

平台竞争中的数据量决策研究

余文诗 乔岳*

摘要: 数据成为重要的生产要素,数据量决策权的归属目前却并不明朗。本文构建了双边市场中平台收集消费者数据的双寡头竞争模型,分析当数据量分别由平台和消费者决定时的均衡结果,并比较了由不同主体决定的均衡数据量以及各个群体的福利。研究发现:当隐私泄露风险较高时,消费者愿意提供的数据量往往低于平台想要的的数据量;从社会福利的角度出发,由平台决定数据量是更好的选择;当由平台决定数据量时,消费者剩余、厂商剩余和社会总福利都随着隐私泄露风险的增加而降低,但如果赋予消费者决定数据量的权利,那么隐私泄露风险对用户剩余和福利的影响并不是单调的。本文的结论可以为数据确权和个人信息保护的相关政策制定提供理论参考。

关键词: 平台竞争;数据收集;隐私泄露;福利分析

中图分类号: F270

一、引言

数字经济的显著特征是数据成为与土地、劳动力、资本和技术同等重要的生产要素。在数字经济背景下,数据不再只是一种战略资源,更能作为生产函数中的要素,对经济产生广泛而深远的影响(王超贤等,2022)。近年来,国家已出台政策文件,如《中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见》《国务院办公厅关于印发要素市场化配置综合改革试点总体方案的通知》《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》等,多次强调要促进数据要素流通,积极探索数据要素参与价值创造与分配的最优配置,数据要素对数字经济发展的的重要性不言自明。数据作为生产要素,在个人、企业和政府等不同主体之间的配置会对经济效率产生不同的影响(陈永伟,2018)。根据科斯定理,要素权属不清会影响整个要素市场的资源配置效率,因此,数据确权是数据要素在市场上进行流通和交易的基础,也是保障数据要素市场健全运行亟需解决的问题。

平台企业在数字经济发展历程中占据非常重要的地位,大数据分析、机器学习和人工智能等先进数字技术的快速发展赋予了平台企业收集、存储和利用海量用户数据的能力。不同于传统单边企业,平台企业具有更强的规模经济效应,双边市场所特有的交叉网络外部性

*余文诗,山东大学经济学院,邮政编码:250100,电子信箱:yuwenshi0707@163.com;乔岳(通讯作者),山东大学国际创新转化学院,邮政编码:266237,电子信箱:qiaoyue@sdu.edu.cn。

本文得到山东省自然科学基金面上项目“数字平台纵向限制的竞争效应研究”(ZR2023MG030)的资助。感谢匿名评审专家的宝贵意见,作者文责自负。

以及价格结构非中性特征对平台企业的运营产生复杂的影响机制(李三希、黄卓,2022),然而鲜有学者在双边市场的理论框架下研究数据确权问题。如果平台企业能够控制用户数据,可能会构筑数据壁垒遏制新平台的进入,滥用市场势力从而导致反竞争行为的产生,因此有必要针对平台企业研究数据确权问题。

在数字经济发展的初期,由于人们隐私意识薄弱以及缺乏相关法律法规的监管,互联网平台在用户不知情的情况下已经收集了大量的用户数据,用户隐私泄露的事件时有发生。然而,随着政府对数据安全和隐私保护监管的加强,各个 APP 和网站等都在陆续发布关于用户协议与隐私政策变更的通知。特别地,《中华人民共和国个人信息保护法》规定个人信息处理者在处理个人信息的各个环节应取得个人的单独同意,明确个人有权限制他人对个人信息的处理,赋予用户是否提供个人数据的决定权。合理界定个人数据上的权利对于人权保护和数字经济健康发展都至关重要(程啸,2018)。基于上述背景,本文试图回答以下几个问题:如果用户有权决定是否提供自己的数据,他们会愿意提供更多数据吗?把数据量决策权交还给用户是否比平台控制数据更好?政府应当如何保护消费者的个人信息?

本文以双边市场为框架,通过建立双寡头平台竞争的 Hotelling 模型,分别计算了由平台决定数据量时和由消费者决定数据量时的均衡结果,分析隐私泄露风险变化对均衡结果的影响,并比较了这两种情形下的均衡数据量和各个群体的福利。研究发现,在隐私泄露风险较高的条件下,由平台决定数据量时均衡数据量更多;在隐私泄露风险较低的条件,由消费者决定数据量时均衡数据量更多。无论平台的收费模式和数据收集模式如何,社会总福利在平台决定数据量的情形下都是更高的,因此从社会福利的角度出发,由平台决定数据量是更好的选择。本文还发现,当隐私泄露风险超过一定阈值时,随着隐私泄露风险的进一步增加,数据量、用户剩余和社会总福利都会降低,因此为了获取更多数据要素,政府应当充分保护个人信息,降低隐私泄露风险。

本文剩余内容安排如下:第二部分是文献综述,第三部分是理论模型的建立,第四部分是两种情形下(由平台决定数据量和由消费者决定数据量)的模型求解,第五部分是均衡结果的比较,第六部分进行模型拓展,第七部分对全文进行总结并提出相关政策建议。

二、文献综述

关于平台竞争的研究历史可以追溯到 Rochet 和 Tirole(2003)、Caillaud 和 Jullien(2003)以及 Armstrong(2006)等文献。在数字经济的背景下,平台企业的快速发展显露了数据在商业策略中的核心地位,平台竞争及其数据驱动的商业模式一直是学术研究的热点。对于企业而言,用户数据具有巨大的商业价值。例如,企业可以利用数据实施价格歧视(Chen and Zhang,2009;Ichihashi,2020;刘坤、喻玲,2022),推送定向广告(De Cornière and De Nijs,2016;Johnson,2013),甚至构建信息壁垒巩固市场地位(汪敏达、李建标,2022)。近年来,有学者通过理论模型研究企业的数据收集策略,数据量成为企业决策的关键变量。Bloch 和 Demange(2018)发现垄断平台的最优数据收集策略是选择合适的数据量从而实现市场覆盖,或者选择收集最多的数据从而使隐私成本较高的用户退出。Choi 等(2019)则指出,垄断市场因数据外部性而导致数据过度收集。Fainmesser 等(2023)发现企业的数据收集策略与收入来源有关,数据驱动型企业会比用户驱动型企业收集更多的数据。Dimakopoulos 和 Sudaric(2018)考虑通过收集数据提高广告定向精度并给用户带来隐私成本时的平台竞争,

发现均衡时平台收集的数据量可能偏离社会最优,依赖于市场的竞争强度和广告的边际价值。王泰茗等(2024)则强调数据用益权对数据收集效率及社会福利的影响。区别于上述文献只考虑了企业的数据收集策略,本文同时分析并比较了当平台决定数据量时和消费者决定数据量时的数据收集策略。

数据一方面可以创造价值,另一方面又会造成隐私侵犯,因此关于个人信息保护的经济后果,学者之间也存在不同的观点(Acquisti et al., 2016)。以 Stigler(1980)、Posner(1981)为代表的学者指出,信息自由流动能提高交易效率,因此隐私保护政策将会使资源配置效率和社会福利都降低。Abowd 和 Schmutte(2019)认为最佳的资源配置方案应该使得增加隐私保护的边际成本等于其边际收益。Taylor 和 Wagman(2014)用四种常见的寡头模型比较了有无隐私保护对企业、消费者和社会福利的影响,发现均衡结果取决于所考虑的特定参数与模型设定。Shy 和 Stenbacka(2016)也研究不同程度的隐私保护对各个群体福利的影响,结果发现弱隐私保护时企业的利润高于强隐私保护或没有隐私保护的企业。李三希等(2021)研究发现一旦引入竞争,消费者总剩余和社会总福利在无个人信息保护时达到最大。实证研究则表明,隐私保护可提升消费者福利(Tsai et al., 2011;汪敏达等,2022)。

由于隐私保护带来的经济后果比较复杂,讨论隐私数据的最优配置和数据确权问题是有必要的。近些年,关于数据权属问题的学术研究已有较为丰富的成果。Jones 和 Tonetti(2020)将数据要素纳入宏观经济增长模型中,发现将数据产权赋予消费者可以导致接近社会最优的配置。Ali 等(2022)发现在竞争市场环境下,消费者控制披露数据的权利总是能改善消费者福利。然而,Acemoglu 等(2022)、Bergemann 等(2022)却得出相反的结论,他们认为让消费者控制自己的数据并不利于效率提升。另有一些学者认为数据权利的最优配置并不是确定的。Dosis 和 Sand-Zantman(2023)认为最优的数据确权取决于数据的价值。李三希等(2023)研究发现把数据产权分配给用户会抑制数据收集者进行数据投资,而把数据产权分配给数据投资者则会阻碍数据共享,需要合理运用政府规制来兼顾二者。Li 和 Zhang(2024)分析了由卖家或买家决定信息收集量的不同影响,研究发现,除非信息的外部价值足够大,否则由买家决定信息收集量可以增加消费者剩余和总体福利。与 Li 和 Zhang(2024)不同,本文在双边市场的框架下讨论数据量应该由消费者还是平台决定。

与以往研究相比,本文的创新之处主要有三点:首先,本文的模型基于双边市场,考虑了由于平台利用数据提高匹配效率,使得厂商与消费者之间的交叉网络外部性增大,即假设交叉网络外部性与数据量正相关;其次,本文探讨的数据量决策权属于数据确权广阔议题中的一个细分领域,引入了消费者的策略行为,即消费者可以决定提供多少数据量,并将之与由平台决定数据量时的情形进行对比,讨论数据量决策权的最优配置;最后,以往基于微观层面上关于数据权属的研究多局限于对已产生的个人数据进行产权分配,而在本文中数据量是内生决定的。

三、模型建立

本文在 Armstrong(2006)模型中引入数据要素,探讨由不同市场主体(平台和消费者)决定数据量时的福利影响。考虑双边市场中存在三类主体,分别是平台、厂商和消费者,下面将依次刻画他们的特征。

(一) 平台

假设有两个对称的平台分别位于线性市场 $[0,1]$ 的两个端点上,平台1的位置为 $y_1=0$,平台2的位置为 $y_2=1$ 。厂商和消费者分别处于平台的两边,平台给厂商和消费者提供交易的场所并向这两个群体收取费用^①,目标是最大化自己的利润。平台 i 对厂商和消费者的定价分别为 p_i^F 和 p_i^C ($i \in \{1,2\}$,上标 F 和 C 分别代表厂商和消费者)。当消费者加入平台后,平台会收集消费者的相关数据(例如购物偏好、浏览记录等)^②,这些数据可以提高厂商和消费者的匹配效率从而促成更多交易。例如,电商平台可以根据消费者的浏览记录和交易记录给消费者推送更符合其偏好的商品,从而提高消费者的购买概率。为了利用消费者数据,平台需要付出一定的成本。实际上,无论是数据的挖掘、存储和分析,还是人员培训、后期维护等,平台都需要投入大量的成本。假设平台处理和利用数据的成本函数为 $f(d) = \frac{\alpha}{2}d^2$,其中 $\alpha > 0$ 代表平台处理数据的效率, α 越大说明平台的效率越低^③;该成本函数满足 $f'(d) > 0$ 和 $f''(d) > 0$,这意味着平台处理数据的成本是单调递增的,且随着数据的增多而边际增加(王申等,2022)。由于平台处理数据的效率不是本文研究的重点,为简化分析,以下假设 $\alpha = 1$ 。 n_i^F 和 n_i^C 表示平台 i 上厂商和消费者的数量,当平台 i 向每个消费者收集的数据量是 d_i 时,收集的总数据量为 $n_i^C d_i$ 。为了便于分析,暂不考虑平台提供服务的成本,于是平台 i 的利润为:

$$\pi_i = n_i^F p_i^F + n_i^C p_i^C - \frac{(n_i^C d_i)^2}{2} \quad (1)$$

(1)式中: $n_i^F p_i^F$ 为平台 i 从厂商群体获得的收益, $n_i^C p_i^C$ 为平台 i 从消费者群体获得的收益, $\frac{(n_i^C d_i)^2}{2}$ 为平台 i 收集和處理数据 $n_i^C d_i$ 所需的成本。

(二) 厂商

厂商的总数标准化为1,均匀分布在线性市场 $[0,1]$ 上,每个厂商只加入一个平台^④。位置为 x^F 的厂商加入平台 i 获得的效用为:

$$u_i^F = \theta^F - p_i^F - t|x^F - y_i| + (1+d_i)\lambda n_i^C \quad (2)$$

(2)式中:右边第一项外生给定,表示厂商加入平台获得的基础保留效用,本文假设其足够大以保证所有厂商都加入平台;第二项为平台 i 对厂商的定价;第三项表示由于厂商自身的偏好与平台提供的服务存在差异而造成的效用损失,参数 $t > 0$ 是Hotelling模型中的单位交通

①现实中平台向商家收取佣金是很常见的现象。虽然很多平台会向消费者免费提供服务,但是平台通过各种增值服务(跳过广告、专属客服等会员定制服务)来获取收益也屡见不鲜。本文的拓展部分会讨论平台只向一边用户收费的情形。

②对于交易性平台来说,消费者的个人信息相对更为重要。本文的拓展部分会讨论平台同时收集两边用户数据的情形。

③实际上,平台进行与数据相关的一系列活动(收集、存储、处理等)还需要一定的固定成本,但忽略固定成本并不影响本文的结论。

④现实中商家可以选择加入多个平台,但是平台可能通过激励性或惩罚性手段实施独家交易改变商家的归属感。为专注于分析数据权属问题,本文假定厂商都是单归属。单归属假设在双边市场文献中被广泛采用,例如Dimakopoulos和Sudanic(2018)、Karle等(2020)、Liu等(2021),等等。

成本,也可以理解为平台间的差异化系数;第四项衡量厂商从消费者群体获得的交叉网络外部性,其中 $(1+d_i)\lambda$ 为消费者对厂商的交叉网络外部性系数,本文假设 $\lambda>0$,表示平台上的消费者越多,交易成功的可能性越大,给厂商带来的效用就越大。与Armstrong(2006)不同,本文假设交叉网络外部性系数由数据量内生决定,厂商拥有的数据越多,交叉网络外部性系数越大,这主要是因为平台可以利用数据提高匹配效率。

(三) 消费者

数量标准化为1的消费者均匀分布在线性市场 $[0,1]$ 上,同样也是单归属^①。位置为 x^c 的消费者加入平台 i 获得的效用为:

$$u_i^c = \theta^c - p_i^c - t|x^c - y_i| + (1+d_i)\lambda n_i^F - \beta d_i^2 \quad (3)$$

(3)式中:右边第一项表示消费者加入平台获得的基础保留效用,本文假设其足够大以保证所有消费者都会加入平台;第二项为平台 i 对消费者的定价;第三项衡量由于消费者自身的偏好与平台提供的服务存在差异而造成的效用损失;第四项表示消费者从厂商群体获得的交叉网络外部性,其中 $(1+d_i)\lambda$ 是厂商对消费者的交叉网络外部性系数,同样地,更多的数据有助于提高交易成功的概率,从而增加交叉网络外部性系数^②;第五项为消费者的隐私成本。实际上,平台收集数据对于消费者来说是一把“双刃剑”:一方面,通过数据分析用户偏好,平台可以为其提供更精准的服务,提升用户体验;另一方面,平台拥有大量数据可能会损害消费者利益,比如“大数据杀熟”,或者将用户数据卖给第三方,导致电信诈骗、密码“撞库”等风险。隐私成本 βd_i^2 衡量隐私泄露给消费者带来的效用损失,其中 $\beta>0$ 既可以解释为消费者的隐私泄露风险,也可以用来衡量个人信息保护水平, β 越大意味着隐私泄露风险越高,个人信息保护水平越低,数据收集给消费者带来的效用损失就越大。

本文分别考虑由平台决定数据量和由消费者决定数据量两种情形下的均衡结果。当由平台决定数据量时,博弈时序如下:第一阶段,两个平台同时决定对两组用户的定价 $p_i^j(j \in \{F, C\})$ 以及向消费者收集的数据量 d_i ;第二阶段,厂商和消费者选择加入哪个平台。当由消费者决定数据量时,博弈时序如下:第一阶段,两个平台同时决定对两组用户的定价 $p_i^j(j \in \{F, C\})$;第二阶段,厂商和消费者选择加入哪个平台;第三阶段,消费者决定向平台提供的数据量 d_i 。本文的均衡为子博弈完美纳什均衡,采用逆向归纳法进行求解。

四、均衡结果

本部分将分别讨论由平台决定数据量和由消费者决定数据量两种情形下的均衡结果。

(一) 由平台决定数据量

1. 均衡结果求解

在博弈的第二阶段,厂商和消费者通过比较加入两个平台的效用大小来决定加入哪个平台。首先分析厂商决策,存在一个位于临界位置 \hat{x}^F 的厂商加入两个平台获得的效用相同,即 $u_1^F = u_2^F$,由此可得:

^①考虑消费者具有单位需求,每次交易只能在一个平台。

^②一般情况下,也可以假设平台两边用户的交叉网络外部性系数不同,本文假设两边的交叉网络外部性系数相同是为了方便计算,不影响结论。

$$\hat{x}^F = \frac{1}{2} + \frac{(1+d_1)\lambda n_1^C - (1+d_2)\lambda n_2^C - (p_1^F - p_2^F)}{2t} \quad (4)$$

位于 $[0, \hat{x}^F)$ 的厂商加入平台 1, 位于 $(\hat{x}^F, 1]$ 的厂商加入平台 2, 即:

$$\begin{aligned} n_1^F &= \hat{x}^F \\ n_2^F &= 1 - \hat{x}^F \end{aligned} \quad (5)$$

从(4)式中可以发现, 对于给定的消费者数量, 当收集的数据量增加时, 平台能够更有效率地实现厂商和消费者之间的匹配, 从而吸引更多的厂商。

同样地, 利用 $u_1^C = u_2^C$ 可以求得无差异消费者的位置:

$$\hat{x}^C = \frac{1}{2} + \frac{(1+d_1)\lambda n_1^F - (1+d_2)\lambda n_2^F - (p_1^C - p_2^C) - \beta(d_1^2 - d_2^2)}{2t} \quad (6)$$

位于 $[0, \hat{x}^C)$ 的消费者加入平台 1, 位于 $(\hat{x}^C, 1]$ 的消费者加入平台 2, 即:

$$\begin{aligned} n_1^C &= \hat{x}^C \\ n_2^C &= 1 - \hat{x}^C \end{aligned} \quad (7)$$

(6)式表明, 对于给定的厂商数量, 当平台收集的数据量增加时, 加入该平台的消费者数量变化趋势依赖于两个作用相反的因素: 一方面, 数据量增加使得厂商对消费者的交叉网络外部性增加, 从而吸引更多的消费者加入; 另一方面, 数据量增加也导致隐私成本增加, 从而降低消费者的加入意愿。当个人信息保护水平较低时, 数据量增加更有可能使加入平台的消费者数量减少。

由(4)—(7)式可求得两个平台的市场份额 n_i^j , 它是关于定价与数据量的函数, 即 $n_i^j \equiv n_i^j(p_1^F, p_1^C, d_1, p_2^F, p_2^C, d_2)$ 。当由平台决定数据量时, 平台的利润最大化问题是 $\max_{p_i^C, p_i^F, d_i} \pi_i = n_i^F p_i^F + n_i^C p_i^C - \frac{(n_i^C d_i)^2}{2}$ 。因为两个平台是对称的, 本文考虑对称均衡, 即 $p_1^F = p_2^F = p_P^F, p_1^C = p_2^C = p_P^C$ 和 $d_1 = d_2 = d_P$ (其中下标 P 代表由平台决定数据量的情形)。通过求解一阶条件 $\frac{\partial \pi_i}{\partial p_i^C} = 0, \frac{\partial \pi_i}{\partial p_i^F} = 0$

和 $\frac{\partial \pi_i}{\partial d_i} = 0$ 可以得出均衡时两个平台的定价以及收集的数据量^①。

具体地, 均衡数据量为:

$$d_P = \frac{2\lambda}{1+4\beta} \quad (8)$$

为了明确均衡数据量的决定因素, 可将 $\frac{\partial \pi_i}{\partial d_i} = 0$ 改写如下:

$$\underbrace{\frac{1}{4}d_P}_{\text{收集数据带来的边际成本}} + \underbrace{\beta d_P}_{\text{隐私泄露带来的边际成本}} = \underbrace{\frac{1}{2}\lambda}_{\text{匹配效率提高带来的边际收益}} \quad (9)$$

由(9)式可以发现, 对于平台来说, 数据量的增加对平台利润有三个方面的影响: (1) 收

^①作者也检验了对应的二阶条件, 要求 $t^2(1+4\beta)^3 - t(1+4\beta+16\beta^2)\lambda^2 - \lambda^2[64\beta^3 + (1+\lambda)^2 + 16\beta^2(3+4\lambda) + 4\beta(3+6\lambda+4\lambda^2)] > 0$ 且 $(t+4\beta)^2 + t\lambda^2 - \lambda^2(1+4\beta+2\lambda)^2 > 0$ 。

集更多的数据需要支付更多的数据处理成本;(2)数据量增加也会加剧对消费者的隐私侵犯,减少消费者数量,从而减少平台收入;(3)数据量增加使得平台更了解消费者的偏好,能吸引更多的消费者加入,从而增加平台收入。因此,当由平台决定数据量时,平台需要权衡收集数据的成本和收益。

平台对厂商和消费者的定价分别为:

$$\begin{aligned} p_p^F &= t - (1 + d_p)\lambda \\ p_p^C &= t - (1 + d_p)\lambda + \frac{d_p^2}{2} \end{aligned} \quad (10)$$

由(10)式可以发现:(1)平台对厂商的定价等于产品差异化系数减去交叉网络外部性系数,这一点与 Armstrong(2006)的结论相同。但与之不同的是,此时交叉网络外部性系数并不是外生给定的,而是依赖于均衡时平台收集的数据量。如果均衡数据量增加,那么交叉网络外部性增大,平台补贴厂商的动机增强。(2)平台对消费者的定价由三部分组成,其中前两项仍然与 Armstrong(2006)相同,即产品差异化系数减去交叉网络外部性系数,但第三项 $\frac{d_p^2}{2}$ 是新增的,这是因为平台收集数据的总成本 $\frac{(n_i^C d_i)^2}{2}$ 与消费者数量正相关,因此为了节省数据成本,平台有动机向消费者收取更高的费用以使消费者数量减少。换言之,成本节约效应导致平台对消费者的定价更高,并且数据量越大成本节约效应越大。

均衡时两个平台平分市场,即 $n_{ip}^j = \frac{1}{2}$,再将 d_p 、 p_p^F 和 p_p^C 代入平台利润表达式(1)可得均衡时每个平台的利润:

$$\pi_p = t - (1 + d_p)\lambda + \frac{d_p^2}{8} \quad (11)$$

厂商剩余为 $\Pi_p = \int_0^{\hat{x}^F} u_1^F dx^F + \int_{\hat{x}^F}^1 u_2^F dx^F$,其中右边第一项表示加入平台1的厂商剩余,第二项表示加入平台2的厂商剩余。类似地,消费者剩余为 $CS_p = \int_0^{\hat{x}^C} u_1^C dx^C + \int_{\hat{x}^C}^1 u_2^C dx^C$,社会总福利为 $W_p = \Pi_p + CS_p + 2\pi_p$ 。同样将 d_p 、 p_p^F 和 p_p^C 代入各式中即可得到均衡时的厂商剩余、消费者剩余以及社会总福利:

$$\Pi_p = \theta^F - \frac{5}{4}t + \frac{3}{2}(1 + d_p)\lambda \quad (12)$$

$$CS_p = \theta^C - \frac{5}{4}t + \frac{3}{2}(1 + d_p)\lambda - \frac{d_p^2}{2} - \beta d_p^2 \quad (13)$$

$$W_p = \theta^F + \theta^C - \frac{t}{2} + (1 + d_p)\lambda - \frac{d_p^2}{4} - \beta d_p^2 \quad (14)$$

2. 隐私泄露风险对均衡数据量和定价的影响分析

由(8)式可得 $\frac{\partial d_p}{\partial \beta} = -\frac{8\lambda}{(1+4\beta)^2} < 0$ 。显然,从(9)式可以看出,数据泄露风险的增加提高了隐私侵犯成本,从而抑制平台收集数据的动机,因此为了增加可利用的数据量,平台需要提

高个人信息保护水平。通过(10)式中 p_p^F 的表达式可得 $\frac{\partial p_p^F}{\partial \beta} = \frac{8\lambda^2}{(1+4\beta)^2} > 0$, 即平台对厂商的定价随着数据泄露风险的增加而增加, 这是因为隐私泄露风险增加时均衡数据量降低, 进而交叉网络外部性系数变小, 于是平台补贴厂商的动机变小, 平台对厂商的定价增加。通过(10)式中 p_p^C 的表达式可得 $\frac{\partial p_p^C}{\partial \beta} = \frac{8(4\beta-1)\lambda^2}{(1+4\beta)^3}$, 当 $\beta > (<) \frac{1}{4}$ 时, 平台对消费者的定价随着隐私泄露风险的增加而增加(降低)。从前面的分析可知, 当隐私泄露风险增加时, 均衡数据量降低, 而数据量降低会对 p_p^C 产生两个方向相反的作用力: 一方面, 数据量降低导致交叉网络外部性降低, 平台补贴消费者的动机减弱, 从而导致 p_p^C 有增加的趋势; 另一方面, 在成本节约效应的作用下数据量降低时, 平台有动机向消费者收取更低费用, 在刺激需求的同时并不会造成数据处理成本的大幅度增加。当隐私泄露风险较低时(β 较小), 成本节约效应起主导作用, 平台对消费者的定价更低; 反之(β 较大), 交叉网络外部性的影响起主导作用, 平台对消费者的定价更高。

3. 隐私泄露风险对各个群体福利的影响分析

下面分析隐私泄露风险的变化对各个群体福利的影响, 从而明确隐私保护的福利效应。

由上文分析可以发现, 隐私泄露风险的变化主要通过影响均衡数据量, 进而间接影响各群体福利。对于平台来说, 当隐私泄露风险增加时, 数据量减少一方面降低平台自身的数据处理成本, 另一方面成本节约效应使得平台有动机向消费者收取更低费用, 二者相比较下, 成本节约效应占主导作用使得平台利润有降低的趋势; 然而数据量减少导致交叉网络外部性变小使得平台支付给厂商和消费者的补贴减少, 在这几个因素综合作用下最终平台利润随着隐私泄露风险的增加而增加, 即 $\frac{\partial \pi_p}{\partial \beta} = \frac{4(1+8\beta)\lambda^2}{(1+4\beta)^3} > 0$ 。

对于厂商剩余, 容易验证 $\frac{\partial \Pi_p}{\partial \beta} = -\frac{12\lambda^2}{(1+4\beta)^2} < 0$, 这是因为当隐私泄露风险增加时, 数据量减少会降低交叉网络外部性从而降低厂商收益, 同时交叉网络外部性减小也会降低厂商从平台获得的补贴, 因此厂商剩余降低。

对于消费者剩余, 从(13)式可以看出, 右边第一项和第二项分别代表保留效用和交通成本, 第三项反映了交叉网络外部性增加给消费者带来的好处^①, 第四项表示平台利用成本节约效应通过价格将数据处理成本转嫁给消费者, 最后一项是消费者的隐私成本。前文指出, 隐私泄露风险的增加会降低均衡数据量, 这一方面会减弱对消费者的隐私侵犯并减少转嫁给消费者的数据成本, 另一方面也导致消费者无法享受更好的匹配服务和更高的价格补贴, 二者中后者对消费者剩余的影响起主导作用, 因而消费者群体的福利随着隐私泄露风险的增加而降低, 即 $\frac{\partial CS_p}{\partial \beta} = -\frac{32\beta\lambda^2}{(1+4\beta)^3} < 0$ 。该结果表明, 当平台利用消费者数据提高匹配效率、增加交叉网络外部性时, 隐私泄露风险的增加会降低消费者群体的福利, 这是符合直觉的, 作

^①它实际上由两部分组成, 交叉网络外部性增加一方面会直接增加消费者的效用, 即 $n_p^f(1+d_p)\lambda = \frac{1}{2}(1+d_p)\lambda$, 另一方面也会增加消费者获得的补贴, 即 $(1+d_p)\lambda$ 。

用机制也比较明确。实际上,如果只关注数据泄露,隐私泄露风险的增加的确会抑制企业收集过多数据的动机,市场机制本身就可以实现保护消费者隐私的目的。但随之而来的问题是,数据量降低不利于提高平台的个性化服务水平,进而导致消费者无法充分享受数字经济发展带来的好处,这一不利因素可能超过隐私保护带来的益处。因此,如何既能有效发挥数据要素的潜在价值又能充分保护消费者个人隐私,是当前数字经济发展面临的重要问题。

对于社会总福利,其各项含义与消费者剩余类似。同样地,当隐私泄露风险增加时,由于数据量降低导致厂商和消费者无法更好地享受数字经济红利(更高的匹配效率),这一主导因素使得社会福利降低,即 $\frac{\partial W_P}{\partial \beta} = -\frac{4\lambda^2}{(1+4\beta)^2} < 0$ 。

下文的数值模拟会更直观地展示隐私泄露风险的变化对各项均衡结果的影响。基于上述分析,本文得到以下命题。

命题1:当由平台决定数据量时,随着隐私泄露风险的增加,

- (1) 均衡数据量降低;
- (2) 平台对厂商的定价单调递增,平台对消费者的定价先增加后降低;
- (3) 平台利润单调递增,消费者剩余、厂商剩余和社会总福利都单调递减。

在现实中,消费者对个人隐私泄露的担忧是显而易见的。例如目前人脸识别公司收集的数据越来越多,这种大数据技术已经实现在各种场景的运用,消费者开始担心自己的人身和财产安全得不到保障。在数字经济时代,数据具有巨大的经济价值,应该注意防止个人数据被滥用。在人脸识别等隐私信息的收集上,用户需要有知情权,对于个人信息的存储、处理等要安全管理,一旦发生问题要有补救的措施。提高个人信息保护水平是为了防止隐私泄露风险的增加,从而为数字经济的健康发展提供保障。

(二) 由消费者决定数据量

1. 均衡结果求解

在博弈的第三阶段,由消费者决定向平台提供的数据量。消费者根据效用最大化原则来决定 d_i ,通过求解 $\frac{\partial u_i^C}{\partial d_i} = 0$ 可得:

$$d_i = \frac{\lambda n_i^F}{2\beta} \quad (15)$$

将(15)式代入效用(2)、(3)式中,然后通过求解 $u_1^F = u_2^F$ 和 $u_1^C = u_2^C$ 可以得到临界处的厂商与消费者的位置:

$$\begin{cases} \hat{x}^F = \frac{1}{2} + \frac{2\beta(p_2^F - p_1^F + \lambda n_1^C - \lambda n_2^C) + \lambda^2(n_1^C n_1^F - n_2^C n_2^F)}{4t\beta} \\ \hat{x}^C = \frac{1}{2} + \frac{4\beta(p_2^C - p_1^C + \lambda n_1^F - \lambda n_2^F) + \lambda^2(n_1^F - n_2^F)(n_1^F + n_2^F)}{8t\beta} \end{cases} \quad (16)$$

再结合 $n_1^i = \hat{x}^i$ 和 $n_2^i = 1 - \hat{x}^i$ 可以得到 n_i^i 关于定价的关系式,即 $n_i^i \equiv n_i^i(p_1^F, p_1^C, p_2^F, p_2^C)$ 。

此时的平台利润最大化问题是 $\max_{p_1^C, p_1^F} \pi_i = n_i^F p_i^F + n_i^C p_i^C - \frac{(n_i^C d_i)^2}{2}$ 。同样考虑对称均衡,当 $p_1^F = p_2^F = p_1^C = p_2^C = p_C^C$ 时(其中下标 C 代表由消费者决定数据量的情形),通过求解一阶条件

$\frac{\partial \pi_i}{\partial p_i^C} = 0$ 和 $\frac{\partial \pi_i}{\partial p_i^F} = 0$ 可以得出均衡时两个平台的定价,进而也可求得其他均衡结果^①。

在由消费者决定数据量的情形下均衡数据量为:

$$d_c = \frac{\lambda}{4\beta} \quad (17)$$

类似地,为了得出均衡数据量的决定因素,将 $\frac{\partial u_i^C}{\partial d_i} = 0$ 改写如下:

$$\underbrace{2\beta d_c}_{\text{隐私泄露造成的效用降低}} = \underbrace{\frac{1}{2}\lambda}_{\text{匹配效率提高带来的效用增加}} \quad (18)$$

对于消费者来说,数据量的变化对其效用有两个方面的影响:其一,数据量增加加剧了对消费者的隐私侵犯,从而造成效用损失;其二,数据量增加也会提高匹配效率,增加消费者从厂商群体获得的交叉网络外部性,从而增加消费者效用。消费者在决定数据量时需要权衡这两个方面的因素。

平台对厂商和消费者的定价分别为:

$$\begin{aligned} p_c^F &= t - (1 + 2d_c)\lambda + \frac{d_c^2}{2} \\ p_c^C &= t - (1 + d_c)\lambda + \frac{d_c^2}{2} \end{aligned} \quad (19)$$

当由消费者决定数据量时,平台对消费者的定价 p_c^C 从形式上看与平台决定数据量时 p_p^C 相同,而平台对厂商的定价 p_c^F 与 p_p^F 有所差别,这是因为平台在博弈第一阶段定价时需要考虑随后消费者对数据量的选择。从(15)式可以看出消费者选择的数据量与厂商数量正相关,所以平台的数据成本不仅依赖于消费者数量,还依赖于厂商数量。前文指出, p_p^F 等于产品差异化系数减去交叉网络外部性系数。一方面,当平台对厂商定价较高时,厂商对平台服务的需求减少,平台的数据成本随之减少,所以平台有动机向厂商收取较高的费用($\frac{d_c^2}{2}$),这也是成本节约效应;另一方面,平台也有动机对厂商进行补贴($-\lambda d_c$),这样可以增加厂商需求,从而增加平台可获取的数据量,进一步提高交叉网络外部性。

根据(17)式和(19)式可求得均衡时每个平台的利润、厂商剩余、消费者剩余以及社会总福利:

$$\pi_c = t - (1 + \frac{3}{2}d_c)\lambda + \frac{3}{8}d_c^2 \quad (20)$$

$$\Pi_c = \theta^F - \frac{5}{4}t + \frac{3}{2}(1 + \frac{5}{3}d_c)\lambda - \frac{d_c^2}{2} \quad (21)$$

$$CS_c = \theta^C - \frac{5}{4}t + \frac{3}{2}(1 + d_c)\lambda - \frac{d_c^2}{2} - \beta d_c^2 \quad (22)$$

^①此时的二阶条件要求 $1024t^3\beta^4 - 16t^2\beta^2(16\beta - 1)\lambda^2 + \lambda^4(4\beta + \lambda)^2 - 16t\beta\lambda^2[4\beta(4\beta + \lambda)^2 - \lambda(4\beta + \lambda)] > 0$ 且 $4096t^2\beta^4 - 128(32\beta^2 - t + 8\beta t)\beta^2\lambda^2 + 256(1 - 8\beta)\beta^2\lambda^3 - (256\beta^2 - 48\beta + 3)\lambda^4 > 0$ 。

$$W_c = \theta^F + \theta^C - \frac{t}{2} + (1+d_c)\lambda - \frac{d_c^2}{4} - \beta d_c^2 \quad (23)$$

2. 隐私泄露风险对均衡数据量和定价的影响分析

由(17)式可得 $\frac{\partial d_c}{\partial \beta} = -\frac{\lambda}{4\beta^2} < 0$, 这主要是因为当给定数据量时, 隐私泄露风险增加会导致消费者效用损失增加, 从而抑制了消费者提供数据的动机。由(19)式可得 $\frac{\partial p_c^C}{\partial \beta} = \frac{(4\beta-1)\lambda^2}{16\beta^3}$ 和 $\frac{\partial p_c^F}{\partial \beta} = \frac{(8\beta-1)\lambda^2}{16\beta^3}$, 可以发现, 随着隐私泄露风险的增加, 平台对两边用户的定价都先降低后增加, 其原因是隐私泄露风险的变化会影响交叉网络外部性与成本节约效应的相对大小。当隐私泄露风险较小时, 成本节约效应起主导作用, 隐私泄露风险增加使得平台对消费者的定价降低; 反之, 交叉网络外部性的影响起主导作用, 平台补贴消费者的动机降低, 从而收取更高的费用。相比较而言, p_c^F 的递增区间更大, 这是因为交叉网络外部性对 p_c^F 的影响更加显著。

3. 隐私泄露风险对各个群体福利的影响分析

对于平台来说, 其均衡利润可表示为 $\pi_c = \frac{1}{2}(p_c^F + p_c^C) - \frac{d_c^2}{8}$ 。由上述分析可知, 隐私泄露风险的提高对定价的影响是不确定的, 虽然数据成本降低, 但最终平台利润变化方向仍然不确定。隐私泄露风险提高对厂商剩余的影响也不确定, 其变化规律与 p_c^F 密切相关, 另外, 因为均衡数据量降低导致交叉网络外部性减小, 所以厂商剩余降低的参数区间 ($\beta > \frac{1}{10}$) 比 p_c^F 增加的参数区间 ($\beta > \frac{1}{8}$) 更大。同样, 消费者剩余的变化规律与 p_c^C 密切相关, 消费者剩余降低的参数区间 ($\beta > \frac{1}{5}$) 比 p_c^C 增加的参数区间 ($\beta > \frac{1}{4}$) 更大。最后, 对于社会总福利, 隐私泄露风险提高虽然会增加交叉网络外部性, 但也会导致数据成本和隐私成本增加, 因此其变化方向是不确定的^①。

通过分析隐私泄露风险对均衡结果的影响, 本文可得命题 2。

命题 2: 当由消费者决定数据量时, 随着隐私泄露风险的增加,

- (1) 均衡数据量降低;
- (2) 平台对厂商和消费者的定价以及平台利润都先降低后增加;
- (3) 消费者剩余、厂商剩余和社会总福利都先增加后降低。

随着信息技术的不断发展, 个人信息保护逐渐成为广大消费者关心的现实问题。由命题 1 和命题 2 可知, 当隐私泄露风险超过某一阈值时, 随着隐私泄露风险进一步增加, 消费

^①根据各个群体福利的均衡结果表达式可计算 $\frac{\partial \pi_c}{\partial \beta} = \frac{3(8\beta-1)\lambda^2}{64\beta^3}$, $\frac{\partial \Pi_c}{\partial \beta} = \frac{(1-10\beta)\lambda^2}{16\beta^3}$, $\frac{\partial CS_c}{\partial \beta} = \frac{(1-5\beta)\lambda^2}{32\beta^3}$,

$\frac{\partial W_c}{\partial \beta} = \frac{(1-6\beta)\lambda^2}{32\beta^3}$, 由此判断隐私泄露风险对其的影响。

者剩余、厂商剩余和社会总福利都降低,但是平台利润增加。这个结果说明在数字经济时代,保护个人信息避免隐私泄露是必要的,否则会使得平台可利用的数据量过少,从而不利于发挥数据作为基础要素的作用。另外,本文还发现,如果赋予消费者决定提供多少数据量的权利,那么对于个人信息的保护并不是越严格越好。命题2表明,当个人信息保护较为严格时(β 很小),消费者愿意提供的数据量会很大,这一方面导致巨大的隐私泄露成本,另一方面也使得平台对厂商和消费者的定价增加,从而导致消费者效用、厂商剩余以及社会总福利都很低,此时适度的隐私保护才是最优策略。

五、由不同主体决定数据量时的均衡结果比较

(一) 均衡数据量比较

由(8)式和(17)式可得, $d_p - d_c = \frac{(4\beta - 1)\lambda}{4\beta(1 + 4\beta)}$ 。当 $\beta < \frac{1}{4}$ 时,由消费者决定数据量时均衡数据量更高;反之,由平台决定数据量时均衡数据量更高。根据(9)式和(18)式分析其经济学含义:增加边际数据量给平台和消费者带来的边际收益相同($\frac{1}{4}\lambda$),当平台决定数据量时,平台需要支付数据收集和隐私侵犯的边际成本($\frac{1}{4}d_p + \beta d_p$),而当消费者决定数据量时,消费者只需要承担隐私侵犯成本($2\beta d_c$)。当 $\beta < \frac{1}{4}$ 时,隐私侵犯成本较小,因此消费者愿意向平台提供更多的数据;反之,隐私侵犯成本较大,消费者愿意提供的数据较少。

(二) 均衡定价比较

比较(10)式和(19)式可得 $p_p^F - p_c^F = \frac{(12\beta - 1)\lambda^2}{32\beta^2(1 + 4\beta)}$, $p_p^C - p_c^C = -\frac{(1 - 4\beta)^2(1 + 8\beta)\lambda^2}{32\beta^2(1 + 4\beta)^2} \leq 0$ 。前文指出,当消费者决定数据量时,他们愿意向平台供给的数据量与厂商数量正相关,因此 p_c^F 比 p_p^F 多出两项与数据量相关的部分。因为二者的作用方向不同,所以两种情形下平台对厂商的定价也存在不确定性。但是当给定数据量时,均衡定价 $p_l^C (l \in \{P, C\})$ 的决定因素完全相同,即 $p_l^C = t - (1 + d_l)\lambda + \frac{d_l^2}{2}$ 。而由上面的分析可知,当隐私泄露风险较小时($\beta < \frac{1}{4}$), $d_c > d_p$,并且两种情形下的数据量都较大,因此 $\frac{d_l^2}{2}$ 起主要作用, p_c^C 更高;而当隐私泄露风险较大时($\beta > \frac{1}{4}$), $d_c < d_p$,并且两种情形下的数据量都较小,此时 p_c^C 仍然更高,但起主要作用的是 $-(1 + d_l)\lambda$ 。

(三) 平台利润比较

比较(11)式和(20)式可得 $\pi_p - \pi_c = \frac{(-3 + 24\beta + 144\beta^2 - 256\beta^3)\lambda^2}{128\beta^2(1 + 4\beta)^2}$ 。当 $\beta < 0.0868$ 或 $\beta > 0.676$ 时,由消费者决定数据量时平台利润更高;当 $0.0868 < \beta < 0.676$ 时,由平台决定数据量时平台利润更高^①。平台利润的构成包括平台从消费者群体和厂商群体获得的收益,再减去处理数据所需的成本,平台利润的比较需综合考虑这些因素。一方面,消费者端或者厂商端

^① $-3 + 24\beta + 144\beta^2 - 256\beta^3 = 0$ 在 $\beta > 0$ 时有近似解 $\beta \approx 0.0868$ 和 $\beta \approx 0.676$ 。

定价越高,平台来自消费者和厂商的收益增加会提升平台利润;另一方面,数据处理成本随数据量的增加而上升,从而降低平台利润。表1给出了在 β 的不同取值范围内,数据量由不同主体决定时均衡数据量、均衡定价以及平台利润的比较结果。

表1 数据量由不同主体决定时均衡数据量、均衡定价以及平台利润的比较

β 取值范围	两种情形下均衡结果比较			
	消费者端定价	厂商端定价	数据量	平台利润
$\beta > 0.676$	$p_p^C < p_c^C$ ※	$p_p^F > p_c^F$	$d_p > d_c$ ※	$\pi_p < \pi_c$
$\frac{1}{4} < \beta < 0.676$	$p_p^C < p_c^C$	$p_p^F > p_c^F$ ※	$d_p > d_c$	$\pi_p > \pi_c$
$0.0868 < \beta < \frac{1}{4}$	$p_p^C < p_c^C$	$p_p^F > p_c^F$ ※	$d_p < d_c$ ※	$\pi_p > \pi_c$
$\frac{1}{12} < \beta < 0.0868$	$p_p^C < p_c^C$ ※	$p_p^F > p_c^F$	$d_p < d_c$	$\pi_p < \pi_c$
$\beta < \frac{1}{12}$	$p_p^C < p_c^C$ ※	$p_p^F < p_c^F$ ※	$d_p < d_c$	$\pi_p < \pi_c$

注:表中的※号指出对平台利润的比较结果起主导作用的因素。

(四) 厂商剩余比较

比较(12)式和(21)式可得 $\Pi_p - \Pi_c = \frac{(1-16\beta+16\beta^2)\lambda^2}{32\beta^2(1+4\beta)}$ 。当 $\beta > \frac{2+\sqrt{3}}{4}$ 或 $\beta < \frac{2-\sqrt{3}}{4}$ 时,由平台决定数据量时厂商剩余更高;当 $\frac{2-\sqrt{3}}{4} < \beta < \frac{2+\sqrt{3}}{4}$ 时,由消费者决定数据量时厂商剩余更高。

厂商剩余主要受厂商端定价以及交叉网络外部性的影响,厂商端定价越高则厂商剩余越低,而数据量的增加,即交叉网络外部性的增强,则会提升厂商剩余。表2给出了在不同参数范围内,数据量由不同主体决定时的均衡数据量、厂商端定价和厂商剩余的比较结果。

表2 数据量由不同主体决定时均衡数据量、厂商端定价以及厂商剩余的比较

β 取值范围	两种情形下均衡结果比较		
	厂商端定价	数据量	厂商剩余
$\beta > \frac{2+\sqrt{3}}{4}$	$p_p^F > p_c^F$	$d_p > d_c$ ※	$\Pi_p > \Pi_c$
$\frac{1}{4} < \beta < \frac{2+\sqrt{3}}{4}$	$p_p^F > p_c^F$ ※	$d_p > d_c$	$\Pi_p < \Pi_c$
$\frac{1}{12} < \beta < \frac{1}{4}$	$p_p^F > p_c^F$ ※	$d_p < d_c$ ※	$\Pi_p < \Pi_c$
$\frac{2-\sqrt{3}}{4} < \beta < \frac{1}{12}$	$p_p^F < p_c^F$	$d_p < d_c$ ※	$\Pi_p < \Pi_c$
$\beta < \frac{2-\sqrt{3}}{4}$	$p_p^F < p_c^F$ ※	$d_p < d_c$	$\Pi_p > \Pi_c$

注:表中的※号指出对厂商剩余的比较结果起主导作用的因素。

(五) 消费者剩余比较

比较(13)式和(22)式可得 $CS_p - CS_c = \frac{(1-4\beta)^2(1+6\beta)\lambda^2}{32\beta^2(1+4\beta)^2} \geq 0$,由平台决定数据量时消费

者剩余更高。消费者剩余主要是由消费者端定价、交叉网络外部性以及消费者隐私成本决定的,消费者端定价越高消费者剩余越低,交叉网络外部性越大消费者剩余越高,隐私成本越大消费者剩余越低。交叉网络外部性和隐私成本都与数据量相关:数据量越多交叉网络外部性越大,隐私成本也越大。而平台对消费者的定价在平台决定数据量时始终是更低的,综合来看消费者剩余在平台决定数据量的情形下更高。

(六) 社会总福利比较

比较(14)式和(23)式可得 $W_p - W_c = \frac{(1-4\beta)^2 \lambda^2}{64\beta^2(1+4\beta)} \geq 0$, 由平台决定数据量时社会总福利更高。社会总福利主要由交叉网络外部性、平台数据成本以及消费者隐私成本决定,归根到底是由数据量决定的。当 $\beta < \frac{1}{4}$ 时, $d_p < d_c$, 虽然此时在由平台决定数据量的情形下交叉网络外部性更小,但平台数据成本以及消费者隐私成本也更小,并且后者占主导地位,因此社会总福利在由平台决定数据量的情形下更高。当 $\beta > \frac{1}{4}$ 时, $d_p > d_c$, 此时在由平台决定数据量的情形下交叉网络外部性更大,并且是影响社会总福利的主导因素,因此社会总福利在由平台决定数据量的情形下也更高。

为了更直观地展示数据量由不同主体决定时的均衡结果比较,本文选定一些参数值对均衡结果进行数值模拟: $\theta^F = \theta^C = 3, t = 1, \lambda = \frac{1}{2}$, 如图 1 所示。

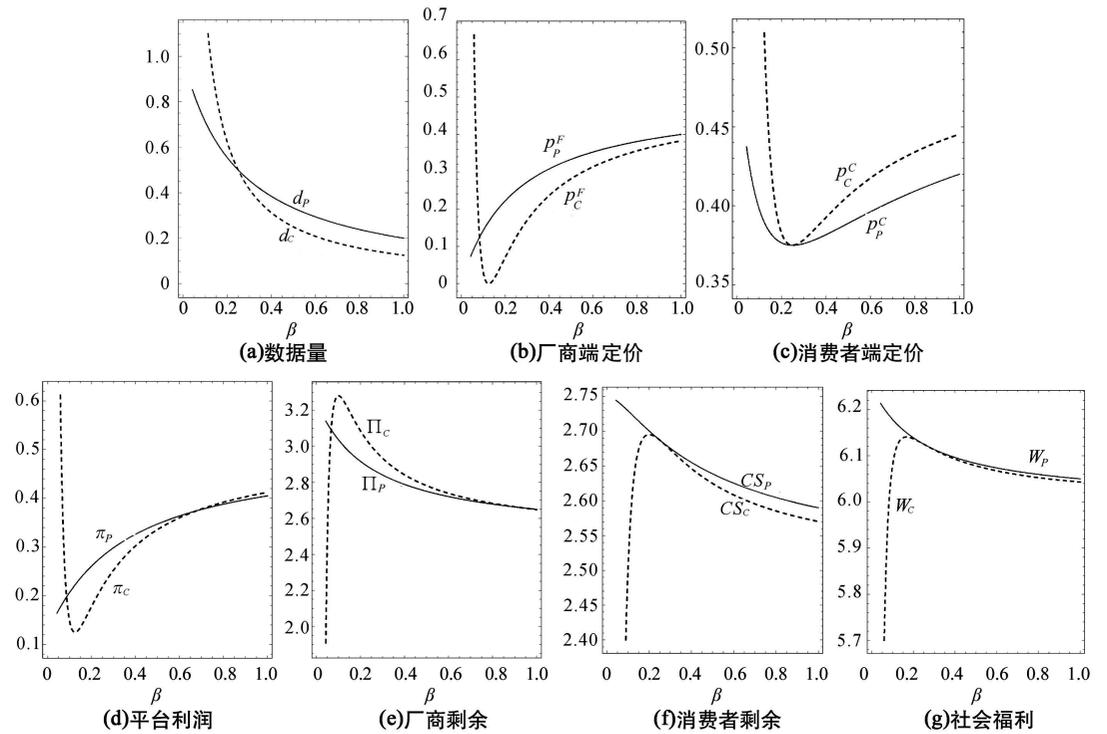


图 1 数据量由不同主体决定时的均衡结果比较

根据以上分析可得命题 3。

命题3:

(1)在隐私泄露风险较高的条件下,由平台决定数据量时均衡数据量更高;在隐私泄露风险较低的条件下,由消费者决定数据量时均衡数据量更高。

(2)在隐私泄露风险很高或者很低的条件下,由平台决定数据量时厂商剩余更高,平台利润更低;在隐私泄露风险相对适中的条件下,由消费者决定数据量时厂商剩余更高,平台利润更低。

(3)无论隐私泄露风险的高低,消费者剩余和社会总福利在由平台决定数据量时都更高。

相较于平台决定数据量,当消费者拥有提供多少数据的决定权时,他们会愿意提供更多的数据吗?从上面的分析可以看出,只有当隐私泄露风险较低时,消费者才愿意提供更多的数据。当隐私泄露风险较高时,消费者愿意提供的数据量往往低于平台想要收集的数据量。那么把数据量决策权交还给消费者是否比企业控制数据更好呢?上文的分析表明,消费者剩余和社会总福利在由平台决定数据量的情形下都优于由消费者决定数据量的情形,因此,从消费者剩余和社会总福利的角度来看,为了保护消费者隐私而将数据量决策权交还给消费者并不是一个更好的选择。而对于厂商和平台而言,数据量决策权交还给消费者是否对他们更有利呢?本文发现结论与隐私泄露风险密切相关,例如在隐私泄露风险较为适中时,由消费者决定数据量时厂商剩余更高,而平台利润更低;而当隐私泄露风险很高或者很低时,由消费者决定数据量时厂商剩余更低,而平台利润更高。这一结论意味着不同经济主体对于个人信息保护政策可能存在不同的偏好,因此,政府在制定个人信息保护相关的法律和政策时,应该明确被保护的目标,并综合考虑各方利益。

六、模型拓展

在基本模型部分,本文假设平台只收集一边用户(消费者)的数据,并且对两边用户都收费。而现实中,不同类型的平台往往会有不同的数据收集模式和不同的收费模式。例如,相亲交友平台需要充分收集两边用户(男方用户和女方用户)的信息以便做出最优匹配,因此平台两边用户的数据同样重要,并且两边用户都会在意个人隐私。另外,现实中的互联网平台通常只对一边用户收费,如消费者可以免费使用搜索引擎,而广告商却需要付费才能在搜索结果界面上展示他们的广告;电商平台往往也只对加入其中的商家收费而对消费者免费。基于此,本部分将考虑平台同时收集两边用户的数据以及平台只向一边用户收费这两种情形,分析本文的核心结论是否会发生变化。

(一)平台收集两边用户的数据

当平台同时收集两边用户的数据并向两边用户收费时,两边用户完全对称,不妨称之为A组用户和B组用户,他们的效用函数是 $u_i^k = \theta - p_i^k - t |x^k - y_i| + (1 + d_i^A + d_i^B) \lambda n_i^m - \beta (d_i^k)^2$, 其中 $k, m \in \{A, B\}, k \neq m$ 。平台利润表达式变为 $\pi_i = n_i^A p_i^A + n_i^B p_i^B - \frac{(n_i^A d_i^A)^2 + (n_i^B d_i^B)^2}{2}$ 。

当由平台决定数据量时,博弈时序如下:第一阶段,两个平台同时决定对两组用户的定价以及数据量;第二阶段,两组用户同时选择加入哪个平台。当由用户决定数据量时,博弈时序如下:第一阶段,两个平台同时决定对两组用户的定价;第二阶段,两组用户同时选择加

入哪个平台;第三阶段,用户决定提供的数据量^①。

同样地,为了更直观地比较平台收集两边用户数据时两种情形下的均衡结果,本文选定 $\theta=3, t=1, \lambda=\frac{1}{2}$ 进行数值模拟,如图 2 所示^②。

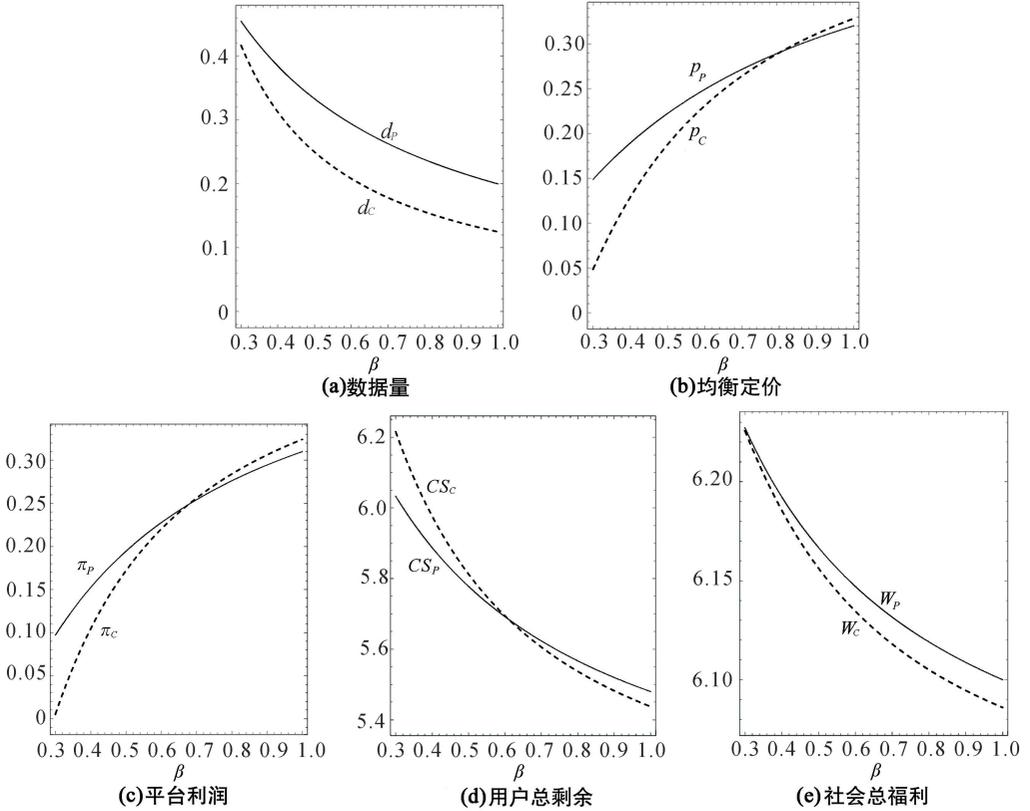


图 2 平台收集两边用户数据时数据量由不同主体决定的均衡结果比较

通过比较两种情形下的均衡数据量、均衡定价、平台利润、用户总剩余和社会总福利,本文得到以下命题。

命题 4: 在平台收集两边用户数据的情形下,

(1) 无论隐私泄露风险的高低,由平台决定数据量时均衡数据量更高,社会总福利也更高。

(2) 在隐私泄露风险较小的条件下,由平台决定数据量时均衡定价和平台利润更高;在隐私泄露风险较大的条件下,由用户决定数据量时均衡定价和平台利润更高。

(3) 在隐私泄露风险较小的条件下,由用户决定数据量时用户总剩余更高;在隐私泄露风险较大的条件下,由平台决定数据量时用户总剩余更高。

命题 4 表明,与平台只收集一边用户的数据相比,当平台收集两边用户的数据时主要结

①分析过程与前文类似,限于文章篇幅,具体均衡结果表达式不在正文中展示,备索。

②此时两种情形下均衡都存在的条件要求 $\beta > 0.296$ 。

论依然成立:社会总福利在平台决定数据量的情形下总是更高的;在隐私泄露风险较大的情况下,均衡数据量和用户总剩余在平台决定数据量的情形下更高,而平台利润在用户决定数据量的情形下更高。

(二) 平台只向一边用户收费

当平台只收集消费者数据,并且向消费者免费提供服务时,消费者的效用函数变为 $u_i^C = \theta^C - t|x^C - y_i| + (1+d_i)\lambda n_i^F - \beta d_i^2$, 厂商的效用函数与正文相同,平台利润表达式变为 $\pi_i = n_i^F p_i^F - \frac{(n_i^C d_i)^2}{2}$ 。为计算简便,以下假设 $t=1, \lambda = \frac{1}{2}$ 。

当由平台决定数据量时,博弈时序如下:第一阶段,两个平台同时决定对厂商的定价 p_i^F 以及向消费者收集的数据量 d_i ;第二阶段,厂商和消费者同时选择加入哪个平台。当由消费者决定数据量时,博弈时序如下:第一阶段,两个平台同时决定对厂商的定价 p_i^F ;第二阶段,两组用户同时选择加入哪个平台;第三阶段,由消费者决定提供的数据量 d_i 。同样地,这里不直接展示均衡结果表达式。

在平台只向厂商收费的情形下,数据量由不同主体决定时的均衡结果比较同样可以利用数值模拟直观地展示。令 $\theta^F = \theta^C = 3$,数值模拟结果见图3。

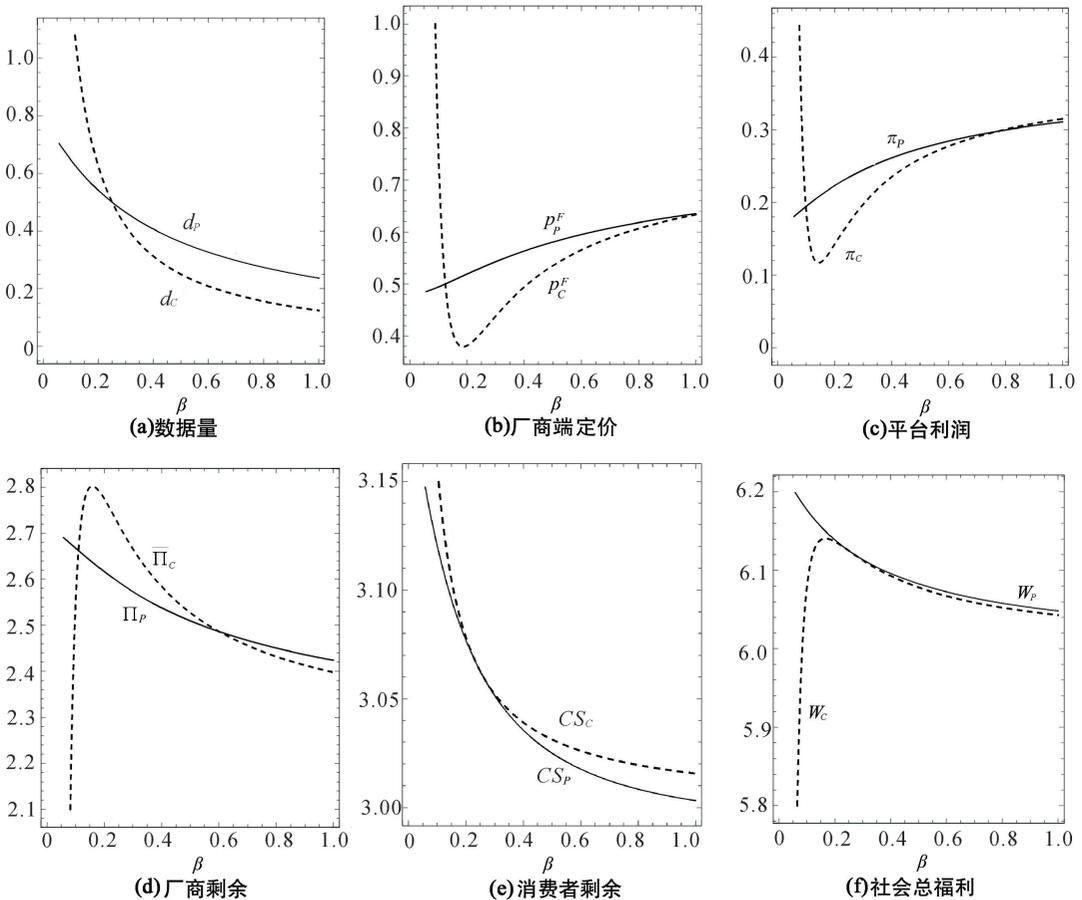


图3 平台只向厂商收费时数据量由不同主体决定的均衡结果比较

通过比较两种情形下的均衡数据量、厂商端定价、平台利润、厂商剩余、消费者剩余和社会总福利,本文得到命题5。

命题5:在平台只向厂商收费的情形下,

(1)在隐私泄露风险较高的条件下,由平台决定数据量时均衡数据量更高;在隐私泄露风险较低的条件下,由消费者决定数据量时均衡数据量更高。

(2)在隐私泄露风险很高或者很低的条件下,由平台决定数据量时厂商剩余更高,平台利润更低;在隐私泄露风险相对适中的条件下,由消费者决定数据量时厂商剩余更高,平台利润更低。

(3)无论隐私泄露风险的高低,消费者剩余在由消费者决定数据量时更高,社会总福利在由平台决定数据量时更高。

当平台只向厂商收费时,消费者剩余在由消费者决定数据量的情形下始终更高,这与平台向两边用户都收费时的结论正好相反。这个结果说明平台不同的收费模式会对经济主体产生不同的影响,因此政府制定政策确定数据权属时也要充分考虑不同行业、不同平台的特性。

命题3、命题4和命题5的共性在于,由平台决定数据量时的社会福利始终高于由消费者决定数据量时的社会福利,因此,无论平台的收费模式和数据收集模式如何,从社会福利的角度来看,赋予消费者决定数据量的权利并不是最优的选择。另外,通过数值模拟可以发现,只要隐私泄露风险超过一定阈值,那么随着隐私泄露风险的进一步增加,平台利润会增加,用户剩余和社会总福利都会降低,这个结论与数据量的决定者、数据收集模式以及收费模式都无关。

七、结论与启示

在数字经济中,数据成为重要的生产要素,企业通过收集、存储、使用大量用户数据牟利,引发了消费者对隐私问题的担忧,同时数据权属不明也会进一步影响数字经济的高质量发展。在此背景下,本文构建了双边市场下平台向两边用户收费并收集消费者数据的双寡头竞争模型,讨论当数据量由不同主体选择时的均衡结果以及隐私泄露风险对均衡结果的影响,并比较了这两种情形下的均衡数据量以及各个群体的福利。在数字经济发展迅速的当代,平台企业收集处理数据的效率越来越高,消费者对隐私的关注度也越来越高,本文发现当由平台决定数据量时,随着隐私泄露风险的增加,均衡数据量会减少,定价和平台利润增加,消费者剩余、厂商剩余和社会总福利都降低;当由消费者决定数据量时,在隐私泄露风险超过一定阈值的条件下,随着隐私泄露风险的进一步增加,均衡数据量、消费者剩余、厂商剩余和社会总福利同样会减少。本文还比较了数据量由平台决定和由消费者决定两种情形下的各个群体福利,发现无论隐私泄露风险的大小,消费者剩余和社会总福利在平台决定数据量的情形下都是更高的;另外,在隐私泄露风险较大的情况下,由平台决定的均衡数据量高于由消费者决定的均衡数据量,由平台决定数据量时的平台利润低于由消费者决定数据量时的平台利润。但在平台只向厂商收费的情形下,消费者剩余在由消费者决定数据量时更高。

基于上述研究结论,本文得到以下政策启示:

(1)政府在制定数据权利相关的法律政策时,应充分考虑不同产业的差异性。具体而言,对于那些平台向两边用户收费的产业,政策应倾向于支持平台拥有数据量决策权,因为本文的研究表明,这有助于提升社会福利和消费者剩余。而对于平台免费向消费者提供服务的行业,若政策目标是保护消费者利益,那么政府应该赋予消费者更多的数据量决策权,以确保消费者剩余不被平台削弱。为此,政府可以引入差异化的数据治理政策,根据行业服务模式的不同,评估数据权利配置对各产业的具体影响并制定相应的数据权利分配指导原则。

(2)在不同数据决策情境下,个人信息保护政策需差异化实施。当平台拥有数据量决策权时,更严格的个人信息保护政策有助于降低隐私泄露的风险,并增加消费者剩余和社会福利。在这种情境下,政府应该推动强有力的监管措施,确保平台在处理个人数据时采取必要的安全和隐私保护措施。而当消费者拥有数据量决策权时,政府应实施更为弹性的个人信息保护政策,避免过度保护导致的福利损失。过于严格的个人信息保护使得消费者更愿意提供数据,从而增加隐私泄露的负效用和平台转嫁给厂商及消费者的数据成本。因此,政府应制定灵活的个人信息保护框架,根据数据量决策权的实际归属进行调整,寻求在尊重个人隐私和促进经济发展之间的平衡点。

(3)随着大数据对企业竞争的影响日益显著,传统的反垄断理论和实践面临新的挑战。反垄断执法机构需要更新其监管工具和方法,以适应数字经济的特点。这包括监控和评估大型互联网平台的数据收集和处理行为,确保它们不会滥用市场支配地位。同时,监管机构应当鼓励这些平台在创新和竞争中发挥积极作用,并防止潜在的数据垄断行为对消费者隐私和市场公平造成损害。此外,政府还应加强与数字平台的合作,制定明确的数据处理和共享标准,共同构建公平竞争、健康发展的数字经济生态。

本文的结论为数据量决策权和个人信息保护的相关政策制定提供了理论参考,然而研究依然有一些不足与局限。比如,本文未考虑平台的差异化竞争以及厂商和消费者多归属等情形,因此为了提高模型的准确性,未来的研究可以引入更细致复杂的消费者和厂商行为模型,如考虑差异化竞争、异质性偏好和多期决策等因素。另外,技术进步和消费者意识变化等长期动态因素未被全面纳入模型,未来的研究还可以进一步探讨这些长期动态因素对数据量决策的重要影响,并考察模型在不同行业 and 不同规模平台企业中的适用性,以更好地反映实际经济环境的复杂性。最后,本文的模型是建立在双边市场的基础上,导致福利减少的关键因素是数据量减少从而交叉网络外部性下降,如果改变双边市场的设定,相应的结论会发生怎样的改变仍需进一步研究与探索。

参考文献:

- 1.陈永伟,2018:《数据产权应划归平台企业还是消费者?》,《财经问题研究》第2期。
- 2.程啸,2018:《论大数据时代的个人数据权利》,《中国社会科学》第3期。
- 3.李三希、黄卓,2022:《数字经济与高质量发展:机制与证据》,《经济学(季刊)》第22卷第5期。
- 4.李三希、王泰茗、刘小鲁,2023:《数据投资、数据共享与数据产权分配》,《经济研究》第7期。

- 5.李三希、武珂璠、鲍仁杰,2021:《大数据、个人信息保护和价格歧视——基于垂直差异化双寡头模型的分析》,《经济研究》第1期。
- 6.刘坤、喻玲,2022:《算法个性化推荐商业应用、消费者损害及其多元救济》,《江南大学学报(人文社会科学版)》第4期。
- 7.汪敏达、李建标,2022:《隐私经济学:研究述评与展望》,《外国经济与管理》第4期。
- 8.汪敏达、李建标、陈志斌,2022:《消费者个人信息保护与厂商广告策略的实验研究》,《中国工业经济》第4期。
- 9.王超贤、张伟东、颜蒙,2022:《数据越多越好吗——对数据要素报酬性质的跨学科分析》,《中国工业经济》第7期。
- 10.王申、许恒、吴汉洪,2022:《数据互操作与知识产权保护竞合关系研究》,《中国工业经济》第9期。
- 11.王泰茗、李三希、刘小鲁,2024:《数据权利界定、数据收集与个性化定价》,《系统工程理论与实践》第1期。
- 12.Abowd, J. M., and I. M. Schmutte. 2019. "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices." *American Economic Review* 109(1): 171–202.
- 13.Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar. 2022. "Too Much Data: Prices and Inefficiencies in Data Markets." *American Economic Journal: Microeconomics* 14(4): 218–256.
- 14.Acquisti, A., C. Taylor, and L. Wagman. 2016. "The Economics of Privacy." *Journal of Economic Literature* 54(2): 442–492.
- 15.Ali, N., G. Lewis, and S. Vasserman. 2022. "Voluntary Disclosure and Personalized Pricing." *The Review of Economic Studies* 90(2): 538–571.
- 16.Armstrong, M. 2006. "Competition in Two-Sided Markets." *The RAND Journal of Economics* 37(3): 668–691.
- 17.Bergemann, D., A. Bonatti, and T. Gan. 2022. "The Economics of Social Data." *The RAND Journal of Economics* 53(2): 263–296.
- 18.Bloch, F., and G. Demange. 2018. "Taxation and Privacy Protection on Internet Platforms." *Journal of Public Economic Theory* 20(1): 52–66.
- 19.Caillaud, B., and B. Jullien. 2003. "Chicken & Egg: Competition among Intermediation Service Providers." *The RAND Journal of Economics* 34(2): 309–328.
- 20.Chen, Y., and Z. J. Zhang. 2009. "Dynamic Targeted Pricing with Strategic Consumers." *International Journal of Industrial Organization* 27(1): 43–50.
- 21.Choi, J. P., Doh-Shin Jeon, and Byung-Cheol Kim. 2019. "Privacy and Personal Data Collection with Information Externalities." *Journal of Public Economics* 173: 113–124.
- 22.De Cornière, A., and R. De Nijs. 2016. "Online Advertising and Privacy." *The RAND Journal of Economics* 47(1): 48–72.
- 23.Dimakopoulos, P. D., and S. Sudaric. 2018. "Privacy and Platform Competition." *International Journal of Industrial Organization* 61: 686–713.
- 24.Dosis, A., and W. Sand-Zantman. 2023. "The Ownership of Data." *Journal of Law, Economics, and Organization* 39(3): 615–641.
- 25.Fainmesser, I. P., A. Galeotti, and R. Momot. 2023. "Digital Privacy." *Management Science* 69(6): 3157–3173.
- 26.Ichihashi, S. 2020. "Online Privacy and Information Disclosure by Consumers." *American Economic Review* 110(2): 569–595.
- 27.Johnson, J. P. 2013. "Targeted Advertising and Advertising Avoidance." *The RAND Journal of Economics*

- 44(1): 128–144.
28. Jones, C. I., and C. Tonetti. 2020. “Nonrivalry and the Economics of Data.” *American Economic Review* 110(9): 2819–2858.
29. Karle, H., M. Peitz, and M. Reisinger. 2020. “Segmentation versus Agglomeration: Competition between Platforms with Competitive Sellers.” *Journal of Political Economy* 128(6): 2329–2374.
30. Li, Y., and J. Zhang. 2024. “Who Should Decide How Much Information to Collect?” *Oxford Economic Papers*, Online.
31. Liu, Q., D. Nedelescu, and J. Gu. 2021. “The Impact of Strategic Agents in Two-Sided Markets.” *Journal of Economics* 134: 195–218.
32. Posner, R. A. 1981. “The Economics of Privacy.” *American Economic Review* 71(2): 405–409.
33. Rochet, J., and J. Tirole. 2003. “Platform Competition in Two-Sided Markets.” *Journal of the European Economic Association* 1(4): 990–1029.
34. Shy, O., and R. Stenbacka. 2016. “Customer Privacy and Competition.” *Journal of Economics & Management Strategy* 25(3): 539–562.
35. Stigler, G. J. 1980. “An Introduction to Privacy in Economics and Politics.” *The Journal of Legal Studies* 9(4): 623–644.
36. Taylor, C., and L. Wagman. 2014. “Consumer Privacy in Oligopolistic Markets: Winners, Losers, and Welfare.” *International Journal of Industrial Organization* 34: 80–84.
37. Tsai, J. Y., S. Egelman, L. Cranor, and A. Acquisti. 2011. “The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study.” *Information Systems Research* 22(2): 254–268.

Research on Data Volume Decision-Making in Platform Competition

Yu Wenshi¹ and Qiao Yue²

(1: School of Economics, Shandong University;

2: School of Innovation and Entrepreneurship, Shandong University)

Abstract: Data has become a critical factor in production. However, the ownership of data volume decision rights is not clear at present. This paper constructs a duopoly model in a two-sided market in which the platforms collect consumers’ data, analyzes the equilibrium outcomes when the data volume is determined by platform enterprises and consumers respectively, and compares the equilibrium data volume and the welfare of each group. It is found that when the privacy disclosure risk is high, the data volume that consumers are willing to provide is often lower than what the platform wants. From the perspective of social welfare, it is better to give platform enterprises the right to decide the data volume. When the data volume is determined by the platform, consumer surplus, manufacturer surplus, and total social welfare all decrease with the increase of privacy disclosure risk. But if consumers are given the right to decide how much data that platform can use, the impact of privacy disclosure risk on users’ surplus and social welfare is not monotonous. The conclusion of this paper provides a theoretical reference for the policymaking of data ownership confirmation and personal information protection.

Keywords: Platform Competition, Data Collection, Privacy Disclosure, Welfare Analysis

JEL Classification: C61, D43, L11

(责任编辑:陈永清)