇.回归分析[M].北京:社会科学文献出版社:2013.
- [58] 徐芳,马丽.国外数字鸿沟研究综述[J].情报学报,2020,39(11):1232-1244.
- [59] 徐越,韵卓敏,王婧媛,景荣杰,黄黎明,沈勤.智能化背景下,老年人数字鸿沟的影响因素及其形成过程分析[J].智能计算机与应用,2020,10(02):75-82.
- [60] 许竹青,郑风田,陈洁.“数字鸿沟”还是“信息红利”?信息的有效供给与农民的销售价格——一个微观角度的实证研究[J].经济学(季刊),2013,12(04):1513-1536.
- [61] 薛伟贤,刘骏.基于技术扩散模型的区域“数字鸿沟”演变阶段划分[J].系统工程,2011,29(01):78-84.
- [62] 薛伟贤,刘骏.区域“数字鸿沟”形成原因分析[J].图书馆建设,2011(01):6-10.
- [63] 薛伟贤,刘骏.数字鸿沟的本质解析[J].情报理论与实践,2010,33(12):41-46.
- [64] 薛伟贤,刘骏.数字鸿沟主要影响因素的关系结构分析[J].系统工程理论与实践,2008(05):85-91.
- [65] 薛伟贤,王江泉.基于投入产出视角的区域数字鸿沟成因分析[J].科学学与科学技术管理,2011,32(02):86-95.
- [66] 薛伟贤,王涛峰.我国“数字鸿沟”的影响因素分析[J].情报杂志,2006(05):108-111.
- [67] 薛伟贤,张飞燕.数字鸿沟的成因、测度、影响及弥合方法[J].软科学,2009,22(01):17-24.
- [68] 闫慧,孙立立.1989年以来国内外数字鸿沟研究回顾:内涵、表现维度及影响因素综述[J].中国图书馆学报,2012,38(05):82-94.
- [69] 杨碧云,吴熙,易行健.互联网使用与家庭商业保险购买——来自 CFPS 数据的证据[J].保险研究,2019(12):30-47.
- [70] 杨焕敏,李新颖,何树娟.我国“数字鸿沟”研究述评[J].情报探索,2010(06):34-36.
- [71] 尹翔硕,刘能华.经济全球化进程中的数字鸿沟——基于跨国面板数据的分析[J].世界经济文汇,2008(02):84-99.

- [72] 尹志超,蒋佳伶,严雨.数字鸿沟影响家庭收入吗[J].财贸经济,2021,42(09):66-82.
- [73] 袁勤俭,黄奇,朱庆华.影响美国数字鸿沟的因素分析[J].情报科学,2005(03):349-354.
- [74] 袁晓东.互联网环境下的“数字鸿沟”影响因素分析[J].采写编,2019(03):129-131.
- [75] 曾凡斌.第二道数字鸿沟的影响因素研究——基于对大学生实证分析[J].江淮论坛,2011(01):134-144.
- [76] 张彬,陈双,李潇.我国数字鸿沟影响因素关系结构模型研究[J].北京邮电大学学报(社会科学版),2009,11(04):28-33.
- [77] 张彬.我国区域信息鸿沟测度的实证研究[J].北京邮电大学学报(社会科学版),2009,11(02):8-13.
- [78] 张丽,黄腾,刘天军.互联网能弥合农产品销售市场的数字鸿沟吗?——基于陕西省苹果种植户的微观数据分析[J].农林经济管理学报,2018,17(06):660-668+737.
- [79] 张新红.数字鸿沟测算方法比较[J].电子政务,2008(11):16-23.
- [80] 张正平,卢欢.数字鸿沟对家庭金融投资的影响——基于 CFPS 数据的实证研究[J].福建论坛(人文社会科学版),2021(03):57-70.
- [81] 张正平,卢欢.数字鸿沟研究进展[J].武汉金融,2020(03):64-71+84.
- [82] 朱建平.应用多元统计分析[M].北京:科学出版社:2016.
- [83] 朱莎,杨浩,冯琳.国际“数字鸿沟”研究的现状、热点及前沿分析——兼论对教育信息化及教育均衡发展的启示[J].远程教育杂志,2017,35(01):82-93.
- [84] 祝建华.数码沟指数之操作定义和初步检验[A].21世纪新闻传播研究[C].汕头:汕头大学出版社,2002:203-211.
- [85] 祝坤子,沙晋明,刘玉琴,徐赛萍,沙昱.中国信息化发展时空格局演变分析[J].中国科技论文,2015,10(21):2574-2579+2584.
- [86] 祝仲坤,冷晨昕.互联网使用对居民幸福感的影响——来自 CSS2013 的经验证据[J].经济评论,2018(01):78-90.

后 记

三年硕士学习经历转瞬即逝，期间由于新冠肺炎疫情，经历线上教学、封校防控，让这段时光尤为波折和难以忘怀。从研一的懵懂迷茫，到研二的充实忙碌，再到研三的收获满满，是一个逐渐认识自我、接受自我、又突破自我的成长过程，回首总结，疑惑焦虑不再，只有满心感慨和感恩。

由衷感谢我的导师孙玉环教授，学术研究上，您卓越深厚的专业素养和认真严谨的治学学风令我钦佩仰慕，无论参加比赛还是撰写论文，无论方法技能还是学术规范，您都言传身教、倾囊相授，是我不断前行的动力源泉和终生受益的宝贵财富；日常生活中，您的殷切关怀和悉心指引让我如沐春风、倍感温暖，能成为您的学生、受您教诲是我学生生涯中最最幸运的事。感谢庞智强教授的亲切教导和耐心指点，您精益求精的工作作风和正直高尚的个人品格对我影响深远，是我为人处世的榜样楷模。同时，感谢统计学院每一位老师的栽培教育，愿老师们学术长青、桃李满园。

感谢室友李妍蓉、许梅、王彤（按学号排序，拒绝踩坑），我们一起学习探索、互相帮助陪伴，你们的存在让寝室像家一样温馨自在；感谢师弟蒙嘉宁和师妹任萍、尤怡卜、李愉航真诚的信任与帮助，你们是最优秀可靠的学习伙伴；感谢统计学院 2019 级所有同学在我担任班长职务时的理解与包容，愿我们都前程似锦、未来可期。

最后，感谢父母家人的呵护培养，朋友的开解陪伴，石岩同学的鼓励支持，以及世界上最可爱的小猫咪咕噜，你们是我永远的后盾和港湾。

我不再是穿梭于校园的学生，但学习和进步之路永无止境，希望可以永远带着严谨的态度和充沛的求知欲面对未知旅途，永远保留善良和童心，不执着于外界的声音，温柔坚定，热情努力，与自己和解。

杨知雨

2022 年 3 月 23 日