

数字平台定价与数据隐私激励

——基于合约变量的结构估计*

王新兴

(安徽新华学院, 安徽 合肥 230088)

摘要: 大数据、人工智能等科技的发展, 激励平台实施价格歧视并加剧了消费者数据隐私损失。研究了消费者数据隐私损失的边际分布状况, 并分析了平台采用非线性合约获取用户数据隐私时的消费者净剩余变化。研究成果有助于为平台反垄断监管与数据隐私保护提供政策支持。根据结构估计以及反事实推断结果, 数据隐私损失的边际分布在高曲率点显示出平台甄别激励强度高, 消费者交易越频繁表示数据隐私损失越高, 当平台采用线性定价时消费者剩余相对较高, 消费者不存在隐藏行为时平台获取全部消费者剩余。

关键词: 数据隐私损失; 数字平台; 非线性定价; 结构估计; 合约

中图分类号: F062.5; F064.1 **文献标志码:** A **DOI:** 10.19358/j.issn.2097-1788.2026.03.009

中文引用格式: 王新兴. 数字平台定价与数据隐私激励——基于合约变量的结构估计 [J]. 网络安全与数据治理, 2026, 45(3): 61-67.

英文引用格式: Wang Xinxing. Digital platform pricing and data privacy incentives: a structural estimation using contracts variation [J]. Cyber Security and Data Governance, 2026, 45(3): 61-67.

Digital platform pricing and data privacy incentives: a structural estimation using contracts variation

Wang Xinxing

(Anhui Xinhua University, Hefei 230088, China)

Abstract: The advancement of big data, artificial intelligence, and other technologies has prompted platforms to implement price discrimination and exacerbate consumer data privacy loss. This paper examines the marginal distribution of consumer data privacy loss and the net residual change of consumers when platforms adopt nonlinear contracts to obtain user data privacy. This study aims to provide policy support for platform anti-monopoly supervision and data privacy protection. Based on the results of structural estimation and counterfactual inference, it is found that the marginal distribution of data privacy loss at points of high curvature indicates a strong incentive for discrimination by the platform, with more frequent consumer transactions leading to higher data privacy loss. Additionally, it is observed that when the platform adopts linear pricing, consumer surplus is relatively high; however, when consumers do not conceal their behavior, the platform captures all consumer surplus.

Key words: data privacy loss; digital platform; nonlinear pricing; structure estimation; contract

0 引言

近年来, 伴随大数据、人工智能等数字化技术快速发展, 双边市场信息不对称和数据隐私问题愈发严重。本文研究不对称信息下双边市场中用户隐私损失

的边际分布状况, 以及平台采用非线性合约时的消费者净剩余变化。平台具有垄断数据实施非线性定价的激励^[1], 例如借助营销活动、页面展示与 Cookie 追踪等手段持续收集用户行为数据, 分析用户隐藏的私人信息, 以推测消费者偏好、评估其购买意愿, 调整定价策略以获取和分配隐私租金^[2-3]。随着消费者隐私偏好增强、数据外部性日益显著以及全球隐私规制趋

* 基金项目: 安徽新华学院校级科研团队项目 (kytdp202504); 安徽新华学院校级青年重点项目 (2025rwqzd04)

严(如 GDPR、CCPA 的相关规定),数据隐私价值及规制正显著影响平台的信息获取成本、收益结构及其定价行为^[4]。

从理论层面看,平台抽取数据隐私租金本质上是信息不对称下的垄断甄别问题。隐私在此可被视为代理人不可观察的行为或特征,代理人采取的最优行动是种混同均衡,如果隐私是可观察的,代理人可以通过控制隐私披露以混淆其他人的感知^[5-6],而平台目标是设计合约以调控信息,比如通过加入噪声混淆价格分布以提高竞争性价格歧视的均衡利润^[7]。垄断甄别情况下当消费者为“数据敏感型”时,其披露信息将承担私人成本^[8],这会反向影响平台的甄别效率与定价策略。Kim 和 Wagman^[9]认为消费者类型差异会影响企业在甄别过程中的信息成本和服务成本,如果企业降低产品价格是以能够允许售卖消费者信息为前提,那么企业会提升甄别集中度并且有助于改进社会福利。

数据隐私规制的影响也呈现出“双刃剑”特性:一方面, Athey 等^[10]认为较弱信息披露激励可解释隐私悖论,搜寻成本对消费者隐私保护选择有重要影响,并且如果消费者事前消费选择是随机的,加强隐私保护还可以增加社会福利^[11];另一方面,规制也可能产生非预期的负面后果,比如 Prince 和 Greenstein^[12]构建模型探究移动互联网用户数据鸿沟与基于用户定价策略间的关系,因为数据鸿沟约束可能会对低收入群体收入产生负面影响。Sabatino 和 Sapi^[13]认为隐私规制会影响线上市场结构,通过测度隐私保护程度和广告信息量关系发现隐私规制对行业领先大企业影响较大。数据隐私规制影响则主要取决于消费者类型分布和个性化定制成本状况^[14]。

尽管现有实证研究已广泛探讨平台价格歧视对用户福利的潜在影响,但多数文献尚未充分纳入平台获取用户隐私时的补偿支付净额,以及用户在加入平台时所承担的信息成本。在数据隐私约束下,平台的双边市场价格结构如何影响平台与用户的福利分配,仍有待深入探讨。在具体研究对象上,本文选取电子商务平台(以淘宝为例)的消费者广告商品交易数据作为样本。由于消费者具有异质性,平台可利用其信息优势,结合用户年龄、性别等硬信息与点击、收藏、上下文内容等软信息,推测消费者类型并评估其购买率以实施甄别定价。

本文的边际贡献主要体现在方法层面:相较于已有关于价格歧视的实证研究,本文融合结构估计方法

与委托—代理理论,识别消费者数据隐私损失的分布特征,并评估平台在信息租金抽取与非线性定价中的效率表现。研究步骤包括:首先构建计量模型确立分析框架;其次采用非参数估计识别信息成本函数与消费者类型分布;最后通过参数估计与反事实分析,测算不同数据隐私状态下的福利剩余差异。

1 模型

1.1 消费者决策

消费者决策是平台最优化决策的均衡约束,主要是对合约中的交易概率 y 建模。假设交易概率 y 服从 $[0, 1]$ 区间的均匀分布,任何消费者在平台的交易概率都是根据相关数据推断得来的。收集信息或数据以及推送广告等行为都需要对消费者进行价格补偿 (δ),可以表示为广告商品价格函数;同时,消费者由于披露隐私或被数据挖掘产生隐私租金或支付信息成本。消费者交易概率 y 取决于两者净额,依赖于不同交易行为与信号,如是否成交、商品收藏量与价格、观察商品评论、店铺评级等内容,平台要激励用户提高交易概率就要综合利用上述信息或数据。为了反映数据隐私和平台补贴行为对消费者决策的影响,假定消费者类型 θ 服从 $F(\theta)$ 累积概率分布,并且消费者类型 θ 会成为隐私成本函数变量,表示为 $C(y, \theta)$,平台价格行为进入到交易收益函数,用 $\delta(\theta) \cdot y$ 表示。类型 θ 消费者在平台上的交易概率选择满足:

$$y(\theta) \in \inf_y \min_y \delta(\theta) \cdot y - C(y, \theta) \quad (1)$$

由于真实交易概率是不可观察的,平台只能根据商品收藏量与价格、观察商品评论、店铺评级等内容进行推断。假定对于任何 $y \in (0, 1)$ 而言成本函数 $C(\cdot, \cdot)$ 是二次连续可微的。

根据上述设定,定义式 (1) 的一阶条件,可得消费者交易概率 $y(\theta)$ 满足:

$$\delta(\theta) = \partial C / \partial y \quad (2)$$

从式 (2) 可知,定义 $\theta(\cdot)$ 为逆函数,消费者类型 θ 下的平台补贴对交易概率 $y(\theta)$ 是严格递减的。

1.2 平台决策

作为知情方的平台设计者,平台在每次试验中从消费者行动获得的私人信息价值记为 $\omega(t)$,每次试验产生信息的社会价值都具有差异性。有理由相信平台会选择非线性合约或价格歧视以最大化目标函数,按照充分信息利用原则,所有会抽取消费者隐私租金的相关数据都会包含在合约中。比如对消费者的数据价值补偿可能与其手机操作端上下文内容相关联,对于商品收藏、点击等增量信号可能会用于三级价格歧

视等。

在消费者参与到平台网络的条件下，每次试验不同类型消费者给平台带来的期望收益表示为 $(\omega(\theta) - \delta(\theta)) \cdot y(\theta, \delta(\theta))$ 。关于 $\delta(\theta)$ ，在所有行为数据都能观察到的情况下，可以从所有类型 θ 的消费者与非线性合约事后推测出来，满足：

$$\delta(\theta) \in \arg \max_{\delta} \mathbb{E}[(\omega(\theta) - \delta(\theta)) \cdot y(\theta, \delta(\theta))] \quad (3)$$

其中 \mathbb{E} 表示期望算子。给定数据价值补偿 δ 下，消费者会选择符合自身类型的市场交易率 $y(\theta, \delta(\theta))$ ，所以式 (3) 包含式 (1) 均衡选择，这是消费者的激励相容约束。 $\delta(\theta)$ 满足式 (3) 的一阶条件为：

$$-\mathbb{E}[y(\theta, \delta(\theta))] + (\omega(\theta) - \delta(\theta)) \mathbb{E}[(\partial y / \partial \delta)(\theta, \delta(\theta))] = 0 \quad (4)$$

最后，给定个人数据或信息的货币补偿方案，当前阶段平台从激励消费者显示真实偏好可获得的净剩余 $\pi(\theta)$ 表示为：

$$\pi(\theta) = (\omega(\theta) - \delta(\theta)) \int y(\theta, \delta(\theta)) dF(\theta) \quad (5)$$

根据假设，可知上述优化问题的充分条件是满足的，并且如果限制在线性合约范围内，净剩余 $\pi(\theta)$ 将不会是最优的。

1.3 反事实合约

在反事实框架下货币补偿计划表示为 t ，消费者在平台交易概率 y 水平下最优合约菜单 $t_h^*(\theta)$ 表示为向量 $(t_{h0}^*(\theta), t_{h1}^*(\theta), \dots, t_{hh}^*(\theta))$ ， $t_{hp}^*(\theta)$ 表示为代理人在交易概率 y 水平选择可观察到商品销量时的最优支付。为了保证实现最优配置，最低参与约束要满足 $t_{h0}^*(\theta) \geq 0$ 。当代理人面临补偿方案 $t = (t_0, t_1, \dots, t_h)$ 时，类型 θ 代理人在交易概率 y 水平的选择满足：

$$y_h(t, \theta) \in \arg \max_y \mathbb{E}[y_h] t_h - C(\theta, y) \quad (6)$$

既定转移支付 t 下，平台从类型 θ 消费者获得平均剩余为 $\omega(\theta) \cdot y_h(t, \theta)$ 。从式 (6) 可以看出平台平均成本是 $\mathbb{E}[y_h] t_h$ 。因此可得，最优合约菜单表示为：

$$t_h^*(\theta) \in \arg \max_t \mathbb{E}[\omega(\theta) \cdot y_h(t, \theta) - \mathbb{E}[y_h] \cdot t_h] \quad (7)$$

现在进一步分析两种反事实场景。第一种是没有道德风险的情况，道德风险表现为故意改变影响交易概率的各种行动，例如频繁点评各类商品评价、刷单等，该情况下平台事后可观察到代理人交易概率，并基于此提供激励补偿计划。如果委托人对不同类型消

费者提供合约 $t_h(y)$ ，类型 θ 的代理人会根据一阶条件选择交易概率 $y_h^{WM}(t, \theta)$ ：

$$t_h'(y_h^{WM}(t_h, \theta)) = C'(y_h^{WM}(t_h, \theta)) \quad (8)$$

此时委托人在没有道德风险情况下时会选择最优合约 $t_h^{WM}(\tau, \cdot)$ ：

$$t_h^{WM}(\theta) \in \arg \max_t \mathbb{E}[\omega(\theta) \cdot y_h^{WM}(t_h, \theta) - \mathbb{E}[y_h^{WM}(t_h, \theta)]] t_h \quad (9)$$

可见，道德风险会诱导平台从交易概率分布中制定货币转移计划。如果转移补偿计划是基于推断的交易概率分布设计的，那么这种货币补偿是可以实施的。在最优合约菜单 $t_h^*(y)$ 中，这类合约可以根据代理人选择主观交易概率予以确定，但是受到数据收集或信息获取质量约束，并且 $t_h^*(y)$ 和 $t_h^{WM}(y, \cdot)$ 两种货币补偿计划可能区别并不是很大。

第二种反事实场景，是考虑没有不对称信息情况下，政策制定者所能获得的剩余。该部分与基准模型分析相似，在不完全信息状态下平台会修正消费者参与平台交易的信念，这种信念是指对交易概率分布的推断。根据消费者信号推断其类型并实施价格歧视，此时类型 θ 代理人最优交易概率 $y^c(\theta)$ 满足：

$$\omega(\theta) = C'(\theta, y^c(\theta)) \quad (10)$$

此时，平台设计者可以获得所有信息租金。基于此，在既定交易概率水平下货币补偿的最优转移补偿函数 $t^c(\cdot)$ 满足：

$$t^c(y^c(\theta)) = C(\theta, y^c(\theta)) \quad (11)$$

由式 (6) ~ (11) 可以看出，即使在反事实框架下委托人采用简单的线性合约限制，依然可以提供基于上下文内容、广告商品收藏量等有限数据集来设计不同的合同菜单，既可以采用简单的线性合同，也可以采用更为复杂的算法定价。通过深入分析不对称信息对平台和消费者决策的影响，如果平台能够获取消费者隐私数据以及估算相应数据的信息价值，则可以比较不同场景下的福利变化。

2 实证策略

2.1 数据与统计

实证数据样本来自于淘宝平台所提供的海量真实交易数据，包括用户特征、搜索广告、商品品质等数据。使用这些数据可以研究消费者在不同试验下（即广告商品组合）的交易概率（或称购买意向）。样本数据类型包括商品特征、行为特征、上下文内容、买家特征、卖家特征 5 大类 14 个变量共 454 836 条数据，其中最终成交的数据为 8 657 条，未成交 446 179 条。数据集经过脱敏处理，价格、品牌等变

量处理为离散的整数型数据类型, 比如商品价格划分为 0 ~ 17 个等级, 等级越高价格越高。具体内容见表 1。

表 1 样本变量及其统计性描述

类别	变量	含义	均值	标准差	最小值	最大值	众数
商品特征	ad_brand	品牌	42.71	1.20	19.10	43.67	43.51
	ad_price	价格	6.78	1.09	1	17	7
	ad_sales	销量	11.14	2.58	0	17	12
行为特征	ad_collected	收藏次数	12.21	2.49	0	17	12
	ad_display	展示次数	17.14	2.15	0	21	18
	trade	是否交易	0.02	0.14	0	1	0
上下文内容	context_page	商品 页面编号	3.36	4.02	1	20	1
	context_id	上下文 信息编号	42.67	1.00	27.04	43.67	40.21
买家特征	user_gender	用户性别	0.21	0.41	0	1	0
	user_age	用户年龄	4.54	1.21	1	8	4
	user_occupation	用户职业	4.00	1.37	2	5	5
卖家特征	shop_Review	店铺 评价数量	15.77	2.83	1	25	17
	shop_Service	店铺 服务态度	0.97	0.01	0.79	1	0.98
	shop_Description	店铺描述评分	0.98	0.01	0.79	1	0.98

对于消费者或平台而言, 双边市场体系中最重要信号就是价格和销量, 价格对消费者是可观察的, 销量则是平台有选择地披露给消费者。一方面, 传统供需理论主要集中在价格和销量(或产量)上, 另一方面, 由于样本数据经过脱敏处理等过程, 假设当价格和销量存在对称的对应关系时, 消费者购买价格等级高于(或低于)观察到的销量等级, 反映了基于消费者类型所披露的信号差异, 这是消费者具有不同私人信息的重要表现。价格信号反映了消费者在平台操作的行为以及对平台广告推动的策略性反应。因此, 根据消费者购买价格等级与观察到的销量等级差异划分为两个子样本, 每个子样本中包括是否成交等关键变量, 显然在消费者购买价格等级高于(或低于)观察到的销量等级情况下, 是否购买或交易概率取决于不可观察的消费者私人信息。子样本 $x = 1$ 表示消费者购买价格等级高于观察到的销量等级的样本, 子样本 $x = 2$ 表示消费者购买价格等级低于观察到的销量等级的样本, $B = 1$ 表示子样本中已购买或交易的消费者,

$B = 0$ 则表示没有交易。

2.2 估计策略

首先, 进行 Logit 回归建立基准模型探究解释变量与被解释变量的相关性, 主体部分的结构估计策略主要借鉴 D'Haultfœuille 和 Février 对工资激励合约的估计方法^[15], 之所以借鉴该模型是因为消费者购买商品进行交易的概率与广告推送时间或广告被展示次数有关。另一方面, 根据交易概率和结构函数可以推断成本函数与类型分布函数的关系, 采用筛最大似然估计方法(Sieve Maximum Likelihood Estimator)可以推导交易概率分布, 以及在样本空间中推断消费者的隐私成本和类型分布的函数形态。

其次, 对于稳健性检验方面, 虽然获取信息越多越好, 但是提高交易概率并不依赖于某些变量, 即过度收集数据有时并非一定对提高预测交易概率有用, 交易概率可能仅仅是某些数据或信息及消费者类型的函数。考虑到样本特性, 主要是评估实证样本是否具有选择效应, 也就是说使用不同样本进行实证计量是是否存在较大差异, 是否能够有效揭示数据隐私租金规模会影响消费者购买决策。如果不具有样本效应, 可以说明估计是稳健的。

再次, 展开不同激励合约下的政策分析, 主要比较最优合约、无信息不对称情况下的剩余状况, 用以说明不同类型消费者选择不同合约的情况。在这部分根据参数估计方法拟合成本分布函数和类型分布函数, 对 Beta 分布、对数正态分布等进行检验, 选择合适参数代入结构函数中以评估不同合约类型的福利剩余或损失。

结构估计遵循如下假设: (1) 成本函数 $C(y, \theta)$ 是可分离的和正则的; (2) 交易概率与数据隐私披露成正比, 无隐私情况交易概率趋近于 1; (3) 消费者类型 θ 保持相对不变, 并且与不同统计试验相对独立。

2.3 基准回归

根据参考文献 [16] 所提方法进行模型选择, 发现上下文内容中的商品展示页面编号、上下文信息编号相关性相对较低, 买家特征中用户职业, 以及卖家特征中店铺评价数量和店铺描述评分的相关性同样较低。基准回归结果如表 2 所示。

从基准回归中可以看出, 选择合适模型变量后所得计量结果相当显著, 其中, 当其他变量保持不变时, 品牌、价格、广告收藏、广告展示次数、店铺服务态度单位变化提升交易概率的相对优势变化低于 1, 或者说市场成交概率要低于不会交易的概率。

表2 Logit 基准回归

	参数	z-值	优势比	优势比的95% 置信区间
截距	7.62*** (1.087)	7.01	2 046.20	[240.32, 17 053.54]
品牌	-0.08*** (0.023)	-3.50	0.92	[0.88, 0.97]
价格	-0.24*** (0.010)	-25.05	0.79	[0.77, 0.80]
销量	0.28*** (0.010)	27.33	1.32	[1.30, 1.35]
性别	0.09*** (0.026)	3.41	1.09	[1.04, 1.15]
年龄	0.13*** (0.009)	14.35	1.14	[1.12, 1.16]
广告收藏	-0.12*** (0.009)	-12.95	0.87	[0.87, 0.90]
广告展示次数	-0.11*** (0.008)	-14.32	0.89	[0.88, 0.91]
店铺服务态度	-10.65*** (1.138)	-9.36	0.00	[0.00, 0.00]

注：括号内为标准误，***表示0.001水平上的统计显著

3 结构估计

3.1 类型分布与成本函数

根据上文理论分析，不同类型消费者参与平台交易的概率是不同的，从平台角度来看，消费者交易概率反映其类型或偏好信息，隐私成本是消费者类型和交易概率的函数。显然，已交易用户（即 trade 变量等于1）相对于未交易用户可能会支付更多隐私成本，并且在概率分布上表现出很大不同。图1反映样本内消费者交易概率累积分布函数情况，从图中可以看出以概率0.9为界限，交易概率累积分布函数曲率存在较大差异，反映出累积分布函数一阶导数大于零，并且概率密度分布在 $[0, 1]$ 区间是单调的，间接说明在区间 $[0, 1]$ 下属于不同子区间的交易概率实际上反映消费者风险偏好和隐私敏感性。从样本实证来看，当交易概率超过0.9时，消费者风险偏好显著增强，消费者会支付更多隐私成本以提高交易概率。

另一方面，图2说明是否成交在交易概率分布上表现出较大差异，虽然不同交易情况的交易概率分布形态与图1相似，但是相对已交易消费者而言，当交易概率低于0.7时，在风险偏好相差不大的情况下，未交易消费者在不确定性决策中是随机占优的，即对各种广告信息总是选择较低概率的行为策略，此时隐私成本支付相对较少，因为平台没有观察到与广告信

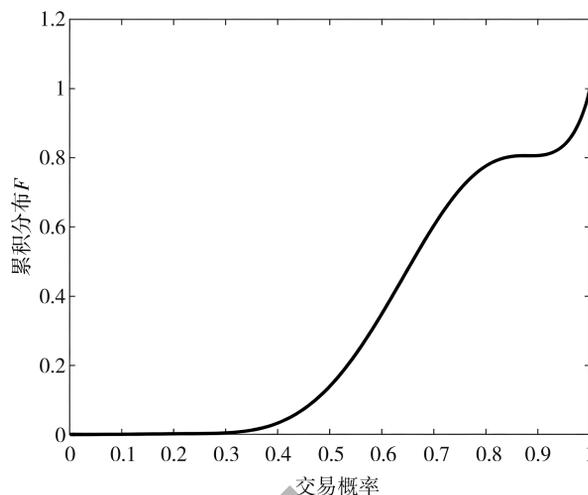


图1 交易概率分布的筛似然估计

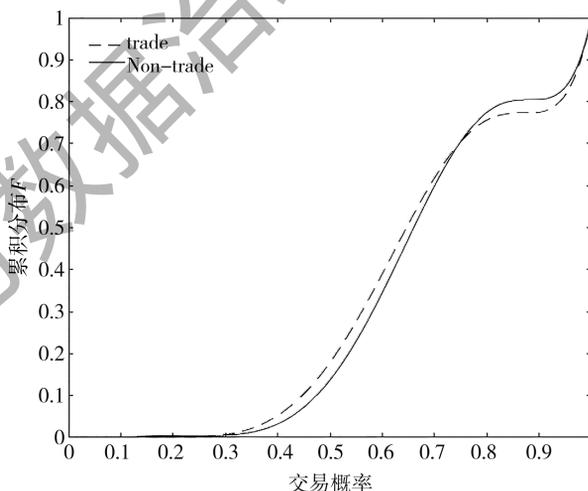


图2 不同交易情况的概率分布

息相匹配的信号，平台无法精准获取相关数据隐私；当未成交消费者在交易概率超过0.7时，必然是已支付较高的隐私成本，平台得以执行个性化定价或广告精准推送。此外，由于交易概率分布主要是受到消费者类型或外界可能状态的影响，不同交易概率分布实际上反映消费者间以及平台与消费者间存在不对称信息。

根据交易概率分布值，图3给出了样本中的消费者类型分布和边际隐私成本分布。一方面，至少在电子商务平台中，消费者对价格和销量等信号相对敏感，用户价值主要集中在较低类型区域，该部分类型消费者更倾向于采用行为策略；另一方面，就边际隐私成本分布而言，非常高的曲率表示提高交易概率必须要增加消费者甄别激励，同时当激励消费者使其交易概率达到0.8时，消费者支付的边际隐私成本至少是其

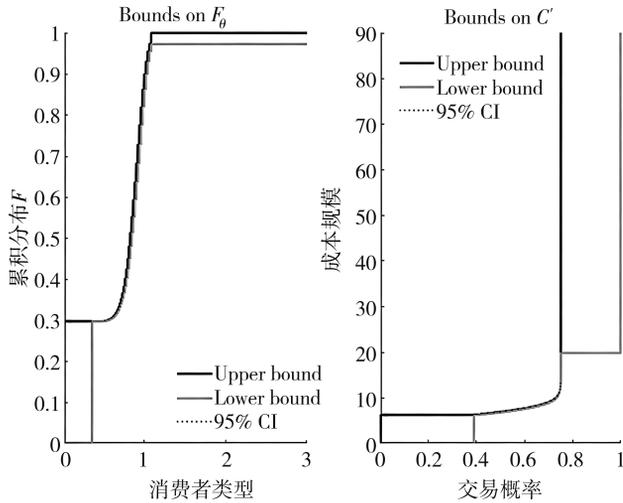


图3 消费者类型分布和边际隐私成本分布

交易价格的 20 倍。

3.2 分布参数估计

已知消费者类型分布和边际隐私成本函数形态，采用参数估计推断分布参数，以用于不对称信息下消费者和平台决策互动的福利变化。结合数据，消费者类型参数推断主要通过 Frechet 分布 ($-\ln(-\ln F_\theta)$)、Log-normal 分布 ($\Phi^{-1}(F_\theta)$)、Weibull 分布 ($\ln(-\ln(1-F_\theta))$) 等分布族进行拟和，这三种分布分别表现出不同尾部性质。边际隐私成本 $\partial C/\partial y = \alpha\varphi(y/(1-y))^\beta$ ，其中 $\varphi(u)$ 分别等于 $\ln(1+u)$ 、 u 、 $\exp(u)-1$ 。考虑到两个子样本之间的比较，假定 $\ln(\theta|x, B=0) \sim \mathcal{N}(a_1, b)$ ， $\ln(\theta|x, B=1) \sim \mathcal{N}(a_2, b)$ ，待估的参数向量表示为 $\eta = (a_1, a_2, a_3, \alpha, \beta)$ 。

通过拟合检验发现，消费者类型参数在 Weibull 分布中拟合效果较好，样本边际隐私成本函数中 $\varphi(u) = \exp(u) - 1$ 拟合效果较好。为了方便看出子样本间的差异，考虑 $a_1 = a_2$ 时，利用筛最大似然估计法获得参数向量 η 各个参数值，可以推断出平台获取信息价值 ω 。

如表 3 所示，从各类参数中可以看到，第一列中参数值 a_3 大于 a_1 和 a_2 ，反映已交易用户中平均交易概率相对更高，并且从消费者获取更高的信息价值，即 $\omega_1 > \omega_2$ 。在限制 $a_1 = a_2$ 情况下，未交易用户的平均交易概率相对较低，平台从消费者获得信息价值并不显著。

3.3 福利分析

本节结合反事实推断研究信息不对称下不同合约的消费者剩余结果以及转移支付情况，表 4 给出相关结果。四种环境下的不同支付补偿方式相比较而言，

表 3 消费者类型分布和边际隐私成本函数参数

参数	无约束估计	附加约束估计 ($a_1 = a_2$)
a_1	1.253 0e+84	12 732
a_2	1.174 3e+07	$a_1 = a_2$
a_3	9.619 3e+17	1 498
b	2.647 2	0.1
α	0.001 3	0.1
β	15.846 0	25
ω_1	3 173.6	1 174
ω_2	457.790 5	1 199

最优合约效率要高于线性合约，但是存在福利分配不合理情况，这也可以解释平台为什么会获取信息或收集数据以提高决策质量。具体来说，首先，完全信息状态下消费者期望剩余最高，这与传统经济理论保持一致；其次，不完全信息状态以及消费者没有隐藏行为时，消费者会受到较高的福利损失，可以说明如果平台掌握数据隐私以实施价格歧视，则可以最大化抽取消费者租金以及转移消费者剩余；最后，如果不完全信息下平台采用线性合约，则可以取得低于完全信息时的福利，不同之处在于，消费者交易概率相对而言更高。

表 4 不同补偿方案下的消费者剩余与交易概率

环境	支付方式	期望剩余	交易概率
完全信息	最优合约	812.450 0	0.676 5
没有道德风险的不完全信息	最优合约	-2.274 7e+10	0.676 2
不完全信息	最优合约	-2.270 4e+10	0.664 2
不完全信息	线性合约	784.809 7	0.999 4

3.4 稳健性检验

本文主要从内生性或样本选择效应评价稳健性。一方面，内生性可能会影响实证稳健性，在本文中主要表现为样本选择效应，因为这部分样本是特殊日期展开大规模广告营销的试验数据。在样本选择效应检验中，即使已交易消费者的分布可能与未交易消费者分布存在不同，依然可以通过检验子样本 1 中没有交易的消费者分布是否与子样本 2 中没有交易的消费者分布相同来评判稳健性。原假设定义为消费者没有交易时的类型分布在两个样本中是相同的，那么备择假设为分布是不同的。根据文献 [15] 提供的思路和计量方法，构建统计量 $T = \text{supp}_{y \in [0,1]} |F_{y|x=2,B=0} - F_{y|x=1,B=0} \circ H^{-1}|$ 判定样本选择偏误，其中 H^{-1} 是关于 $F_{y|x,B}$ 函数的分位数-分位数逆变换。通过检验可得 $T =$

0 及其对应 p -值等于 1, 所以接受原假设, 即子样本 1 中没有交易的消费者分布与子样本 2 中没有交易的消费者分布相同, 所以实证不会因为样本选择而产生较大偏差。

4 结论

消费者数据隐私对平台定价效率与消费者剩余的影响是非常复杂的, 从实证结果可以看出, 根据不同消费者数据预测平台交易概率存在较大差异, 平台定价更可能采取基于反映需求的信号 (三级价格歧视) 和提供不同的价格—产量组合诱导自选择 (二级价格歧视) 等混合定价策略, 以最大化抽取消费者隐私租金。虽然本文实证反映消费者数据隐私边际损失与交易概率呈现正相关, 但是考虑到不同双边市场类型或样本, 两者因果关系仍是不明确的。

参考文献

- [1] ACQUISTI A, VARIAN H R. Conditioning prices on purchase history [J]. *Marketing Science*, 2005, 24 (3): 367–381.
- [2] GOMES R. Optimal auction design in two-sided markets [J]. *The RAND Journal of Economics*, 2014, 45 (2): 248–272.
- [3] KIRPALANI R, PHILIPPON T. Data sharing and market power with two-sided platforms [J/OL]. NBER Working Paper, No. 28023, 2020. DOI: 10.3386/w28023.
- [4] ARIDOR G, CHE Y K, SALZ T. The effect of privacy regulation on the data industry: empirical evidence from GDPR [J]. *The RAND Journal of Economics*, 2023, 54 (4): 695–730.
- [5] DAUGHETY A F, REINGANUM J F. Public goods, social pressure, and the choice between privacy and publicity [J]. *American Economic Journal: Microeconomics*, 2010, 2 (2): 191–221.
- [6] RAMBACHAN A, KLEINBERG J, LUDWIG J, et al. An economic perspective on algorithmic fairness [J]. *AEA Papers and Proceedings*, 2020, 110: 91–95.
- [7] ELLISON G, ELLISON S F. Search, obfuscation, and price elas-

ticities on the Internet [J]. *Econometrica*, 2009, 77 (2): 427–452.

- [8] KRÄHMER D, STRAUSS R. Optimal nonlinear pricing with data-sensitive consumers [J]. *American Economic Journal: Microeconomics*, 2023, 15 (2): 80–108.
- [9] KIM J H, WAGMAN L. Screening incentives and privacy protection in financial markets: a theoretical and empirical analysis [J]. *The RAND Journal of Economics*, 2015, 46 (1): 1–22.
- [10] ATHEY S, CATALINI C, TUCKER C. The digital privacy paradox: small money, small costs, small talk [J/OL]. NBER Working Paper, No. 23488, 2017. DOI 10.3386/w23488.
- [11] BHASKAR V, ROKETSKIY N. Consumer privacy and serial monopoly [J]. *The RAND Journal of Economics*, 2021, 52 (4): 917–944.
- [12] PRINCE J, GREENSTEIN S. Mobile Internet usage and usage-based pricing [J]. *Journal of Economics & Management Strategy*, 2021, 30 (4): 760–783.
- [13] SABATINO L, SAPI G. Online privacy and market structure: theory and evidence [J]. *Information Economics and Policy*, 2022, 60, 100985.
- [14] COLOMBO S, GARELLA P G. Customer-based customization and price competition [J]. *The Journal of Industrial Economics*, 2022, 70 (4): 939–961.
- [15] D’HAULTFÈUILLE X, FÉVRIER P. The provision of wage incentives: a structural estimation using contracts variation [J]. *Quantitative Economics*, 2020, 11: 349–397.
- [16] BREHENY P, HUANG J. Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors [J]. *Statistics and Computing*, 2015, 25 (2): 173–187.

(收稿日期: 2025-10-17)

作者简介:

王新兴 (1989-), 男, 博士, 讲师, 主要研究方向: 信息经济学与数字治理。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.com