# 数据产权初始分配、私有数据共享与社会福利

## ——基于“谁投入、谁贡献、谁受益”的产权分配原则

朱丹丹　郑智杰　潘士远　何怡瑶*

**摘　要**：本文构建一个动态一般均衡模型，探究基于数据来源结构进行产权分配的数据制度——消费者和企业的初始产权比例等于各自贡献的数据来源比例——如何影响社会福利水平。研究发现，不同的数据来源结构将影响两类私有数据的共享程度，并在经济产出效益和隐私成本的权衡之中影响社会福利。若以社会计划者均衡为基准，则分散均衡的社会福利效率与消费者数据占比呈非线性关系。数据脱敏保护政策和企业数据交易补贴政策将提升该产权制度的实施效果。

**关键词**：数据产权；数据共享；社会福利
**DOI**：10.13821/j.cnki.ceq.2025.05.02

## 一、引　　言

数据要素是国家经济增长强劲持续的关键生产要素和战略性资源，完善数据市场制度是充分释放数据价值的前提基础。2023 年 12 月，国家数据局等 17 部门联合印发《“数据要素×”三年行动计划（2024—2026 年）》，强调充分发挥数据要素的放大、叠加、倍增作用，发挥数据要素报酬递增、低成本复用等特点。我国数据市场潜藏着庞大的“共同使用、共享收益”的社会福利增益，2018—2024 年数据产量保持每年约 20%至 30%的增速，全球占比约 10%且基本长期位居世界第二。① 同时，我国数据市场的制度演进已从夯实数据信息安全保障基石②，深化推进到构建发挥市场配置作用的数据产权、流通交易、收益分配、安全治理等全链条制度体系。2022 年 12 月，中共中央、国务院发布《关于构建数据基础制度更好发挥数据要素作用的意见》（以下简称“数据二十条”），提出探索建立数据资源持有权、数据加工使用权、数据产品经营权的结构性分置制度。2023 年 12 月财政部《关于加强数据资产管理的指导意见》和 2024 年 7 月党的二十届三中全会进一步确立

* 朱丹丹，中山大学岭南学院、中山大学管理学院（创业学院）；郑智杰，北京师范大学湾区国际商学院；潘士远，浙江大学民营经济研究中心、浙大城市学院商学院；何怡瑶，杭州师范大学经济学院。通信作者及地址：潘士远，浙江省杭州市西湖区余杭塘路 866 号浙江大学民营经济研究中心，310058；电话：0571-88981985；E-mail：shiyuanpan@zju.edu.cn。本文受到国家自然科学基金专项项目（72141305）、国家自然科学基金青年项目（72203243）、广东省基础与应用基础研究基金自然科学基金面上项目（2024A1515012622）、浙江省自然科学基金资助项目（LMS25G030004）的资助。感谢编辑和审稿专家提出的宝贵建议，文责自负。

① 摘自国家数据局和国家互联网信息办公室发布的《数字中国发展报告》（2021—2024）。

② 例如《网络安全法》《数据安全法》《个人信息保护法》《促进和规范数据跨境流动规定》等。

中国知网　https://www.cnki.net

起基于市场化的“谁投入、谁贡献、谁受益”的原则，健全由市场评价贡献、按贡献决定报酬的机制，鼓励数据来源者依法依规分享数据并获得相应收益。

从数据来源主体的角度，私有数据是消费和生产的副产品（Varian，2018；Farboodi and Veldkamp，2022；Veldkamp and Chung，2024），故其初始产权应当归属于参与数据生成过程的消费者和企业。调查报告显示[①]，数据资源主要由个人及行业机构持有的各类设备产生，这意味着通过识别和测算不同数据产出渠道及对应设备的数据产量，可以有效界定各经济主体的数据来源构成。具体而言，个人数据的产生渠道包括生活消费（手机、私人电脑、智能家居、可穿戴设备）和交通运输（火车、飞机、船、智能汽车）等，企业数据的产生渠道包括物联感知（监控设备、仪器仪表、传感器）和生产制造（边缘智能设备、工业机器人、数控机床）等。此时，在每个行业“一刀切式”地讨论数据产权单独属于企业还是消费者（Jones and Tonetti，2020）将不符合“谁投入、谁贡献、谁受益”的产权分配原则。数据产权制度应充分依据不同行业的经济活动特征，重点考虑数据来源结构及其对经济生产的相对贡献。故本文的研究问题是，基于消费者和企业的数据来源结构进行初始产权分配的数据制度如何影响社会福利水平及其理论机制和政策启示。

因此，本文构建了一个包含数据驱动生产的动态一般均衡模型。在该理论模型中，消费者和企业的数据产权初始分配比例取决于数据来源结构，也即等于两类主体各自贡献的数据来源比例。私有数据经由数据市场交易实现共享，数据共享使其能够非竞争性地促进所有企业生产。在分散经济的微观决策层面，家庭将权衡出售消费者数据的收益与隐私泄露的成本，企业将权衡出售企业数据的收益与因创造性破坏而被潜在新企业取代的风险，两者决策共同决定数据共享的均衡水平。研究结果表明，不同数据来源结构将影响实施该数据产权制度时两类私有数据的共享程度，进而从宏观层面上在经济产出效益和消费者隐私成本的权衡之中影响社会福利水平。具体而言，当消费者数据相对于企业数据的占比越大时，消费者数据共享量越多，这将既通过要素深化效应而促进经济总产出，从而提升社会福利，又通过恶化消费者隐私泄露程度而降低社会福利。与此同时，企业数据共享量越少，这将通过要素深化效应弱化、熟练劳动力再配置效应抑制新产品创新而降低经济总产出，进而降低社会福利。为判断该数据产权制度是否能够有效促进社会福利，本文进一步比较分散经济均衡和社会计划者均衡结果。当以社会计划者均衡为基准时，分散经济均衡的社会福利效率与消费者数据占比呈现非线性关系。此外，数据脱敏保护政策和企业数据交易补贴政策将能显著提升该数据产权制度的实施效果，且在消费者和企业的数据来源比例相近时叠加配套政策的提升效果更显著。

本文与论证数据要素促进经济产出的文献密切相关。已有研究将数据视为企业生产经营和创新研发过程的新型投入要素（Jones and Tonetti，2020；Cong et al.，2021；蔡继明等，2022；杨俊等，2022；徐翔等，2023；刘涛雄等，2024；Veldkamp and Chung，

① 摘自中国信息通信研究院和中国网络空间研究院联合发布的《国家数据资源调查报告（2020）》、全国数据资源调查工作组发布的《全国数据资源调查报告》（2023—2024）。其中，《国家数据资源调查报告（2020）》显示 2019 年个人数据与行业机构（包括政府、企业、事业单位、社会团体等）的数据产量的相对比例是 0.231∶0.769。

中国知网　https://www.cnki.net

2024），或者认为数据在信息摩擦环境中有助于降低预测不确定性、缓解双边信息不对称问题（Farboodi and Veldkamp，2020，2022；Acemoglu et al.，2022；Bergemann et al.，2022；Parlour et al.，2022；He et al.，2023）。Jones and Tonetti（2020）通过构建数据驱动生产的增长理论模型，论证数据非竞争性的规模报酬递增效应。Cong et al.（2021）通过构建数据驱动研发的增长理论模型，考察数据要素对技术创新和经济增长的影响。Farboodi and Veldkamp（2022）认为由于预测误差最多降至零，因此数据积累降低预测误差的边际收益递减且其增长效应存在上限。

本文还与讨论数据产权归属的文献密切相关。已有文献认为将消费者数据的产权配置给消费者优于配置给企业。一方面，消费者在披露数据时会直接考虑数据的隐私成本和社会福利影响（Acquisti et al.，2016；Miller and Tucker，2017；Abowd and Schmutte，2019；Jones and Tonetti，2020；程华等，2023）。Jones and Tonetti（2020）认为消费者拥有数据产权比企业拥有数据产权的数据共享程度更高，更加接近社会计划者最优水平。李三希等（2023）则关注了数据产权分配在促进用户共享与激励收集者投资之间的权衡。另一方面，在企业拥有消费者数据的普遍现状下，人工智能技术推动了行业集中度的提高和“超级明星”企业的崛起（Farboodi et al.，2019；Babina et al.，2024a），此时将数据所有权从企业转移给消费者有助于打破大公司的数据优势并激励新企业进入市场（Babina et al.，2024b）。国内经济学界还积极从理论层面探索建立“谁投入、谁贡献、谁受益”的数据产权制度体系，强调其特色创新在于淡化数据所有权、强调数据使用权（龚强等，2022；刘涛雄等，2023；张新宝，2023）。

本文的创新点主要包括三方面：第一，从研究主题来看，已有文献缺乏对以“数据二十条”、《关于加强数据资产管理的指导意见》为代表的中国特色数据产权制度的经济学理论研究。本文将结合对“谁投入、谁贡献、谁受益”产权分配原则的解读与提炼，为衡量该产权制度的有效性提供经济学理论框架。第二，在理论创新方面，已有文献对数据要素作用于宏观经济效益的微观理论机制仍处于初步探索阶段，且尚未从数据来源结构的角度关注数据异质性。本文创新性地将数据来源结构纳入数据驱动生产的动态一般均衡模型之中，从微观主体的数据共享机制来探讨数据要素如何影响宏观经济产出和社会福利。第三，在政策启示方面，本文充分融合中国数字经济高质量发展和惠利民生福祉的战略需求，着重关注如何推动数据产权制度有效增进社会福利，研究提出了配套数据脱敏保护政策和数据交易补贴政策的宏观政策工具思路。

## 二、理论模型构建

本部分将在 Jones and Tonetti（2020）基础上构建一个包含数据驱动生产的动态一般均衡模型，分析分散经济的市场均衡条件和社会计划者经济的最优均衡条件。

中国知网 https://www.cnki.net

（一）分散经济

1. 模型基本环境

在 $t$ 期，一个经济体（将一个行业视为经济体）中存在 $N_t$ 种产品，产品之间互相替代且替代弹性为 $\sigma>1$，通过 CES（Constant Elasticity of Substitution，常数替代弹性）形式加总得到最终总产出。令第 $i$ 种产品的数量为 $Y_{it}$，$i\in\{1,2,3,\ldots,N_t\}$，最终总产出的数量为 $Y_t$。为简化计算与分析，假设 $N_t$ 种产品企业的生产决策是对称的：

$$Y_t=\left[\int_0^{N_t}Y_{it}^{(\sigma-1)/\sigma}di\right]^{\sigma/(\sigma-1)}=N_t^{\sigma/(\sigma-1)}\cdot Y_{it}.$$

令劳动力总量为 $\bar{L}_t=\bar{L}_0\cdot e^{n_{\bar{L}}\cdot t}$，增长率为外生常数 $g_{\bar{L}}=n_{\bar{L}}$。[①] 非熟练劳动力的数量为 $L_t=(1-l_H)\bar{L}_0\cdot e^{n_{\bar{L}}\cdot t}$，主要投入 $N_t$ 种产品的生产经营环节。熟练劳动力的数量为 $H_t=l_H\bar{L}_0\cdot e^{n_{\bar{L}}\cdot t}$，要么选择投入 $N_t$ 种产品的生产经营环节，要么选择投入新产品的研发创新环节。外生常数 $l_H$ 满足 $0<l_H<1$。

第 $i$ 种产品对应的数据产量 $R_{it}=Y_{it}$ 是家庭消费和企业生产的副产品。数据来源主体差异和数据共享成本异质性是本文理论模型不同于 Jones and Tonetti（2020）的核心假设之一。[②] Jones and Tonetti（2020）的数据生成过程可视为一个“黑箱”，且消费者和企业共同承担同一份私有数据被全社会共享使用的成本。而本文尝试从划分数据来源主体的角度打开这一“黑箱”，认为消费者数据的共享会恶化消费者的隐私泄露问题，而企业数据的共享会增大在位企业被新企业取代的风险。数据产权可分割性是本文理论模型不同于 Jones and Tonetti（2020）的核心假设之二。本文假设数据产权初始分配取决于两类主体的数据来源结构：令消费者数据与企业数据的来源比例为 $\lambda^c:(1-\lambda^c)$，则消费者持有初始产权的数据量为 $\lambda^c\cdot R_{it}$，企业持有初始产权的数据量为 $(1-\lambda^c)\cdot R_{it}$。

外生参数 $\lambda^c$ 由数据生成过程的行业异质性决定并且满足 $0\leqslant\lambda^c\leqslant1$。例如，零售电商、社交媒体、智能家居、交通出行等行业主要产生消费者行为数据（$\lambda^c$ 更高）；而工业制造、能源矿产、B2B 金融服务、物流仓储等行业则更侧重于积累企业生产经营数据（$\lambda^c$ 更低）。同时，参数 $\lambda^c$ 也与数字经济的发展阶段密切相关，能够反映数据生成模式的演进特征。在初期阶段，数据来源以企业传统采集方式为主。随着数字化进程加速，消费者数据和企业复杂数据（如物联网数据等）的占比呈现上升趋势。调查报告显示[③]，在智能技术创新和硬件迭代更新的双重驱动下，智能网联车促使消费者出行数据在 2023 年同比增长 49%，而智能生产设备升级推动企业生产制造数据在 2023 年同比增长 20%。

接下来，介绍数据市场交易规则。假设经济体中存在两个数据中介搭建社会数据池，第一家数据中介专门收集消费者数据，第二家数据中介专门收集企业数据。数据交

① 为统一表达，$g_X$ 表示经济变量 $X$ 的增速，$\dot{X}_t$ 表示经济变量 $X$ 在 $t$ 期的单位时间变化量。

② 本文将 Jones and Tonetti（2020）理论模型应用于中国经济研究的合理性包括以下三方面：第一，从中国数字经济的超大市场基础来看，该理论模型所重点刻画的数据非竞争性与我国数字经济发展的庞大规模效应相契合。第二，从中国数字经济的制度体系建设来看，该理论模型所引入的数据产权分配研究契合我国数据市场制度初级建设阶段的进步需求。第三，从中国数字经济的宏观政策工具来看，该理论模型所引入的分散经济均衡效率测度、政府补贴提升效率等内容符合我国“有效市场与有为政府更好结合”的治理理念。

③ 摘自全国数据资源调查工作组发布的《全国数据资源调查报告（2023）》。

中国知网　https://www.cnki.net

易情形包括：第一，消费者可选择将所持有数据量 $\lambda^c \cdot R_{it}$，以每单位内部使用比例 $x_{it}^c \in [0,1]$、数据价格 $p_{Dit}^c$ 定向出售给第 $i$ 种产品企业，并以每单位共享使用比例 $\tilde{x}_{it}^c \in [0,1]$、数据价格 $\tilde{p}_{Dit}^c$ 出售给第一家数据中介。第二，企业可选择将所持有数据 $(1-\lambda^c) \cdot R_{it}$ 以每单位共享使用比例 $\tilde{x}_{it}^f \in [0,1]$、数据价格 $\tilde{p}_{Dit}^f$ 出售给第二家数据中介。第三，两家数据中介将所收集购入的消费者数据和企业数据，分别以数据价格 $\tilde{p}_{Dt}^{mc}$ 和 $\tilde{p}_{Dt}^{mf}$ 出售给需要使用数据开展生产的所有产品企业。因此，整体而言，平均每单位私有数据的社会共享比例为 $\tilde{x}_{it} = \lambda^c \cdot \tilde{x}_{it}^c + (1-\lambda^c) \cdot \tilde{x}_{it}^f$，包括消费者数据共享值 $\lambda^c \cdot \tilde{x}_{it}^c$ 和企业数据共享值 $(1-\lambda^c) \cdot \tilde{x}_{it}^f$。为简化数据定价，假设在任意情形的数据交易中，数据卖家始终具有数据的定价权，数据买家是数据价格的接受者，不具备协商议价的能力（Jones and Tonetti，2020）。令两家数据中介是垄断者。家庭将权衡数据出售的收益与隐私泄露的成本，决策消费者数据的共享比例 $\tilde{x}_{it}^c$（伴随决策数据价格 $\tilde{p}_{Dit}^c$）。企业将权衡数据出售的收益与被潜在新企业取代的风险，决策企业数据的共享比例 $\tilde{x}_{it}^f$（伴随决策数据价格 $\tilde{p}_{Dit}^f$）。同时，不失一般性地假设数据在每一期都完全折旧，不存在数据存量。

2. 代表性家庭的最优化问题

代表性家庭的人均消费量 $c_t = \left[\int_0^{N_t} c_{it}^{(\sigma-1)/\sigma} di\right]^{\sigma/(\sigma-1)}$ 是每种产品人均消费量 $c_{it} = Y_{it}/\bar{L}_t$ 的 CES 加总形式。此处假设代表性家庭包含所有熟练劳动力和非熟练劳动力，且家庭内部成员中两类劳动力的消费水平均等化。[①] 易知消费者持有第 $i$ 种产品的数据量为 $\lambda^c R_{it} = \lambda^c Y_{it} = \lambda^c c_{it} \cdot \bar{L}_t$，家庭的终生效用最大化问题可写成：

$$\max_{\{c_{it}, x_{it}^c, \tilde{x}_{it}^c, a_t\}} U = \bar{L}_0 \cdot \int_0^{\infty} e^{-(\rho - n_{\bar{L}}) \cdot t} \cdot \left\{ \ln\left( \int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} - \left[ \frac{\lambda^c}{N_t^2} \cdot \kappa \cdot \int_0^{N_t} (x_{it}^c)^2 di + \frac{\lambda^c}{N_t} \cdot \tilde{\kappa} \cdot \int_0^{N_t} (\tilde{x}_{it}^c)^2 di \right] \right\} dt, \tag{1}$$

$$\text{s.t. } \dot{a}_t = (r_t - n_{\bar{L}}) a_t + w_{Lt}(1 - l_H) + w_{Ht} l_H + \int_0^{N_t} (p_{Dit}^c \cdot x_{it}^c + \tilde{p}_{Dit}^c \cdot \tilde{x}_{it}^c) \cdot \lambda^c c_{it} di - \int_0^{N_t} p_{it} c_{it} di, \tag{2}$$

$$0 \leqslant x_{it}^c \leqslant 1,\ 0 \leqslant \tilde{x}_{it}^c \leqslant 1.$$

式(1)是家庭终生效用函数，$\rho$ 是时间贴现率，满足 $\rho > n_{\bar{L}}$ 以保证终生效用收敛。人均隐私成本关于每单位数据使用比例 $x_{it}^c$ 和 $\tilde{x}_{it}^c$ 是边际递增的二次函数形式。所有企业使用内部数据，将造成消费者隐私“小范围”泄露，负向作用强度为参数 $\kappa$，规模权重系数为 $\lambda^c/N_t^2$；所有企业非竞争性地使用社会数据池，将造成消费者隐私“大范围”泄露，负向作用强度为参数 $\tilde{\kappa}$，规模权重系数为 $\lambda^c/N_t$，且是“小范围”泄露的 $N_t$ 倍（即市场规模）。式(2)是家庭的人均预算函数，$r_t$ 是人均资产 $a_t$ 的利率水平，$p_{it}$ 是第 $i$ 种产品的销售价格。人均消费支出为 $\int_0^{N_t} p_{it} c_{it} di$。人均收入包括资产利息、劳动薪酬（非熟练劳动力和

① 为简化理论模型的推导过程，本文假设家庭内部成员中两类劳动力的消费水平均等化。若修改此假设，令每类劳动力的消费水平分别取决于各自的工资水平。由于消费者隐私成本是在消费水平基础上的百分比效用损失，因此这一修改将不影响本文基于数据共享决策结果的核心逻辑和主要结论。

中国知网 https://www.cnki.net

熟练劳动力的工资水平分别为 $w_{Lt}$ 和 $w_{Ht}$）、所持有消费者数据的出售收入。

家庭的最优决策结果为①：

$$p_{it}-p_{Dit}^{c}\cdot\lambda^{c}x_{it}^{c}-\widetilde{p}_{Dit}^{c}\cdot\lambda^{c}\widetilde{x}_{it}^{c}=(c_t/c_{it})^{1/\sigma}=(Y_t/Y_{it})^{1/\sigma}. \tag{3}$$

3. 数据中介的最优化问题

数据中介是垄断者，但令其同时受到潜在进入者以接近零成本进入市场的竞争威胁。因此，数据中介的最优决策结果将导致其实现零利润以阻止潜在进入者（Jones and Tonetti，2020）。② 第一家数据中介所搭建的社会数据池总量为 $B_t^c$，且向每家企业成功售出的数据量为 $\widetilde{D}_t^c\leqslant B_t^c$，故其利润最大化问题可写成：

$$\max_{\{\widetilde{p}_{Dt}^{mc}\}}\widetilde{p}_{Dt}^{mc}\cdot\int_0^{N_t}\widetilde{D}_t^c di-\int_0^{N_t}\widetilde{p}_{Dit}^{c}\cdot\widetilde{x}_{it}^{c}\cdot\lambda^{c}R_{it}di,$$

$$\text{s.t. }\widetilde{D}_t^c\leqslant B_t^c=\int_0^{N_t}(\widetilde{x}_{it}^{c}\cdot\lambda^{c}R_{it})di.$$

同理，令第二家数据中介所搭建的社会数据池总量为 $B_t^f$，且向每家企业成功售出的数据量为 $\widetilde{D}_t^f\leqslant B_t^f$，故其利润最大化问题可写成：

$$\max_{\{\widetilde{p}_{Dt}^{mf}\}}\widetilde{p}_{Dt}^{mf}\cdot\int_0^{N_t}\widetilde{D}_t^f di-\int_0^{N_t}\widetilde{p}_{Dit}^{f}\cdot\widetilde{x}_{it}^{f}\cdot(1-\lambda^{c})R_{it}di,$$

$$\text{s.t. }\widetilde{D}_t^f\leqslant B_t^f=\int_0^{N_t}[\widetilde{x}_{it}^{f}\cdot(1-\lambda^{c})R_{it}]di.$$

4. 代表性企业的最优化问题

对第 $i$ 种产品的代表性企业而言，其所使用的内部数据包括企业自身拥有的持有规模比例为 $1-\lambda^{c}$、每单位使用比例为 $x_{it}^{f}\in[0,1]$ 的数据量 $(1-\lambda^{c})x_{it}^{f}\cdot R_{it}$，以及从消费者处定向购买的内部数据量 $D_{it}^{c}=\lambda^{c}x_{it}^{c}\cdot R_{it}$；其所使用的社会外部数据包括从两家数据中介处购买的外部数据量 $\widetilde{D}_t^c+\widetilde{D}_t^f$。令参数 $\alpha$ 和 $1-\alpha$ 分别表示内部数据和外部数据对加总数据的权重，参数 $\alpha$ 满足取值范围 $(0,1)$，则数据投入函数可写成：

$$D_{it}=\alpha[D_{it}^{c}+(1-\lambda^{c})x_{it}^{f}\cdot R_{it}]+(1-\alpha)(\widetilde{D}_t^c+\widetilde{D}_t^f).$$

同时，企业雇用 $L_{yit}$ 单位的非熟练劳动力、$H_{yit}$ 单位的熟练劳动力开展产品的生产经营工作。$z_L$ 和 $z_H$ 分别表示非熟练劳动力和熟练劳动力的劳动效率，且满足 $z_L<z_H$。参数 $\zeta$ 表示两类劳动力的产出替代弹性。参数 $\eta$ 和 $1-\eta$ 分别表示数据要素和劳动要素的产出弹性，参数 $\eta$ 满足取值范围 $(0,1)$，可刻画数据要素对经济生产的重要性。所以，产品生产函数可写成如下柯布-道格拉斯（Cobb-Douglas）的函数形式③：

$$Y_{it}=D_{it}^{\eta}\cdot\{[(L_{yit}\cdot z_L)^{(\zeta-1)/\zeta}+(H_{yit}\cdot z_H)^{(\zeta-1)/\zeta}]^{\zeta/(\zeta-1)}\}^{1-\eta}.$$

代表性企业面临垄断竞争市场，且结合式（3）可得第 $i$ 种产品的销售价格为 $p_{it}=p_{Dit}^{c}\cdot\lambda^{c}x_{it}^{c}+\widetilde{p}_{Dit}^{c}\cdot\lambda^{c}\widetilde{x}_{it}^{c}+(Y_t/Y_{it})^{1/\sigma}$。同时，由于社会数据池的企业信息泄露问题，每期在位企业被潜在新企业取代的概率服从泊松到达率 $\delta(1-\lambda^{c})\cdot(\widetilde{x}_{it}^{f})^{2}$。该概率关于每单位数据共享比例 $\widetilde{x}_{it}^{f}$ 是边际递增的二次函数形式，创造性破坏强度是参数 $\delta$，规模系数为

① 本节分散经济均衡的数理推导过程和完整结果参见附录Ⅰ，感兴趣的读者可在《经济学》（季刊）官网（https://ceq.ccer.pku.edu.cn）下载。

② 这将使得数据中介仅仅发挥私人部门间“数据转移”的作用，是对数据交易市场的简化设定。

③ 本文将创新活动构造为产品种类的增加（也即横向创新），而不是企业生产率的提升（也即纵向创新）。

中国知网　https://www.cnki.net

$(1-\lambda^c)$。因此，企业市场价值 $V_{it}$ 的最大化问题满足如下贝尔曼(Bellman)方程：

$$r_t V_{it} = \max_{\substack{(L_{it},H_{it},\tilde{D}_{it}^f,\\ D_{it}^c,x_{it}^f,\tilde{x}_{it}^f)}} \left\{ \begin{aligned} &[p_{Dit}^c \cdot \lambda^c x_{it}^c + \tilde{p}_{Dit}^c \cdot \lambda^c \tilde{x}_{it}^c + (Y_t/Y_{it})^{1/\sigma}] \cdot Y_{it} - w_{Lt} L_{yit} - w_{Ht} H_{yit} - \tilde{p}_{Dt}^{mc} \cdot \tilde{D}_t^c \\ &- \tilde{p}_{Dt}^{mf} \cdot \tilde{D}_t^f - p_{Dit}^c \cdot D_{it}^c + \tilde{p}_{Dit}^f \cdot \tilde{x}_{it}^f \cdot (1-\lambda^c) R_{it} + \dot{V}_{it} - \delta(1-\lambda^c) \cdot (\tilde{x}_{it}^f)^2 \cdot V_{it} \end{aligned} \right\},$$

$$\text{s.t. } R_{it} = Y_{it} = \{\alpha [D_{it}^c + (1-\lambda^c) x_{it}^f \cdot R_{it}] + (1-\alpha)(\tilde{D}_t^c + \tilde{D}_t^f)\}^{\eta} \cdot$$

$$\{[(L_{yit} \cdot z_L)^{(\zeta-1)/\zeta} + (H_{yit} \cdot z_H)^{(\zeta-1)/\zeta}]^{\zeta/(\zeta-1)}\}^{1-\eta},$$

$$0 \leqslant x_{it}^f \leqslant 1,\ 0 \leqslant \tilde{x}_{it}^f \leqslant 1.$$

5. 企业的进入退出

潜在的市场进入企业将雇用熟练劳动力进行研发创新，平均 $\chi$ 单位的熟练劳动力能够成功研发出一种新产品。新增产品种类 $\dot{N}_t$ 的动态方程可以表示为：

$$\dot{N}_t = H_{et}/\chi.$$

成功进入企业的收益包括新产品价值，以及与其他成功进入企业平均瓜分被取代且退出市场的原在位企业的价值。故产品市场的自由进出条件为：

$$\chi w_{Ht} = V_{it} + \int_0^{N_t} \delta(1-\lambda^c) \cdot (\tilde{x}_{it}^f)^2 \cdot V_{it} di / \dot{N}_t = V_{it} + \delta(1-\lambda^c)(\tilde{x}_{it}^f)^2 \cdot V_{it}/g_N.$$

6. 分散经济均衡配置条件

分散经济达到均衡时需满足市场出清条件。其中，产品市场出清条件是 $\bar{L}_t c_{it} = Y_{it} = R_{it}$；劳动力市场出清条件包括 $\int_0^{N_t} L_{yit} di = N_t \cdot L_{yit} = L_t = (1-l_H)\bar{L}_t$，$\int_0^{N_t} H_{yit} di = N_t \cdot H_{yit} = H_{yt}$，$H_{yt} + H_{et} = H_t = l_H \bar{L}_t$；数据市场出清条件包括 $D_{it}^c = x_{it}^c \cdot \lambda^c Y_{it}$，$\tilde{D}_t^c = B_t^c = N_t \cdot \tilde{x}_{it}^c \cdot \lambda^c Y_{it}$，$\tilde{D}_t^f = B_t^f = N_t \cdot \tilde{x}_{it}^f \cdot (1-\lambda^c) Y_{it}$。

（二）社会计划者经济

为判断该数据产权制度是否能够有效促进社会福利，本文考虑将分散经济均衡的社会福利水平与社会计划者最优水平进行对比。社会计划者的最优化问题为[①]：

$$\max_{\{x_{it}^c, x_{it}^f, \tilde{x}_{it}^c, \tilde{x}_{it}^f, H_{yt}, N_t\}} U = \bar{L}_0 \cdot \int_0^{\infty} e^{-(\rho-n\bar{L})t} \cdot \{\ln(Y_t/\bar{L}_t) - \kappa \cdot \lambda^c (x_{it}^c)^2/N_t - \tilde{\kappa} \cdot \lambda^c (\tilde{x}_{it}^c)^2\} dt,$$

$$\text{s.t. } Y_t = N_t^{\frac{1}{\sigma-1}+\frac{\eta}{1-\eta}} \cdot \left\{ \begin{aligned} &\alpha [\lambda^c x_{it}^c + (1-\lambda^c) x_{it}^f] \cdot N_t^{-1} \\ &+ (1-\alpha)[\lambda^c \tilde{x}_{it}^c + (1-\lambda^c)\tilde{x}_{it}^f] \end{aligned} \right\}^{\frac{\eta}{1-\eta}} \cdot$$

$$\{[(1-l_H)\bar{L}_t z_L]^{\frac{\zeta-1}{\zeta}} + (H_{yt} z_H)^{\frac{\zeta-1}{\zeta}}\}^{\frac{\zeta}{\zeta-1}},$$

$$\dot{N}_t = H_{et}/\chi = (l_H \bar{L}_t - H_{yt})/\chi,$$

$$0 \leqslant x_{it}^c \leqslant 1,\ 0 \leqslant x_{it}^f \leqslant 1,\ 0 \leqslant \tilde{x}_{it}^c \leqslant 1,\ 0 \leqslant \tilde{x}_{it}^f \leqslant 1.$$

## 三、经济均衡结果的理论分析

（一）分散经济的均衡结果

当市场规模足够大（$N_t \to \infty$）时，分散经济的数据共享和劳动力配置均衡将无限收

① 本节社会计划者最优均衡的数理推导过程和完整结果参见附录Ⅱ。

中国知网 https://www.cnki.net

敛于如下稳态结果(以右上标 * 标示)：

$$\underbrace{2\tilde{\kappa}\cdot\tilde{x}_i^{c*}}_{\text{消费者数据共享的边际成本}}=\frac{2\delta\cdot\tilde{x}_i^{f*}}{\underbrace{\sigma\cdot[\delta(1-\lambda^c)(\tilde{x}_i^{f*})^2+\rho]}_{\text{企业数据共享的边际成本}}}=\frac{\eta(\sigma-1)}{\underbrace{(1-\eta)\sigma\cdot[\lambda^c\tilde{x}_i^{c*}+(1-\lambda^c)\tilde{x}_i^{f*}]}_{\text{数据共享的边际收益}}},\tag{4}$$

$$H_{yi}^{*}\cdot\left\{\left[\frac{(1-l_H)z_L}{l_Hz_H}\cdot\left(\frac{n_{\bar{L}}\chi}{H_{yi}^{*}}+1\right)\right]^{(\zeta-1)/\zeta}+1\right\}=n_{\bar{L}}\chi(\sigma-1)\cdot\left[\frac{\rho-n_{\bar{L}}}{n_{\bar{L}}+\delta(1-\lambda^c)(\tilde{x}_i^{f*})^2}+1\right],\tag{5}$$

$$L_{yi}^{*}=[(1-l_H)/l_H]\cdot(n_{\bar{L}}\chi+H_{yi}^{*}),\tag{6}$$

$$H_{et}^{*}:H_{yt}^{*}=n_{\bar{L}}\chi/H_{yi}^{*},\tag{7}$$

$$N_t^{*}=N_0^{*}\cdot e^{n_{\bar{L}}\cdot t}=[l_H\bar{L}_0/(n_{\bar{L}}\chi+H_{yi}^{*})]\cdot e^{n_{\bar{L}}\cdot t}.\tag{8}$$

由上述结果可知，在微观层面上，式(4)刻画了数据共享的边际成本等于边际收益。消费者数据共享的负向影响是家庭因隐私泄露而产生效用损失，边际成本为 $2\tilde{\kappa}\tilde{x}_i^{c*}$。企业数据共享的负向影响是在位企业被取代的概率边际提高 $2\delta\tilde{x}_i^{f*}$，同时平均存活寿命降低也导致企业市场价值降低($V_i$ 正比于 $1/[\sigma\delta(1-\lambda^c)(\tilde{x}_i^{f*})^2+\sigma\rho]$)。两类数据共享的正向影响是非竞争性地促进所有企业生产，边际收益均为 $\eta(\sigma-1)/\{(1-\eta)\sigma\cdot[\lambda^c\tilde{x}_i^{c*}+(1-\lambda^c)\tilde{x}_i^{f*}]\}$。理论结果显示①，分散经济均衡的消费者数据共享值 $\lambda^c\tilde{x}_i^{c*}$ 以及隐私成本值 $\tilde{\kappa}\lambda^c(\tilde{x}_i^{c*})^2$ 关于消费者数据占比 $\lambda^c$ 递增，企业数据共享值 $(1-\lambda^c)\tilde{x}_i^{f*}$ 以及被取代概率值 $\delta(1-\lambda^c)(\tilde{x}_i^{f*})^2$ 关于企业数据占比 $(1-\lambda^c)$ 递增，即关于 $\lambda^c$ 递减。由式(5)至式(8)可继续得到企业数据共享将产生熟练劳动力再配置效应：随着 $\lambda^c$ 增加，企业共享数据值 $(1-\lambda^c)\tilde{x}_i^{f*}$ 越少，潜在进入者开展创新活动后成功取代在位企业的概率 $\delta(1-\lambda^c)\cdot(\tilde{x}_i^{f*})^2$ 越低。故熟练劳动力选择投入新产品创新环节的比例越少($H_{et}^{*}/H_{yt}^{*}$ 越低)，产品种类的数量越少($N_t^{*}$ 越低)，同时将使得单个企业的生产规模增大($H_{yi}^{*}$ 和 $L_{yi}^{*}$ 均越高)。

分散经济的宏观总量均衡将无限收敛于如下稳态结果：

$$Y_t^{*}=\underbrace{e^{\left(\frac{1}{\sigma-1}+\frac{1}{1-\eta}\right)\cdot n_{\bar{L}}\cdot t}}_{\text{长期增长}}\cdot\underbrace{\{(1-\alpha)[\lambda^c\tilde{x}_i^{c*}+(1-\lambda^c)\tilde{x}_i^{f*}]\}^{\frac{\eta}{1-\eta}}}_{\text{水平增加：两类数据共享的要素深化效应}}\cdot\underbrace{N_0^{*\frac{1}{\sigma-1}+\frac{1}{1-\eta}}\cdot[(L_{yi}^{*}z_L)^{\frac{\zeta-1}{\zeta}}+(H_{yi}^{*}z_H)^{\frac{\zeta-1}{\zeta}}]^{\frac{\zeta}{\zeta-1}}}_{\text{水平增加：企业数据共享的熟练劳动力再配置效应}},\tag{9}$$

$$g_Y^{*}=\dot{Y}_t/Y_t=[1/(\sigma-1)+1/(1-\eta)]\cdot n_{\bar{L}},\tag{10}$$

$$U^{*}=\frac{\bar{L}_0}{\rho-n_{\bar{L}}}\cdot\left[\ln\left(\frac{Y_0^{*}}{\bar{L}_0}\right)-\tilde{\kappa}\lambda^c(\tilde{x}_i^{c*})^2+\frac{1}{\rho-n_{\bar{L}}}\cdot(g_Y^{*}-n_{\bar{L}})\right].\tag{11}$$

由经济总产出的式(9)和式(10)可得，数据共享促进经济总产出长期增长的作用机制是数据要素非竞争性的规模报酬递增效应，该增长效应 $n_{\bar{L}}/(1-\eta)$ 随着数据要素越重

① 本节对经济均衡结果进行理论分析的详细推导过程参见附录Ⅲ。

中国知网　https://www.cnki.net

要（参数 $\eta$ 越大）而加强。[①] 数据共享影响经济总产出水平提高的作用机制包括直接的要素深化效应和间接的熟练劳动力再配置效应。当 $\lambda^c$ 越大时，一方面，消费者数据共享值 $\lambda^c\tilde{x}_i^{c*}$ 越多，其将作为生产要素直接促进企业生产，发挥要素深化效应且作用方向为正：$\frac{\partial Y_t^*}{\partial[\lambda^c\tilde{x}_i^{c*}+(1-\lambda^c)\tilde{x}_i^{f*}]}\cdot\frac{\partial\lambda^c\tilde{x}_i^{c*}}{\partial\lambda^c}\geqslant 0$。另一方面，企业共享数据值 $(1-\lambda^c)\tilde{x}_i^{f*}$ 越少，要素深化效应弱化且作用方向为负：$\frac{\partial Y_t^*}{\partial[\lambda^c\tilde{x}_i^{c*}+(1-\lambda^c)\tilde{x}_i^{f*}]}\cdot\frac{\partial(1-\lambda^c)\tilde{x}_i^{f*}}{\partial\lambda^c}\leqslant 0$；同时其将通过熟练劳动力再配置效应降低产品种类数量水平值 $N_0^*$，并提高单个企业的生产规模 $H_{yi}^*$ 和 $L_{yi}^*$，作用方向为负：$\left[\frac{\partial Y_t^*}{\partial N_0^*}\cdot\frac{\partial N_0^*}{\partial H_{yi}^*}+\frac{\partial Y_t^*}{\partial L_{yi}^*}\cdot\frac{\partial L_{yi}^*}{\partial H_{yi}^*}+\frac{\partial Y_t^*}{\partial H_{yi}^*}\right]\cdot\frac{\partial H_{yi}^*}{\partial\lambda^c}\leqslant 0$。

由社会福利的式(11)可得，社会福利取决于经济总产出的水平值 $Y_0^*$ 和长期增速 $g_Y^*$，以及消费者的隐私泄露程度（$-\tilde{\kappa}\lambda^c(\tilde{x}_i^{c*})^2$）。因此，数据共享将在经济产出效益和消费者隐私成本的权衡之中影响社会福利水平：当 $\lambda^c$ 越大时，消费者数据共享越多，这将通过要素深化效应而提升经济总产出，进而提升社会福利，并通过恶化消费者隐私泄露程度而降低社会福利。企业数据共享越少，这将通过要素深化效应弱化、熟练劳动力再配置效应抑制新产品创新而降低经济总产出，进而降低社会福利。

此处需要注意的是，每单位数据共享比例 $\tilde{x}_i^{c*}$ 和 $\tilde{x}_i^{f*}$ 关于 $\lambda^c$ 的变动关系式，与两类数据共享的成本强度参数之比 $\tilde{\kappa}/\delta$ 密切相关。为聚焦本文的研究结论并更符合后文数值模拟的参数取值范围，后续理论分析将令 $\tilde{\kappa}/\delta$ 的取值范围为 $\tilde{\kappa}/\delta\leqslant[1-\eta(\sigma-1)/(1-\eta)]/(\rho\sigma)$。这将使得①分散经济均衡结果更符合数据市场上消费者数据的交易多于企业数据的现实，也即 $\tilde{x}_i^{c*}\geqslant\tilde{x}_i^{f*}$：企业出售可能涉及商业机密的私有数据后，将面临更加劣势的产品市场竞争和更加严峻的存亡危机，故企业共享数据的意愿比较低。②每单位数据共享比例 $\tilde{x}_i^{c*}$ 和 $\tilde{x}_i^{f*}$ 关于 $\lambda^c$ 递减：数据共享的边际收益关于 $\lambda^c$ 递减。

（二）分散经济均衡的效率：比较分散经济均衡和社会计划者均衡

当市场规模足够大（$N_t\to\infty$）时，社会计划者的数据共享和劳动力配置均衡将无限收敛于如下稳态结果（以右上标 $sp$ 标示，其他结果表达式与式(6)至式(11)相同）：

$$2\tilde{\kappa}\cdot\tilde{x}_i^{c,sp}=\frac{\eta}{(1-\eta)\cdot[\lambda^c\tilde{x}_i^{c,sp}+(1-\lambda^c)\tilde{x}_i^{f,sp}]},\tag{12}$$

$$\tilde{x}_i^{f,sp}=1,\tag{13}$$

$$H_{yi}^{sp}\cdot\left\{\left[\frac{(1-l_H)z_L}{l_Hz_H}\cdot\left(\frac{n_{\bar{L}}\chi}{H_{yi}^{sp}}+1\right)\right]^{(\zeta-1)/\zeta}+1\right\}=\frac{n_{\bar{L}}\chi}{1/(\sigma-1)+\eta/(1-\eta)}.\tag{14}$$

比较易得，分散经济均衡与社会计划者均衡的增长效应完全等价，两种配置结果的差距主要集中于水平效应。可构造变量 $\Omega(*,sp)$ 表示 $U^*(c_t^*)=U^{sp}(c_t^{sp}\cdot\Omega)$：当两种

① 经济总产出的长期增速正比于人口增速 $n_{\bar{L}}$，这源自本文的动态一般均衡模型属于半内生经济增长模型。同时，经济总产出的增长源泉还包括产品种类多样性偏好的规模报酬递增效应，也即为 $n_{\bar{L}}/(\sigma-1)$。

中国知网 https://www.cnki.net

配置结果实现等量社会福利时，社会计划者均衡中的人均消费量需乘以 $\Omega(*,sp)\leqslant 1$ 的比例值进行折损。易得，变量 $\Omega(*,sp)$ 的数值越大或越接近 1，则分散经济均衡的社会福利越接近社会计划者最优水平，也即分散经济均衡的社会福利效率越高。可得：

$$\Omega(*,sp)=\frac{Y_0^*}{Y_0^{sp}}\cdot\exp\{-\widetilde{\kappa}\left[\lambda^c\left(\widetilde{x}_i^{c*}\right)^2-\lambda^c\left(\widetilde{x}_i^{c,sp}\right)^2\right]\}, \tag{15}$$

$$\frac{\partial(Y_0^*/Y_0^{sp})}{\partial\lambda^c}=\frac{Y_0^*}{Y_0^{sp}}\Bigg[\underbrace{\frac{\eta\widetilde{x}_i^{sp}}{(1-\eta)\widetilde{x}_i^*}\cdot\frac{\partial(\widetilde{x}_i^*/\widetilde{x}_i^{sp})}{\partial\lambda^c}}_{\text{两类数据共享的要素深化效应(+)}}$$

$$\underbrace{-\frac{\dfrac{\eta}{1-\eta}+\dfrac{(\rho-n_{\overline{L}})/(\sigma-1)}{\rho+\delta(1-\lambda^c)\left(\widetilde{x}_i^{f*}\right)^2}}{n_{\overline{L}}\chi+H_{yi}^*}\cdot\frac{\partial H_{yi}^*}{\partial(1-\lambda^c)\left(\widetilde{x}_i^{f*}\right)^2}\cdot\frac{\partial(1-\lambda^c)\left(\widetilde{x}_i^{f*}\right)^2}{\partial\lambda^c}}_{\text{企业数据共享的熟练劳动力再配置效应(-)}}\Bigg]. \tag{16}$$

分析式(15)和式(16)可得，分散经济均衡的社会福利效率取决于两种均衡的数据共享程度差异以及其导致的经济产出效益和消费者隐私成本的差异。第一，比较两类数据共享的要素深化效应及其导致的经济产出效益。比较式(4)和式(12)易得，由于产品市场垄断竞争且企业的垄断加价系数是 $\sigma/(\sigma-1)$，因此企业垄断加价的扭曲效应导致两类数据平均共享程度 $\widetilde{x}_i^*=\lambda^c\widetilde{x}_i^{c*}+(1-\lambda^c)\widetilde{x}_i^{f*}$ 低于社会计划者的最优水平 $\widetilde{x}_i^{sp}=\lambda^c\widetilde{x}_i^{c,sp}+(1-\lambda^c)\widetilde{x}_i^{f,sp}$。同时，随着消费者数据相对企业变得更加丰富即 $\lambda^c$ 增加，分散经济均衡(共享消费者数据的意愿高于企业)的 $\widetilde{x}_i^*$ 逐渐增大，而社会计划者均衡(共享消费者数据的意愿低于企业[①])的 $\widetilde{x}_i^{sp}$ 逐渐减小。换言之，从两类数据共享的要素深化效应来看，分散经济均衡的整体数据平均共享不足问题随着 $\lambda^c$ 增大而减弱：$\widetilde{x}_i^*\leqslant\widetilde{x}_i^{sp}$，$\partial(\widetilde{x}_i^*/\widetilde{x}_i^{sp})/\partial\lambda^c\geqslant 0$，促进经济总产出水平更加接近社会计划者最优水平。

第二，比较企业数据共享的熟练劳动力再配置效应及其导致的经济产出效益。比较式(5)至式(8)与式(14)，易得分散经济均衡中熟练劳动力配置到新产品创新环节的比例、产品种类数量均小于社会计划者最优水平($H_{yi}^*>H_{yi}^{sp}$，$H_{et}^*/H_{yt}^*<H_{et}^{sp}/H_{yt}^{sp}$，$N_t^*<N_t^{sp}$)。其中，社会计划者的熟练劳动力最优配置水平与数据共享程度无关，主要与产品种类增加的规模报酬递增效应相关。而分散经济均衡将随着 $\lambda^c$ 增大而产生抑制新产品创新的熟练劳动力再配置效应，这致使经济总产出水平更加偏离社会计划者最优水平。

第三，比较消费者数据共享的隐私成本。一方面，比较式(4)与式(13)可得，由于分散经济均衡的企业数据共享程度低于社会计划者最优水平，$(1-\lambda^c)\cdot(\widetilde{x}_i^{f*}-\widetilde{x}_i^{f,sp})\leqslant 0$，数据市场处于企业数据供给不足的状态，因此消费者共享数据的边际收益高于社会计划者水平，则消费者存在过度共享数据的动机。另一方面，比较式(4)与式(12)可得，分散经济中企业垄断加价的扭曲效应导致消费者共享数据的意愿低于社会计划者水平。结合上述的作用方向相反的两方面，可推导证得当企业数据共享成本强度参数满足条件

① 由式(13)易得，社会计划者共享企业数据的意愿高于消费者数据($\widetilde{x}_i^{c,sp}\leqslant\widetilde{x}_i^{f,sp}=1$)。对社会计划者而言，新旧企业更替仅仅是在位企业价值向新企业价值的“无损”转移，故其将选择完全共享企业数据，并相应挤出消费者数据的共享程度。

中国知网　https://www.cnki.net

$\delta>\rho\eta\sigma^2/[2(1-\eta)(\sigma-1)-\eta(\sigma-1)^2]$ 时，分散经济均衡的企业数据共享不足问题更突出，第一方面的影响较强。此时分散经济均衡存在消费者数据共享过多问题，且消费者隐私过度泄露程度随着 $\lambda^c$ 增大而先恶化后缓解，也即 $\partial\lambda^c[(\widetilde{x}_i^{c*})^2-(\widetilde{x}_i^{c,sp})^2]/\partial\lambda^c$ 先为正数、后为负数。反之，第二方面的影响较强，此时分散经济均衡存在消费者数据共享不足问题，且消费者隐私泄露程度随着 $\lambda^c$ 增大而减弱，也即 $\partial\lambda^c[(\widetilde{x}_i^{c*})^2-(\widetilde{x}_i^{c,sp})^2]/\partial\lambda^c\leqslant 0$。

第四，最终比较社会福利水平。基于经济产出效益和消费者隐私成本的权衡可得，若以社会计划者最优均衡作为比较基准，则数据来源结构的变化将对分散经济均衡的社会福利效率 $\Omega(*,sp)$ 产生正向和负向的影响作用。可证得命题 1。

**命题 1** 当两类数据共享的成本强度参数满足条件 $\widetilde{\kappa}/\delta\leqslant[1-\eta(\sigma-1)/(1-\eta)]/(\rho\sigma)$ 时，若以社会计划者最优均衡作为比较基准，则分散经济均衡的社会福利效率 $\Omega(*,sp)$ 与消费者数据占比 $\lambda^c$ 呈现非线性关系。

## 四、数 值 模 拟

### （一）参数设定

广泛受认可且合理科学的参数取值是数值模拟的必要前提。本文主要依据理论模型属性和中国经验事实设定合适的参数取值，旨在使数值模拟结果尽可能贴合中国经济现实。同时，附录Ⅳ将对参数进行敏感性分析，以检验参数选取和主要结论的稳健性。

第一步，部分参数取值将依据理论模型属性，沿用重要文献的合理设定或测算结果。在加总参数中，第一，时间贴现率。Jones and Tonetti(2020)令时间贴现率为 $\rho=2.5\%$。本文选取 $\rho=2.5\%$，并在附录Ⅳ中设定 $\rho=2\%$ 和 $\rho=3\%$。第二，产品替代弹性。已有量化研究估算得到同部门或同行业内不同种类产品之间的替代弹性的取值范围约为[5，10](Anderson and Van Wincoop，2004；Atkeson and Burstein，2008；Edmond et al.，2015；Gaubert and Itskhoki，2021)。本文选取 $\sigma=7$，并在附录Ⅳ中设定 $\sigma=5$ 和 $\sigma=10$。第三，两类劳动力的产出替代弹性。已有量化研究估算得到熟练劳动力和非熟练劳动力的产出替代弹性的取值范围约为[1.4，2](Katz and Murphy，1992；Angrist，1995；Krusell et al.，2000；Autor et al.，2008)，通常取值 1.4，并通常将两类劳动力分别定义为大学本科学历及以上、大学专科学历及以下的劳动者。本文选取 $\zeta=1.4$，并在附录Ⅳ中设定 $\zeta=2$。

在数据要素的相关参数中，第一，数据要素重要性。Jones and Tonetti(2020)初步判断数据要素对 GDP 产出的贡献份额大约在 5%至 10%区间内，据此认为企业生产函数中数据重要性参数 $\eta$ 的取值范围是[0.03，0.12]。郭凯明等(2024)结合中国经济事实认为在与数字经济直接相关的新兴行业中，数据要素(不考虑外部性)的产出弹性为 0.125。

中国知网 https://www.cnki.net

因此，本文选取 $\eta=0.06$[①]，并在附录Ⅳ中设定 $\eta=0.03$ 和 $\eta=0.12$。第二，令内部数据和外部数据对加总数据的权重相等，也即参数 $\alpha=1-\alpha=0.5$。[②]

在劳动力要素的相关参数中，对于劳动力生产效率，徐悦等（2024）根据 2000 年至 2010 年中国人口普查数据的技能工资溢价，测算得到高技能劳动力（大学本科及以上）对中技能劳动力（高中和大学专科）和低技能劳动力（初中及以下）的相对生产效率比值是 1.34 和 2.13。因此，本文令非熟练劳动力的生产效率参数标准化为 $z_L=1$，熟练劳动力的生产效率参数为 $z_H=1.7$，并在附录Ⅳ中设定 $z_H=1.34$ 和 $z_H=2.13$。

第二步，部分参数取值将依据于中国可获取最新年份或者过去二十年左右的实际经济数据。在数据要素的相关参数中，令企业被创造性破坏的成本强度参数的基准值 $\delta=0.2$，这意味着当行业内全部是企业数据时泄露所有数据会增加 20% 的被取代概率，则在位企业在没有其他冲击时的平均寿命约为 5 年。这是较为保守的取值，实际上中国企业的平均寿命低于 5 年。[③] 本节也将开展较低值 $\delta=0.1$ 和较高值 $\delta=0.3$ 的比较分析。

在劳动力要素的相关参数中，令初始劳动力总量标准化为 $\bar{L}_0=100$，令熟练劳动力和非熟练劳动力的数量比值为 $H_t:L_t=l_H:(1-l_H)=8:92$，也即 $l_H=0.08$。参考依据为：2020 年全国人口普查数据显示，大学本科及研究生学历人数占所有 6 岁及以上人口总数的比例约为 8%。令市场进入的劳动力成本参数为 $\chi=0.001$。参考依据为：国家统计局数据显示，2023 年研究与试验发展人员（以全时当量计算）约占劳动力总量的 0.8%，近似占大学本科学历及以上劳动力总量的 10%。因此，假设数字经济相关产业的劳动力总量为 1 亿（标准化为 $\bar{L}_0=100$），并设定新产品进入市场的劳动力成本为 1 000 个（同比例换算为 $\chi=0.001$）熟练劳动力且包含 100 个科研人员。[④] 令劳动力数量增速为 $n_{\bar{L}}=0.5\%$。参考依据为：国家统计局数据显示，中国 2000—2023 年的人口数量年均增速约为 0.5%。

第三步，部分参数没有量化文献或者中国实际经济数据的支撑，故本文将自主设定合理的取值或取值范围。在数据要素的相关参数中，令消费者隐私成本强度参数的基准值为 $\tilde{\kappa}=-\ln(1-0.2)=0.223$，这意味着当行业内全部是消费者数据时泄露所有数据等价于消费量下降 20% 的效用损失。并且，两类数据共享的成本强度参数容易满足条件 $\tilde{\kappa}/\delta=1.115<[1-\eta(\sigma-1)/(1-\eta)]/(\rho\sigma)\approx3.526$ 和 $\delta>\rho\eta\sigma^2/[2(1-\eta)(\sigma-1)-\eta(\sigma-1)^2]\approx0.008$。本节也将开展较低值 $\tilde{\kappa}=-\ln(1-0.1)=0.105$ 和较高值 $\tilde{\kappa}=-\ln(1-0.3)=0.357$ 的比较分析。

---

① 参数 $\eta=0.06$ 是较为保守的参数设定。具有参考意义的是，国家工业信息安全发展研究中心等机构联合发布的《中国数据要素市场发展报告（2021—2022）》结合国家统计局投入产出表和统计年鉴数据，测算了中国 14 个行业的数据要素产出弹性。结果显示，14 个行业的数据产出弹性估算值最低值为 0.0016、最高值为 3.0440。其中，信息传输、软件和信息技术服务业，科学研究和技术服务业，制造业的数据产出弹性依次为 3.0440、1.5699、0.4643。

② 参数 $\alpha$ 的取值仅影响模拟结果的数量级，不影响数值模拟的主要结论。

③ 目前仅能收集到年份较早的相关统计数据，故选取较为保守的取值。2012 年人力资源领域综合服务机构 CHINA HRKEY 的《中国中小企业人力资源管理白皮书》公布数据显示，中国中小企业的平均寿命仅为 2.5 年。2016 年 6 月 1 日《经济日报》文章《中国企业平均寿命为什么短》指出，中国民营企业平均寿命仅 3.7 年，中小企业平均寿命更是只有 2.5 年（详见 http://paper.ce.cn/jjrb/html/2016-06/01/content_302495.htm，访问时间：2025 年 9 月 5 日）。考虑到中国民营经济约占据 90% 以上的企业数量，故可大体认为中国企业的平均寿命低于 5 年。

④ $\bar{L}_0$ 与 $\chi$ 的取值数量级仅影响模拟结果的数量级，不影响数值模拟的主要结论。回顾 Jones and Tonetti（2020）的参数设定：在假设数字经济相关产业的劳动力（同质）总量为 1 亿、科研人员占比 1% 的基础上，设置新产品进入市场的劳动力成本为 1 万个劳动力且包含 100 个科研人员。

中国知网　https://www.cnki.net

（二）数值模拟结果

1. 分散经济均衡的效率

图 1 的数值模拟结果绘制了分散经济均衡的效率。图 1(1)至图 1(3)、图 1(4)至图 1(7)、图 1(8)的纵轴依次是数据共享程度、劳动力配置、经济总产出水平值在分散经济均衡与社会计划者均衡的差值。该差值越接近数值 0，则分散经济均衡越接近社会计划者最优水平。图像结果与第三部分的理论分析结果一致。随着消费者数据占比 $\lambda^c$ 越高，分散经济均衡的消费者数据共享过多问题先增强后减弱，企业数据共享不足问题和两类数据平均共享不足问题逐渐减弱。图 1(8)表明两类数据共享的经济产出效益先降低后提升，图 1(1)表明消费者数据共享的隐私成本先恶化后缓解。图 1(9)是综合经济产出效益和消费者隐私成本后的社会福利比较，可将命题 1 进一步拓展为模拟结果 1。

**模拟结果 1** 若以社会计划者最优均衡作为比较基准，则分散经济均衡的社会福利效率 $\Omega(*,sp)$ 与消费者数据占比 $\lambda^c$ 呈现 U 形关系。①

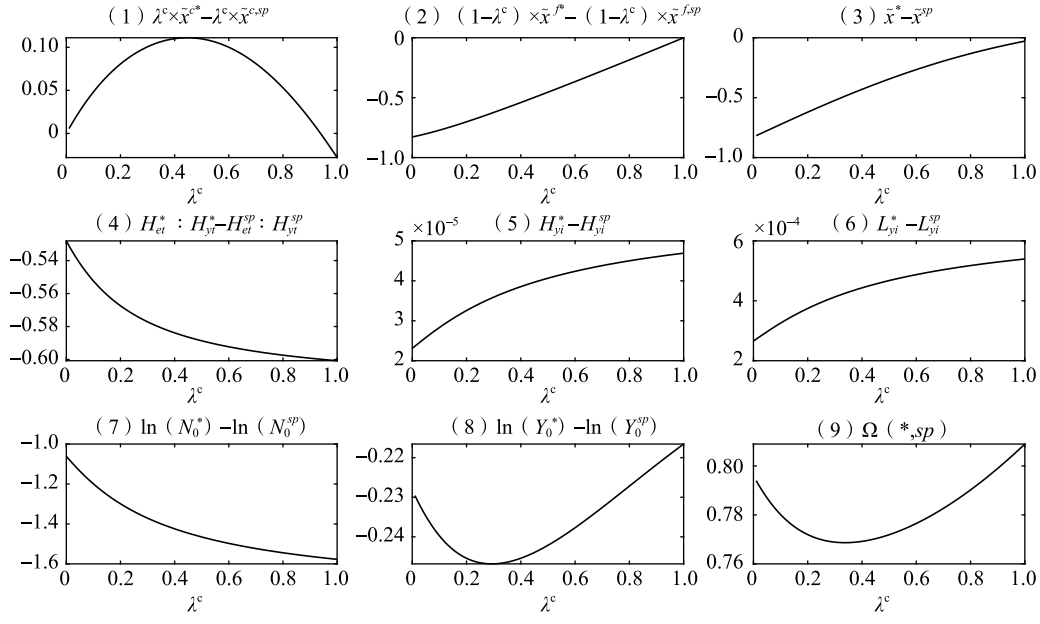



**图 1 分散经济均衡的效率：分散经济均衡与社会计划者均衡的比较**

2. 数据脱敏保护政策与分散经济均衡效率

数据脱敏保护政策旨在保障消费者和企业的信息安全，在确保数据可用性的同时降低消费者隐私泄露风险和企业被创造性破坏风险，从而降低两类数据的共享成本。图 2 和图 3 分别绘制了数据共享成本强度参数 $\tilde{\kappa}$ 和参数 $\delta$ 的变化如何影响分散经济均衡的

① 正文理论模型中假设在位企业被新企业取代的概率只与企业数据有关。但是，借鉴 Jones and Tonetti (2020)的思路，一种可拓展的合理假设是，在位企业被新企业取代的概率同时与企业数据泄露和消费者数据泄露有关。据此，附录Ⅴ将对原假设进行修改，令新假设是：每期在位企业被新企业取代的概率服从泊松到达率 $\delta\cdot[\gamma\cdot\lambda^c\tilde{x}_{it}^{c}+(1-\lambda^c)\tilde{x}_{it}^{f}]^2$。新增引入参数 $\gamma$ 满足取值范围 $0\leqslant\gamma\leqslant1$，表示消费者数据泄露对企业被创造性破坏的作用力度，应当不高于企业数据泄露的作用力度。附录Ⅴ的数值模拟结果依然支持本文结论：若以社会计划者最优均衡作为比较基准，则实施该数据产权制度时分散经济均衡的社会福利效率与消费者数据占比呈现非线性关系。

中国知网 https://www.cnki.net

效率，实线、短虚线、长虚线分别对应着成本强度参数的基准值、较低值、较高值。

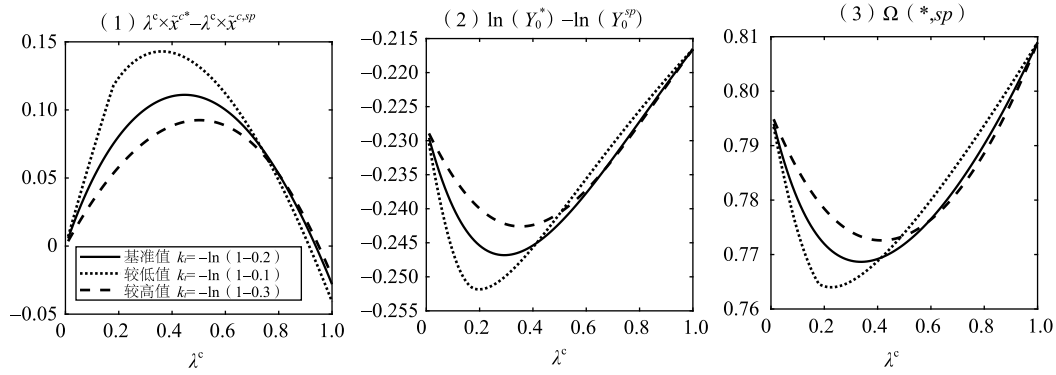



**图 2　消费者数据脱敏保护与分散经济均衡效率**

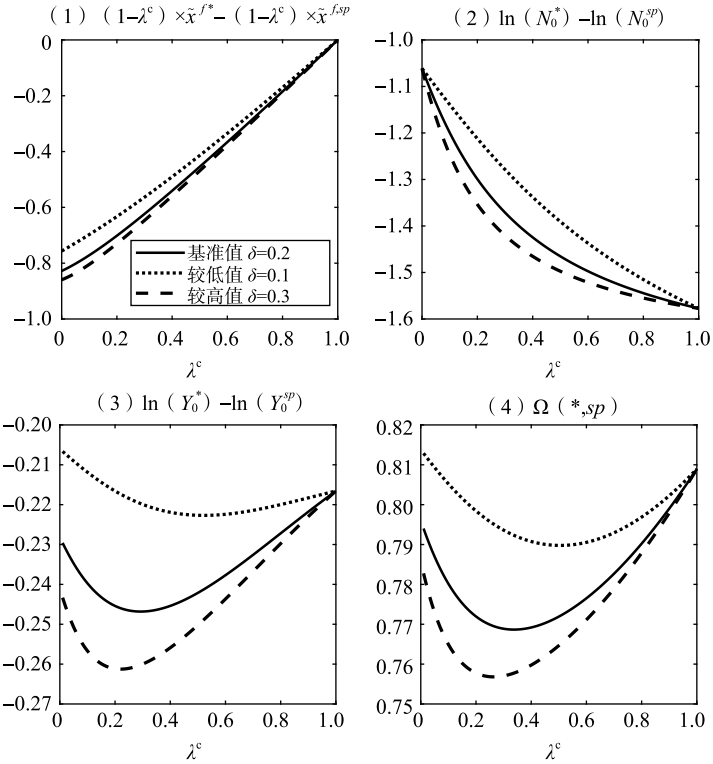



**图 3　企业数据脱敏保护与分散经济均衡效率**

分析图 2 可得，当消费者数据加强脱敏保护（参数 $\tilde{\kappa}$ 越小）时，一方面，消费者隐私泄露影响效用的负向作用强度越低，这直接改善社会福利水平。另一方面，消费者共享私有数据的意愿加强，在原本已过度共享私有数据的基础上可能进一步偏离社会计划者最优水平（图 2(1)）。综合来看，当 $\lambda^c$ 较大（较小）时，分散经济均衡的社会福利效率与消费者数据脱敏保护力度呈现正（负）相关关系（图 2(3)）。分析图 3 可得，当企业数据加强脱敏保护（参数 $\delta$ 越小）时，企业共享私有数据的意愿加强，在原本共享不足的基础上更加接近社会计划者最优水平（图 3(1)），促进产品种类数量增多、提高经济总产出水平。故分散经济均衡的社会福利效率与企业数据脱敏保护力度呈现正相关关系（图 3(4)）。

中国知网　https://www.cnki.net

3. 数据交易补贴政策与分散经济均衡效率

此处将在第二部分的理论模型基础上新增数据交易补贴和政府收支平衡条件。假设政府将对数据中介向企业出售社会数据池的过程进行价格补贴，消费者数据池和企业数据池的交易补贴比例分别为 $s_D^c$ 和 $s_D^f$。同时，假设政府向产品企业的产品销售过程征收 $\tau_y$ 比例的税收，用于平衡财政预算和支付数据交易补贴。政府收支平衡条件为：

$$s_D^c \cdot \tilde{p}_{Dt}^{mc} \cdot \int_0^{N_t} \tilde{D}_t^c di + s_D^f \cdot \tilde{p}_{Dt}^{mf} \cdot \int_0^{N_t} \tilde{D}_t^f di = \int_0^{N_t} \tau_Y \cdot \begin{bmatrix} p_{Dit}^c \cdot \lambda^c x_{it}^c + \\ \tilde{p}_{Dit}^c \cdot \lambda^c \tilde{x}_{it}^c + (Y_t / Y_{it})^{1/\sigma} \end{bmatrix} \cdot Y_{it} di.$$

基于新增数据交易补贴后的分散经济均衡结果[①]，图 4 和图 5 分别绘制了消费者数据交易补贴比例 $s_D^c$、企业数据交易补贴比例 $s_D^f$ 的变化如何影响分散经济均衡的效率，短虚线、实线、长虚线分别对应着补贴比例 10%、20%和 30%。

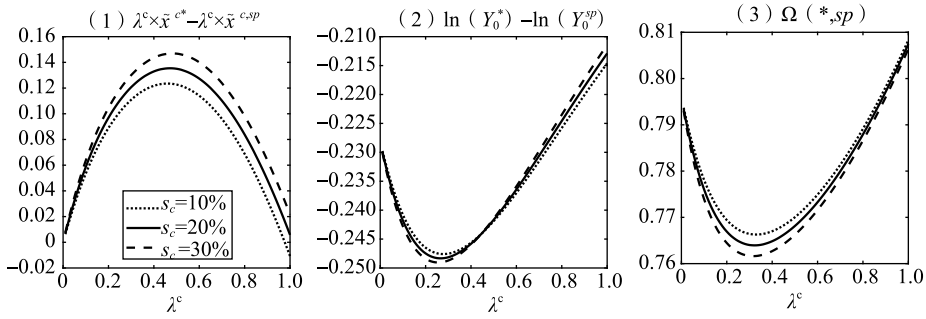



**图 4 消费者数据交易补贴与分散经济均衡效率**

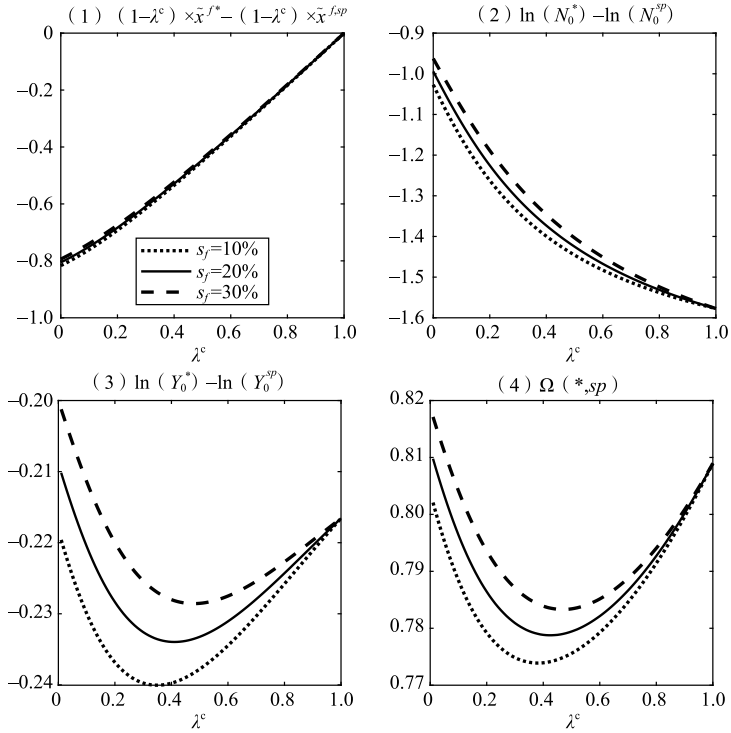



**图 5 企业数据交易补贴与分散经济均衡效率**

① 在新增数据交易补贴后，分散经济均衡的数理推导过程和完整结果参见附录Ⅰ。

中国知网 https://www.cnki.net

分析图 4 可得，当消费者数据交易的补贴力度 $s_D^c$ 越大时，这越激励消费者共享私有数据，进一步加剧过度共享问题（图 4(1)）。不同于数据脱敏保护政策还可以直接优化消费者效用（图 2 的分析），此处分散经济均衡的社会福利效率与 $s_D^c$ 呈现负相关关系（图 4(3)）。分析图 5 可得，当企业数据交易的补贴力度 $s_D^f$ 越大时，这越激励企业共享私有数据，改善共享不足问题。分散经济均衡的社会福利效率与 $s_D^f$ 呈现正相关关系（图 5(4)）。

此外，一方面，对比图 2(3)和图 3(4)、对比图 4(3)和图 5(4)可知，对企业数据进行脱敏保护或交易补贴的社会福利提升效果将高于消费者数据。另一方面，观察图 3(4)和图 5(4)易得，当消费者数据和企业数据比例相近（即 $\lambda^c \in [0,1]$ 的中间区域）时，分散经济均衡的社会福利效率处于低值水平，在此中间区域叠加配套政策的提升效果更显著。

结合图 2 至图 5 的分析，可得到模拟结果 2。

**模拟结果 2**　数据脱敏保护政策和企业数据交易补贴政策将有助于提升数据产权制度的实施效果，且在消费者和企业的数据来源比例相近时叠加配套政策的提升效果更显著。

## 五、政策建议和未来研究方向

本文对数据要素市场建设提出如下政策建议：第一，加快建立和完善数据产权制度的相关法律法规，明确数据初始产权在消费者和企业等不同经济主体之间的分配规则。考虑基于行业内各经济主体的数据来源结构特征来确定初始产权分配，同时推动发展对各行各业的数据来源结构的测算技术。第二，继续完善消费者信息和企业信息的安全监管制度，在理论与实践中探索实现有效筛选生产与消费信息、兼顾信息安全保护的数据提取模式，深度开发“原始数据不出域、数据可用不可见、数据可控可计量”的交易模式。第三，重点关注消费者和企业数据来源比例相近的行业，优先配套数据交易补贴政策；重点关注企业数据共享严重不足的行业，优先配套鼓励企业共享的数据交易补贴政策。

值得继续研究的是，由于现实中企业能够利用数字技术系统性地获取消费者生成的数据，发挥数据收集和数据质量提升等支撑数据价值转化的实质贡献，这使得企业主张对消费者数据享有部分所有权可能具有一定的合理性。因此，可将本文研究拓展为更一般化的消费者和企业共同拥有数据产权（也即考虑企业拥有部分的消费者数据），亦可以进一步在数字经济各发展阶段下对比不同产权分配制度的社会福利表现。

## 参 考 文 献

[1] Abowd, J. M., and I. M. Schmutte, “An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices”, *American Economic Review*, 2019, 109(1), 171-202.

[2] Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar, “Too Much Data: Prices and Inefficiencies in Data Markets”, *American Economic Journal: Microeconomics*, 2022, 14(4), 218-256.

中国知网　https://www.cnki.net

[3] Acquisti, A., C. Taylor, and L. Wagman, "The Economics of Privacy", *Journal of Economic Literature*, 2016, 54(2), 442-492.

[4] Anderson, J. E., and E. van Wincoop, "Trade Costs", *Journal of Economic Literature*, 2004, 42(3), 691-751.

[5] Angrist, J. D., "The Economic Returns to Schooling in the West Bank and Gaza Strip", *American Economic Review*, 1995, 85(5), 1065-1087.

[6] Atkeson, A., and A. Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices", *American Economic Review*, 2008, 98(5), 1998-2031.

[7] Autor, D. H., L. F. Katz, and M. S. Kearney, "Trends in US Wage Inequality: Revising the Revisionists", *Review of Economics and Statistics*, 2008, 90(2), 300-323.

[8] Babina, T., A. Fedyk, A. He, and J. Hodson, "Artificial Intelligence, Firm Growth, and Product Innovation", *Journal of Financial Economics*, 2024a, 151, 103745.

[9] Babina, T., S. A. Bahaj, G. Buchak, F. De Marco, A. K. Foulis, W. Gornall, F. Mazzola, and T. Yu, "Customer Data Access and Fintech Entry: Early Evidence from Open Banking", NBER Working Paper, w32089, 2024b.

[10] Bergemann, D., A. Bonatti, and T. Gan, "The Economics of Social Data", *RAND Journal of Economics*, 2022, 53(2), 263-296.

[11] 蔡继明、刘媛、高宏、陈臣，"数据要素参与价值创造的途径——基于广义价值论的一般均衡分析"，《管理世界》，2022 年第 7 期，第 108—121 页。

[12] 程华、武珺璠、李三希，"数据交易与数据垄断：基于个性化定价视角"，《世界经济》，2023 年第 3 期，第 154—178 页。

[13] Cong, L. W., D. Xie, and L. Zhang, "Knowledge Accumulation, Privacy, and Growth in a Data Economy", *Management Science*, 2021, 67(10), 6480-6492.

[14] Edmond, C., V. Midrigan, and D. Y. Xu, "Competition, Markups, and the Gains from International Trade", *American Economic Review*, 2015, 105(10), 3183-3221.

[15] Farboodi, M., R. Mihet, T. Philippon, and L. Veldkamp, "Big Data and Firm Dynamics", *AEA Papers and Proceedings*, 2019, 109, 38-42.

[16] Farboodi, M., and L. Veldkamp, "Long-run Growth of Financial Data Technology", *American Economic Review*, 2020, 110(8), 2485-2523.

[17] Farboodi, M., and L. Veldkamp, "A Model of the Data Economy", NBER Working Paper, w28427, 2022.

[18] Gaubert, C., and O. Itskhoki, "Granular Comparative Advantage", *Journal of Political Economy*, 2021, 129(3), 871-939.

[19] 龚强、班铭媛、刘冲，"数据交易之悖论与突破：不完全契约视角"，《经济研究》，2022 年第 7 期，第 172—188 页。

[20] 郭凯明、王钰冰、杭静，"数据要素规模效应、产业结构转型与生产率提升"，《中国工业经济》，2024 年第 8 期，第 5—23 页。

[21] He, Z., J. Huang, and J. Zhou, "Open Banking: Credit Market Competition When Borrowers Own the Data", *Journal of Financial Economics*, 2023, 147(2), 449-474.

[22] Jones, C. I., and C. Tonetti, "Nonrivalry and the Economics of Data", *American Economic Review*, 2020, 110(9), 2819-2858.

[23] Katz, L. F., and K. M. Murphy, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", *Quarterly Journal of Economics*, 1992, 107(1), 35-78.

[24] Krusell, P., L. E. Ohanian, J. V. Ríos-Rull, and G. L. Violante, "Capital Skill Complementarity and Inequality: A Macroeconomic Analysis", *Econometrica*, 2000, 68(5), 1029-1053.

中国知网 https://www.cnki.net

[25] 李三希、王泰茗、刘小鲁，“数据投资、数据共享与数据产权分配”，《经济研究》，2023 年第 7 期，第 139—155 页。
[26] 刘涛雄、李若菲、戎珂，“基于生成场景的数据确权理论与分级授权”，《管理世界》，2023 年第 2 期，第 22—39 页。
[27] 刘涛雄、张亚迪、戎珂、周迪，“数据要素成为中国经济增长新动能的机制探析”，《经济研究》，2024 年第 10 期，第 19—36 页。
[28] Miller, A. R., and C. Tucker, “Privacy Protection, Personalized Medicine, and Genetic Testing”, *Management Science*, 2017, 64(10), 4648-4668.
[29] Parlour, C. A., U. Rajan, and H. Zhu, “When FinTech Competes for Payment Flows”, *Review of Financial Studies*, 2022, 35(11), 4985-5024.
[30] Varian, H. R., “Artificial Intelligence, Economics, and Industrial Organization”, NBER Working Paper, w24839, 2018.
[31] Veldkamp, L., and C. Chung, “Data and the Aggregate Economy”, *Journal of Economic Literature*, 2024, 62(2), 458-484.
[32] 徐恺、彭晴、李杰，“中国城市间不同技能劳动力的分布合理吗？——基于高、低技能劳动力互补的视角”，《经济学》(季刊)，2024 年第 2 期，第 554—569 页。
[33] 徐翔、赵墨非、李涛、李帅臻，“数据要素与企业创新：基于研发竞争的视角”，《经济研究》，2023 年第 2 期，第 39—56 页。
[34] 杨俊、李小明、黄守军，“大数据、技术进步与经济增长——大数据作为生产要素的一个内生增长理论”，《经济研究》，2022 年第 4 期，第 103—119 页。
[35] 张新宝，“论作为新型财产权的数据财产权”，《中国社会科学》，2023 年第 4 期，第 144—163 页。

# Initial Allocation of Data Property Rights, Data Sharing and Social Welfare: Based on the Principle of “Who Inputs, Who Contributes, and Who Benefits”

ZHU Dandan
(Sun Yat-sen University)
ZHENG Zhijie
(Beijing Normal University)
PAN Shiyuan*
(Zhejiang University)
HE Yiyao
(Hangzhou Normal University)

**Abstract**: A dynamic general equilibrium model is developed to explore how a data property rights system, where initial rights are allocated according to the ratio of consumer-generated and firm-generated

* Corresponding Author: PAN Shiyuan, Center for Research of Private Economy, Zhejiang University, No. 866 Yuhangtang Road, Xihu District, Hangzhou, Zhejiang 310058; Tel: 86-571-88981985; E-mail: shiyuanpan@zju.edu.cn.

中国知网　https://www.cnki.net

data, affects the level of social welfare. Data composition is found to affect the sharing degrees for both types of private data, ultimately affecting social welfare in the trade-off between economic output gains and consumer privacy costs. Compared to the social planner's optimum, the efficiency of the decentralized equilibrium exhibits a non-linear relationship with the share of consumer-generated data. And data anonymization policies and firm-side data transaction subsidies enhance the effectiveness of this property rights system.

**Keywords**: data property rights; data sharing; social welfare

**JEL Classification**: C80, D60, K11

中国知网 https://www.cnki.net