

数据经济学

第六章：数据市场均衡

陈希路

暨南大学经济学院

2026 年春

章节目录

- 1 第一节：数据市场均衡理论
- 2 第二节：数据一级市场定价
- 3 第三节：数据二级市场定价
- 4 第四节：数据三级市场定价

市场均衡

是指商品或服务的供给和需求达到平衡的一种状态。

理论基础:

- 核心：供需理论（亚当·斯密）
- 动态性：随供需条件变化，价格变化以重新实现均衡
- 分类：一般均衡 vs. 局部均衡

数据市场均衡的特点:

- 复杂性：涉及确权、隐私、合规、价值链
- 去中心化：以平台为中心，涉及供给方、需求方、平台方

第一节

数据市场均衡理论

数据市场均衡的前提

结合数据要素特征与传统经济理论，提出以下前提：

① 数据确权：

- 基础环节（劳动价值论）
- 产权不明导致交易无法律依据，收益分配无据

② 数据脱敏：

- 保护隐私，降低泄露风险
- 提高信任度，确保交易合法合规

③ 数据标准化：

- 统一格式、结构和质量
- 提高可比性，降低协商成本，促进价格公正

④ 信息完全：

- 解决“信息悖论”（买方了解信息后就不需购买）
- 消除信息不对称（如保险公司 vs. 个体）

数据市场均衡的前提

5 竞争性市场：

- 参与者地位对等，相互制约
- 推动价格达到均衡

6 无外部性：

- 交易仅影响双方（理想假设）
- 若有外部性，价格无法反映真实供需

7 无交易成本：

- 假设不存在手续费、税费等
- 过高成本会阻碍交易发生

8 有效监管：

- 规范行为，保障权益
- 提高交易的可靠性和可信度

数据市场均衡的机制与多样性

● 一般市场均衡的基本内涵

- **核心定义**：存在特定的均衡价格，使得市场供给量完全等于需求量
- **价格调节机制**：
 - 超额供给 → 导致价格下降
 - 超额需求 → 导致价格上涨
 - 均衡状态 → 价格停止波动，保持稳定

● 数据市场均衡的特征

- **供需决定**：数据市场的均衡价格由数据的供给与需求共同决定
- **模型多样性**：由于不同的市场结构具有不同的供需关系与价格形成机制，因此数据市场的均衡模型呈现出高度的多样性

案例分析：上海数据交易所

案例背景：如何为数据市场均衡创造前提？

数据作为新型要素，其非标准化和权属复杂的特征导致场外交易极易失败。上海数据交易所通过制度设计，努力满足市场均衡的前提条件。

理论与现实的映射：

- **数据标准化与确权**：首创数据产品登记凭证，将杂乱的原始数据转化为可流通、标准化的数据产品，解决交易客体不清的问题
- **解决信息悖论**：提供数据产品说明书和质量评估报告，买方在不看到底层原始数据（防泄露）的情况下，也能准确预估数据价值（信息完全）
- **降低交易成本与合规监管**：引入合规评估、质量评估等第三方机构，提供“即插即用”的交易环境，大幅降低买卖双方面议的交易成本

不同市场结构的数据市场均衡

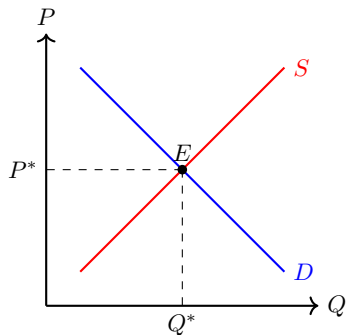
完全竞争市场

- 市场结构特点:

- 单个买者和卖者都无法控制价格，是价格的接受者

- 市场均衡点表现:

- 供给曲线与需求曲线共同决定市场均衡
- 在均衡价格下，市场出清，即需求量完全等于供给量



不同市场结构的数据市场均衡

垄断竞争市场

- **市场结构特点：**

- 市场中存在大量的卖方和买方
- 每个卖方都占据一定的市场份额，并且产品之间存在差异化，它们通过价格与非价格竞争来争夺市场

- **市场均衡点表现：**

- 在短期内，企业可能获得超额利润
- 但在长期来看，新企业的自由进入会导致市场回归到零超额利润的均衡状态

不同市场结构的数据市场均衡

寡头垄断市场

- **市场结构特点：**

- 少数几家大型数据供应商占据了市场的主导地位
- 市场中仍存在多个竞争对手，企业之间存在很强的相互依存性

- **市场均衡点表现：**

- 与垄断竞争类似，短期内可能存在超额利润
- 但长期市场同样会趋向于回归到零超额利润的均衡状态

案例分析：彭博终端（Bloomberg）的寡头垄断

案例背景：金融数据市场的寡头定价

彭博终端为全球金融机构提供实时的全球市场数据、新闻和分析工具。单台终端年费高达约 2.4 万美元，且极少打折。

寡头垄断数据市场的特征体现：

- **数据供给壁垒高**：整合全球数百万个数据源，建立专用闭环网络，目前仅路孚特 Refinitiv 等极少数寡头能与之抗衡
- **需求高度缺乏弹性**：对于基金经理和交易员而言，错过一秒钟的数据可能损失惨重，这种刚需使得买方对价格不敏感
- **市场均衡表现**：买方作为价格接受者，而彭博作为价格制定者。均衡点高于完全竞争市场，企业获得高额的超额利润

不同市场结构的数据市场均衡

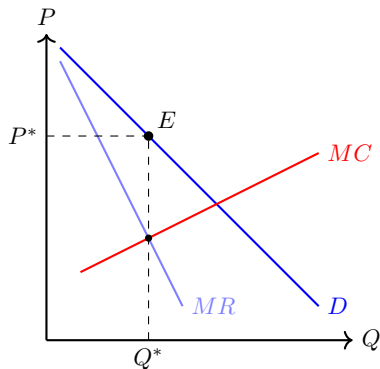
完全垄断市场

- 市场结构特点：

- 市场上只有一个供应商占主导地位

- 市场均衡点表现：

- 完全垄断企业是价格的制定者
- 企业根据边际收益等于边际成本 ($MR = MC$) 的原则确定产量，并在需求曲线上获得更高的垄断定价



不同市场结构的数据市场均衡

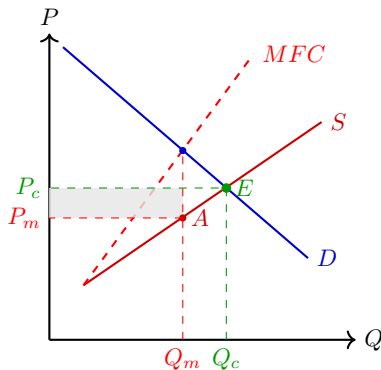
单一买方市场

- 市场结构特点:

- 只有一个买方，但供给方有多个

- 市场均衡点表现:

- 由于买方在市场上的垄断地位，市场需求曲线相对更加垂直
- 买方可以影响价格，使得价格随着需求变化而变化
- 买方通过将购买量限制在边际价值等于边际要素成本 ($D = MFC$) 的水平，并支付供给曲线上对应的更低价格（相对于完全竞争价格），从而获得垄断租金



不同市场结构的数据市场均衡

双边市场

- **市场结构特点：**

- 市场供给曲线和需求曲线都不是单一的曲线
- 而是由两个群体的需求和供给曲线组成的

- **市场均衡点表现：**

- 市场上的均衡点不仅取决于供给和需求的均衡
- 还取决于两个群体之间的网络外部性

自由数据市场的市场均衡机制：模型

假定自由数据交易市场的供需关系如下：

- 供给函数： $Q_s = c + d \times P$
- 需求函数： $Q_d = a - b \times P$

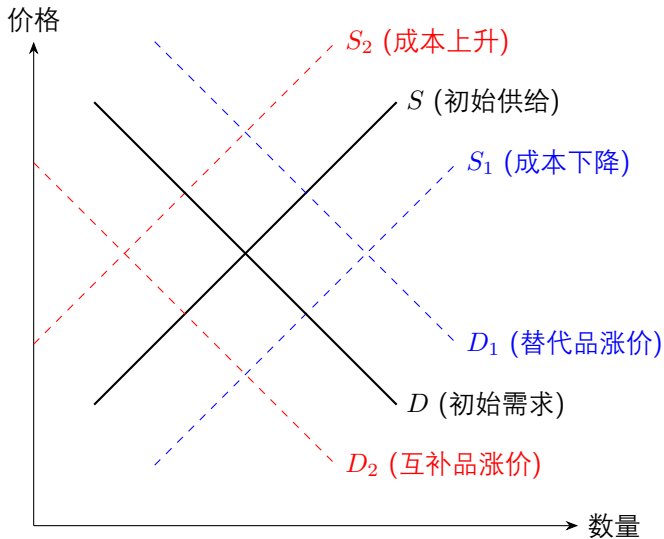
市场均衡点 ($Q_s = Q_d$) 的均衡价格为：

$$P = \frac{a - c}{d + b}$$

其中：

- P ：价格
- a, c ：截距项
- b, d ：弹性系数（价格变动引起数量变动的程度）

自由数据市场的市场均衡机制：图示



自由数据市场的市场均衡机制：动态变动

供给曲线移动

- 向左上 (S_2): 收集成本上升 → 均衡价格上升, 数量下降。
- 向右下 (S_1): 开发成本下降 → 均衡价格下降, 数量上升

需求曲线移动

- 向左上 (D_1): 替代品价格上升 → 均衡价格、数量均上升
- 向左下 (D_2): 互补品价格上升 → 均衡价格、数量均下降

价格调节机制

- 超额供给 → 价格下降
- 超额需求 → 价格上升

案例分析：AI 数据标注市场的动态均衡

案例背景：Scale AI 与自动驾驶/大模型数据

独角兽企业 Scale AI 早期依赖大量人工对图像和文本进行标注，供给给大模型训练。如今，他们引入了 AI 预标注和自动化工具。

供需曲线的动态变动分析：

- **阶段一：高成本时期（均衡点在左上方）**

- 纯人工标注导致**开发成本高**，数据供给曲线 S 处于较高位置，导致高质量训练数据价格高昂，交易量受限

- **阶段二：技术进步引发的动态变动 ($S \rightarrow S_1$)**

- AI 辅助工具介入，大幅降低了数据的单位处理成本
- **供给曲线向右下方移动**，在相同价格下，供应商愿意提供更多数据
- **新的市场均衡**：高质量标注数据的**均衡价格下降**，总体的**均衡数量增加**

第二节

数据一级市场定价

数据一级市场定价概述

定价基础

在数据一级市场中，企业从事授权、采集和归集工作，将原始数据转化为可交易资源。

- **核心考量：** 转化的成本成为企业定价主要考虑的因素
- **准则：** 以最小的成本实现利润最大化
- **主要方法：** 会计学成本法

数据一级市场定价策略：成本结构

核心视角——供给：生产成本

1. 企业生产函数

- 数据挖掘成本
- 脱敏成本
- 标准化成本
- 运储成本
- 隐私风险

2. 数据产品特性

- 可变成本
 - 平均成本体系规模效益
 - 复制边际成本 ≈ 0
 - 使用者的单位成本体现
- 固定成本
 - 存储成本

成本法定价的优劣与方法

优势

- **客观性**：不受市场价格波动影响
- **易操作**：定价成本较低，简单直接

劣势

- 基于历史成本，无法体现数据的**未来发展潜力**
- 往往会**低估**数据的价值

具体方法分类：

- ① **直接成本法**：将劳动力、授权费用等直接成本作为计算基础
- ② **间接成本法**：将存储、销售、折旧费用等间接成本分摊到各个数据产品
- ③ **标准成本法**：以标准成本为基础计算单个数据产品的生产成本
- ④ **作业成本法**：将成本分配到流程作业环节，再分摊到各个数据产品

案例分析：AI 数据标注平台

案例背景

数据服务平台（如 Scale AI/百度众包）需要将原始的图像、文本等数据经人工 + 机器清洗标注，转化为 AI 大模型训练可用的一级市场数据资源。

成本法定价的具体商业体现：

- **直接成本法的应用**：按标注员的计件工资或工时（如拉框、语义打标数量）计算最直接的人力支出，这直接构成了数据包的保底价格
- **作业成本法的应用**：将流程拆分为数据采集 → 自动预处理 → 人工标注 → 质检验收四个作业环节，分别核算每一步的服务器算力耗费与各环节管理成本，最终加总形成对外报价

数据一级市场定价的实践应用

案例：加拿大统计局

采用成本法估计数据资产的价值。

估算步骤：

- 1 在国家职业分类中筛选与数据、数据库、数据科学相关的职业
- 2 假设各职业耗费在生产数据上的时间占比，作为**直接劳动力成本**
- 3 **计算附加成本：**
 - 间接成本：设定为总工资成本的 50%
 - 资本服务加成：设定为 3%
- 4 加总计算得出投入价值

加拿大 2018 年数据资产估算结果

- **投入价值**：295 亿至 400 亿美元
 - 数据：94 亿至 142 亿
 - 数据库：80 亿至 116 亿
 - 数据科学：120 亿至 142 亿
- **存量价值**：1570 亿至 2170 亿美元
 - 占知识产权净资产存量的 **68.9% 至 95.2%**

案例分析：商业查询平台的数据归集

案例背景

商业查询平台（如天眼查/企查查）通过技术手段从公开渠道采集工商数据，经过清洗、关联并建立知识图谱后，以 API 或会员账号形式向外出售。

成本结构与规模效益的体现：

- **高昂的固定与初始化成本：** 初期需要投入巨额资金构建庞大爬虫集群、开发数据清洗模型、架构知识图谱
- **趋近于零的复制边际成本：** 一旦底层结构化数据库建立，多服务一个 API 调用客户或增添一位 VIP 会员的边际新增成本几乎为零
- **盈利模式归宿：** 前期基于归集与清洗历史成本设定底价，后期则完全依赖 B 端/C 端用户基数摊薄平均成本，展现了数据产品的**规模经济效应**

第三节

数据二级市场定价

数据二级市场定价概述

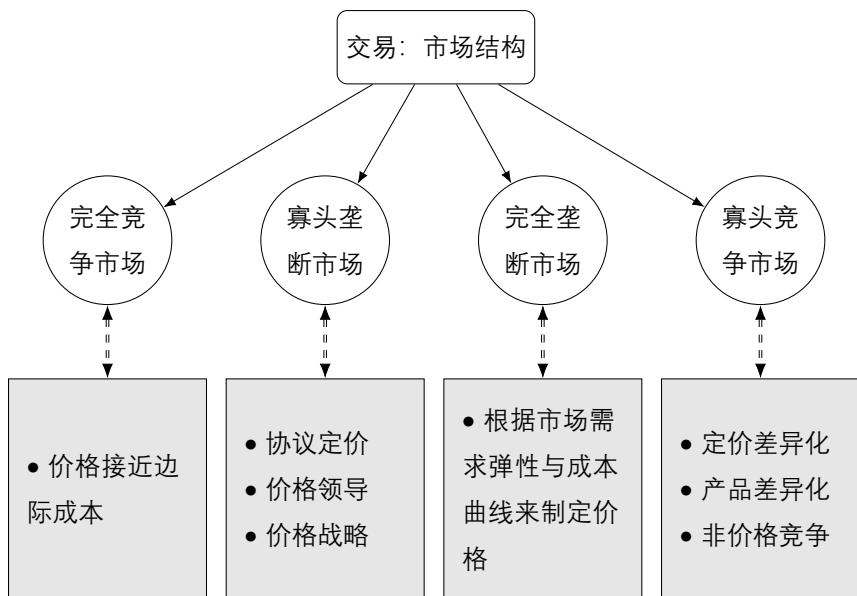
市场定位

二级市场是数据要素市场，介于一级市场（授权/采集）和三级市场（产品/服务）之间。

定价考量的特殊性：

- 相对一级市场：需额外考虑要素市场结构。
- 相对三级市场：需充分考虑数据获取成本。

数据二级市场定价策略：结构决定策略



案例分析：API 中介平台的差异化定价

案例背景：API 数据中介平台的寡头竞争策略

API 数据中介平台（如聚合数据、阿里云数据市场）通过向上游集中采购天气、交通、工商等底层数据，打包成标准化 API 出售给下游开发者。

理论与现实的映射：

- **充分考虑获取成本**：平台定价的底线锚定向上游（一级市场）采购授权的固定成本，叠加自身接口维护的边际成本
- **产品差异化**：即使是同样的天气数据，平台也会根据时效性、每秒并发量、服务可用性区分为免费体验版、开发者版和企业高可用版
- **定价差异化（阶梯定价）**：根据调用次数（如 1 万次/10 万次/不限量包年）实施非线性定价，调用量越大，单次调用价格越低

定价策略的深化与现状

- **博弈论视角：**

- 需考虑买卖双方互动策略（如 Stackelberg 模型）
- 关注消费者异质性及数据产品的协同性

- **中国市场现状：**

- 定价仍以**卖方报价**为主
- 受市场势力影响大：
 - 平台型头部企业既是“上游”又是“对手”，往往成为**价格领导者**

案例分析：Wind（万得）的定价权与市场势力

案例背景：中国金融数据市场的“价格领导者”

Wind（万得）在国内金融终端和金融底层结构化数据市场占据绝对的领先份额，是券商、基金等金融机构不可或缺的二级市场数据源。

理论与现实的映射：

- **卖方报价主导**：由于金融核心数据的收集清洗壁垒，以及习惯了 Wind 的指标体系的金融从业者高昂的转换成本，Wind 在定价中拥有绝对话语权
- **价格领导者**：作为寡头垄断/寡头竞争市场中的绝对头部，Wind 的涨价行为往往成为行业标杆，如东方财富、同花顺 iFinD 等同类竞争者通常采取对标 Wind 稍微折价的跟随定价策略
- **利用消费者异质性进行博弈**：针对大型头部券商和小型私募机构，Wind 在实际商业谈判中，会利用其不可替代性进行价格歧视与强硬博弈

案例：贵阳大数据交易所

全国第一家大数据交易所，截至 2023 年 3 月交易额达 6.3 亿元。

三种定价模式：

- ① 协议定价：协商确定
- ② 固定定价：明码标价
- ③ 实时定价：取决于样本量和单一样本指标项价值，系统自动浮动

创新工具：数据产品交易价格计算器

- **背景**：国家发改委价格监测中心指导
- **模型**：建立多因子修正成本模型
- **考量因子**：
 - 准确性、质量
 - 稀缺性、时效性
- **功能**：自动化计算建议价格
- **意义**：补全“报价—估价—议价”路径，促进公平贸易

第四节

数据三级市场定价

数据三级市场定价概述

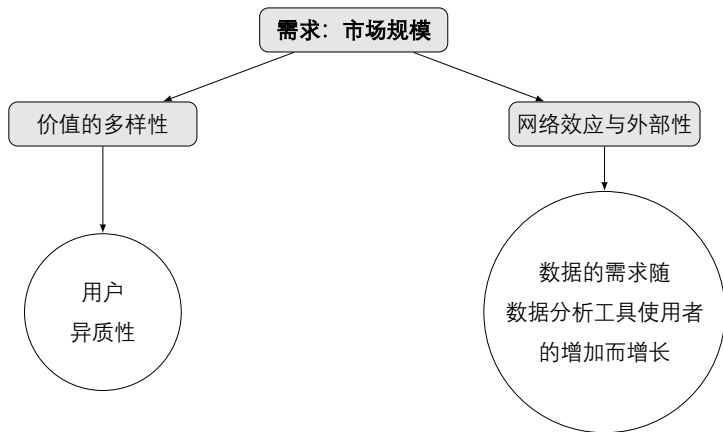
市场特征

三级市场是数据产品/服务的交易市场。与传统市场相似，但具有**极强的买方异质性**和复杂的市场结构。

核心策略：

- 主要是以**需求**为导向
- 需通过多种定价机制促进多主体参与

数据三级市场定价策略：需求驱动



具体定价机制：静态与动态

1. 静态定价机制

- 固定定价：
 - 优势：交易成本低
 - 局限：难以应对市场波动
- 差别定价：
 - 针对不同需求者设不同价格
 - 因买方异质性，具有一定合理性

2. 动态定价机制

- 自动实时定价：系统计算撮合，效率高但实施难
- 拍卖式定价：适合一卖多买、价值难定场景，激励转让
- 协商定价：双方接受，但时间成本高

案例分析：字节跳动（巨量引擎）的动态定价

案例背景：数字广告市场的“拍卖与实时定价”

数字广告是典型的数据三级市场产品。字节跳动通过巨量引擎，将用户标签与注意力转化为数据服务，出售给异质性极强的广告企业。

定价机制的现实映射：

- **拍卖式定价（RTB 实时竞价）**：广告企业针对特定目标受众进行竞价，出价高且预估点击率高的买方获得展示机会，有效解决一卖多买的定价难题
- **自动实时定价**：毫秒级的系统自动撮合，数据底价会随着用户的活跃度、特定时段的竞争激烈程度（如“双十一”大促）动态浮动
- **差别定价**：不同行业的买方（如高利润的游戏公司 vs. 传统餐饮），由于对流量转化价值的预期不同，针对特定目标受众的最终成交价差异巨大

具体定价机制：信息熵定价法

理论基础：香农信息论 (Shannon, 1948)

信息熵表示信息中排除冗余后的平均信息量。

核心思想：

- 事件的不确定性越大，信息熵越大，价值越高
- 公式： $H = -\sum_{i=1}^k p_i \ln p_i$ ，其中 H 为信息熵， p_i 为各种可能结果的概率

评价：

- 优点：充分考虑稀缺性，强调信息量
- 局限：操作难度大，适用范围有限

案例分析：Orbital Insight 与另类数据定价

案例背景：利用“信息熵”给卫星数据定价

Orbital Insight 是一家利用 AI 分析卫星图像（如统计沃尔玛停车场车辆数、监测原油储罐阴影）的数据服务商，核心客户是对冲基金和跨国企业。

理论与商业的结合（信息熵与异质性）：

● 消除不确定性（降低信息熵）：

- 对冲基金在上市零售商财报发布前，面临极大的业绩不确定性
- 卫星数据分析越准确，排除了越多的冗余猜测，其定价就越高

● 买方异质性决定价格上限：

- 同样是一张特定地区的卫星照片，对于普通地图应用而言只是基础图像
- 对于急需评估该地区供应链断裂风险的跨国企业，它消除了高额损失的不确定性，因此能以高协商定价成交

数据三级市场定价的实践应用

- **主要模式：** 智能定制化服务（通过谈判和合作完成）
- **典型案例：** 某互联网平台为跨国快餐企业提供选址服务
- **合作细节：**
 - **供给：** 利用定位数据分析分布密度、消费能力
 - **需求：** 快餐企业获得市场洞察和选址依据
 - **定价考量：** 数据价值、分析复杂度、人力成本、竞争情况

本章要点总结

数据市场均衡的核心逻辑

数据市场均衡是供需力量相互作用的动态结果。由于数据的特殊属性（非竞争性、可复制性等），其均衡过程比传统市场更依赖制度支撑。

1. 均衡的八大前提：

- **制度保障**：数据确权（基础）、数据脱敏（安全）、有效监管（秩序）
- **交易效率**：标准化（降低成本）、无交易成本（理想状态）
- **市场质量**：信息完全（消除悖论）、竞争性（地位对等）、无外部性

2. 动态调节机制：

- 超额供给 → 价格下降；超额需求 → 价格上升
- 技术进步推动供给增加（ S 右移），替代品涨价拉动需求增加（ D 右移）

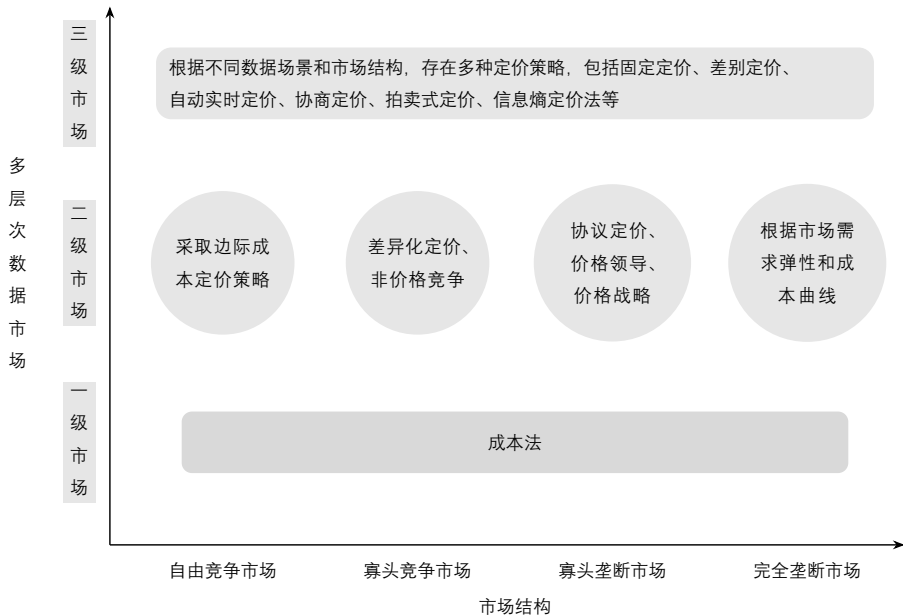
本章要点总结

市场层次	核心逻辑	主要定价策略与工具
一级市场	成本导向	会计学成本法（直接/间接/作业成本）、加拿大统计局估算法
二级市场	结构导向	价格领先制（如 Wind）、边际成本定价（完全竞争）、博弈定价（Stackelberg 模型）
三级市场	需求导向	智能定制化服务、动态定价（RTB 拍卖、实时算法）、信息熵定价法

学习建议：从“结构”看“定价”

- 一级看投入：理解数据的“资源化”过程。
- 二级看对手：关注企业在要素市场中的市场势力。
- 三级看场景：关注买方异质性如何驱动价值的最终释放。

总结：多层次数据市场定价策略汇总



谢谢大家!