

重庆市数字经济发展对城乡收入差距的影响研究

摘要

数字经济正日益成为促进中国社会经济发展的新动力，但“数字鸿沟”使地区间数字经济发展失衡、城乡收入差距等问题日益凸显。因此，本文建立数字经济发展评价模型，并从泰尔指数的视角出发，研究数字经济发展与城乡差距的关系，探索数字经济发展在缩小城乡差距中的瓶颈，并提出相应建议。

对于问题 1，首先，聚焦于数字产业化、产业数字化和发展环境三大领域，结合移动电话普及率、互联网普及率、电信业务总量、数字普惠金融指数及 $R\&D$ 研发支出五大视角，构建重庆市及其区县数字经济发展水平评估体系。其次，建立基于灰色理论的 $GM(1,1)$ 模型，对数据缺口进行填补。然后，采用泰尔指数测算近九年的城乡差距。最后，对结果可视化和对比分析，当前数字经济发展态势良好，但城乡之间存在较大差距。

对于问题 2，首先，使用熵权法，对各个数字经济发展评价指标进行客观定权，得到的权重结果为：移动电话普及率 (0.0296)，互联网普及率 (0.0866)，电信业务总量 (0.4143)，数字普惠金融指数 (0.0003)，以及 $R\&D$ 投入支出水平 (0.4652)。然后，通过权重建立数字经济得分公式，得到重庆市及其区县的数字经济得分，其中 2023 年重庆市数字经济得分为：535.2861。

对于问题 3，首先，基于前文建立的数字经济得分公式，计算 2015 年至 2023 年重庆市的年度数字经济得分，并通过标准化处理消除量纲差异，构建数字经济发展水平与城乡收入差距的回归模型。其次，利用 spss 进行曲线估算，发现通过二次函数拟合时 R^2 达到最大，证明数字经济发展与城乡差距之间为 U 型关系。

对于问题 4，首先，计算二次函数关系式的对称轴，得出当标准化后的数字经济得分为 21.831 时，数字经济发展对城乡差距的缩小效应达到瓶颈。然后，猜想瓶颈产生的原因与城乡之间“数字鸿沟”有关，并对数字经济发展与城乡居民人均可支配收入进行回归分析，发现数字经济对城市居民收入增长的推动力明显超过农村，证明了“数字鸿沟”的存在。

对于问题 5，首先，根据前文研究总结出，在瓶颈 (21.831) 前后，数字经济发展对城乡差距先存在缩小效应后存在增大效应。因此，基于本文发现提出针对性建议：避免发展瓶颈的出现，解决数字鸿沟问题；政府加强政策引导，推动城乡经济协调发展；从宏观调控出发，完善收入分配机制。

关键字： 数字经济 城乡差距 灰色理论 泰尔指数 熵权法 回归分析

一、问题重述

1.1 问题背景

数字经济是当前世界创新最活跃、增长最迅猛的领域，具有高创新性、强渗透性和广覆盖性。数字技术打破了时空限制，在推动制造业转型升级、实现农业现代化、改善营商环境和民生方面发挥了重要作用，成为我国经济增长的新引擎。

近年来，我国政府工作报告中多次提到数字经济的发展规划，表明国家已经开始在战略层面高度重视数字经济发展。2019年，重庆获批建设国家数字经济创新发展实验区，集中力量打造“智造重镇”和“智慧名城”，逐渐成为我国数字经济发展的的重要支撑力量，并在缩小城乡“数字鸿沟”、推动乡村经济发展方面取得显著成效。在我国大力发展数字经济和实施乡村振兴战略的背景下，分析我国数字经济发展现状并研究其对城乡收入差距的影响，可以更准确地把握数字经济的运行规律和发展趋势，为缩小城乡差距、协调区域发展、实现共同富裕提供新思路和新对策。

1.2 问题重述

为促进重庆市数字经济发展、缩小城乡收入差距，结合实际情况与题目信息建立数学模型，分析以下问题：

问题一：收集重庆市及其区县的数字经济发展数据和城乡收入差距数据，进行描述性分析，了解重庆市及各区县在数字经济发展和城乡收入方面的现状。

问题二：确定衡量数字经济发展程度的指标，建立数字经济发展评价模型，并根据收集的数据计算重庆市及各区县的数字经济得分，以量化各区域的数字经济发展水平，为进一步分析提供基础。

问题三：分析数字经济发展对城乡收入差距的影响，探讨数字经济在缩小城乡收入差距方面的作用，并得出相应结论。

问题四：识别重庆市及其区县在通过数字经济发展缩小城乡收入差距过程中遇到的主要瓶颈和障碍，并研究这些瓶颈产生的原因。

问题五：基于研究结果，提出针对性的政策建议，帮助进一步推动数字经济发展、缩小城乡收入差距，为实现共同富裕提供支持。

二、问题分析

2.1 问题一的分析

为全面剖析 2023 年重庆市及其区县的数字经济发展现状与城乡收入差距问题，首先选取具有代表性和数据可得性的重庆市及其区县作为分析样本，并依据灰色理论对缺失值进行插补。其次，筛选确定反映数字经济发展水平的指标，构建数字经济发展水平评估指标体系。然后，依据指标对重庆市及其区县的数字经济发展水平进行可视化，并采用泰尔指数表示并分析城乡收入差异的现状。

2.2 问题二的分析

本问题要求建立综合评价模型，旨在量化重庆市及其区县的数字经济发展水平。针对该问题，首先依据数字经济发展综合指标体系，使用熵权法，对各个数字经济发展评价指标进行客观定权。然后，通过权重建立数字经济得分公式，进而确定重庆市及其各个区县的数字经济发展得分。

2.3 问题三的分析

为得出数字经济发展对城乡差距的影响，首先，基于数字经济得分公式，计算 2015 年至 2023 年重庆市的年度数字经济得分。然后，对数据进行标准化处理，并构建数字经济发展水平与城乡收入差距的回归模型。最后，利用 spss 等软件对该关系进行曲线估算，结合现有研究结果，确定拟合程度最佳的函数模型，并得到函数解析式。

2.4 问题四的分析

通过查阅资料可知，数字经济发展与城乡差距之间存在着 U 型关系 [5]，因此数字经济发展在缩小城乡收入差距过程中的瓶颈可以通过开口向上的二次函数的对称轴横坐标来表示，表明数字经济发展对城乡收入差距先有积极的缩小效应，达到一定水平后，数字经济发展对城乡收入差距产生消极的扩大效应，即出现了发展瓶颈。

首先，结合现有研究成果以及计算得到的二次函数关系式，计算其对称轴横坐标，得出数字经济发展对城乡差距的缩小效应达到瓶颈的位置。然后，对数字经济发展与城乡居民人均可支配收入进行回归分析，探讨发展瓶颈产生的原因。

2.5 问题五的分析

综合重庆市及其区县数字经济发展水平和城乡差距状况、数字经济发展对城乡差距的影响因素以及发展鸿沟产生的位置与原因等分析结果，提出针对性的政策建议，推动数字经济发展，缩小城乡收入差距，弥合“信息鸿沟”。

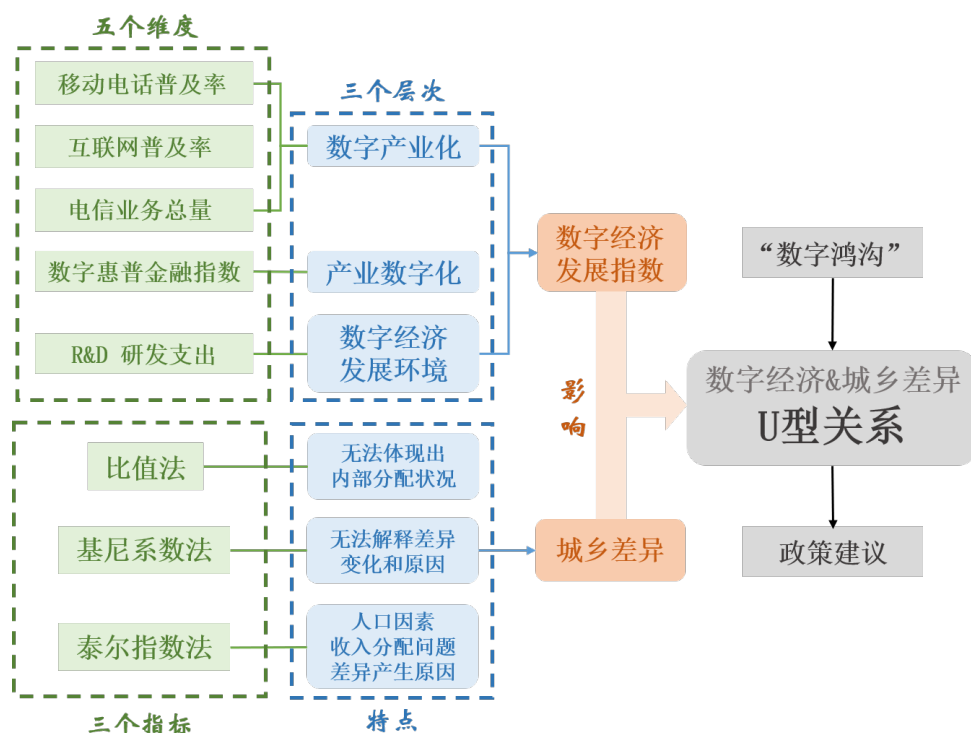


图 1 模型建立与分析流程图

三、模型假设

本文提出以下合理假设：

- 查找与插补的 2015 年至 2023 年重庆市及其区县各项数据均准确。

本文所使用的数据均来自各区县的统计公报和统计年鉴，具有极高的权威性与准确性；重庆市社会经济发展稳定，可认为其数字经济指标在短时间内不会发生剧烈变化，使得数字经济指标的超短短期预测具有可行性。

- 数字经济发展水平与城乡居民收入差距之间的关系呈现 U 型关系。

研究指出，数字经济发展初期可以缩小城乡居民收入差距，但随着数字经济的进一步推进，城乡数字鸿沟效应变得明显，可能导致收入差距扩大。因此，数字经济发展水平与城乡居民收入差距之间的关系呈现 U 型关系。

四、符号说明

符号	意义
$X^0(n)$	数字经济指标在某一年的数值
$Theil_t$	区分高峰和低峰的一个临界值
t	年份
I_{1t}	t 年份城镇居民人均可支配收入与当地常住人口的乘积
I_{2t}	t 年份农村居民人均可支配收入与当地常住人口的乘积
I_t	城乡居民总收入之和
P_{1t}	t 年份城镇地区常住人口总数
P_{2t}	t 年份农村地区常住人口总数
$r(i, j)$	第 i 个地区在第 j 个指标上的统计值
$H_{(j)}$	第 j 个指标的熵
$w_{(j)}$	第 j 个指标的熵权
$S_{(i)}$	第 i 个区县数字经济得分

五、问题一的模型建立与求解

数字经济是基于数字化技术、以信息作为核心生产要素，通过数字化技术和实体经济的深度融合，以实现为传统产业赋能并催生出新业态的新型经济形态 [2]。为深入研究数字经济对城乡收入差距的影响，本章以 2023 年重庆市及其 27 个区县（剔除部分数据缺失较为严重的区县数据）为研究对象，选取部分指标，对这些地区数字经济发展水平和城乡收入差距状况进行描述性统计分析，旨在全面反映其发展趋势与现状，为后文的进一步分析打下基础。

5.1 数字经济发展水平的统计与分析

5.1.1 指标选择与数据处理

本文从数字产业化、产业数字化、数字经济发展环境三个方面，选取了移动电话普及率、互联网普及率、电信业务总量、数字普惠金融指数和 $R\&D$ (Research and Development)

投入支出水平等五个指标 [1]，来考察重庆市及其区县的数字经济发展水平。针对研究对象中存在的部分缺失数据，依据灰色理论进行预测填补 [7]。

模型构建

灰色模型能够对缺失的数字经济指标进行预测的基本原理在于两个方面：

(1) 由于当前重庆市社会经济发展相对平稳，其数字经济指标在短时间内难以发生剧烈变化，这使得数字经济指标的超短短期预测具有可行性；

(2) 灰色理论尤其适合于在样本数据很少、信息量极少的条件下进行预测，且预测结果有效。

基于此，本文采用传统的灰色模型 $GM(1,1)$ ，预测 2023 年重庆市及其区县的某些数字经济指标，具体方法为：

时序数字经济指标为：

$$X^0 = [X^0(1), X^0(2), \dots, X^0(n)] \quad (1)$$

其中， $X^0(n)$ 表示该数字经济指标在某一年的值。

对数列按照传统 $GM(1,1)$ 模型的算法进行一次累加，生成数列：

$$\begin{aligned} X^{(1)} &= [X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n)] \\ X^{(1)}(k) &= \sum_{i=1}^k X^{(0)}(i) \quad \forall k = 1, \dots, n \end{aligned} \quad (2)$$

构造 $GM(1,1)$ 模型的一阶微分方程：

$$\frac{dX^{(1)}(t)}{dt} + aX^{(1)} = b \quad (3)$$

在上式中，模型系数 a 和 b 可用最小二乘法求得，即：

$$A = \begin{bmatrix} a \\ b \end{bmatrix} = (\beta^T \beta)^{-1} \beta^T Y \quad (4)$$

其中，

$$\beta = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \vdots & 1 \\ -Z^{(1)}(n) & 1 \end{bmatrix}; \quad Y = \begin{bmatrix} -X^{(0)}(2) \\ -X^{(0)}(3) \\ \vdots \\ -X^{(0)}(n) \end{bmatrix} \quad (5)$$

$$Z^{(1)}(i) = \frac{X^{(1)}(i-1) + X^{(1)}(i)}{2} \quad (6)$$

预测生成序列 \hat{X} 可由下述公式得到：

$$\hat{X}^{(1)}(i+1) = \left(X^{(0)}(1) - \frac{b}{a} \right) e^{-ai} + \frac{b}{a} \quad (7)$$

预测的结果按照下式进行还原：

$$\hat{X}^{(0)}(i+1) = \hat{X}^{(1)}(i+1) - \hat{X}^{(1)}(i), i = 1, \dots, n \quad (8)$$

由此可以得到预测结果。

实例分析

在本节中，以江津区的互联网用户数为例，来操作上述灰色预测步骤。

通过查阅资料得到，2017 至 2021 年江津区的互联网用户人数，形成历史时序互联网用户人数数据，并生成预测数列：

$$X(0) = [31.9, 38.7, 45.0, 49.3, 54.6] \quad (9)$$

对数列进行一次累加生成数列：

$$X(1) = [31.9, 70.6, 115.6, 164.9, 219.5] \quad (10)$$

构造 $GM(1,1)$ 模型的一阶微分方程为：

$$\frac{dX^{(1)}(t)}{dt} + aX^{(1)}(t) = b \quad (11)$$

然后按照公式（4）公式（5）的方法计算模型的参数向量矩阵 A ，解得 $a=-0.11$ ， $b=337.7$ 。

再根据公式（6）计算预测生成数列，得到预测值为：[31.9,39.4,44.0,49.1,54.8]。

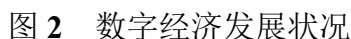
由此可知，2022 年江津区互联网用户数为 68.4 万。

5.1.2 描述性统计与分析

对已填补完整的 2023 年重庆市及其 27 个区县各个数字经济指标的数据进行可视化处理，结果如下所示。由于我们的研究对象中仅包含北碚区一个主城区，因此，综合分析图表内容，可以得出以下结论：

（1）重庆市数字经济发展水平整体较好：由图 2 知，对于选定的五个指标，重庆市的平均水平均高于全国平均水平，说明在数字产业化、产业数字化和数字经济发展环境三个方面，重庆都取得了较程度的发展，这得益于我国“西部大开发战略”的实施与“成渝双城经济圈”的建立 [4]。

（2）重庆市区县的数字经济发展存在短板：由图 2a、图 2c、图 2d 可知，在移动电话普及率、电信业务总量和数字普惠金融指数三个方面，区县的发展水平均低于重庆市



(3) 主城与区县的数字经济发展不平衡：由图 2c 可知，北碚区的电信业务总量接近 70，而区县的电信业务总量大部分不到 10，说明主城与区县在数字产业化方面发展水平悬殊。此外，在重庆市各区县研发投入方面，半数区县的研发投入不足 5，仅为北碚区研发投入的七分之一，说明区县在科学研究与试验发展方面的经费投入不足，地方部门对科学技术和自主创新领域的关注度不够。总之，主城和区县之间的数字经济发展不平衡，存在“数字鸿沟”。

5.2.1 城乡收入差距的测度

(1) 比值法: 即用城乡居民的收入比值表示城乡收入差距水平, 城乡居民的收入比

值越高，表明居民收入差距越大。该指标计算简便且反映直观，但只能反映城市和农村人均收入的相对变动而无法体现出内部分配状况。

(2) 基尼系数法：即用基尼系数数值表示城乡收入差距水平，基尼系数越大，表明居民收入差距越大。该指标能够利用数值简化洛伦兹曲线所描述的居民收入分配情况，但不能解释差异变化和原因。

(3) 泰尔指数法：即用泰尔指数数值表示城乡收入差距水平，泰尔指数越大，表明居民收入差距越大。该指标综合考虑了人口因素、收入分配问题以及差异产生原因，弥补了前两种方法的不足 [3]。因此，本文借鉴程名望等（2019）的做法，采用泰尔指数测算城乡收入差距，计算公式如下所示：

$$Theil_t = \sum_{i=1}^2 \left(\frac{I_{it}}{I_t} \right) \ln \left(\frac{\frac{I_{it}}{P_{it}}}{\frac{I_t}{P_t}} \right) \quad (12)$$

其中， $Theil_t$ 表示用泰尔指数表征的 t 年份城乡收入差距； I_{1t} 和 I_{2t} 分别表示 t 年份城市和乡村居民的人均可支配收入与当地常住人口的乘积； I_t 表示城乡居民总收入之和； P_{1t} 和 P_{2t} 分别表示 t 年份城镇和乡村地区常住人口。

5.2.2 城乡收入差距的可视化分析

根据前文所述，将整理的数据代入公式，得出重庆市 2015 年 2023 年的泰尔指数。重庆市 2015 年 2023 年的城镇居民人均可支配收入、乡村居民人均可支配收入和泰尔指数的变化如下图所示：

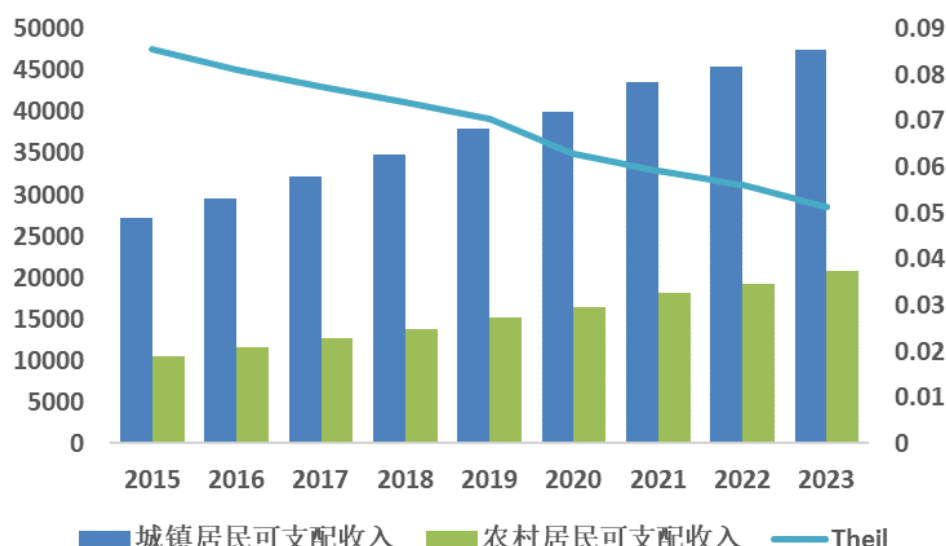


图 3 城乡收入差距情况

从图中可以看出，重庆市城乡居民的人均可支配收入都呈上升趋势，城镇人均可支配收入从 2015 年 27239 元增至 2023 年的 47435 元；农村人均可支配收入从 2015 年

10505 元增至 2023 年的 20820 元。由此可知，我国城镇居民的人均可支配收入一直高于农村居民，且这种差异已经持续了相当长一段时间，且城镇居民的收入水平往往是农村居民的 2 倍以上。这表明重庆市城乡发展仍存在明显差异性。但是，表征城乡收入差距水平的泰尔指数一直呈明显的下降趋势，从 2015 年的 0.086 下降至 2023 年的 0.051。这表明重庆市城乡发展的差异性正在逐渐缩小。

六、问题二的模型建立与求解

数字经济能够创造经济价值和社会经济效益，对经济和社会的影响是多方面的，因此需要具有合理性、代表性的指标来衡量。本文将已有的数字经济研究维度进行整合，并对其进行了完善与扩展，形成了“三层五维”的数字经济发展综合指标体系，并利用熵权法得出每个指标的客观权重，从而建立数字经济发展评价模型，计算重庆市及其区县数字经济得分。

6.1 数字经济发展综合指标体系的构建

如前文所述，本文主要从数字产业化、产业数字化、数字经济发展环境三个方面来构建数字经济的指标体系，共选取了移动电话普及率、互联网普及率、电信业务总量、数字普惠金融指数和 *R&D*（Research and Development）投入支出水平等五个指标。

一方面，数字经济的重要表现形式具有双重特征，既包括推动数据要素和数字技术转化为可创造经济效益的产业，实现数字产业化的过程，也包括将数字化技术与各行各业深度融合，推动传统产业的数字化升级，实现产业数字化的过程。本文通过移动电话普及率、互联网普及率、电信业务总量观测数字产业化状况，从数字普惠金融指数角度衡量产业数字化状况。另一方面，数字经济发展环境决定着数字经济的可持续发展能力，可以从 *R&D* 投入支出水平来考虑。

本文根据数字经济的内涵，参照已有的研究成果，对指标的全面性、合理性、可获得性和可测量的时间进行全面的考量，构建数字经济指数体系。该指标体系的一级指标共涵盖 3 个二级指标下 5 个三级指标，具体指标见表 1。

6.2 数字经济得分的测算

6.2.1 测算方法的选择：熵权法

本文所使用的方法为熵权法。该方法属于客观赋权法，是一种多指标权重确定方法。通过归纳已有文献选取数字经济指标。先对选定指标进行数据收集和处理，通过熵权法得出每个指标的客观权重，从而得到数字经济发展水平的综合性指标 [8]。

熵权法的计算过程如下：

表 1 数字经济指标体系

一级指标	二级指标	三级指标
数字经济发展指数	数字产业化	移动电话普及率
		互联网普及率
		电信业务总量
	产业数字化	数字普惠金融指数
	数字经济发展环境	R&D 研发支出

首先，对指标进行标准化处理，以消除消除指标数据单位和意义的差异对计算结果的影响。标准化公示如下：

$$r_{(i,j)} = \frac{r_{(i,j)}}{\sqrt{\sum_{i=1}^{25} (r_{(i,j)})^2}} \quad (i = 1, 2, \dots, 27; j = 1, 2, \dots, 5) \quad (13)$$

其中, $r_{(i,j)}$ 表示第 i 个地区在第 j 个指标上的统计值。

其次，构造评价矩阵：

$$R = (r_{ij})_{27 \times 5} \quad (i = 1, 2, \dots, 27; j = 1, 2, \dots, 5) \quad (14)$$

然后，计算各指标的信息熵。第 j 个指标的熵 $H_{(j)}$ 公式如下：

$$H_{(j)} = -k \sum_{i=1}^{27} f_{ij} \ln f_{ij} \quad (15)$$

其中, $f_{(i,j)} = r_{(i,j)} / \sum_{i=1}^{27} r_{(i,j)}$, $k = 1 / \ln 27$ 。当 $f_{(i,j)} = 0$ 时, $\ln f_{(i,j)} = 0$ 。

最后，确定第 j 个指标的熵权 $w_{(j)}$ ，公式如下：

$$W_{(j)} = (1 - H_{(j)}) / 5 - \sum_{j=1}^5 H_{(j)} \quad (16)$$

综上，第 i 个区县数字经济得分的计算公式为：

$$S_{(i)} = \sum_{j=1}^5 w_{(j)} \times (r_{(i,j)})' \quad (17)$$

6.2.2 测算结果分析

将 2023 年重庆市及其 27 个区县五个关于数字经济发展水平的指标代入计算，得出各指标权重如下表所示：

表 2 数字经济指标权重

一级指标	二级指标	三级指标	权重
数字经济发展指数	数字产业化	移动电话普及率	0.0296
		互联网普及率	0.0866
		电信业务总量	0.4183
	产业数字化	数字普惠金融指数	0.0003
	数字经济发展环境	R&D 研发支出	0.4652

根据表 2 的测算结果可知，数字经济的发展环境中的 *R&D* 投入支出水平所占的比重最大，达到了 46.52%。这说明数字经济的发展环境尤其是 *R&D* 投入支出水平很大程度上影响数字经济发展水平。数字产业化权重占比过半，在二级指标中为第一高，表明了数字经济发展过程中数字产业化建设起到了重要的支撑作用。相比之下，产业数字化中的数字普惠金融指数占比小到几乎可以忽略不计，可以推断其对数字经济发展的影响不如其他指标显著。

将 2023 年重庆市及其 27 个区县五个关于数字经济发展水平的指标数据代入数字经济得分的计算公式，可得到重庆市及每个区县的数字经济发展水平得分，结果如下图所示：

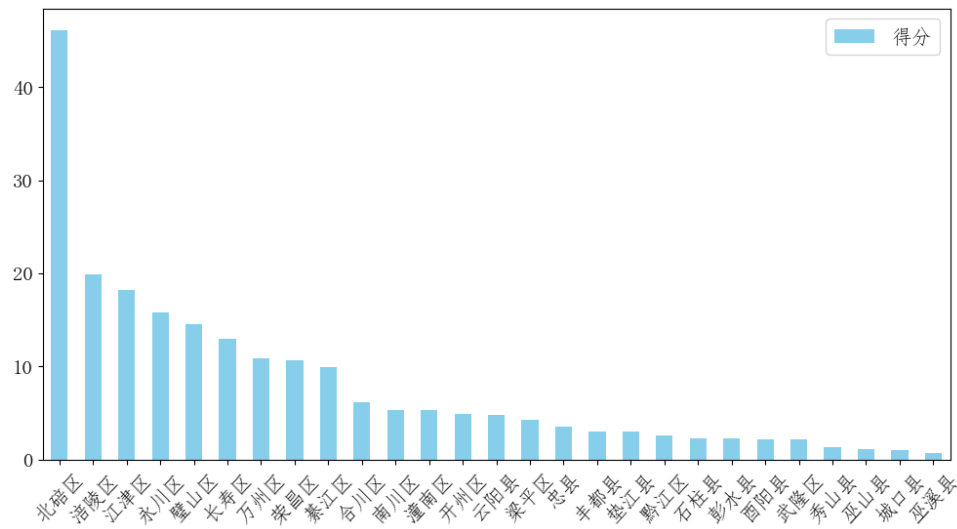


图 4 数字经济得分

结合计算出来的权值，可得 2023 年重庆市数字经济得分为 535.2861。

七、问题三的模型建立与求解

本部分利用 spss 软件，对数字经济发展和城乡收入差距之间的关系进行曲线估算，得到拟合程度最佳的关系为二次函数关系，进而得到表达式城乡收入差距（以泰尔指数表示）关于数字经济发展的二次表达式，并对数字经济发展对城乡收入差距的影响进行分析。

7.1 U 型关系的理论来源

根据陈文、吴赢等人的研究 [5]，数字经济的发展主要能从以下方面影响城乡居民收入差距：

- 数字经济可以通过影响城镇化进程来影响城乡居民收入差距。

首先，数字经济可以提高进城务工农民工的收入，缓解农村地区人地紧张的状况，提高农业生产率，进而增加农民收入。同时，数字经济促进的城镇化能够抑制城镇居民工资水平的过快增长，有利于城乡要素报酬的均等化，缩小收入差距。此外，数字经济的发展使得城镇化更容易获得金融资源支持，通过提升金融资源配置效率和财政透明度，支持地方政府扩展城市基础设施投资，加快城镇化进程。数字经济还促进中小企业的发展，为中小企业提供新的销售渠道，增加对劳动力的需求，提升农村剩余劳动力的就业机会。最后，数字经济的发展打破劳动力市场的信息壁垒，使得农村劳动力能够便捷低成本地找到适合的工作，加快就业，从而进一步促进城镇化。

- 数字经济能够影响农民的创业水平，进而影响城乡居民收入差距。

在资金方面，数字金融利用大数据获取信用积分，不需要借款人提供抵押物，降低了贷款门槛，缓解了农村居民创业的融资约束，激发了他们的创业意愿。同时，国家对通信基础设施的巨大投资，使得移动互联网在农村得到普及，农村居民可以使用手机轻易地获取信息，学习创业所需的知识，降低创业的门槛和风险。在数字经济初期，这些因素有效地刺激了农村居民的创业行为，提高了农村地区的收入，从而缩小了城乡居民的收入差距。

- 数字经济发展产生的城乡数字鸿沟会影响城乡居民收入差距。

数字经济要求高技能劳动力，而由于农村居民受教育水平普遍低于城市居民，人力资本储备不足，农村剩余劳动力难以满足城市对于数字经济时代新型人才的需求，导致了逆城镇化问题。此外，数字经济发展依赖于人工智能、大数据、云计算等技术，低技能劳动力的失业率上升和收入下降，使得城乡居民收入差距扩大。即使互联网的使用普及了，农村居民在信息的鉴别和利用能力方面仍存在不足，影响了创业收益。随着数字经济的发展，人力资本差异导致的城乡居民在应用数字技术能力上的差距变得日益显著，农村地区创业带来的收入增加效应下降，使得城乡居民收

入差距再次扩大。

因此，数字经济发展和城乡收入差距的关系呈现为一个 U 型曲线。数字经济发展的初期有助于缩小城乡收入差距，但随着发展深入，数字鸿沟和技能差异问题导致城乡收入差距扩大。

7.2 相关性检验

首先，利用问题二中得到的权值，计算得到重庆市在 2015-2023 年的数字经济得分，分别为：[126.583730 , 159.446515 , 269.885273 , 287.869033 , 385.446006 , 438.718866 , 483.280304 , 535.286125]。

然后，对 2015-2023 年重庆市的数字经济发展得分和泰尔指数进行相关性分析，计算得到二者的相关系数为-0.997，且显著性水平 $P < 0.01$ ，表明二者存在极强的负相关性且在 1% 的水平上显著。

7.3 曲线估算与拟合

利用 spss 软件，用多种模型对数字经济发展和城乡收入差距之间的关系进行拟合，探索二者之间的函数关系，得到的结果如下所示：

表 3 曲线拟合结果

模型摘要						参数估计值			
方程	R 方	F	自由度 1	自由度 2	显著性	常量	b1	b2	b3
线性	0.995	1383.268	1	7	0	0.095	-8.08E-05		
对数	0.968	211.414	1	7	0	0.199	-0.023		
逆	0.874	48.704	1	7	0	0.048	5.326		
二次	0.996	628.716	2	6	0	0.096	-8.91E-05	1.27E-08	
三次	0.995	369.744	3	5	0	0.098	0	1.19E-07	-1.08E-10
复合	0.993	1061.087	1	7	0	0.1	0.999		
幂	0.943	115.36	1	7	0	0.458	-0.337		
S	0.83	34.221	1	7	0.001	-2.993	77.26		
增长	0.993	1061.087	1	7	0	-2.306	-0.001		
指数	0.993	1061.087	1	7	0	0.1	-0.001		

由表可知，当二者之间的关系为二次函数时， $R^2 = 0.996$ 达到最大值，且显著性水平 $P < 0.001$ ，说明该结果在 1% 的水平上显著。因此，认为数字经济发展水平和城乡收入差距之间的关系为二次函数。

因此，选取“二次”作为函数模型，对数字经济发展水平和城乡收入差距之间的函数表达式进行拟合。为了排除量纲的影响，先对数字经济发展指数和泰尔指数的值进行标准化，再使用 spss 软件进行曲线估计，得到泰尔指数关于数字经济发展指数的函数表达式为：

$$\text{Theil}_{(i)} = 0.023 S_{(i)}^2 - S_{(i)} - 0.02 \quad (18)$$

其中， $R^2 = 0.996$ ，说明数字经济发展得分在二次关系上对泰尔指数有极强的解释能力；显著性水平 $P < 0.001$ ，说明该函数表达式在 1% 的水平上显著。因此，可以认为该模型的拟合效果非常好。

由此可知，数字经济发展水平与城乡居民收入差距之间存在 U 型关系，即随着数字经济发展水平的提高，城乡居民收入差距先缩小后增大。

基于已有文献可知，数字经济可以通过影响城镇化进程和农民创业水平来影响城乡居民收入差距 [6]。具体而言，数字经济发展的初期可以提高城镇化进程和农村居民创业水平，从而增加农村居民收入，缩小城乡居民收入差距。然而，随着数字经济的发展，城乡数字鸿沟效应变得比较明显，会出现“逆城镇化”现象以及数字经济对于农村居民创业支持的有效性下降，导致城乡居民收入差距扩大。

八、问题四的模型建立与求解

8.1 瓶颈位置的确定

由公式（17）计算可得，二次曲线的对称轴的横坐标为 21.831，即当标准化后的数字经济发展得分为 21.831 时，数字经济对城乡差距的缩小作用就已经达到了瓶颈。即在标准化后得分 < 21.831 时，城乡收入差距随数字经济发展得分的增加而缩小；在标准化后得分 > 21.831 时，城乡收入差距随数字经济发展得分的增加而扩大。

因此，可将标准化后的数字经济发展得分 $= 21.831$ 作为缩小城乡差距的瓶颈，当数字经济发展超过这一得分所代表的经济发展水平后，其不仅不会缩小城乡差距，反倒适得其反。

8.2 瓶颈来源的回归分析

通过前文的实证分析发现，数字经济超过一定限度的发展会引起城乡收入差距的扩大。为进一步验证这一结论，并寻找导致这一现象的原因，本文对数字经济发展水平与城市居民人均可支配收入、农村居民人均可支配收入分别进行回归分析，结果报告于表 4。

表 4 数字经济发展与人均可支配收入回归分析

		未标准化系数		标准化系数	t	显著性
		B	标准错误	Beta		
城镇居民人均	常量	22074.965	762.185		28.963	0
可支配收入	数字经济得分	48.346	2.181	0.993	22.166	0
农村居民人均	常量	7575.16	302.746		25.022	0
可支配收入	数字经济得分	24.242	0.866	0.996	27.982	0

回归结果显示，数字经济发展水平对城市和农村的居民人均可支配收入的回归系数均为正，验证了数字经济在一定程度内的发展对城乡居民收入增长均具有促进效应。从城乡差别来看，数字经济发展水平对城镇居民可支配收入的回归系数为 48.346，大于其对农村居民可支配收入的回归系数为 24.242，且两个回归系数都在 1% 的水平上显著。

这说明虽然数字经济水平的发展对城乡居民的收入增长均具有促进作用，但在城市地区，这一促进作用更加明显。这说明数字经济发展虽然显著促进了我国城乡居民收入的增长，但由于数字鸿沟的存在，其对城市居民增收的作用要大于农村居民，即数字经济的红利在城乡之间没有实现均衡分配，导致在数字经济发展的后期，城乡收入差距不断扩大，使得二者的关系总体呈现出“先减后增”的 U 型关系。

九、问题五的结论与建议

9.1 主要结论

本文借助灰色理论描述了 2023 年重庆市及其区县的数字经济发展和城乡收入差距的发展现状，建立了包括三个二级指标、五个三级指标的数字经济发展评价模型，采用熵权法测度了 2015-2023 年重庆市的数字经济发展水平，探讨了数字经济发展对城乡收入差距的影响机制并找到了发展瓶颈，进一步通过回归分析分析了产生这一瓶颈的原因，主要得到以下几点结论：

第一，本文构建了囊括数字产业化、产业数字化、数字经济发展环境三个维度、移动电话普及率、互联网普及率、电信业务总量、数字普惠金融指数和 R&D 投入支出水平五个指标的数字经济发展综合评价体系。二级指标层面，数字产业化权重最高，三级指标层面，数字经济的发展环境中的 R&D 投入支出水平的占比最大。这说明数字产业化和 R&D 投入支出水平在数字经济发展过程中起到了重要的支撑作用。

第二，结合数字经济发展综合评价体系，通过熵权法的测算得到我国数字经济发展的整体水平，发现重庆市数字经济发展态势良好，但区县发展存在短板，主城与区县间数字经济发展水平差异悬殊，区域间数字鸿沟明显，且存在不断增大的趋势。

第三，通过数字经济发展影响城乡收入差距的理论分析，发现数字经济发展可能对城乡收入差距产生正负两个方向的影响，使得数字经济发展水平与城乡居民收入差距之间存在 U 型关系。通过对数字经济发展水平和城乡人均可支配收入的回归分析发现，“数字鸿沟”是导致数字经济发展扩大城乡收入差距的重要原因。

9.2 政策建议

本文研究发现，数字经济的发展在后期对城乡收入差距产生了扩大效应，数字经济的发展伴随着数字鸿沟问题。基于此，本文提出如下几点政策建议：

第一，解决数字鸿沟问题。重视数字经济发展过程中产生的数字鸿沟问题，依据数字鸿沟的不同类别提出解决数字鸿沟问题的具体办法。弥合数字鸿沟能够改善其对城乡收入差距的不良影响，缓解城乡收入差距。与此同时，我们也要关注数字鸿沟对经济社会发展产生的影响，如“就业鸿沟”“利润分配鸿沟”，从机会和结果公平的角度去关注数字经济的发展。

第二，注重经济的协调发展。在发展数字经济的同时，也要协调好地区、城乡之间的关系，使经济发展水平不同的地区享受平等的数字红利的机会。在这个过程中，政府应该发挥主导力量，在经济比较落后的地区发展数字经济的先导性数字产业，助力数字经济的区域间协调发展。与此同时，也要注重农村地区数字经济的发展，大力推进新型城镇化建设，积极发展数字农业，实现乡村振兴。

第三，完善收入分配制度。如今我国的城乡收入差距正在不断缩小，但数字经济发展却未在城乡收入差距减小的过程中发挥积极作用，政府应该从宏观调控的角度出发，引导数字经济发展与城乡收入差距的缩小相协调，充分实现财政支出的再分配功能，制定合理的分配政策来解决数字红利在城乡居民之间分配不平衡的问题，弥合数字鸿沟。

十、模型检验

对基于灰色原理的 $GM(1,1)$ 模型的预测结果进行残差检验、关联度检验和后残差检验，结果如下。

10.1 残差检验

令残差 $\epsilon^{(0)}(k)$ 为：

$$\begin{aligned}\epsilon^{(0)}(k) &= x^{(0)}(k) - \hat{x}^{(0)}(k), k = 1, 2, 3, \dots, n \\ \epsilon^{(0)} &= (\epsilon^{(0)}(1), \epsilon^{(0)}(2), \dots, \epsilon^{(0)}(n))\end{aligned}\quad (19)$$

令相对误差 $\Delta^{(0)}(k)$ 为:

$$\begin{aligned}\Delta^{(0)}(k) &= \left| \frac{\epsilon^{(0)}(k)}{x^{(0)}(k)} \right| = \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right|, k = 1, 2, 3, \dots, n \\ \Delta^{(0)} &= (\Delta^{(0)}(1), \Delta^{(0)}(2), \dots, \Delta^{(0)}(n))\end{aligned}\quad (20)$$

对于 $k \leq n$, 定义 k 点模拟相对误差为:

$$\Delta_k = \left| \frac{\epsilon^{(0)}(k)}{x^{(0)}(k)} \right| \quad (21)$$

定义平均相对误差为:

$$\bar{\Delta} = \frac{1}{n} \sum_{k=1}^n \Delta_k \quad (22)$$

称 $1 - \bar{\Delta}$ 为平均相对精度, 称 $1 - \Delta_k$ 为 k 点模拟精度; 给定一个阈值 α , 如果 $\bar{\Delta} < \alpha$ 且 $\Delta_k < \alpha$ 成立, 称模型为残差合格模型。

10.2 关联度检验

以 x_0 为参考序列:

$$x_0 = \{x_0(k) \mid k = 1, 2, \dots, n\} = \{x_0(1), x_0(1), \dots, x_0(n)\} \quad (23)$$

其中, k 表示时刻。定义 m 个比价数列:

$$x_i = \{x_i(k) \mid k = 1, 2, \dots, n\} = \{x_i(1), x_i(1), \dots, x_i(n)\} \quad i = 1, 2, 3, \dots, m \quad (24)$$

比较数列 x_i 对参考数列 x_0 在 k 时刻的关联系数定义为:

$$\zeta_1(k) = \frac{\min_t |x_0(t) - x_1(t)| + \rho \max_t |x_0(t) - x_1(t)|}{|x_0(k) - x_1(k)| + \rho \max_t |x_0(t) - x_1(t)|} \quad (25)$$

其中, 分辨系数 $\rho \in [0, 1]$ 且合分辨率正相关。

序列 x_1 对参考序列 x_0 的关联度定义为:

$$r_1 = \frac{1}{n} \sum_{k=1}^n \zeta_1(k) \quad (26)$$

关联度为序列 x_1 对参考序列 x_0 所有 k 时刻关联系数的平均值, 因此可以分析序列 x_i 与参考序列 x_0 之间的关联程度。

10.3 后残差检验

定义 $\mathbf{X}^{(0)}$ 的均值和方差分别为：

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{k=1}^n x^{(0)}(k) \\ S_1^2 &= \frac{1}{n} \sum_{k=1}^n \left(x^{(0)}(k) - \bar{x} \right)^2\end{aligned}\tag{27}$$

定义的 $\epsilon^{(0)}$ 的均值和方差分别为：

$$\begin{aligned}\bar{\epsilon} &= \frac{1}{n} \sum_{k=1}^n \epsilon^{(0)}(k) \\ S_2^2 &= \frac{1}{n} \sum_{k=1}^n \left(\epsilon^{(0)}(k) - \bar{\epsilon} \right)^2\end{aligned}\tag{28}$$

定义均方差比值为：

$$C = \frac{S_2}{S_1}\tag{29}$$

对于给定的 $C_0 > 0$ ，当 $C < C_0$ 时，称方程为均方差比合格模型。

在此基础上，定义小概率为：

$$p = P \left\{ \left| \epsilon^{(0)}(k) - \bar{\epsilon} \right| < 0.6745 S_1 \right\}\tag{30}$$

对于给定的 $P_0 > 0$ ，当 $P > P_0$ 时，称模型为小误差概率合格模型。

10.4 结果检验

对江津区的预测结果：[31.9,39.4,44.0,49.1,54.8] 进行检验可得，残差检验的相对精度为 99%，故通过残差检验；关联度检验的关联度在 $pho = 0.5$ 时等于 $0.63 > 0.55$ ，故通过关联度检验；后残差检验中 $C = 0.0466 < 0.35$ ， $P > 0.95$ ，则 $GM(1,1)$ 预测精度等级为“好”，故通过后验差检验。在通过三个检验的基础上，结合原始序列与预测序列的对比，可认为预测结果是准确的，即 2022 年江津区互联网用户数为 68.4 万。

十一、优缺点分析

11.1 优点

(1) 使用基于灰色理论的 $GM(1,1)$ 模型，在指标变化平稳、样本量和信息较少的情况下进行缺失值的插补，具有较高的合理性与准确性。

(2) 数字经济发展评价指标体系的建立综合考虑了数字产业化、产业数字化和发展环境三大领域，移动电话普及率、互联网普及率、电信业务总量、数字普惠金融指数及 $R\&D$ 研发支出五大视角，使得数字经济发展评价模型具有较强的准确性和解释力。

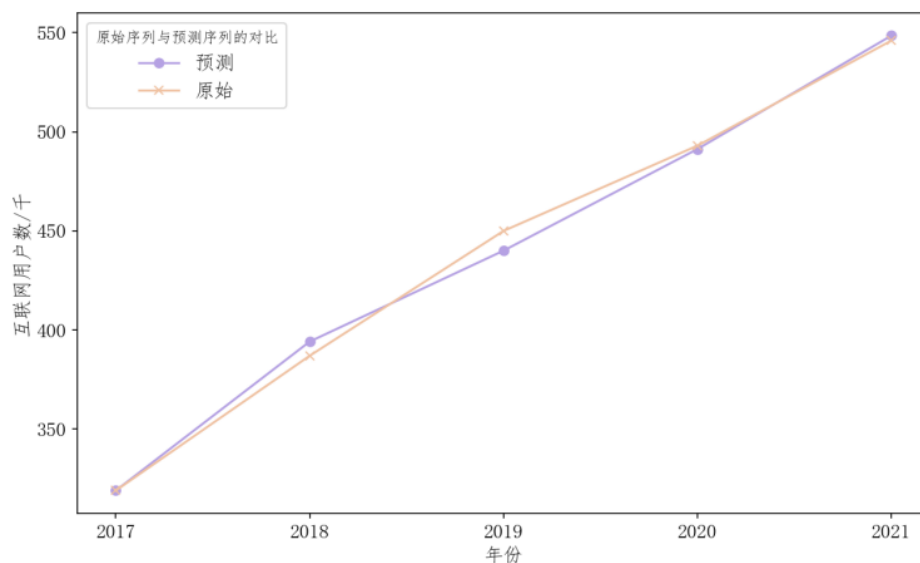


图 5 原始序列与预测序列的对比

(3) 数字经济发展和城乡差距的二次函数关系的 $R^2 = 0.996$, 说明在此模型中数字经济发展对城乡差距有极强的解释力, 拟合效果非常理想。

11.2 缺点与改进

(1) 由于重庆市区县的部分数据难以获取, 本文分析的样本量相对较少, 可能会增大结论的随机误差。因此, 增大样本量会使数据分析和结论更加准确。

(2) 本文仅考察了数字经济发展水平对城乡差距产生的影响, 但不排除其他因素以及其他因素与数字经济发展的交互项也会产生影响。适当引入控制变量会使得对数字经济发展水平与城乡差距关系的分析更准确。

参考文献

- [1] 郭峰等. 测度中国数字普惠金融发展: 指数编制与空间特征. 经济学 (季刊) ,19(4):1401-1418
- [2] 张力丹. 重庆市数字经济发展水平评价研究. 中国西部 (6):95-107
- [3] 李晓钟等. 数字经济发展对城乡收入差距的影响研究. 农业技术经济 (2):77-93
- [4] 李蕊等. 数字经济发展能缩小城乡收入差距吗?——基于中国省级面板数据的实证研究. 新金融 (11):47-55
- [5] 陈文等. 数字经济发展、数字鸿沟与城乡居民收入差距. 南方经济,50(11):1-17

[6] 王军等. 中国数字经济发展水平及演变测度. 数量经济技术经济研究,38(7):26-42

[7] 冯源. 基于随机模型预测算法的风电场智能运行控制

[8] 贾艳红等. 基于熵权法的草原生态安全评价——以甘肃牧区为例. 生态学杂志,25(8):1003-1008

附录

表 5 支撑材料文件列表

文件列表名	
问题一	重庆市各区县数字经济统计数据.csv
	重庆市各年城乡收入差距.csv
	灰色预测.py
	重庆市各区县数字经济发展情况.py
	重庆市城乡收入差距现状.py
问题二	重庆市各区县数字经济得分.csv
	重庆市数据最终版.csv
	熵权法.py
	重庆市各区县数字经济得分可视化.py
问题三	相关性分析.py
	SPSS 标准化后回归数据.sav
	数字经济与城乡差异曲线估计.spv
问题四	数字经济发展与人均可支配收入回归分析.spv

附录 A 灰色预测

```
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
```

```

mpl.rc("font",family='FangSong')

def GM11(X0, t, k):
    # 第一步，一次累加生成序列 X(1)
    print("\n为弱化原始时间序列的随机性，采用累加的方式处理数据\n")
    X1 = np.cumsum(X0)
    print("原始序列X(0)为:", X0)
    print("累加序列X(1)为:", X1, '\n')

    # 第二步，开始构造矩阵 B 和数据向量 Yn
    b = -(X1[:-1] + X1[1:]) / 2
    D = np.ones(t-1)
    B = np.column_stack((b, D))
    YN = X0[1:]

    # 第三步，开始计算待估参数向量 alpha2，利用最小二乘法求解，以便获得发展灰数 a 和内生控制灰数 u
    alpha = np.linalg.inv(B.T @ B) @ B.T @ YN
    a = alpha[0]
    u = alpha[1]
    print("GM(1,1)参数估计值：发展系数a =", a, "灰色作用量u =", u, '\n')

    # 第四步，开始计算 X(0) & X(1) 的预测序列
    Xhat1 = np.zeros(t)
    Xhat1[0] = X1[0]
    for i in range(1, t):
        Xhat1[i] = (X0[0] - u / a) * np.exp(-a * i) + u / a

    Xhat0 = np.zeros(t)
    Xhat0[0] = Xhat1[0]
    for o in range(1, t):
        Xhat0[o] = Xhat1[o] - Xhat1[o-1]

    print("X(1)的模拟值:", Xhat1)
    print("X(0)的模拟值:", Xhat0, '\n')

    # 第五步，开始进行残差检验
    print("开始进行残差检验: ", '\n')
    e = X0 - Xhat0
    r_e = np.abs(e) / X0
    e_average = np.mean(r_e)

    print("绝对残差: ", '\n', " ", np.round(np.abs(e), 6))
    print("相对残差:", '\n', " ", np.round(r_e, 8), '\n')
    print("残差平方和 =", np.sum(e**2), '\n')
    print("平均相对误差 =", e_average * 100, "%", '\n')
    print("相对精度 =", (1 - e_average) * 100, "%", '\n')

```

```

if (1 - e_average) * 100 > 90:
    print("模型精确度较高, 通过残差检验", '\n')
else:
    print("模型精确度未高于90%, 残差检验未通过", '\n')
return

# 第六步, 开始进行关联度检验
print("开始进行关联度检验: ", '\n')
pho = 0.5
eta = (np.min(np.abs(e)) + pho * np.max(np.abs(e))) / (np.abs(e) + pho * np.max(np.abs(e)))
r = np.round(np.mean(eta), 2)

print("关联度为:r =", r, '\n')
if r > 0.55:
    print("满足 pho=0.5 时的检验准则 r>0.55, 通过关联度检验", '\n')
else:
    print("未通过关联度检验", '\n')
return

# 第七步, 开始进行后验差比值检验
print("开始进行后验差检验: ", '\n')
eavge = np.mean(np.abs(e))
se = np.std(np.abs(e), ddof=1)
X0avge = np.mean(X0)
sx = np.std(X0, ddof=1)
c_value = se / sx
S0 = np.sum((np.abs(e) - eavge) < 0.6745 * sx) / len(e)

print("原始序列X0的标准差=", sx, '\n')
print("绝对误差的标准差=", se, '\n')
print("C值=", c_value, '\n')
print("小误差概率:P值=", S0, '\n')

if c_value < 0.35 and S0 > 0.95:
    print("C<0.35, P>0.95,GM(1,1)预测精度等级为: 好, 通过后验差检验", '\n')
elif c_value < 0.5 and S0 > 0.80:
    print("C值属于[0.35,0.5), P>0.80,GM(1,1)模型预测精度等级为: 合格, 通过后验差检验", '\n')
elif c_value < 0.65 and S0 > 0.70:
    print("C值属于[0.5,0.65), P>0.70,GM(1,1)模型预测精度等级为: 勉强合格, 通过后验差检验", '\n')
else:
    print("C值>=0.65, GM(1,1)模型预测精度等级为: 不合格, 未通过后验差检验", '\n')
return

# 第八步, 画出输入预测序列Xhat0与原始序列X0的比较图像
# 更改横坐标表示的数据, 为2017-2021年而不是0-4, 且显示的是整数
plt.figure(figsize=(10, 6))

```

```

x = np.arange(2017, 2022, 1)
plt.plot(x, Xhat0, color='#B5A1E3', marker='o', linestyle='-', label='预测')
plt.plot(x, X0, color='#F0C2A2', marker='x', linestyle='-', label='原始')
plt.xlabel('年份', fontsize=13)
plt.ylabel('互联网用户数/千', fontsize=13)
plt.xticks(x, fontsize=13)
plt.yticks(fontsize=13)
plt.legend(loc='best', title='原始序列与预测序列的对比', fontsize=13)
plt.show()

# 第九步, 求出第k期的预测值 Xhat0[k]
Xhat1_k = (X0[0] - u / a) * np.exp(-a * (k-1)) + u / a
Xhat1_k_minus_1 = (X0[0] - u / a) * np.exp(-a * (k-2)) + u / a
Xhat0_k = Xhat1_k - Xhat1_k_minus_1

print("第", k, "期的预测值为:", Xhat0_k)

x = np.array([319, 387, 450, 493, 546])

x = np.array([319, 387, 450, 493, 546])
GM11(x, len(x), 7)

```

附录 B 重庆市各区县经济发展情况

```

import pandas as pd

data = pd.read_csv('重庆市各区县数字经济统计数据.csv')

# 相关性分析, 计算各个指标之间可能存在的相关性
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns

matplotlib.rc("font", family='FangSong')
numeric_data = data.select_dtypes(include=['float64', 'int64'])

correlation = numeric_data.corr()

plt.figure(figsize=(14, 10))
sns.heatmap(correlation, annot=True, cmap='coolwarm', linewidths=0.5, annot_kws={"size": 15})
plt.xticks(rotation=45, ha='right', fontsize=12)
plt.yticks(rotation=0, fontsize=12)
plt.show()

# 重庆市各区县指标数据可视化

```



```

columns_of_interest = ['重庆市各区县/2023', '移动电话普及率', '电信业务总量/亿元',
                        '数字惠普金融指数/2022', '研发投入/亿元/2022', '互联网普及率']
data_filtered = data[columns_of_interest]

data_filtered.set_index('重庆市各区县/2023', inplace=True)
data_filtered.index.name = None

# data_sorted = data_filtered.sort_values(by=['研发投入/亿元/2022'], ascending=False)

# 移动电话普及率
plt.figure(figsize=(12, 6))
data_filtered['移动电话普及率'].plot(kind='bar', color='#bc5559')
plt.axhline(y=1.19, color='#FEDC5E', linestyle='--', label='全国移动电话普及率')
plt.axhline(y=1.34, color='#FF6F48', linestyle='--', label='重庆市移动电话普及率')
plt.xticks(fontsize=15, rotation=45, ha='right')
plt.yticks(fontsize=15)
plt.legend(fontsize=14)
plt.tight_layout()
plt.show()

# 电信业务总量/亿元
plt.figure(figsize=(12, 6))
data_filtered['电信业务总量/亿元'].plot(kind='bar', color='#fae385')
plt.xticks(fontsize=15, rotation=45, ha='right')
plt.yticks(fontsize=15)
plt.legend(['电信业务总量/亿元'], fontsize=14)
plt.tight_layout()
plt.show()

# 数字惠普金融指数/2022
plt.figure(figsize=(12, 6))
data_filtered['数字惠普金融指数/2022'].plot(kind='bar', color='#fcc3c5')
plt.axhline(y=116.76, color='#FEDC5E', linestyle='--', label='全国数字惠普金融指数')
plt.xticks(fontsize=15, rotation=45, ha='right')
plt.yticks(fontsize=15)
plt.legend(fontsize=14)
plt.tight_layout()
plt.show()

# 研发投入/亿元/2022
plt.figure(figsize=(12, 6))
data_filtered['研发投入/亿元/2022'].plot(kind='bar', color='#d2d4de')
plt.legend(['研发投入/亿元/2022'], fontsize=14)
plt.xticks(fontsize=15, rotation=45, ha='right')
plt.yticks(fontsize=15)

```

```
plt.tight_layout()
plt.show()

# 互联网普及率
plt.figure(figsize=(12, 6))
data_filtered['互联网普及率'].plot(kind='bar', color='#5f6fbc')
plt.axhline(y=0.775, color='#FEDC5E', linestyle='--', label='全国互联网普及率')
plt.axhline(y=1.639, color='#FF6F48', linestyle='--', label='重庆市互联网普及率')
plt.xticks(fontsize=15, rotation=45, ha='right')
plt.yticks(fontsize=15)
plt.legend(fontsize=14)
plt.tight_layout()
plt.show()
```

附录 C 熵权法

```
import pandas as pd
import numpy as np
from factor_analyzer.factor_analyzer import calculate_kmo, calculate_bartlett_sphericity

# 熵权法函数
def entropy_weight_method(matrix):
    # m, n = matrix.shape
    # print(m,n)
    # normalized_matrix = matrix / np.sum(matrix, axis=0)
    # log_matrix = -np.log(normalized_matrix + 1e-10)
    # entropy = np.sum(normalized_matrix * log_matrix, axis=1)
    # entropy_weight = (1 - entropy) / np.sum(1 - entropy)
    # return entropy_weight
import numpy as np # 导入numpy库, 并简称为np

# 定义一个自定义的对数函数mylog, 用于处理输入数组中的零元素
def mylog(p):
    n = len(p) # 获取输入向量p的长度
    lnp = np.zeros(n) # 创建一个长度为n, 元素都为0的新数组lnp
    for i in range(n): # 对向量p的每一个元素进行循环
        if p[i] == 0: # 如果当前元素的值为0
            lnp[i] = 0 # 则在lnp中对应位置也设置为0, 因为log(0)是未定义的, 这里我们规定为0
        else:
            lnp[i] = np.log(p[i]) # 如果p[i]不为0, 则计算其自然对数并赋值给lnp的对应位置
    return lnp # 返回计算后的对数数组
csv_file='重庆市数据最终版.csv'
data=pd.read_csv(csv_file)
#print(data)
indicators=data[['移动电话普及率','电信业务总量/亿元','数字普惠金融指数/2022','研发投入/亿元/2022','互联网普及率']]
```

```

# print(indicators)
# 相关性分析
X = indicators.values
result2 = np.corrcoef(X, rowvar=False)
# print(result2)
# 相关性分析分析结果较强, 可以考虑使用主成分分析

# kmo_all, kmo_model = calculate_kmo(X)
# print(kmo_all)
# print(X)
# # 计算权重
# weights = entropy_weight_method(matrix)
# print("权重:", weights)
Z = X / np.sqrt(np.sum(X*X, axis=0))
# print("标准化矩阵 Z = ")
# print(Z) # 打印标准化矩阵Z
# 计算熵权所需的变量和矩阵初始化
n, m = Z.shape # 获取标准化矩阵Z的行数和列数
D = np.zeros(m) # 初始化一个长度为m的数组D, 用于保存每个指标的信息效用值
# 计算每个指标的信息效用值
for i in range(m): # 遍历Z的每一列
    x = Z[:, i] # 获取Z的第i列, 即第i个指标的所有数据
    p = x / np.sum(x) # 对第i个指标的数据进行归一化处理, 得到概率分布p
    # 使用自定义的mylog函数计算p的对数。
    e = -np.sum(p * mylog(p)) / np.log(n) # 根据熵的定义计算第i个指标的信息熵e
    D[i] = 1 - e # 根据信息效用值的定义计算D[i]
# 根据信息效用值计算各指标的权重
W = D / np.sum(D) # 将信息效用值D归一化, 得到各指标的权重W
# print("权重 W = ")
print(W) # 打印得到的权重数组W
weighted_average = Z @ W
# len_w = len(weighted_average)
# sum_w = sum(weighted_average)
# for i in range(len_w):
#     weighted_average[i] /= sum_w
print("加权平均值为:")
print(weighted_average)

```

附录 D 重庆市各区县数字经济得分

```

import pandas as pd

data = pd.read_csv('重庆市各区县数字经济得分.csv')

import matplotlib.pyplot as plt

```

```

import matplotlib

matplotlib.rc("font",family='FangSong')

data.set_index('重庆市各区县/2023', inplace=True)
data.index.name = None

plt.figure(figsize=(12, 6))
data['得分'].sort_values(ascending=False).plot(kind='bar', color='skyblue')
plt.xticks(rotation=45, fontsize=15)
plt.yticks(fontsize=15)
plt.legend(['得分'], fontsize=15)
plt.show()

```

附录 E 重庆市各年城乡收入差距

```

import pandas as pd

urban_rural_data = pd.read_csv('重庆市各年城乡收入差距.csv')

import numpy as np

def calculate_theil(row):
    I1 = row['城镇居民可支配收入']
    I2 = row['农村居民可支配收入']
    P1 = row['城镇常住人口']
    P2 = row['农村常住人口']

    I_total = I1 * P1 + I2 * P2
    P_total = P1 + P2

    I1_ratio = (I1 * P1) / I_total
    I2_ratio = (I2 * P2) / I_total
    P1_ratio = P1 / P_total
    P2_ratio = P2 / P_total

    Theil_1 = I1_ratio * np.log(I1_ratio / P1_ratio)
    Theil_2 = I2_ratio * np.log(I2_ratio / P2_ratio)

    return Theil_1 + Theil_2

urban_rural_data['Theil'] = urban_rural_data.apply(calculate_theil, axis=1)
urban_rural_data['常住人口城镇化率'] = urban_rural_data['城镇常住人口'] /
    urban_rural_data['重庆市总人口']

```

```
print(urban_rural_data)
```

附录 F 相关性分析

```
# 相关性分析
import numpy as np
a = [0.051360, 0.056074, 0.059327, 0.062796, 0.070551]
b = [535.2861248, 483.2803043, 438.7188661, 385.4460061, 287.8690332]

correlation_coefficient = np.corrcoef(a, b)[0, 1]
print("相关系数:", correlation_coefficient)
```