

城镇化发展新视域： 数据要素的创新驱动与信息牵动*

王理¹ 廖祖君² 贾男³

摘要：本文通过估算 2006—2020 年中国 272 个地级行政单元或直辖市的新增数据要素价值，进而构建理论与实证模型，揭示了数据要素对城镇化的影响及其机制。结果显示：第一，在传统工业化对城镇化带动作用减弱的背景下，数据要素能够通过创新驱动机制和信息牵动机制推动城镇化，为城镇化发展提供新视域；第二，创新驱动机制体现在数据要素通过促进技术创新，从生产方式、生活方式、治理方式三个层面增强城镇吸纳乡村劳动力的能力；第三，信息牵动机制表现为数据要素凭借其承载的有效事实信息，一方面降低生产不确定性，助力城镇企业扩大生产规模并创造更多就业机会，另一方面提供转移决策所需信息，助力乡村劳动力向城镇转移；第四，目前中国数字基础设施的规模滞后于数据要素的产生规模；第五，数据要素不仅能提升城镇化率，还能改善城镇化区域失衡、促进城乡融合发展和以县城为载体的城镇化，是现阶段加快形成新质生产力的重要资源。

关键词：数据要素 城镇化 数据价值 数字基础设施 新质生产力

中图分类号：F291.1；F061.5 **文献标识码：**A

一、引言

党的二十大报告、2023 年中央经济工作会议和 2024 年中央“一号文件”，相继阐述了城镇化与中国式现代化、乡村振兴的关系。习近平总书记曾深刻指出：“如果城镇化目标正确、方向对头，能走出一条新路，将有利于释放内需巨大潜力，有利于提高劳动生产率，有利于破解城乡二元结构，有利于促进社会公平和共同富裕，而且世界经济和生态环境也将从中受益。”^①城镇化是中国实现乡

*本文研究得到国家社会科学经济一般项目“‘数字+文化+价值’赋能框架下农业产业化联合体协同发展研究”（编号：23BJY190）、“增强内生动力促进农村低收入群体稳定增收的理论机理与政策模拟”（编号：23BJY009）和“‘市场+政府+社群’协同治理框架下数据要素高质量供给的系统规制研究”（编号：22BJL033）的支持。感谢匿名审稿专家提出的宝贵意见，同时感谢四川大学胡珊在本文修改过程中提供的帮助，但文责自负。本文通讯作者：贾男。

^①中共中央文献研究室，2017：《习近平关于社会主义经济建设论述摘编》，北京：中央文献出版社，第 160 页。

村全面振兴、推动共同富裕的关键所在，更是迈向中国式现代化的必由之路。因此，如何持续推动城镇化发展成为一个兼具理论和实践意义的重要议题。依赖高投资的传统工业化，曾被认为是城镇化发展的根本动力（夏柱智和贺雪峰，2017）。传统工业化一方面能够提升城镇的“拉力因素”，另一方面也会加剧乡村的“推力因素”（Haris and Todaro, 1970），二者合力促进了中国乡村劳动力的快速转移^①，使中国常住人口城镇化率在短短 32 年间（1978 年末至 2011 年末）从 17.92% 跃升至 51.27%^②。“十二五”以后，随着调结构、转动能、提质量逐渐成为中国经济发展的主基调（佟家栋等，2017），继续通过传统工业化来带动城镇化的模式已不可行，一个重要原因在于中国长期发展资本密集型部门的战略阻碍了城镇工业化对劳动力的持续吸纳（陈斌开和林毅夫，2013）。2023 年，中国常住人口城镇化率虽已达到 66.16%^③，但离发达国家 80% 的平均水平仍有较大差距。鉴于此，跳出依赖高投资传统工业化带动的“旧模式”并积极探寻城镇化发展的“新视域”，便成为当务之急。

城镇化的传统内涵体现为市场作用下土地、劳动力和资本等生产要素在城乡间流动配置的结果（万广华等，2022），该内涵对应了传统工业化带动的城镇化。现阶段，数据要素已正式成为与土地、劳动力、资本和技术并列的基础性生产要素，其非竞争性、虚拟可复制性、零边际成本等新型特性，使数据要素能够突破资源稀缺性约束，并拓展传统经济理论的内涵与边界（陈晓红等，2022）。此时，城镇化的新内涵应延伸为市场作用下土地、劳动力、资本和数据等生产要素在城乡间流动配置的结果。数据要素能够与传统生产要素相融合，通过放大、叠加、倍增作用持续提升各要素的边际报酬和配置效率。那么，城镇化的上述新内涵是否预示了城镇化发展的新视域？换言之，在新的生产要素体系下，是否可以跳出对传统工业化的过度依赖，转而依靠数据要素推动城镇化持续发展？

数据要素虽是数字经济的底层逻辑和核心资源，且已有研究显示数字经济能提升城镇化水平（Wang et al., 2021），但这却不能直接断定数据要素能推动城镇化发展。原因有三：一是概念有别。数字经济是宏观的经济形态，数据则是微观生产要素，尽管数字经济是数据要素发挥作用的外在表征，但两者本质不同。二是测度不足。数字经济涵盖数据要素、数字技术和数字基础设施三方面，但在衡量数字经济时，现有研究通常构建的“数字经济指数”往往仅聚焦于数字技术和数字基础设施的指标，例如通信技术、互联网端口、移动电话基站等，鲜有涉及数据要素本身。三是认知偏误。数据要素和人工智能、物联网等属于数字经济的细分领域，与数字经济本身并不等同，如果笼统地使用数字经济指数来衡量这些细分领域，可能会导致结果偏误。例如，有学者在分别研究数字经济和人工智能对区域经济收敛的影响时，得到了完全相悖的结论。事实上，缺少官方的数据要素统计核算，也制约了学

^① 推力因素来自乡村内部，包括乡村自然资源枯竭、劳动力过剩、低收入等；拉力因素主要由城镇产生，包括更多就业机会、更高收入、优良的教育环境、完善的基础设施等。

^② 资料来源：《城镇化水平不断提升 城市发展阔步前进——新中国成立 70 周年经济社会发展成就系列报告之十七》，https://www.stats.gov.cn/sj/zxfb/202302/t20230203_1900425.html。

^③ 资料来源：《中华人民共和国 2023 年国民经济和社会发展统计公报》，https://www.stats.gov.cn/sj/zxfb/202402/t20240228_1947915.html。

界关于数据要素的研究，导致尚无经验证据能有效论证数据要素对城镇化的影响。此外，也有学者尝试构建“数据要素集聚”“数据要素配置”等指数，但其与数字经济指数并无本质区别。

综上所述，本文直接以数据要素为研究视角，深入剖析其与城镇化的内在关系，以揭示数据要素能否以及如何为城镇化发展提供新视域。相较于现有研究，本文有以下三方面的边际贡献：第一，利用“中国时间利用调查”（CTUS）和“中国家庭金融调查”（CHFS）数据，估算了2006—2020年中国272个地级行政单元或直辖市的年度新增数据要素价值，可以缓解当前学界在数据要素测度上的不足，为进一步讲好数字经济的“中国故事”提供数据支撑和方法参考。第二，将数据要素引入经典的城乡二元经济模型，从理论和实证相统筹的角度分析数据要素对城镇化的影响，论证了数据要素能够通过“创新驱动”和“信息牵动”双重机制为城镇化发展提供新视域，在新生产要素体系下拓展了城镇化理论的边界。第三，对数字基础设施规模与数据要素产生规模的匹配性、数据要素改善城镇化区域失衡与促进城乡融合发展以及以县城为载体的城镇化等重要问题进行了初步探讨，从现实层面增强了对如何有效利用数据要素推动城镇化的理解和认识。

二、理论阐释

为探究数据要素与城镇化的关系，一个自然的出发点是立足于数据要素的经济功能，由此建构其与城镇化的理论联系。根据现有文献，数据要素主要具备两类经济功能：促进技术创新和消弭不确定性。前者体现为数据要素能够促进知识和想法的产生与积累（Jones and Tonetti, 2020; 徐翔等, 2023）。以ChatGPT为例，一方面，它基于大规模通用语料库（例如新闻报道、电影对白）、知识数据库（例如维基百科）等被开发，这是知识和想法的产生；另一方面，全球用户每天产生的海量交互数据，使得ChatGPT实现了持续快速的进化，这属于知识和想法的积累。消弭不确定性的功能源自数据要素的本质——有效事实信息（杨艳等, 2023），数据要素所承载的有效事实信息能够帮助主体减少决策中的不确定性。例如，企业通过分析生产、运营和销售数据，能更精准地把握并预测消费者偏好变动，从而优化投资与生产。接下来，基于Lewis（1954）、Matsuyama（1992）和Barrios et al.（2006）的理论框架，本文将构建一个包含数据要素的城乡二元经济模型，旨在刻画数据要素在城镇化中的作用。

假设存在一个连续无限期界的封闭经济体，由乡村和城镇两部门构成。在乡村，唯一的经济活动是农业生产，农业生产函数具有如下线性形式：

$$Y_R(t) = A_R L_R(t) \quad (1)$$

（1）式中：下标 R 表示乡村部门； t 表示时间； Y_R 表示乡村产出； $A_R > 0$ ，表示不变的农业技术水平； L_R 表示乡村劳动力。城镇部门投入资本、劳动力和数据要素进行生产。参考Jones and Tonetti（2020）、杨艳等（2023）的做法，将城镇生产函数及其技术创新边界分别设定为如下形式：

$$Y_U(t) = A_U(t) D(t)^{1-\alpha-\beta} K(t)^\alpha L_U(t)^\beta \quad (2)$$

$$A_U(t) = D(t)^\mu \quad (3)$$

(2) 式中：下标 U 表示城镇部门； Y_U 表示城镇产出； A_U 代表城镇技术水平； K 、 L 、 D 分别表示资本、劳动力和数据要素； α 、 $\beta \in (0,1)$ ， α 、 β 、 $1-\alpha-\beta$ 分别为资本、劳动力和数据要素的产出份额。(3) 式中： $\mu \in (0,1)$ ，衡量技术创新过程中数据要素的利用效率。

为简化分析，本文做了如下假设：第一，忽略乡村部门的技术进步^①；第二，城镇生产仅雇用城镇内部劳动力，且城镇劳动力充分就业；第三，劳动力市场具有竞争性；第四，人口总数标准化为 1，此时 $L_R(t)+L_U(t)=1$ ， $L_U(t)$ 可以代表城镇化率；第五，城镇产品与农产品可以完全替代。

此处还需强调三点。首先，本文未区分生产和创新中数据要素的类别。例如，实时路况数据既可以用于“滴滴出行”产品研发，也能优化订单匹配并预测行程时间和费用。此外，数据要素凭借非竞争性 and 虚拟可复制性，能同时无损地被引入 (2) 式、(3) 式。其次，(2) 式中数据要素的经济功能是消弭不确定性。生产中使用数据要素旨在缓解不确定性的负面影响，产出包含了数据要素的价值。值得注意的是，数据要素在此处并不具有文献广泛提到的具有边际收益递增规律，而是递减规律，因为数据要素最多只能将不确定性减至 0 (Farboodi and Veldkamp, 2021)。最后，(3) 式中数据要素的经济功能是促进技术创新。联立 (2) 式、(3) 式会得到 $Y_U(t) = D(t)^{1+\mu-\alpha-\beta} K(t)^\alpha L_U(t)^\beta$ ，当 $0 < \mu < \alpha + \beta$ 时，数据要素的边际收益依然递减；但当 $\alpha + \beta \leq \mu < 1$ 时，数据要素具有边际收益递增规律。因此，数据要素的边际收益递增可能是受其非竞争性、虚拟可复制性以及促进技术创新三者共同作用的结果。

在竞争性的劳动力市场中，乡村和城镇的工资率分别为：

$$W_R(t) = \frac{\partial Y_R(t)}{\partial L_R(t)} = A_R \quad (4)$$

$$W_U(t) = \frac{\partial Y_U(t)}{\partial L_U(t)} = \beta A_U(t) D(t)^{1-\alpha-\beta} K(t)^\alpha L_U(t)^{\beta-1} \quad (5)$$

不失一般性，假设城镇的工资率高于乡村，即使 $t=0$ 时期所有劳动力都位于城镇时也是如此。这意味着：

$$\lim_{\substack{L_U(t) \rightarrow 1 \\ t \rightarrow 0}} \beta A_U(t) D(t)^{1-\alpha-\beta} K(t)^\alpha L_U(t)^{\beta-1} > A_R \quad (6)$$

在城镇化进程中，乡村劳动力并非都愿意向城镇转移，因为转移过程中涉及交通、生活、制度、心理、信息等多重成本，直接影响了乡村劳动力的转移决策，其中，信息对劳动力转移往往具有决定性作用 (万广华和张琰, 2021)。广泛且有效的信息不仅有助于降低转移中寻求就业、住房、子女入学等不确定性，还能缓解背井离乡、人际交往不适等心理压力 (Munshi, 2003)。数据要素的本质是有效事实信息，随着信息技术发展和数字基础设施在城乡间普及，乡村劳动力可以便捷地获取海量数据要素，从而为转移决策提供所需信息。

^①若与城镇部门一致，即乡村部门的技术创新边界设定为 $A_R(t) = D(t)^\zeta$ ，并不影响本文理论模型的最终结论。这是因为，从中国数字经济的发展现状来看，乡村对数据的利用效率普遍低于城镇，故而有 $\zeta < \mu$ 。

乡村劳动力向城镇转移的成本函数可以表示为：

$$C(t) = C_I(t) + \sum C_i(t) \quad (7)$$

(7) 式中： $C_I(t)$ 、 $\sum C_i(t)$ 分别表示信息成本和其他成本，转移成本 $C(t)$ 关于所有成本递增。进一步定义 $C_I(t)$ ：

$$C_I(t) = \begin{cases} \hat{\kappa} - \frac{1}{2}\eta(\theta D(t))^2, & \text{如果 } D(t) < \sqrt{2\hat{\kappa}/\eta\theta^2} \\ 0, & \text{如果 } D(t) \geq \sqrt{2\hat{\kappa}/\eta\theta^2} \end{cases} \quad (8)$$

(8) 式中： $\hat{\kappa} > 0$ ，表示乡村劳动力向城镇转移所必须掌握的全部信息； $\theta \in (0,1)$ 表示，数据要素承载的有效事实信息中适用于转移决策的部分； $\eta > 0$ ，衡量了 $\theta D(t)$ 相对 $\hat{\kappa}$ 的权重。当乡村劳动力在转移前获得的数据要素超过了 $\sqrt{2\hat{\kappa}/\eta\theta^2}$ ，此时已掌握转移所需的全部信息，转移过程将不再产生信息成本。综上所述，乡村劳动力的动态转移函数可以设为如下形式：

$$\dot{L}_R(t) = \begin{cases} -\lambda(t)L_R(t), & \text{如果 } W_U(t) > W_R(t) + C(t) \\ \in [-\lambda(t)L_R(t), 0], & \text{如果 } W_U(t) = W_R(t) + C(t) \\ 0, & \text{如果 } W_U(t) < W_R(t) + C(t) \end{cases} \quad (9)$$

(9) 式中： $\lambda(t) > 0$ ，反映了乡村劳动力的转移速度，转移成本越高， $\lambda(t)$ 越小。(9) 式表明，只要城镇工资率高于乡村工资率与转移成本之和，劳动力就会持续从乡村向城镇转移。因此，乡村劳动力转移的均衡条件为：

$$W_U(t) = W_R(t) + C(t) \quad (10)$$

将 (4) 式、(5) 式、(7) 式、(8) 式代入 (10) 式，可求得：

$$L_U(t) = \left[\frac{\beta A_U(t) D(t)^{1-\alpha-\beta} K(t)^\alpha}{A_R + \hat{\kappa} - \frac{\eta\theta^2}{2} D(t)^2 + \sum C_i(t)} \right]^{\frac{1}{1-\beta}} \quad (11)$$

(11) 式中，将 $L_U(t)$ 分别对 $K(t)$ 和 $D(t)$ 求二阶导，进一步得到（省略时间 t ）：

$$\frac{\partial^2 L_U}{\partial K^2} = -\frac{\alpha(1-\alpha-\beta)}{(1-\beta)^2} \left[\frac{\beta A_U(t) D^{1-\alpha-\beta} K^{\alpha+2\beta-2}}{\psi(D)} \right]^{\frac{1}{1-\beta}} \quad (12)$$

$$\begin{aligned} \frac{\partial^2 L_U}{\partial D^2} = & (\beta K^\alpha)^{\frac{1}{1-\beta}} \left\{ \frac{\beta \left[\frac{D^\varphi}{\psi(D)} \right]^{\frac{2\beta-1}{1-\beta}} \left[\frac{\eta\theta^2 D^{\varphi+1}}{\psi(D)^2} + \frac{\varphi D^{\varphi-1}}{\psi(D)} \right]^2}{(1-\beta)^2} \right. \\ & \left. + \frac{\left[\frac{D^\varphi}{\psi(D)} \right]^{\frac{\beta}{1-\beta}} \left[\frac{2\eta^2\theta^4 D^{\varphi+2}}{\psi(D)^3} + \frac{\eta\theta^2(2\varphi+1)D^\varphi}{\psi(D)^2} + \frac{\varphi(\varphi-1)D^{\varphi-2}}{\psi(D)} \right]}{1-\beta} \right\} \end{aligned} \quad (13)$$

(12) 式和 (13) 式中： $\psi(D) = A_R + \hat{\kappa} + \sum C_i(t) - \eta\theta^2 D(t)^2 / 2$ ， $\varphi = 1 + \mu - \alpha - \beta > 0$ 。

联合 (11) 式~ (13) 式, 本文可以得到四点核心推论。

推论一: $L_U(t)$ 与 $K(t)$ 正相关, 但 $\partial^2 L_U / \partial K^2 < 0$ 。经济学含义为: 在传统工业化带动的模式下, 高投资虽然能提升城镇化水平, 却无法维系城镇化的持续发展。因此, 依赖高投资的传统工业化来带动城镇化发展的“旧模式”并非长久之计。

推论二: $L_U(t)$ 与 $D(t)$ 正相关。这表明, 在新生产要素体系下, 城镇化的新内涵可以体现城镇化发展的新视域, 即除了传统工业化的带动之外, 数据要素也能推动城镇化发展。此外, 在 (13) 式中, 仅最后一项系数 $\varphi(\varphi-1)$ 的符号尚无法确定, 其余项均为正数。正如前文所述, 当数据要素具有边际收益递增规律 ($\alpha+\beta \leq \mu < 1$) 时, $\varphi-1 \geq 0$ 、 $\partial^2 L_U / \partial D^2 > 0$, 数据要素能够推动城镇化持续发展, 然而, 受限于 $L_U(t)$ 不会超过 1 且 $\mu \in (0, 1)$, 数据要素推动的城镇化增长率最终也会下降; 当数据要素具有规模报酬递减规律 ($\mu < \alpha+\beta$) 时, $\varphi-1 < 0$, 暂时无法明确 $\partial^2 L_U / \partial D^2$ 的符号。因此, 除了非竞争性、虚拟可复制性等数据要素固有特性之外, 还需要充分保障数据要素促进技术创新这一经济功能的发挥, 只有这样, 才能维系其边际收益递增, 从而不断推动城镇化发展。

推论三: $D(t)$ 能够通过促进 $A_U(t)$ 来推动城镇化发展。本文称之为数据要素对城镇化的“创新驱动”机制, 其内涵可以体现在生产方式、生活方式和治理方式三个层面。在生产方式上, 数据要素通过促进技术创新, 可以发展新质生产力, 催生新产业、新模式, 倒逼传统产业的数字化转型, 实现产业结构和就业结构的并行优化 (王理, 2023), 从而为城镇化创造更多的就业机会。在生活方式上, 数据要素通过促进技术创新, 能实现个性化、差异化、价格低廉化产品和服务的持续供给 (王海燕和郑秀梅, 2017), 助力城乡人民群众追求美好生活, 推动乡村生活方式向城镇生活方式收敛。在治理方式上, 数据要素通过促进技术创新, 能使城镇治理方式更加数智化 (蔡跃洲, 2021), 增强城市各部门合作的横向协调与纵向贯通, 提升城镇治理效率, 有效缓解“城市病”。事实上, 数据要素创新驱动的三方面都在于提升城镇的“拉力因素”, 强化城镇吸纳乡村劳动力的能力。

推论四: $D(t)$ 能够直接通过 $D(t)^{1-\alpha-\beta}$ 和 $\hat{\kappa} - \eta\theta^2 D(t)^2 / 2$ 推动城镇化发展, 它们分别从城镇部门生产函数和乡村劳动力转移成本函数进入了 (11) 式。这说明数据要素所承载的有效事实信息发挥了作用, 本文称之为数据要素对城镇化的“信息牵动”机制。该机制主要有两类实现途径: 一方面, 在城镇部门生产过程中, 当前大规模产生的数据要素所承载的海量信息, 能够有效减少企业面临的生产不确定性 (Farboodi and Veldkamp, 2021), 进而降低生产成本、提升利润, 从而促使企业扩大生产规模并创造更多就业机会。另一方面, 在乡村劳动力向城镇转移过程中, 充足的信息是乡村劳动力转移发生的关键基础, 数据要素的高流动性可以确保信息高效广泛地传输 (李海舰和赵丽, 2023), 使得乡村劳动力能够便捷、及时地获取信息, 同时在一定程度上减轻了转移过程中的心理成本, 更好地促进乡村劳动力向城镇转移。

进一步地, 本文将城镇化对传统工业化与年份虚拟变量 (2006—2020 年) 的交互项进行了回归。结果表明, 发现自 2011 年起^①, 传统工业化对城镇化的带动作用整体上呈衰减态势, 如图 1 (a) 所示。

^①2011 年以后中国工业化水平进入了工业化后期 (黄群慧, 2021)。

推论一得到初步证实。然后，在图 1（a）基础上加入数据要素，回归结果显示，在传统工业化带动作用衰减的背景下，数据要素与城镇化高度正相关，如图 1（b）所示。推论二得到初步验证。接下来，本文将对上述四点推论进行实证检验，以揭示数据要素能否以及如何为城镇化的发展提供新视域。

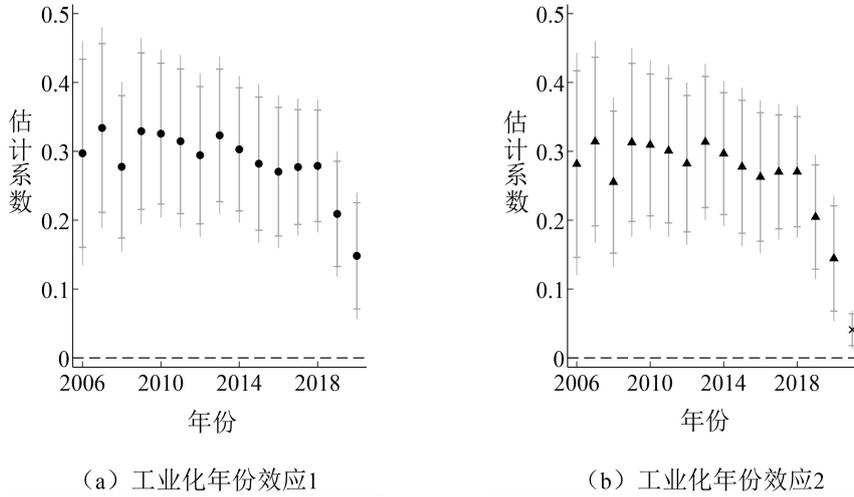


图 1 传统工业化对城镇化带动作用衰减态势

注：①“●、▲”表示工业化年份效应，“×”表示数据要素的估计系数。②控制变量与基准回归一致。③误差线代表 95%的置信区间。

三、实证策略与数据来源

（一）变量定义与数据来源

1.被解释变量：城镇化水平。本文的被解释变量为城镇化水平，城镇化反映了乡村人口向城镇地区转移流动与集聚的动态过程。借鉴大多数文献的做法，选择常住人口城镇化率作为城镇化水平的衡量指标，其计算公式为：

$$Urban_{it} = \frac{People_{it}^1}{People_{it}^2} \times 100\% \quad (14)$$

（14）式中：下标*i*表示城市^①，*t*表示年份；*Urban*表示城镇化水平；*People*¹和*People*²分别为常住城镇居民人口数量和常住人口总数量，数据来源于 2007—2021 年各城市统计年鉴。

2.核心解释变量：数据要素。本文将估算城市层面的年度新增数据要素价值，作为数据要素的衡量指标。杨艳等（2023）估算了省级层面的新增数据要素价值，其思路为：首先，将数据要素的价值形成过程拆解为数据收集、数据要素整理和数据要素使用三个环节，并将数据活动的相关职业归入其中。其次，通过调研访谈获取从事数据活动相关职业的日均有效工作时间^②，以及非直接工资和其他

^①在本文的实证研究部分，城市包括地级行政单元或直辖市。

^②有效工作时间是指能够直接创造数据要素价值的工作时间。

成本占工资的平均比重。最后，根据成本法构建年度新增数据要素价值估算公式，由此估算省级新增数据要素价值。然而，由于统计核算的缺位，上述估算方法虽逻辑清晰，但存在误差。对此有两种改进措施。一是明确劳动力市场中所有数据活动职业的分类及其从业者数量；二是扩大调研访谈的规模，使日均有效工作时间更趋近真实值。前者能较好消除数据活动职业划分不准的误差，但难度较高且涉及个人隐私；后者可借助涉及工作时间调查的微观数据库加以改进。

本文基于 2017 年 CTUS 数据库^①，对杨艳等（2023）中日均有效工作时间的精度进行改进。具体做法如下：第一，根据《2017 中国时间利用调查日志表》中的问题“您是就业人员，还是学生或上幼儿园，还是两者都不是？”，筛选就业人员并剔除超出正常工作时间的个体。第二，将上述样本与 2017 年 CHFS 数据库匹配，筛选与表 1 数据活动相关的就业人员，并分类至数据收集、数据要素整理和数据要素使用三个环节，共得到 1924 个数据活动就业人员。第三，进一步去除工作间隙、工间休息、工作等待、交通通勤等时间，得到数据活动就业人员的“专职工作时间”^②。第四，分别计算 3 个环节中就业人员专职工作时间的均值，作为创造数据要素价值的“新日均有效工作时间”。如表 1 所示，（1）列为原日均有效工作时间，（2）列~（5）列展示了数据活动就业人员专职工作时间的统计特征。

表 1 数据活动、日均有效工作时间和专职工作时间

数据活动	杨艳等（2023）	本文中的专职工作时间			
	（1） 原日均有效工作时间 （小时）	（2） 新日均有效工作时间 （小时）	（3） 最小值 （小时）	（4） 最大值 （小时）	（5） 观测值 （个）
数据收集	5.500	6.788	3	8	1521
数据要素整理	5	7.650	7.483	7.817	2
数据要素使用	7	6.592	3	8	401

注：在 2017 年 CTUS 数据中，符合本文从事数据要素整理活动的从业人员仅 2 人。

虽然 1924 个样本相比杨艳等（2023）的 20 家机构访谈样本已得到极大丰富，但不同地区的工作强度并不一致，那么，表 1 中的新日均有效工作时间可靠吗？为检验该时间均值的稳健性，本文进一步从“东部、中部、西部”“省会城市与非省会城市”“一线城市（北、上、广、深）与非一线城市”3 个层面，对 1924 个样本在区域上进行划分，以考察不同区域下的专职工作时间均值相较于表 1（2）列全样本均值的偏差。结果显示^③，不同区域划分下，参与数据收集、数据要素整理和数据要素使用的就业人员平均专职工作时间均非常接近，并且与全样本均值的偏差控制在-4.475%~2.640%。因此，本文认为表 1 中的新日均有效工作时间具有稳健性。

^①2017 年“中国时间利用调查”（CTUS）采用时间日志问卷形式，针对抽样家庭中所有 3 周岁及以上成员的全天时间配置信息进行调研，范围覆盖全国 29 个省份（除新疆、西藏和港、澳、台地区外）。

^②专职工作时间是指为获得劳动报酬、利润等收入而花费的工作时间，与直接创造数据要素价值的有效工作时间相对应。

^③因篇幅所限，此处的稳健性检验结果可在《中国农村经济》网站（zgncjj.ajcass.com）查看本文附录。

综上，本文在杨艳等（2023）的基础上，进一步利用新日均有效工作时间以及 2021 年 CHFS 数据库，估算 2006—2020 年中国 272 个地级行政单元或直辖市的新增数据要素价值^①，作为数据要素的直接衡量指标。估算结果如图 2 所示（因篇幅限制，图中仅标注 7 个城市名）。进一步地，本文分析城市新增数据要素价值的现实贴合性：一是对比新增数据要素价值的均值和中位数并分析在 2013 大数据元年前后的差异。表 2 所示结果表明，新增数据要素价值在 2013 年后出现了显著性攀升。二是考察 2015—2020 年全国数字经济规模^②和新增数据要素价值的对数变化趋势。如图 3 所示，尽管数字经济规模与新增数据要素价值同向增长，但二者趋势有明显差异。鉴于数字经济涵盖了数据要素、数字技术和数字基础设施，因此进一步引入新增数据要素与数字技术和数字基础设施之和^③交互项的对数变化趋势。可以看到，该交互项的对数变化趋势更接近数字经济规模的发展趋势，揭示了数据要素价值需通过数字技术和数字基础设施才能转化为数字经济规模，而缺少数据要素的数字基础设施无法真实反映数字经济。

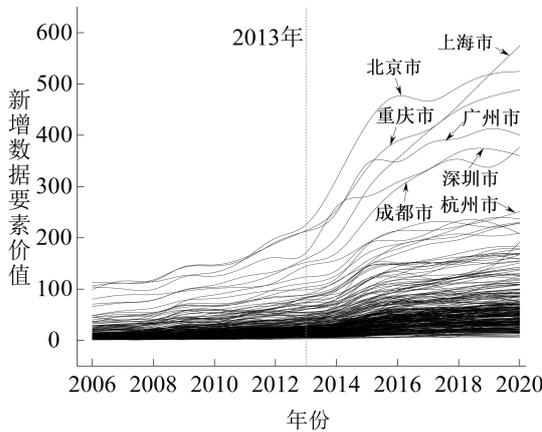


图2 2006—2020年城市新增数据要素价值（亿元）

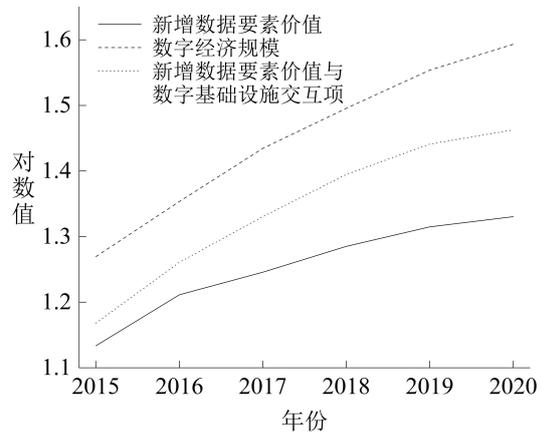


图3 对数变化趋势

表2 2013年前后新增数据要素价值的均值、中位数差异t检验

数据要素价值	2006—2012年	2013—2020年	差异显著性
均值	1.875	5.605	3.730***
中位数	1.187	3.828	2.641***

注：***表示1%的显著性水平。

3.控制变量。本文分别从城市特征和乡村劳动力两个层面进行回归控制。参考万广华和张琰（2021）的做法，选取的城市特征控制变量包括：基础教育设施，用小学数量度量；政府财政支出，用政府一般公共预算支出占地区生产总值比重度量；教育投入水平，用政府教育支出占地区生产总值比重度量；道

^①因篇幅所限，新增数据要素价值估算公式可在《中国农村经济》网站（zgncjj.ajcass.com）查看本文附录。

^②资料来源：《中国数字经济发展白皮书》，<http://www.caict.ac.cn/kxyj/qwfb/bps/202104/P020210424737615413306.pdf>。

^③数字基础设施用互联网宽带接入端口、移动电话基站和长途光缆线路长度来衡量，数字技术水平用移动电话交换机容量和局用交换机容量来衡量。数据由笔者从中华人民共和国工业和信息化部网站“工信数据”一栏手工收集并整理得到。

路密度，用城市道路面积与行政区域土地面积之比度量；人口密度，用每平方千米土地面积上的人口数量度量；人口增长，用人口自然增长率度量；人均地区生产总值，用连续一致校正的 DMSP/OLS 和 NPP/VIIRS 城市夜间平均灯光亮度衡量；传统工业化，用第二产业产值占地区生产总值比重度量。在乡村劳动力层面，参考张吉鹏等（2020）的做法，个体特征的控制变量包括：年龄，取值为 16~64 的整数；性别，男性取值为 1，女性取值为 0；健康状况，身体健康或非常健康取值为 1，否则为 0；婚姻状况，已婚取值为 1，其他情况取值为 0；受教育年限，按照没上过学、小学、初中、高中、中专、大专、本科、硕士和博士，依次取值为 0、6、9、12、13、15、16、19 和 22；家庭资产，用家庭人均资产度量；家庭收入，用家庭人均年收入度量。此外，本文还控制了城市和年份双固定效应。DMSP/OLS 和 NPP/VIIRS 数据来源于美国国家海洋和大气管理局^①，本文参考杨艳等（2021a）的做法，得到 2006—2020 年连续一致校正的夜间灯光数据，其余城市特征控制变量的数据来源于 2007—2021 年《中国城市统计年鉴》，乡村劳动力的个体特征数据来源于 2017 年、2019 年 CHFS 数据库。

（二）基准模型设定

可以严格证明在忽略高阶项后，理论部分（11）式的泰勒展开是关于 $\ln D(t)$ 的一阶多项式。因此，为检验数据要素对城镇化的影响，将（11）式转化为如下实证模型的估计：

$$Urban_{it} = \alpha_0 + \alpha_1 \ln D_{it} + \alpha_2 C_{it} + \mu_t + \lambda_i + \varepsilon_{it} \quad (15)$$

（15）式中：下标 i 表示城市， t 表示年份； $Urban$ 表示城镇化水平； D 是数据要素； C 为一列城市特征控制变量； μ 、 λ 分别是城市和年份固定效应； ε 是扰动项。

（三）内生性处理策略

本文基准回归模型可能存在内生性问题。一方面，数据要素多源自社会经济活动，城镇的社会经济活动往往比乡村更活跃，其数据要素在规模和价值上通常高于乡村，并且城镇化水平提升也会带动数据要素的规模增长，从而导致城镇化与数据要素的反向因果关系。另一方面，可能遗漏了与城镇化和数据要素同时相关的变量，例如社会网络、宗族文化、心理因素等会影响乡村劳动力转移决策和社会经济活动水平。为缓解以上潜在的内生性问题，本文采用两种应对策略：①在两阶段最小二乘法（2SLS）框架内，利用 Bartik 工具变量重新识别数据要素对城镇化的影响；②基于 Kinky 最小二乘法，直接推断数据要素对城镇化的影响。

借鉴 Bartik（2009）的思想构造工具变量。首先将城市层面的新增数据要素价值滞后一期得到 D_{it-1} ，然后对全国层面的新增数据要素价值在时间上进行一阶差分得到 $\Delta D_{t,t-1}$ ，最终得到 Bartik 工具变量^②：

$$Bartik_iv = D_{it-1} \times \Delta D_{t,t-1} \quad (16)$$

^①两类数据由美国国防气象卫星搭载的可见光成像线性扫描业务系统(DMSP/OLS)和国家极轨卫星搭载的可见光近红外成像辐射仪(NPP/VIIRS)采集得到，可通过美国国家海洋和大气管理局网站 (<https://ngdc.noaa.gov>) 下载提取。

^②本文将估算的 2006—2020 年 272 个城市新增数据要素价值加总，近似得到 2006—2020 年全国层面的新增数据要素价值，该数值不会明显地受到某个城市城镇化率的影响，即全国新增数据要素价值的变化对具体某城市而言是相对外生的。

2SLS 需要找到合适的工具变量,当恰好识别时,排他性在标准的工具变量法框架内难以得到有效检验。相比之下,Kinky 最小二乘法(Kinky Least Squares, KLS)不依赖工具变量,而是基于内生变量与误差项相关系数的区间假设进行参数估计和统计推断(Kiviet, 2020),当存在弱工具变量时,KLS 能比 2SLS 提供更为精确的推断信息。本文用 KLS 估计作为 2SLS 估计的补充。

(四) 机制分析策略

1.数据要素的“创新驱动”机制。根据前文理论部分,本文拟从产业、行业、城市三个维度,分别选择城市创新指数^①(赵涛等, 2020)、专利获取量以及智慧城市作为技术创新的代理变量,并采用如下实证策略检验创新驱动机制:①产业维度:数据要素→城市创新指数(专利获取量)↑→非农产业从业人员比重↑→城镇化率↑^②。数字经济能够显著促进非农就业(田鸽和张勋, 2022),其本质是数据要素在催生新产业、新模式中创造了丰富的就业机会,能吸纳大量乡村劳动力进入城镇就业,从而推动城镇化发展。②行业维度:数据要素→城市创新指数(专利获取量)↑→高新技术服务业从业人数↑→城镇化率↑。事实上,数据要素创造的就业机会大量集中于高新技术服务业,例如滴滴出行、美团外卖等,因此将行业维度作为产业维度的补充。③城市维度:数据要素×智慧城市→城镇化率↑。依托数据要素运行的智慧城市,是技术创新促进生产、生活以及治理方式变革的充分体现,建设智慧城市有助于引导城镇化从侧重投资的传统模式向强调创新的新型城镇化转变(辜胜阻和刘江日, 2012)。基于上述分析,本文建立如下创新驱动机制的实证检验模型:

$$\begin{cases} \ln Z_{it} = \beta_0^1 + \beta_1^1 \ln D_{it-1} + \beta_2^1 C_{it} + \beta_3^1 \ln Z_{it-1} + \mu_i + \lambda_t + \varepsilon_{it} \\ \ln Emp_{it}^1 = \beta_0^2 + \beta_1^2 \ln Z_{it} + \beta_2^2 C_{it} + \mu_i + \lambda_t + \varepsilon_{it} \\ Urban_{it} = \beta_0^3 + \beta_1^3 \ln Emp_{it}^3 + \beta_2^3 C_{it} + \mu_i + \lambda_t + \varepsilon_{it} \end{cases} \quad (17)$$

$$Urban_{it} = \xi_0^1 + \xi_1^1 \ln D_{it} \times T_{it} + \xi_2^1 \ln D_{it} + \xi_3^1 C_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (18)$$

(17)式用于验证上述实证策略①和实证策略②,(18)式用于验证实证策略③。其中:下标*i*表示城市,*t*表示年份;*Z*表示技术创新变量,包括城市创新指数与专利获取量, Z_{it-1} 用来控制知识与技术积累的影响; Emp^1 表示创新驱动的从业人员变量,包括非农产业从业人员比重和高新技术服务业从业人数,前者使用二三产业从业人员比重来度量,后者则参考宣烨等(2019)的研究,选取“金融业”“信息传输、计算机服务和软件业”“租赁和商业服务业”“科研综合技术服务业”4类行业从业总人数来表征; T_{it} 为智慧城市虚拟变量,表征*i*城市在*t*年是否进入国家智慧城市试点名单^③;其余变量的含义同基准模型一致。城市层面的专利获取量、二三产业从业人员比重以及4类高新技术服务业从业人数数据,均源自2007—2021年《中国城市统计年鉴》。

^①资料来源:《中国城市和产业创新力报告 2017》, <https://www.waitang.com/report/15662.html>。

^②为了简化表述,本文用“→”表示驱动或牵动,“↑”表示提升,“A×B→C↑”表示A通过B驱动了C提升。

^③2013—2015年,中华人民共和国住房和城乡建设部共公布了第三批国家智慧城市试点名单,本文根据第三批名单确定样本城市是否进入智慧城市试点。第三批国家智慧城市试点名单可在中华人民共和国住房和城乡建设部网站查阅。

2.数据要素的“信息牵动”机制。对应有有效事实信息牵动城镇化发展的两类理论途径，本文分别从宏观城市和微观个体双维度实证检验数据要素的信息牵动机制：①宏观城市：数据要素→企业规模↑→企业从业人数↑→城镇化率↑。以城市层面规模以上的工业企业为考察对象，检验数据要素能否通过提升企业规模来促进从业人数增加，进而助力城镇化发展。②微观个体：数据要素×数字素养→乡村劳动力转移概率↑。具备数字素养是个人获取并使用数据要素的重要前提，本文利用2017年CHFS数据库，考察数据要素能否通过数字素养这一渠道提升乡村劳动力的转移概率。

数字素养通过2017年CHFS问卷中的两个问题来衡量，一是针对问题“目前，您家拥有下列哪些类型耐用品？”，若回答手机或电脑则记为1，若同时包含手机和电脑则为2；二是针对问题“您使用过互联网吗？”，若回答“是”则记为1。本文将上述答案综合起来，作为乡村劳动力数字素养的量化指标，其取值范围为0~3。至于乡村劳动力是否发生转移，主要依据户籍城市与常住城市来识别。若乡村劳动力流入了常住城市的城镇地区，则视为发生转移；若其常住城市与户籍城市一致且现居住地为乡村，则视为未发生转移。此外，进一步通过问题“在哪一年来到家庭成员常住省市？”，确定劳动力发生转移的具体时间^①。最终，将数字素养、转移时间和转移城市进行匹配，得到关于乡村劳动力转移的2014—2017年非平衡微观面板数据，由此建立信息牵动机制的实证模型：

$$\begin{cases} \ln X_{it} = \gamma_0^1 + \gamma_1^1 \ln D_{it-1} + \gamma_2^1 C_{it} + \mu_i + \lambda_t + \varepsilon_{it} \\ \ln Emp_{it}^2 = \gamma_0^2 + \gamma_1^2 \ln X_{it} + \gamma_2^2 C_{it} + \mu_i + \lambda_t + \varepsilon_{it} \\ Urban_{it} = \gamma_0^3 + \gamma_1^3 \ln Emp_{it}^2 + \gamma_2^3 C_{it} + \mu_i + \lambda_t + \varepsilon_{it} \end{cases} \quad (19)$$

$$Prob_mig_{ijt}^1 = \xi_0^2 + \xi_1^2 \ln D_{it} \times S_{ijt} + \xi_2^2 \ln D_{it} + \xi_3^2 C_{it} + \xi_4^2 K_j + \nu_i + \lambda_t + \varepsilon_{it} \quad (20)$$

(19)式、(20)式分别用于验证宏观城市维度和微观个体维度的信息牵动机制。其中：下标*i*表示城市；*j*表示乡村劳动力；*t*表示年份；*X*表示企业规模，用规模以上工业企业的固定资产与流动资产之和作为表征（Rajan and Zingales, 1995）；*Emp*²表示信息牵动的从业人员变量，此处为企业从业人数，用第二产业采矿业与制造业的从业总人数衡量^②；*Prob_mig*¹为转移虚拟变量，表征乡村劳动力是否向城镇转移；*S*代表乡村劳动力具备的数字素养；*K*是一系列乡村劳动力个体特征的控制变量；*ν*为户籍城市固定效应；其余变量的含义同基准模型一致。规模以上工业企业的固定资产与流动资产、第二产业采矿业与制造业从业人数等数据，均源自2007—2021年《中国城市统计年鉴》；乡村劳动力转移虚拟变量和数字素养的相关数据来源于2017年、2019年CHFS数据库。

（五）样本与变量描述性统计

在基准回归、创新驱动机制、城市层面的信息牵动机制等实证分析中，样本为2006—2020年中国

^①因篇幅所限，该转移时间的具体确定方法可在《中国农村经济》网站（zgncjj.ajcass.com）查阅本文附录。

^②《中国城市统计年鉴》关于地级市层面规模以上工业企业从业人员年均数的记录截至2010年，故用城市采矿业与制造业从业人员总数近似替代。

272 个地级行政单元或直辖市的平衡面板数据；关于微观层面信息牵动机制的实证分析，样本为 2014—2019 年非平衡微观面板数据。

本文主要的变量定义及描述性统计如表 3 所示。

表 3 主要变量定义及描述性统计结果

变量类型	变量名称	变量定义或赋值	观测值	均值	标准差
被解释变量	城镇化水平	常住人口城镇化率 (%)	4080	52.32	0.16
核心解释变量	数据要素	新增数据要素估算价值 (亿元)	4080	38.69	51.08
城市特征 控制变量	基础教育设施	小学数量 (所)	4080	168.34	188.79
	政府财政支出	政府一般公共预算支出占地区生产总值比重 (%)	4080	17.81	9.06
	教育投入水平	政府教育支出占地区生产总值比重 (%)	4080	3.12	1.59
	道路密度	城市道路面积与行政区域土地面积之比 (%)	4080	1.18	1.38
	人口密度	每平方千米土地面积人口数 (人/平方千米)	4080	0.42	0.68
	人口增长	人口自然增长率 (%)	4080	5.60	5.38
	人均地区生产总值	连续一致校正的 DMSP/OLS 和 NPP/VIIRS 城市夜间平均灯光亮度 (DN 值)	4080	7.27	8.94
	传统工业化	第二产业产值占地区生产总值比重 (%)	4080	47.23	10.85
乡村劳动力 个体特征 控制变量	年龄	实际年龄 (岁), 取值范围为 16~64 的整数	16192	44.42	12.81
	性别	性别: 男性=1, 女性=0	16192	0.51	0.50
	健康	健康状况: 健康或非常健康=1, 其他=0	16192	0.50	0.50
	婚姻	婚姻状况: 已婚=1, 其他=0	16192	0.87	0.34
	受教育年限	受教育年限 (年), 范围为 0~22	16192	8.04	3.55
	家庭资产	家庭人均资产 (万元)	16192	12.32	33.63
	家庭收入	家庭人均年收入 (万元)	16192	1.82	4.60
工具变量	Bartik 工具变量	基于 (16) 式计算	4080	0.07	0.16
机制变量	城市创新指数	参考《中国城市和产业创新力报告 2017》计算	4080	17.26	86.73
	专利获取量	发明专利、实用新型专利、绿色发明专利和绿色实用新型专利 4 类专利总量 (百件)	4031	37.55	107.23
	非农产业从业人员比重	二三产业从业人员比重 (%)	4080	98.75	3.94
	高新技术服务业从业人数	金融, 信息传输、计算机服务和软件, 租赁和商业服务, 以及科研综合技术服务 4 类行业从业总人数 (万人)	3808	4.47	14.08
	智慧城市虚拟变量	是否为智慧城市试点城市: 是=1, 否=0	4080	0.19	0.39
	企业规模	规模以上工业企业的资产总计 (亿元)	4063	2300.81	3565.18
	企业从业人数	第二产业采矿业与制造业从业总人数 (万人)	4080	16.94	25.92
	转移虚拟变量	乡村劳动力是否向城镇转移: 是=1, 否=0	16192	0.05	0.23
	数字素养	手机、电脑或互联网使用情况, 取值范围为 0~3	16192	1.70	1.21

四、实证结果

（一）基准回归结果

表4报告了基准回归结果。(1)列、(2)列的混合回归结果表明,数据要素与城镇化水平在1%的统计水平上显著正相关。(3)列是加入了城市固定效应和年份固定效应的回归结果,数据要素的估计系数在5%的统计水平上显著为正。(4)列在(3)列基础上加入传统工业化变量,可以看到数据要素与传统工业化均具有正估计系数。(5)列在(4)列基础上进一步加入控制变量,数据要素的估计系数为0.0123,且在5%的统计水平上显著。基准回归结果初步表明,数据要素与城镇化之间具有正向关系,但还需通过一系列稳健性检验来确认数据要素对城镇化发展的推动作用。

表4 基准回归结果

变量	被解释变量：城镇化水平				
	(1)	(2)	(3)	(4)	(5)
数据要素	0.0724*** (0.0027)	0.0376*** (0.0025)	0.0165** (0.0051)	0.0152** (0.0051)	0.0123** (0.0051)
基础教育设施		-0.0269*** (0.0025)			0.0047* (0.0028)
政府财政支出		0.1498*** (0.0107)			0.0056 (0.0069)
教育投入水平		-0.1808*** (0.0103)			0.0054 (0.0067)
道路密度		0.0179*** (0.0025)			-0.0005 (0.0016)
人口密度		-0.0829*** (0.0065)			-0.0083** (0.0035)
人口增长		0.0009*** (0.0003)			0.0006*** (0.0002)
人均地区生产总值		0.0610*** (0.0027)			0.0237*** (0.0047)
传统工业化		-0.0170** (0.0083)		0.0588*** (0.0090)	0.0510*** (0.0095)
城市固定效应	未控制	未控制	已控制	已控制	已控制
年份固定效应	未控制	未控制	已控制	已控制	已控制
观测值	4080	4080	4080	4080	4080
R ²	0.1995	0.5562	0.9189	0.9202	0.9214
组内 R ²			0.0038	0.0199	0.0352

注：①***、**和*分别表示1%、5%和10%的显著性水平。②括号内为稳健标准误。

(二) 稳健性检验

1. 内生性处理。①基于工具变量的 2SLS 回归。关于工具变量的有效性，一方面，从 (16) 式的构造来看，它与数据要素相关，且二者间回归系数的显著性也能证明；另一方面，直接将 Bartik 工具变量对基准模型误差项回归，其估计系数高度不显著，从侧面佐证了 Bartik 工具变量的排他性。详细结果如表 5 (1) 列和表 5 (2) 列所示。

表 5 内生性处理结果

变量	(1) 误差项 IV 排他性佐证	(2) <i>Bartik_iv</i> IV 相关性检验	(3) 数据要素 2SLS 第一阶段	(4) 城镇化水平 2SLS 第二阶段	(5) 城镇化水平 KLS 回归
工具变量	0.0113 (0.0127)		0.4993*** (0.0496)		
数据要素		0.1041*** (0.0123)		0.0518** (0.0238)	0.0520*** (0.0025)
控制变量	已控制	已控制	已控制	已控制	已控制
城市固定效应	已控制	已控制	已控制	已控制	已控制
年份固定效应	已控制	已控制	已控制	已控制	已控制
第一阶段 F 值			910.6400		
内生相关系数					-0.0750
内生相关系数置信 区间					[-0.0280, 0.0189]
排他性结果					-0.0049
观测值	4080	4080	3808	3808	3808
R ²	0.0002	0.8496	0.9763	0.9272	

注：①***和**分别表示 1%和 5%的显著性水平。②圆括号内为稳健标准误。③方括号中为内生相关系数的 95%置信区间。

为进一步削弱反向因果偏误，本文将所有控制变量滞后一期，并基于 2SLS 框架重新识别数据要素对城镇化的影响，结果呈现于表 5 (3) 列和表 5 (4) 列。可以看到，数据要素对城镇化的影响在方向和显著性上均与基准回归保持一致。

②基于无工具变量的 KLS 推断。图 4 展示了 KLS 与 2SLS 的估计结果。在内生相关系数区间内，KLS 与 2SLS 均能获得一致估计。如表 5 (5) 列所示，当内生相关系数取值-0.0750 时，KLS 中数据要素的估计系数为 0.0520，与 2SLS 中的估计系数 0.0518 非常接近。此外，本文进一步使用 KLS 对工具变量的排他性进行检验。根据表 5 (5) 列的排他性结果，Bartik 工具变量满足排他性条件。

2. 其他稳健性检验。①替换核心解释变量。参考万广华和张琰 (2021)，使用互联网普及率和移动电话普及率替换数据要素。考虑到 2013 年中国正式发放 4G 牌照后，移动端数据要素才实现了快速增长，因此，移动电话普及率的样本选择在 2013 年及以后。此外，本文还利用原日均有效工作时间，

重新估算了 2006—2018 年城市新增数据要素价值，作为数据要素的衡量指标。估计结果表明^①，替换核心解释变量后，本文研究结论基本不受影响。

②收缩样本。中国当前推进的是新型城镇化，数据要素在绿色发展、社会治理、公共服务均等化、共同富裕等方面持续发挥着重要作用（沈文玮，2022）。这体现了由数据要素提供的城镇化发展新视域，与新型城镇化本质相契合。鉴于此，剔除三批国家新型城镇化综合试点地区共计 124 个城市，以减缓新型城镇化政策冲击带来的干扰。收缩样本后的估计结果显示，数据要素的回归系数始终显著为正，进一步证实了本文结论的稳健性。

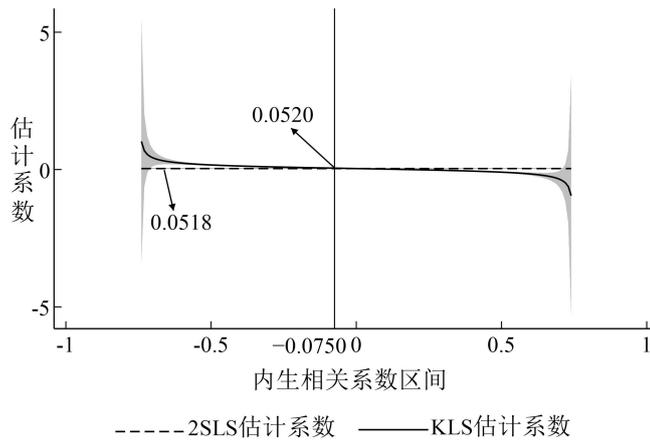


图4 2SLS 与 KLS 的估计结果

（三）异质性分析

本文从经济区域和城市规模两方面，探讨数据要素对城镇化的异质性影响。

1.经济区域异质性。将经济区域划分聚焦于长江经济带以内和以外城市，长江经济带内部的人口和生产总值超过全国的 40%，其更加活跃的社会经济活动能产生比带外地区更多的数据要素^②。此外，地方政府设立的大数据交易中心大多沿长江经济带分布，助力数据要素更好发挥对经济增长的驱动效应（杨艳等，2023）。根据图 5 所示结果，数据要素对城镇化的推动作用在长江经济带内部更为显著。

2.城市规模异质性。不同规模城市不仅对乡村劳动力吸纳与承载的能力存在区别，在数字基础设施建设程度上的落差，也会导致数据要素体量悬殊（见图 2）。因此，数据要素对城镇化的影响会因城市规模而异，图 6 报告了相关结果。可以发现，从中小城市到特大城市^③，数据要素对城镇化的推动作用逐渐增强，但在超大城市中却呈现负向影响，原因将在延伸性讨论中初步解析。

^①因篇幅所限，替换核心解释变量和收缩样本的稳健性检验结果可在《中国农村经济》网站（zgnccj.ajcass.com）查阅本文附录中的附表 2。

^②据本文估算，长江经济带内城市的新增数据要素价值比长江经济带外城市平均高出约 8.4%。

^③参见《国务院关于调整城市规模划分标准的通知》，https://www.gov.cn/zhengce/content/2014-11/20/content_9225.htm。本文进一步将小城市与中等城市合并为中小城市，得到中小城市、大城市、特大城市、超大城市四类样本。

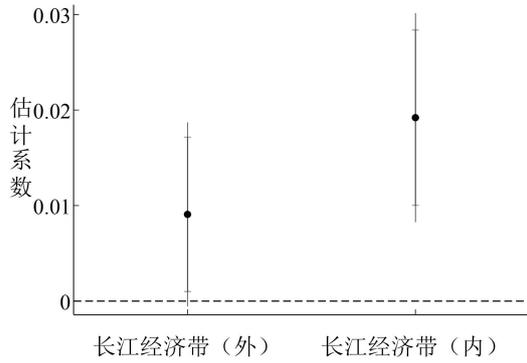


图5 经济区域带异质性估计结果

注：误差线表示 95%的置信区间。

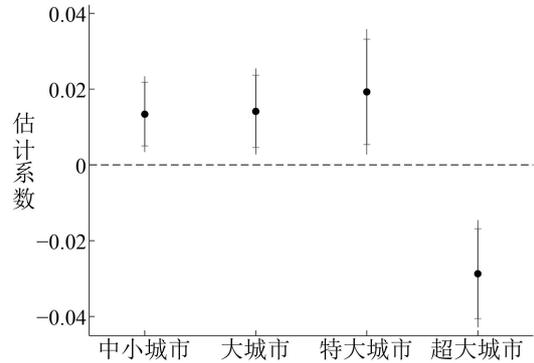


图6 城市规模异质性估计结果

注：误差线表示 95%的置信区间。

(四) 机制检验结果

1.创新驱动机制检验。一方面，基于（17）式，利用城市创新指数与专利获取量，从产业和行业双维度考察数据要素创新驱动机制的有效性，结果汇总至表6（1）列~（6）列。

表6 数据要素对城镇化的创新驱动机制

变量	(1) 城市 创新指数	(2) 专利 获取量	(3) 非农产业 从业人员比重	(4) 高新技术服务业 从业人数	(5) 城镇化 水平	(6) 城镇化 水平	(7) 城镇化 水平
数据要素	0.0370*** (0.0125)	0.0798*** (0.0266)					0.0120*** (0.0050)
城市创新指数			0.2835*** (0.0918)				
专利获取量				0.0904*** (0.0116)			
非农产业从业人员比重					0.0014** (0.0006)		
高新技术服务业从业人数						0.0063* (0.0038)	
数据要素×智慧城市							0.0033* (0.0035)
控制变量	已控制	已控制	已控制	已控制	已控制	已控制	已控制
城市固定效应	已控制	已控制	已控制	已控制	已控制	已控制	已控制
年份固定效应	已控制	已控制	已控制	已控制	已控制	已控制	已控制
观测值	3808	3734	4080	4031	4080	3808	4080
R ²	0.9967	0.9848	0.8192	0.9604	0.9215	0.9288	0.9215
组内 R ²	0.8778	0.6079	0.0263	0.0549	0.0356	0.0313	0.0360

注：①***、**和*分别表示 1%、5%和 10%的显著性水平。②括号内为稳健标准误。

其中：表6（1）列、（3）列和（5）列报告了产业维度的检验结果，表明数据要素通过促进技术创新，扩大了非农产业从业人员比重，进而推动城镇化水平提升；表6（2）列、（4）列和（6）列是关于行业维度的检验结果，数据要素通过技术创新促进了高新技术服务业的从业人数增长，从而助力城镇化发展。另一方面，基于（18）式，从智慧城市的角度进一步验证数据要素创新驱动机制在城市维度的有效性。表6（7）列的估计结果显示，数据要素通过智慧城市显著推动了城镇化发展。

2.信息牵动机理检验。表7（1）列~（3）列是基于（19）式的城市维度检验结果，可以看到，数据要素承载的有效事实信息能助力企业规模扩张，从而创造更多就业机会，促进乡村劳动力向城镇转移。在个体维度检验方面，使用Logit模型对（20）式进行估计，表7（4）列的检验结果表明，在个人数字素养的渠道下，数据要素承载的有效事实信息能够显著提升乡村劳动力向城镇转移的概率。

表7 数据要素对城镇化的信息牵动机理

变量	(1) 企业规模	(2) 企业从业人数	(3) 城镇化水平	(4) 转移虚拟变量
数据要素	0.1618*** (0.0256)			3.0656*** (0.3218)
企业规模		0.2207*** (0.0276)		
企业从业人数			0.0061** (0.0027)	
数据要素×数字素养				0.2374*** (0.0527)
控制变量	已控制	已控制	已控制	已控制
城市固定效应	已控制	已控制	已控制	未控制
户籍城市固定效应	未控制	未控制	未控制	已控制
年份固定效应	已控制	已控制	已控制	已控制
观测值	4063	4063	4080	16192
R ²	0.9638	0.9351	0.9214	0.9215
组内R ²	0.1696	0.1422	0.0345	

注：①***和**分别表示1%和5%的显著性水平。②括号内为稳健标准误。

五、延伸性讨论

前文已证实，随着传统工业化带动作用的减弱，数据要素凭借创新驱动与信息牵动双重机制，为城镇化发展提供了新视域。然而，该新视域并不仅仅止于城镇化率提升，为深理解，本文将开展三项延伸性讨论。首先，聚焦数字基础设施规模与数据要素规模的匹配性，揭示数字基础设施规模在数据要素推动城镇化发展中的重要作用；其次，通过后续两项讨论，进一步展现新视域的丰富内涵。

（一）数字基础设施规模与数据要素规模的匹配性

数字基础设施是数据要素采集、存储和流动的关键基石，那么，当前数字基础设施规模与数据要素

规模是否匹配？这一问题的答案将直接关系到数据要素对城镇化发展的推动效果。为此，本文从两个方面进行初步探讨：①逆向分析。考虑到2013年后数据要素的爆炸式增长，将2006—2012年和2013—2020年两个时间段的数字基础设施变量进行逆向互换，而数据要素变量保持不变，以观察在不同数字基础设施规模和相同数据要素规模下，数据要素对城镇化影响的差异。②门限分析。以数字基础设施为门限变量，考察数据要素对城镇化的推动作用是否会受到数字基础设施规模的影响。本文分别采用互联网宽带接入用户数的对数和2013—2020年地级市政府工作报告中数字基础设施关键词词频的对数（钞小静和刘亚颖，2023），作为数字基础设施的两个代理变量。

在逆向分析中，始终使用互联网宽带接入用户数，以确保互换过程中的变量口径一致；而在门限分析中，分别使用互联网宽带接入用户数和数字基础设施关键词词频，通过交叉验证来增强结论的可靠性。两方面分析中的城镇化水平与数据要素在时间年限上保持一致，估计结果汇总至表8。表8（4）列和（5）列的门限估计结果表明，随着数字基础设施规模的扩大，数据要素对城镇化的推动作用逐渐增强。表8（1）列~（3）列展示了逆向分析的估计结果，在2013—2020年的数字基础设施作用下，2006—2012年数据要素的估计系数为0.0025；在2006—2012年数字基础设施作用下，2013—2020年爆炸式增长的数据要素的估计系数仅为0.0016；但当数字基础设施与数据要素均处于2013—2020年水平时，数据要素的估计系数又升至0.0019。上述结论初步表明，数字基础设施的规模对于数据要素能否有效推动城镇化发展至关重要，然而，现有数字基础设施的规模尚无法匹配2013年后数据要素规模的急剧增长，阻碍了数据要素价值的充分发挥，存在“数据要素等待数字基建”的现实窘境。

表8 基础设施规模与数据要素规模的匹配性估计结果

变量	被解释变量：城镇化水平				
	逆向分析			门限分析	
	(1)	(2)	(3)	(4)	(5)
数据要素（2006—2012年）×互联网宽带接入用户数（2013—2020年）	0.0025* (0.0014)				
数据要素（2013—2020年）×互联网宽带接入用户数（2006—2012年）		0.0016** (0.0008)			
数据要素（2013—2020年）×互联网宽带接入用户数（2013—2020年）			0.0019** (0.0010)		
数据要素×互联网宽带接入用户数 (门限≤3.9616)				0.0531*** (0.0059)	
数据要素×互联网宽带接入用户数 (门限>3.9616)				0.0694*** (0.0049)	
数据要素×数字基础设施关键词词频 (门限≤3.1355)					0.0422*** (0.0072)
数据要素×数字基础设施关键词词频 (门限>3.1355)					0.0495*** (0.0059)

表 8 (续)

控制变量	已控制	已控制	已控制	已控制	已控制
城市固定效应	已控制	已控制	已控制	已控制	已控制
年份固定效应	已控制	已控制	已控制	已控制	已控制
观测值	1904	2176	2176	4080	2128
R ²	0.9329	0.9649	0.9649	0.3238	0.2229
组内 R ²	0.0245	0.0253	0.0253	0.6205	0.5158

注：①***、**和*分别表示 1%、5%和 10%的显著性水平。②括号内为稳健标准误。③门限分析中仅单门限显著。

(二) 数据要素跨越空间限制对城镇化的两类影响

现有文献常提及“数据要素能够跨越空间限制”（于施洋等，2020；杨艳等，2021b），但鲜少探讨其如何影响城镇化，本文将尝试对此进行初步阐释。数据要素的非竞争性和虚拟可复制性，使其可以在不同空间被同时使用，这是数据要素能够跨越空间限制的本质与动力。

数据要素跨越空间限制，会对城镇化产生两类积极影响：一是改善城镇化区域失衡，相同的数据要素能同时被发达和欠发达地区投入城镇化发展，并对后者产生更大边际贡献，由此改善传统工业化带动下的城镇化区域失衡；二是促进城乡融合发展，数据要素在城乡治理、公共服务均等化、共同富裕等方面持续发挥着重要作用，有助于城乡公平分享经济发展红利，进而弱化城乡二元边界，促进城乡融合发展。

接下来，本文将为上述两类积极影响寻求经验证据。针对第一类积极影响，构建如下实证模型：

$$G_Urban_{it} = \varphi_0 + \varphi_1 \ln D_{it-1} + \varphi_2 Urban_{it-1} + \varphi_3 \ln D_{it-1} \times Urban_{it-1} + \varphi_4 C_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (21)$$

(21) 式中： G_Urban_{it} 表示城镇化水平的增长率；其余变量的含义与基准模型一致，估计结果汇报至表 9 (1) 列。可以看到，估计系数 φ_1 显著为正且 φ_2 、 φ_3 显著为负，表明数据要素在推动城镇化发展过程中，能够促使城镇化水平收敛，逐步改善城镇化区域失衡。

针对第二类积极影响，构建如下实证模型：

$$Border_{it}^a = \eta_0 + \eta_1 \ln D_{it} + \eta_2 C_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (22)$$

(22) 式中： $Border^a$ 是反映城乡融合发展水平的变量，本文用城乡接合区内部夜间平均灯光亮度值的对数进行衡量；其余变量的含义同基准模型一致。一方面，城乡接合区作为连接高人口密度区与农业生产活动区之间的过渡地带，促进了城镇与乡村的相互融合；另一方面， $Border^a$ 的大小能反映城乡接合区的经济活跃程度，其值越大，意味着城乡间要素流动越畅通。参考张志刚等（2016）的做法，通过灯光亮度、灯光亮度起伏度以及特征值组合值的三维数据来识别城乡接合区的空间范围，其中具有“中灯光亮度、高灯光亮度起伏度、中特征值组合值”特征的空间点属于城乡接合区。根据表 9 (2) 列的初步结果，数据要素能够促进城乡融合发展。

表 9 数据要素跨越空间限制对城镇化影响的估计结果

变量	(1) 城镇化水平		(2) 城乡融合发展水平	
	系数	标准误	系数	标准误
数据要素滞后一期	0.0299***	0.0104		
城镇化水平滞后一期	-0.9451***	0.0746		
数据要素滞后一期×城镇化水平 滞后一期	-0.0631***	0.0174		
数据要素			0.0939***	0.0142
控制变量	已控制		已控制	
城市固定效应	已控制		已控制	
年份固定效应	已控制		已控制	
观测值	3808		4065	
R ²	0.3210		0.9765	
组内 R ²	0.2575		0.5958	

注：***表示 1%的显著性水平。

(三) 数据要素促进以县城为载体的城镇化

党的二十大报告明确指出：“推进以县城为重要载体的城镇化建设。”^①县城是连接城乡的重要纽带，将县城作为中国城镇化的主要空间载体，是缓解“大城市病”、促进区域协调发展、推动乡村振兴等的关键着力点。推进以县城为重要载体的城镇化建设，关键在于畅通城乡要素流动，而数据要素跨越空间限制的特性，可为此提供助力。例如，电子商务能借助数据要素向更高效、更智能、更简单的方向演进，“电商下乡”为县城带来了无技术障碍的新发展模式，助力广大县城能依托自身资源发展特色产业，为县城承载更多的乡村劳动力就业提供支撑。此外，县城实施数字化升级改造也能提振县城的经济发展能力（曾亿武等，2021），而数据要素是数字化升级改造所需的核心资源。

为检验数据要素能否促进以县城为载体的城镇化，本文建立如下实证模型：

$$Prob_mig_{ijt}^2 = \theta_0 + \theta_1 \ln D_{it} + \theta_2 C_{it} + \theta_3 K_j + v_i + \lambda_t + \varepsilon_{it} \quad (23)$$

(23) 式中： j 表示乡村劳动力； $Prob_mig^2$ 代表乡村劳动力向县城转移的虚拟变量；其余变量含义同(20)式一致。借助 CHFS 数据库中户籍地与常住地信息，识别了乡村劳动力向县城的转移情况。若乡村劳动力流入常住县城，则表示发生转移， $Prob_mig^2=1$ ；若常住地与户籍地相同且现居地为乡村，则表示未出现转移， $Prob_mig^2=0$ ；此外，进一步通过问题“在哪一年来到家庭成员常住省市？”明确了转移时间^②。最终，将转移时间与转移县城匹配，构建 2014—2019 年乡村劳动力

^①习近平，2022：《高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告》，北京：人民出版社，第 32 页。

^②因篇幅所限，该转移时间的具体确定方法可在《中国农村经济》网站（zgncjj.ajcass.com）查阅本文附录。

向县城转移的非平衡微观面板数据。表 10（1）列的 Logit 估计结果显示，数据要素显著提高了乡村劳动力向县城转移的概率，能够促进以县城为载体的城镇化。

进一步，推进以县城为重要载体的城镇化，关键在于引导超大城市的过度集聚人口有序迁移至其他城市，而数据要素可以为此做出贡献。一方面，数据要素跨越空间限制的特性能为所有城市提供经济增长的动力（杨艳等，2023）；另一方面，数据要素还能提供有效事实信息，降低人口向外迁移过程中的不确定性和心理成本。鉴于此，将前文 *Border^a* 重新定义为除超大城市外的各城市平均夜间灯光亮度与所有超大城市平均夜间灯光亮度之比，由于夜间灯光亮度变化可以较好地反映人口在空间上的集聚和流动（王贤彬和黄亮雄，2018），因而该比值能在一定程度上反映人口从超大城市向其他城市的流动情况。基于（22）式，从中小城市、大城市和特大城市三个维度进行回归。表 10（2）列的估计结果显示，数据要素对人口从超大城市向其他城市迁移具有积极的影响，这也解释了图 6 中数据要素与超大城市城镇化之间的负向关系。

表 10 数据要素促进以县城为载体的城镇化估计结果

变量	(1)		(2)	
	乡村劳动力向县城转移		超大城市人口向其他城市迁移	
	系数	标准误	系数	标准误
数据要素	3.8825***	0.4293		
中小城市数据要素			1.0821***	0.0598
大城市数据要素			1.2647***	0.0675
特大城市数据要素			1.1953***	0.1009
控制变量	已控制		已控制	
城市固定效应	未控制		已控制	
户籍地固定效应	已控制		未控制	
年份固定效应	已控制		已控制	
观测值	9928		3930	
R ²	0.7016		0.7549	
组内 R ²			0.1228	

注：***表示 1%的显著性水平。

六、结论与政策启示

本文深入探讨了数据要素与城镇化的关系，通过理论与实证分析，揭示了数据要素能否以及如何为城镇化发展提供新视域。研究发现：第一，在传统工业化对城镇化带动作用减弱的背景下，数据要素能通过创新驱动机制和信息牵动机制推动城镇化发展。第二，创新驱动机制体现在数据要素通过促进技术创新，从生产方式、生活方式、治理方式三个层面提升城镇的“拉力因素”，增强城镇吸纳乡村劳动力的能力。第三，信息牵动机制表现为数据要素凭借承载的有效事实信息，一方面降低生产不确定性，助力城镇企业扩大生产规模，为乡村劳动力创造更多就业机会；另一方面通过提供转移决策

所需信息，助力乡村劳动力向城镇转移，但需要乡村劳动力具备一定的数字素养。第四，完善的数字基础设施是发挥数据要素对城镇化推动作用的前提保障，但当前中国数字基础设施规模却滞后于数据要素的产生规模。第五，数据要素不仅能提升城镇化率，还能改善城镇化区域失衡、促进城乡融合发展和以县城为载体的城镇化，是现阶段加快形成新质生产力的重要资源。

上述结论在数据要素推动城镇化发展方面具有现实且具体的政策启示：

首先，加快推进数据要素市场体系建设。数据要素市场体系涉及三类主体：一是供给主体，主要为数据要素型企业；二是需求主体，包括各级数据交易中心、数据经纪人、场外交易等；三是服务性数据机构，例如数据中介机构。推进数据要素市场体系建设，既能持续壮大三类市场主体，提升全社会数据要素的规模与质量，也能保障数据要素在城市间、城乡间自由有序流动，助力城镇化发展长期向好。

其次，加强数字基础设施的研发和建设力度。当前，全球数据正以“井喷式”的速度生产，数字基础设施不仅是数据能够被大规模生成与采集的基石，更是将数据转化为数据要素并存储、管理和传输的关键。当前，中国正处于一个“数据要素等待数字基建”的阶段，由于数字基础设施规模不足、建设滞后，许多有价值的数字要素难以被有效采集和使用，制约了数据要素对城镇化发展的推动作用。因此，应积极研发先进的数字基础设施，并不断加强城市间、城乡间的数字基建力度，只有这样，才能提升数据要素的产生、采集和使用效率，也才能确保数据要素更好地为城镇化发展提供新视域。

最后，着力提升全体社会公民的数字素养。当今时代是依靠数据驱动的数字经济时代，数据要素已经渗透到生产生活中的方方面面，每个人都面临着掌握并用好数据要素的时代挑战。具备数字素养，不仅关系到乡村劳动力的转移决策，也是超大城市人口获取迁移信息的关键。因此，着力提升全体社会公民的数字素养，将其从最初的科学素养转变为全民基本素养，不仅能确保全体公民共享数据红利，还能保证数据要素在推动城镇化水平提升和以县城为载体的城镇化建设中发挥更大价值。

参考文献

1. 蔡跃洲，2021：《中国共产党领导的科技创新治理及其数字化转型——数据驱动的新型举国体制构建完善视角》，《管理世界》第8期，第30-46页。
2. 陈斌开、林毅夫，2013：《发展战略、城市化与中国城乡收入差距》，《中国社会科学》第4期，第81-102页。
3. 陈晓红、李杨扬、宋丽洁、汪阳洁，2022：《数字经济理论体系与研究展望》，《管理世界》第2期，第208-224页。
4. 钞小静、刘亚颖，2023：《新型数字基础设施建设与中国经济高质量发展——基于“条件—过程—结果”协同联动的视角》，《贵州财经大学学报》第4期，第1-11页。
5. 辜胜阻、刘江日，2012：《城镇化要从“要素驱动”走向“创新驱动”》，《人口研究》第6期，第3-12页。
6. 黄群慧，2021：《中国共产党领导社会主义工业化建设及其历史经验》，《中国社会科学》第7期，第4-20页。
7. 李海舰、赵丽，2023：《数据价值理论研究》，《财贸经济》第6期，第5-20页。
8. 沈文玮，2022：《数字技术促进共同富裕的内在机理与实践路径研究》，《政治经济学评论》第6期，第175-191页。
9. 田鸽、张勋，2022：《数字经济、非农就业与社会分工》，《管理世界》第5期，第72-84页。

10. 佟家栋、谢丹阳、包群、黄群慧、李向阳、刘志彪、金碚、余淼杰、王孝松，2017：《“逆全球化”与实体经济转型升级笔谈》，《中国工业经济》第6期，第5-59页。
11. 万广华、江葳蕤、赵梦雪，2022：《城镇化的共同富裕效应》，《中国农村经济》第4期，第2-22页。
12. 万广华、张琰，2021：《信息与城市化——基于跨国数据的实证研究》，《经济学（季刊）》第2期，第465-492页。
13. 王海燕、郑秀梅，2017：《创新驱动发展的理论基础、内涵与评价》，《中国软科学》第1期，第41-49页。
14. 王理，2023：《数据要素驱动经济发展研究》，四川大学博士学位论文。
15. 王贤彬、黄亮雄，2018：《夜间灯光数据及其在经济学研究中的应用》，《经济学动态》第10期，第75-87页。
16. 夏柱智、贺雪峰，2017：《半工半耕与中国渐进城镇化模式》，《中国社会科学》第12期，第117-137页。
17. 徐翔、赵墨非、李涛、李帅臻，2023：《数据要素与企业创新：基于研发竞争的视角》，《经济研究》第2期，第39-56页。
18. 宣烨、陆静、余泳泽，2019：《高铁开通对高端服务业空间集聚的影响》，《财贸经济》第9期，第117-131页。
19. 杨艳、王理、李雨佳、廖祖君，2023：《中国经济增长：数据要素的“双维驱动”》，《统计研究》第4期，第3-18页。
20. 杨艳、王理、廖祖君，2021a：《数据要素市场化配置与区域经济发展——基于数据交易平台的视角》，《社会科学研究》第6期，第38-52页。
21. 杨艳、王理、廖祖君，2021b：《数据要素：倍增效应与人均产出影响——基于数据要素流动环境的视角》，《经济问题探索》第12期，第118-135页。
22. 于施洋、王建冬、郭巧敏，2020：《我国构建数据新型要素市场体系面临的挑战与对策》，《电子政务》第3期，第2-12页。
23. 张吉鹏、黄金、王军辉、黄勳，2020：《城市落户门槛与劳动力回流》，《经济研究》第7期，第175-190页。
24. 张志刚、张安明、郭欢欢，2016：《基于DMSP/OLS夜间灯光数据的城乡结合部空间识别研究——以重庆市主城区为例》，《地理与地理信息科学》第6期，第37-42页。
25. 曾亿武、宋逸香、林夏珍、傅昌鑫，2021：《中国数字乡村建设若干问题刍议》，《中国农村经济》第4期，第21-35页。
26. 赵涛、张智、梁上坤，2020：《数字经济、创业活跃度与高质量发展——来自中国城市的经验证据》，《管理世界》第10期，第65-76页。
27. Barrios, S., L. Bertinelli, and E. Strobl, 2006, “Climatic Change and Rural–Urban Migration: The Case of Sub-Saharan Africa”, *Journal of Urban Economics*, 60(3): 357-371.
28. Bartik, T., 2009, “How Do the Effects of Local Growth on Employment Rates Vary with Initial Labor Market Conditions?”, *Upjohn Working Papers & Journal Articles*, 64(18): 9-148.
29. Farboodi, M., and L. Veldkamp, 2021, “A Growth Model of the Data Economy”, NBER Working Paper 28427, https://papers.ssm.com/sol3/papers.cfm?abstract_id=3781314.
30. Haris, J., and M. Todaro, 1970, “Migration, Unemployment and Development: A Two-Sector Analysis”, *American Economic Review*, 60(1): 126-142.

31. Jones, C., and C. Tonetti, 2020, "Nonrivalry and the Economics of Data", *American Economic Review*, 110(9): 2819-2858.
32. Kiviet, J., 2020, "Testing the Impossible: Identifying Exclusion Restrictions", *Journal of Econometrics*, 218(2): 294-316.
33. Lewis, W., 1954, "Economic Development with Unlimited Supplies of Labour", *The Manchester School*, 22(2): 139-191.
34. Matsuyama K., 1992, "Agricultural Productivity, Comparative Advantage and Economic Growth", *Journal of Economic Theory*, 58(2): 317-334.
35. Munshi K., 2003, "Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market", *Quarterly Journal of Economics*, 118(2): 549-599.
36. Rajan R., and L. Zingales 1995, "What Do We Know about Capital Structure? Some Evidence from International Data", *Journal of Finance*, 50(5): 1421-1460.
37. Wang, D., T. Zhou, and M. Wang, 2021, "Information and Communication Technology (ICT), Digital Divide and Urbanization: Evidence from Chinese Cities", *Technology in Society*, Vol.64, 101516.

(作者单位：¹成都师范学院经济与管理学院；

²四川省社会科学院；

³四川大学经济学院)

(责任编辑：小林)

New Horizons for Urbanization: Innovation-Driven and Information-Pulled Mechanisms of Data Factors

WANG Li LIAO Zujun JIA Nan

Abstract: This paper estimates the added value of data factors in 272 prefecture-level administrative units or municipalities of China between 2006-2020, and constructs theoretical and empirical models to reveal the impacts and mechanisms of data factors on urbanization. The results are as follows: First, under the background of the weakening role of traditional industrialization in driving urbanization, data factors can promote urbanization through innovation-driven mechanism and information-pulled mechanism, offering new horizons for urbanization. Second, the innovation-driven mechanism is reflected in data factors boosting urban areas' ability to absorb rural labor, through technological advancements in production, lifestyle, and governance. Third, the information-pulled mechanism leverages data factors' inherent information to reduce production uncertainties on the one hand, enabling urban enterprises to expand and create jobs, and to guide migration decisions on the other hand, promoting rural-to-urban labor shifts. Fourth, the current scale of China's digital infrastructure lags behind the scale of data factors generation. Fifth, data factors not only increase the urbanization rate, but also play a pivotal role in rectifying regional imbalances in urbanization, promoting integrated urban-rural development and facilitating urbanization with county towns as the carrier. This demonstrates that data factors are important resources for accelerating the formation of new quality productive forces at the current stage.

Keywords: Data Factors; Urbanization; Data Value; Digital Infrastructure; New Quality Productive Forces