产业组织学

Industrial Organization

自学讲义

2026 年 5 月 22 日

市场势力从何而来？

我们应该如何治理它？

全书共 15 章正文 + 2 个附录，覆盖从 1838 年 Cournot 模型到 2025 年 AI 反垄断的完整智识旅程。

目录

第一章 什么是产业组织学？ 1

1.1 核心问题 . . . 1
1.2 不完全竞争的普遍性 . . . 2
1.2.1 教科书世界 vs. 现实世界 . . . 2
1.2.2 市场势力的来源 . . . 2
1.2.3 一个经典的定量快照 . . . 3
1.3 一个永恒的张力：结构 vs. 行为 . . . 3
1.3.1 结构论 . . . 3
1.3.2 行为论 . . . 4
1.3.3 为什么这个张力如此重要？ . . . 4
1.4 产业组织学的三根支柱 . . . 5
1.4.1 理论（Theory） . . . 5
1.4.2 实证（Empirics） . . . 5
1.4.3 政策（Policy） . . . 6
1.5 全书路线图 . . . 6
1.6 方法论预览 . . . 8
1.6.1 博弈论 . . . 8
1.6.2 计量经济学 . . . 9
1.6.3 政策分析框架 . . . 9
1.7 关键概念预登记 . . . 10
1.8 本章小结 . . . 10
1.9 延伸阅读 . . . 11

第二章 古典奠基：Cournot, Bertrand, Hotelling 13

2.1 核心问题 . . . 13
2.2 历史背景：1838–1933 . . . 14
2.2.1 Cournot 之前的经济学世界 . . . 14
2.2.2 Cournot 为什么被忽视了？ . . . 14
2.2.3 1883–1933：回应、反驳与综合 . . . 14

2.3 Cournot 模型 (1838) . . . 15
2.3.1 设定 . . . 15
2.3.2 均衡推导：$n = 2$ 情形 . . . 16
2.3.3 一般化到 n 个企业 . . . 16
2.3.4 比较静态与 Cournot 极限定理 . . . 17
2.3.5 Lerner 指数与 Cournot . . . 17
2.3.6 图形直觉 . . . 18
2.4 Bertrand 模型 (1883) . . . 18
2.4.1 设定 . . . 19
2.4.2 均衡推导：削价逻辑 . . . 19
2.4.3 Cournot vs. Bertrand：根本分歧 . . . 19
2.4.4 Bertrand 悖论的五种化解方案 . . . 20
2.4.5 不对称成本的 Bertrand . . . 20
2.5 Edgeworth 模型 (1897) 与产能约束 . . . 20
2.5.1 核心创新 . . . 21
2.5.2 三个区域 . . . 21
2.5.3 Edgeworth 价格周期 . . . 21
2.5.4 Kreps-Scheinkman (1983)：Cournot 与 Bertrand 之间的桥梁 . . . 21
2.5.5 现代应用 . . . 22
2.6 Hotelling 模型 (1929)：空间竞争 . . . 22
2.6.1 为什么需要 Hotelling？ . . . 22
2.6.2 线性城市的设定 . . . 22
2.6.3 基准：固定价格的均衡 . . . 23
2.6.4 完整的两阶段博弈 . . . 23
2.6.5 d'Aspremont 修正：从最小差异化到最大差异化 . . . 24
2.6.6 比较静态 . . . 24
2.6.7 Hotelling 的现代身影 . . . 24
2.7 Chamberlin (1933) 与垄断竞争 . . . 25
2.7.1 从寡头到" 众多企业" . . . 25
2.7.2 设定 . . . 25
2.7.3 dd 曲线与 DD 曲线——Chamberlin 的创新 . . . 25
2.7.4 长期均衡与过剩产能定理 . . . 26
2.7.5 Chamberlin vs. Cournot/Bertrand/Hotelling . . . 26
2.7.6 从遗忘到复兴：Dixit-Stiglitz (1977) . . . 26
2.8 Joan Robinson (1933) 与不完全竞争 . . . 27
2.8.1 与 Chamberlin 的分野 . . . 27
2.8.2 Robinson 的两个主要分析贡献 . . . 27

2.9 本章小结 28
2.10 习题 29
2.11 延伸阅读 30

第三章 结构-行为-绩效：一个范式的诞生 33
3.1 核心问题 33
3.2 历史背景：为什么是 1930 年代的哈佛？ 34
3.2.1 四种力量的汇聚 34
3.2.2 Mason (1939)：SCP 的创立文献 34
3.2.3 Mason 1949：SCP 进入反垄断 35
3.3 Bain (1956)：进入壁垒的系统理论 36
3.3.1 从 Mason 到 Bain：一个博士生将愿景变为纲领 36
3.3.2 Bain (1951)：第一次系统检验——42 个制造业 36
3.3.3 Bain (1956)：《新竞争者的壁垒》——IO 的” 创世记” 37
3.3.4 Bain 的进入壁垒三分类 37
3.3.5 Bain 的 20 个行业数据：壁垒与利润 37
3.3.6 Bain 的方法论遗产与早期局限 40
3.4 SCP 范式的完整表述 40
3.4.1 Bain 的单向模型 (1959) 40
3.4.2 Scherer (1970)：SCP 的” 圣经” 41
3.4.3 SCP 流程图 42
3.5 测量工具：集中度与进入壁垒的操作化 43
3.5.1 集中度的测量：从 CR_n 到 HHI 43
3.5.2 进入壁垒的测量：Bain 之后 45
3.5.3 利润的测量问题（预览 Ch.4 的批判） 45
3.6 结构主义的政策含义 46
3.6.1 “大就是坏”：从分析到行动 46
3.6.2 标志性判例：SCP 逻辑写入法律 46
3.6.3 1968 年 DOJ 合并指南：结构主义的制度化 47
3.6.4 德国平行发展：秩序自由主义（Ordoliberalism） 47
3.7 从古典模型到 SCP：一个认知跃迁 48
3.7.1 Cournot 的知识生产方式 vs. Bain 的知识生产方式 48
3.7.2 全书的关键知识链 49
3.7.3 从 Mason 的” 泥泞” 到 Tirole 的” 清晰” 49
3.8 本章小结 49
3.9 习题 50
3.10 延伸阅读 53

第四章 SCP 的实证遗产与衰落 **55**

4.1 核心问题 . . . 56
4.2 SCP 经验纲领的黄金时代（1960s–1970s） . . . 56
4.2.1 从 Bain 的手工分析到跨行业回归 . . . 56
4.2.2 关键经验发现 . . . 57
4.2.3 PIMS 数据库：SCP 的咨询业变体 . . . 60
4.2.4 1968 年 DOJ 合并指南：SCP 的政策巅峰 . . . 61
4.3 Demsetz 效率假说：芝加哥学派的第一枪 . . . 61
4.3.1 Demsetz (1973) 的核心论证 . . . 61
4.3.2 一个数值例子 . . . 62
4.3.3 Demsetz 的经验检验策略 . . . 63
4.3.4 Peltzman (1977) 的价格-成本分解 . . . 64
4.3.5 其他芝加哥学派的早期批评 . . . 65
4.4 测量问题：会计利润、市场界定与变量误差 . . . 65
4.4.1 Fisher-McGowan (1983)：会计利润的致命缺陷 . . . 65
4.4.2 市场界定问题 . . . 67
4.4.3 变量误差：集中度作为” 市场势力” 的代理变量 . . . 68
4.5 内生性问题：因果方向的根本不确定性 . . . 68
4.5.1 为什么相关性 $\neq$ 因果关系在这里特别严重 . . . 68
4.5.2 反向因果（Reverse Causality） . . . 69
4.5.3 遗漏变量（Omitted Variables） . . . 69
4.5.4 联立性（Simultaneity） . . . 70
4.5.5 Schmalensee (1985, 1989)：权威的” 验尸报告” . . . 70
4.5.6 集中度-利润关系的时变：DHP (1986–87) . . . 71
4.6 SCP 衰落的多重原因 . . . 72
4.6.1 五线并进的解释框架 . . . 72
4.6.2 各战线的相互作用 . . . 72
4.7 SCP 的遗产：哪些东西被保留？ . . . 75
4.7.1 遗产 1：HHI——从学术公式到全球监管标准 . . . 75
4.7.2 遗产 2：进入壁垒分析 . . . 76
4.7.3 遗产 3：市场界定的核心地位 . . . 76
4.7.4 遗产 4：结构性推定的政策直觉 . . . 77
4.7.5 遗产 5：SCP 在邻域的存续 . . . 78
4.7.6 遗产清单 . . . 80
4.8 从 SCP 到 NEIO：一个研究纲领的嬗变 . . . 80
4.8.1 Bresnahan (1989) 的诊断与处方 . . . 80
4.8.2 行为参数 θ：度量市场势力的新方法 . . . 80

4.8.3 SCP 跨行业回归 vs. NEIO 结构估计：全面的方法论对比 . . . 81
4.8.4 总结性时间线：IO 经验研究的范式转换 . . . 81
4.9 本章小结 . . . 81
4.10 习题 . . . 83
4.11 延伸阅读 . . . 86

第五章 效率 vs. 市场势力：芝加哥学派的颠覆 89
5.1 核心问题 . . . 90
5.2 “桌子旁的经济学”：芝加哥口述传统与 Aaron Director . . . 90
5.2.1 Aaron Director：没有论文的创始人 . . . 90
5.2.2 “一个垄断利润” 定理 . . . 91
5.2.3 反垄断项目 (The Antitrust Project, 1953–) . . . 92
5.2.4 芝加哥反垄断的” 神圣三位一体” . . . 93
5.2.5 芝加哥学派的两代人 . . . 93
5.3 Stigler 的进入壁垒重定义：一场定义之战 . . . 93
5.3.1 Bain 的原定义（简要回顾） . . . 93
5.3.2 Stigler (1968) 的精确措辞与逻辑 . . . 94
5.3.3 为什么规模经济不是壁垒 . . . 94
5.3.4 为什么资本需求不是壁垒——Modigliani-Miller 连接 . . . 95
5.3.5 幸存者原则（1958 年） . . . 96
5.3.6 在 Stigler 的定义下什么仍算作壁垒？ . . . 97
5.3.7 关键歧义：McAfee、Mialon 和 Williams (2004) . . . 97
5.3.8 Demsetz (1982)：Nirvana 谬误应用于壁垒 . . . 97
5.4 Demsetz 效率假说：从批判到世界观 . . . 99
5.4.1 在 Ch.4 基础上构建 . . . 99
5.4.2 “Nirvana 谬误”：芝加哥的方法论签名 . . . 99
5.4.3 Demsetz 论产权与效率 . . . 100
5.5 Posner 与 Bork：反垄断的重塑 . . . 101
5.5.1 Posner (1976)：反垄断的经济学重构 . . . 101
5.5.2 Posner (2001) 第二版：什么被改变、什么没有 . . . 102
5.5.3 Posner 作为第七巡回法院法官：关键判例 . . . 102
5.5.4 Bork (1978)：《反垄断悖论》 . . . 103
5.5.5 关键芝加哥反垄断概念 . . . 104
5.6 芝加哥方法论武器：Friedman、Coase 与 Telser . . . 106
5.6.1 Friedman (1953) 与” 好像” 方法论 . . . 106
5.6.2 Coase 的连接 . . . 107
5.6.3 Telser (1960) 与纵向约束的免费搭车者逻辑 . . . 109
5.7 政策的芝加哥化：里根革命与放松管制 . . . 110

5.7.1 里根政府的反垄断革命 . . . 110
5.7.2 1982 年合并指南 . . . 111
5.7.3 放松管制运动 . . . 111
5.7.4 四代合并指南：一个世纪的演变 . . . 112
5.8 胜利的限度：后芝加哥综合与当前的挑战 . . . 112
5.8.1 芝加哥的胜利有多完整？ . . . 112
5.8.2 后芝加哥综合的关键思想 . . . 112
5.8.3 三校对比：一个分析框架 . . . 114
5.8.4 错误成本辩论的继续 . . . 114
5.8.5 2023 年合并指南：钟摆的回摆 . . . 116
5.8.6 欧洲的平行道路："更加经济学的方法" . . . 116
5.9 本章小结 . . . 117
5.10 习题 . . . 119
5.11 延伸阅读 . . . 121

第六章 可竞争市场与进入威胁 125
6.1 核心问题 . . . 125
6.2 可竞争市场理论的核心框架 . . . 126
6.2.1 完全可竞争市场的定义 . . . 126
6.2.2 完全可竞争 $\neq$ 完全竞争 . . . 127
6.2.3 "打了就跑" 进入：可竞争性的约束引擎 . . . 127
6.2.4 三个核心定理 . . . 128
6.2.5 自然垄断的重新思考：次可加性 . . . 130
6.2.6 Baumol 的智识轨迹：AT&T 联系 . . . 130
6.3 政策的可竞争性化：放松管制运动 . . . 131
6.3.1 来自 Demsetz (1968) 的谱系 . . . 131
6.3.2 航空放松管制：可竞争性的伟大实验 . . . 131
6.3.3 1982 年合并指南：进入分析框架的制度化 . . . 133
6.4 枢纽-辐射式批判：可竞争性在经验上的失败 . . . 133
6.4.1 Borenstein (1989)：枢纽与高票价 . . . 133
6.4.2 为什么可竞争性在航空枢纽失灵？ . . . 134
6.4.3 Morrison & Winston (1987)：正式检验 . . . 134
6.4.4 Southwest 效应：实际 $>$ 潜在 . . . 136
6.5 理论批判：为什么完全可竞争性在逻辑上失效 . . . 137
6.5.1 Shepherd (1984)：同义反复批判 . . . 137
6.5.2 Schwartz & Reynolds (1983)：微小沉没成本问题 . . . 137
6.5.3 其他理论批判 . . . 138
6.6 博弈论批判：策略行为为什么重要 . . . 139

6.6.1 可竞争性的核心盲点：在位者是被动的 . . . 139
6.6.2 Dixit (1980)：产能作为承诺工具 . . . 140
6.6.3 Fudenberg & Tirole (1984)：策略行为的分类学 . . . 140
6.6.4 可竞争性 → 博弈论的完美过渡 . . . 141
6.7 Sutton 与内生沉没成本 . . . 142
6.7.1 外生 vs. 内生沉没成本 . . . 142
6.8 可竞争性的遗产：什么留下了，什么没有 . . . 143
6.8.1 持久的概念贡献 . . . 143
6.8.2 “基准”辩护：Baumol & Willig (1986) . . . 143
6.8.3 “可竞争性”一词的制度化 . . . 144
6.8.4 当代相关性与前言 . . . 144
6.9 本章小结 . . . 145
6.10 习题 . . . 147
6.11 延伸阅读 . . . 149

第七章 博弈论重塑产业组织：Tirole 的统一框架 153
7.1 核心问题 . . . 153
7.2 博弈论以前的 IO：困境与缺口 . . . 154
7.2.1 SCP 的”没有理论的实证” . . . 154
7.2.2 芝加哥的”没有策略互动的价格理论” . . . 155
7.2.3 博弈论填补的三个空白 . . . 155
7.3 博弈论工具箱：核心概念与 IO 应用 . . . 156
7.3.1 Nash 均衡 → Cournot 与 Bertrand . . . 157
7.3.2 子博弈完美均衡 → Stackelberg 与 Dixit 承诺 . . . 160
7.3.3 Fudenberg-Tirole 策略动物分类学 . . . 161
7.3.4 Bayesian Nash 均衡 → 拍卖与不完全信息寡头 . . . 162
7.3.5 重复博弈与无名氏定理 → 合谋 . . . 163
7.3.6 完美 Bayesian 均衡 → 信号传递与声誉 . . . 165
7.3.7 Markov 完美均衡 → 动态寡头 . . . 166
7.4 Tirole (1988) 的体系结构 . . . 167
7.4.1 总体结构 . . . 167
7.4.2 为什么”博弈论用户手册”在最后一章？ . . . 167
7.5 深层统一性：不同的 IO 问题，相同的博弈结构 . . . 170
7.5.1 统一性对比表 . . . 170
7.5.2 三条统一原理 . . . 170
7.6 方法论对比：旧 IO vs. 新 IO . . . 172
7.7 博弈论 IO 的批评与回应 . . . 172
7.7.1 Fisher (1989)：”任何事都可以发生”批判 . . . 172

7.7.2 Sutton (1991) 的边界方法：一个建设性回应 174
7.7.3 NEIO 的经验回应（预览第 ~10~ 章） 174
7.8 博弈论 IO 的中国脉络 175
7.8.1 核心中文教材与翻译 175
7.8.2 张维迎与博弈论 IO 的中国引入 175
7.9 本章小结 176
7.10 习题 177
7.11 延伸阅读 179

第八章 策略行为的微观基础：进入遏制、掠夺与合谋 183
8.1 核心问题 183
8.2 进入遏制 I：产能承诺与限价信号 184
8.2.1 进入遏制的博弈论骨架：一个统一的七步框架 184
8.2.2 Dixit (1980)：产能承诺模型 186
8.2.3 战略替代品与战略互补品 187
8.2.4 Fudenberg-Tirole (1984)：策略动物分类学 188
8.2.5 Milgrom-Roberts (1982)：限价作为信号 190
8.3 进入遏制 II：产品增殖、提高对手成本与排他性合同 192
8.3.1 Schmalensee (1978)：品牌增殖（Brand Proliferation） 192
8.3.2 Salop & Scheffman (1983)：提高对手成本（Raising Rivals' Costs, RRC） 192
8.3.3 Aghion & Bolton (1987)：排他性合同（Exclusive Contracts） . 193
8.3.4 Gilbert & Newbery (1982)：先占专利（Preemptive Patenting） 194
8.4 掠夺性定价 I：芝加哥批判与连锁店悖论 194
8.4.1 什么是掠夺性定价？ 194
8.4.2 芝加哥批判：掠夺”从不理性” 195
8.4.3 Selten (1978)：连锁店悖论——博弈论”支持”了芝加哥 195
8.5 掠夺性定价 II：后芝加哥复兴 197
8.5.1 声誉机制：”四人帮”（Kreps-Milgrom-Roberts-Wilson 1982） 197
8.5.2 深口袋机制：Bolton & Scharfstein (1990) 198
8.5.3 信号模型 199
8.5.4 法律标准：从 Areeda-Turner 到 AKZO 200
8.6 合谋 I：囚徒困境、重复博弈与无名氏定理 200
8.6.1 囚徒困境作为合谋的原始结构 201
8.6.2 无限重复博弈中的触发策略 201
8.6.3 无名氏定理（The Folk Theorem） 202
8.6.4 最优惩罚：Abreu (1986) 203
8.6.5 Green & Porter (1984)：不完全监测下的价格战 203

8.6.6 Rotemberg & Saloner (1986)：繁荣期的价格战 204
8.7 合谋 II：实证识别、便利因素与瓦解 205
8.7.1 Porter (1983)：联合执行委员会（JEC）——结构估计的先驱 . 205
8.7.2 便利合谋的因素 206
8.7.3 合谋的实证检测 207
8.7.4 宽大计划（Leniency Programs）——用博弈论瓦解合谋 207
8.8 策略行为的中国脉络 208
8.8.1 中国《反垄断法》中的策略行为 208
8.8.2 标志性案例 208
8.8.3 中国合谋执法的特征 209
8.9 本章小结 209
8.10 习题 211
8.11 延伸阅读 214

第九章 不对称信息与规制：Laffont-Tirole 激励理论 217
9.1 核心问题 217
9.2 规制面临的根本问题：信息不对称 218
9.2.1 自然垄断与规制的必要性 218
9.2.2 逆向选择与道德风险：两种信息不对称的交织 219
9.2.3 公共资金的影子成本 λ：连接规制与宏观财政 220
9.3 Laffont-Tirole (1986) 基准模型 221
9.3.1 机制设计五步框架 221
9.3.2 模型设定——两种类型情形 222
9.3.3 完全信息基准（First-Best） 223
9.3.4 不对称信息下的最优规制 224
9.3.5 哪些约束是紧的？——机制设计的标准逻辑 224
9.3.6 最优解与核心结果 225
9.3.7 最优线性成本报销规则 226
9.3.8 比较静态与实证含义 227
9.4 菜单规制与显示原理 227
9.4.1 从两种类型到连续类型：直觉过渡 227
9.4.2 显示原理（Revelation Principle, Myerson 1979） 228
9.4.3 单交叉条件与分离的可行性 229
9.4.4 与 Ch.8 信号模型的对比 229
9.5 机制设计基础：Baron-Myerson 与最优定价 229
9.5.1 Baron-Myerson (1982) 模型 230
9.5.2 最优定价公式 231
9.5.3 单交叉条件与单调性 232

9.6 防合谋原理与规制俘获 . . . 232
9.6.1 从两层到三层：谁来监督监督者？ . . . 232
9.6.2 防合谋原理（Collusion-Proofness Principle） . . . 233
9.6.3 规制俘获：从 Stigler 描述到 Laffont-Tirole 微观基础 . . . 233
9.7 不完备合同与 Maskin-Tirole 定理 . . . 234
9.7.1 不完备合同问题 . . . 234
9.7.2 Maskin-Tirole (1999) 定理 . . . 235
9.8 应用与实践：全球激励规制改革 . . . 236
9.8.1 价格上限规制（RPI-X）vs. 回报率规制 . . . 236
9.8.2 标尺竞争（Yardstick Competition, Shleifer 1985） . . . 237
9.8.3 接入定价（Access Pricing） . . . 238
9.8.4 全球规制改革浪潮 . . . 238
9.9 中国语境：从国企承包制到现代规制体系 . . . 239
9.9.1 SOE 改革的三个阶段：一个激励规制的视角 . . . 239
9.9.2 自然垄断行业的规制改革 . . . 240
9.9.3 张维迎-林毅夫辩论：一个激励规制理论的解读 . . . 241
9.10 现代前沿：数字平台、数据与气候规制 . . . 242
9.10.1 数字平台的规制 . . . 242
9.10.2 AI 与算法规制 . . . 243
9.10.3 气候规制中的激励设计 . . . 243
9.11 章末总结与通往 NEIO 的桥梁 . . . 243
9.11.1 核心概念回顾 . . . 243
9.11.2 核心公式 . . . 244
9.11.3 方法论总结：机制设计五步法 vs. Tirole 七步建模法 . . . 244
9.11.4 通往 Ch.10 的桥梁 . . . 245

第十章 从 SCP 到结构估计：Bresnahan 纲领与新实证产业组织 251
10.1 核心问题 . . . 251
10.2 SCP 的黄昏：为什么需要 NEIO . . . 252
10.2.1 SCP 跨行业回归的逻辑与根本局限 . . . 252
10.2.2 从“行业间比较”到“行业内估计”：NEIO 的方法论转向 . . . 253
10.2.3 芝加哥学派对 SCP 的挑战与 NEIO 的回应 . . . 254
10.3 Bresnahan (1982, 1989)：行为参数与需求旋转 . . . 255
10.3.1 从企业 FOC 到行为参数 θ：一个统一的定价方程 . . . 255
10.3.2 识别问题的精确陈述：为什么 θ 和 MC 不可分离？ . . . 258
10.3.3 解决方案：需求旋转（Demand Rotation） . . . 258
10.3.4 识别条件的正式陈述 . . . 260
10.3.5 估计方法：GMM 与极大似然 . . . 260

10.3.6 嵌套检验与非嵌套检验：从 $\hat{\theta}$ 到行为推断 261
10.4 经典应用 I：Porter (1983) 与铁路卡特尔 262
10.4.1 制度背景：19 世纪末 JEC 铁路卡特尔 262
10.4.2 识别策略：五大湖通航状态作为需求旋转变量 263
10.4.3 Regime-Switching 模型：合谋与价格战的切换 263
10.4.4 核心结果：合谋 $\theta \approx 0.9$，价格战 $\theta \approx 0.1$ 264
10.4.5 Porter 的方法论遗产 265
10.5 经典应用 II：更多识别策略 265
10.5.1 Bresnahan (1987)：美国汽车业——区分 Bertrand、Cournot 与合谋 266
10.5.2 Genesove-Mullin (1998)：糖业卡特尔的“黄金标准”验证 . . . 267
10.6 识别问题的深化：从需求旋转到 Corts 批判 269
10.6.1 识别条件的正式证明（简化版） 269
10.6.2 需求旋转变量的经典选择 269
10.6.3 Corts (1999) 批判：当 MC 也是弹性的 270
10.7 生产函数方法 I：生产率估计 271
10.7.1 生产函数估计的核心问题：未观测的生产率 ω 272
10.7.2 Olley-Pakes (1996)：用投资作为生产率代理 272
10.7.3 Levinsohn-Petrin (2003)：用中间投入替代投资 274
10.7.4 Ackerberg-Caves-Frazer (2015)：函数依赖性批判 274
10.8 生产函数方法 II：加成率估计（De Loecker-Warzynski 2012） 275
10.8.1 核心公式：$\mu = \theta^V / \alpha^V$ 275
10.8.2 实施细节与挑战 276
10.8.3 De Loecker-Eeckhout-Unger (2020)：全球加成率上升趋势 . . . 277
10.9 结构估计的方法论反思 278
10.9.1 “结构”到底意味着什么？ 278
10.9.2 结构估计的优势与代价 279
10.9.3 GMM、嵌套/非嵌套检验与模型选择 280
10.9.4 Corts 批判的再审视与“MC 弹性”问题 281
10.10 中国语境：NEIO 在中国的应用与挑战 282
10.10.1 中国的 NEIO 实证研究概览 282
10.10.2 中国市场对 NEIO 方法的独特挑战 283
10.10.3 中国 NEIO 研究的最佳实践建议 284
10.11 章末总结与通往 BLP 的桥梁 284
10.11.1 核心概念回顾 284
10.11.2 核心公式回顾 285
10.11.3 本章的方法论框架：NEIO 结构估计四步法 286

10.11.4 通往 Ch.11 的桥梁 286

第十一章 BLP 与需求估计：差异化产品市场的结构需求估计 293

11.1 问题的背景：为什么差异化产品需求估计是一个难题？ 295
11.1.1 从同质产品到差异化产品：一个世界的过渡 295
11.1.2 维度灾难：$J(J-1)/2$ 个弹性的问题 296
11.1.3 价格内生性：不可观测质量的挑战 297
11.2 BLP (1995) 的解决方案：随机系数 Logit 模型 298
11.2.1 消费者效用设定 299
11.2.2 均值效用分解与线性参数集中化 300
11.2.3 外部商品（Outside Good）的关键作用 301
11.2.4 个体选择概率与市场份额 302
11.2.5 随机系数如何打破 IIA 302
11.3 关键方法论创新 304
11.3.1 收缩映射——Berry (1994) 的反解定理 304
11.3.2 BLP 工具变量：从寡头定价理论到价格工具变量 306
11.3.3 模拟积分与数值方法 308
11.4 GMM 估计与嵌套不动点（NFP）算法 308
11.4.1 矩条件与 GMM 框架 308
11.4.2 参数分区与线性参数集中化 309
11.4.3 完整的 NFP 算法伪代码 310
11.4.4 MPEC 替代方法 312
11.4.5 假设检验 313
11.5 识别理论 313
11.5.1 Berry & Haile (2014, 2016) 的核心结果 313
11.5.2 弱工具变量问题 314
11.6 供给端：Bertrand-Nash 定价与联合估计 315
11.6.1 多产品企业 Bertrand-Nash FOC 315
11.6.2 需求-供给联合 GMM 316
11.7 经典实证应用全景 316
11.7.1 BLP (1995) 原始应用——美国汽车市场 316
11.7.2 Nevo (2001)——即食麦片市场 317
11.7.3 Petrin (2002)——微型面包车的福利 317
11.8 Nevo (2000) 实践指南：从理论到实现 318
11.8.1 数据准备与市场定义 319
11.8.2 人口统计互动与完整异质性 319
11.8.3 从估计到政策分析 319
11.9 拓展与变体 320

11.9.1 嵌套 Logit 与 BLP 的比较 . . . 320
11.9.2 纯特征模型（Pure Characteristics Model） . . . 321
11.9.3 动态 BLP . . . 321
11.9.4 微观与宏观数据的整合 . . . 321
11.10 现代工具与前沿 . . . 322
11.10.1 PyBLP (Conlon & Gortmaker 2020) . . . 322
11.10.2 其他实现工具比较 . . . 322
11.10.3 当前研究前沿 . . . 322
11.11 局限、反思与通往 Ch.12 的桥梁 . . . 323
11.11.1 BLP 的主要局限 . . . 323
11.11.2 BLP 的方法论遗产 . . . 324
11.11.3 通往 Ch.12 的桥梁 . . . 325
练习题 . . . 326
推荐阅读 . . . 327

第十二章 并购模拟、市场界定与反垄断损害量化 331
12.1 引言：反垄断为什么需要量化分析？ . . . 331
12.1.1 反垄断的经济学逻辑 . . . 331
12.1.2 来自 Ch.10–11 的工具箱 . . . 333
12.1.3 本章结构 . . . 334
12.2 市场界定与 SSNIP 测试 . . . 335
12.2.1 SSNIP 测试的逻辑结构 . . . 335
12.2.2 临界损失分析 . . . 336
12.2.3 从临界损失到分流比：SSNIP 的现代版本 . . . 339
12.2.4 Cellophane Fallacy（玻璃纸谬误） . . . 339
12.2.5 市场界定的实证方法概览 . . . 340
12.3 并购模拟：单边效应 . . . 340
12.3.1 多产品 Bertrand-Nash 定价：从 FOC 到矩阵形式 . . . 341
12.3.2 并购前 → 并购后：五步模拟过程 . . . 343
12.3.3 一个完整的数值例子：两产品 logit 并购 . . . 345
12.3.4 分流比：单边效应的充分统计量 . . . 345
12.4 UPP 与 GUPPI：向上定价压力 . . . 348
12.4.1 UPP 的推导：从 FOC 到定价压力 . . . 348
12.4.2 GUPPI：标准化指数 . . . 349
12.4.3 Net UPP：纳入并购效率 . . . 349
12.4.4 UPP vs. 完整并购模拟：比较 . . . 350
12.4.5 Farrell–Shapiro (2010) 的范式转变 . . . 350
12.5 协调效应与并购效率 . . . 351

12.5.1 协调效应：并购如何促进合谋 . . . 351
12.5.2 并购效率 . . . 352
12.5.3 Williamson (1968) 权衡 . . . 353
12.6 反垄断损害量化 . . . 354
12.6.1 过度要价与 But-For 价格 . . . 354
12.6.2 损害估计的四种方法 . . . 354
12.6.3 伞形效应 . . . 355
12.7 计算实现：PyBLP 与数值方法 . . . 355
12.7.1 求解并购后均衡的数值方法 . . . 356
12.7.2 PyBLP 中的并购模拟 . . . 356
12.7.3 常见数值问题 . . . 357
12.8 中国反垄断的制度语境 . . . 358
12.8.1 中国反垄断法的制度架构 . . . 358
12.8.2 中国并购审查中的量化分析 . . . 358
12.8.3 中美欧并购审查比较 . . . 359
12.9 总结、局限与通往 Ch.13 的桥梁 . . . 360
12.9.1 本章的核心方法论文档 . . . 360
12.9.2 局限与未解决的问题 . . . 360
12.9.3 通往 Ch.13 的桥梁：平台与双边市场 . . . 361

第十三章 双边市场与平台经济学 367

13.1 什么是双边市场？ . . . 367
13.1.1 从 Ch.12 到 Ch.13：为什么传统工具需要重装系统？ . . . 367
13.1.2 科斯定理为什么失败了？ . . . 368
13.1.3 经典例子与概念区分 . . . 369
13.1.4 双边市场的正式定义：价格结构非中性 . . . 371
13.2 倾斜定价：为什么一侧免费？ . . . 372
13.2.1 Rochet-Tirole (2003) 垄断平台模型 . . . 372
13.2.2 机会成本解释：最直观的理解 . . . 373
13.2.3 谁被补贴？——决定因素框架 . . . 374
13.2.4 为什么不是永远最大化倾斜？ . . . 375
13.2.5 Armstrong (2006) 会员费模型的补充视角 . . . 376
13.3 竞争性瓶颈：单归属 vs. 多归属 . . . 376
13.3.1 Armstrong (2006) 的三个模型 . . . 376
13.3.2 实证匹配：App Store 与新闻媒体 . . . 379
13.3.3 部分多归属与扩展 . . . 379
13.4 鸡-蛋问题与临界规模 . . . 380
13.4.1 协调失灵的形式化 . . . 380

13.4.2 Caillaud-Jullien (2003) 模型的核心结构 380
13.4.3 分而治之策略的数学逻辑 381
13.4.4 解决鸡-蛋问题的策略工具箱 382
13.4.5 Jullien (2011) 的 Stackelberg 扩展 384
13.5 价格歧视与平台设计 384
13.5.1 Rochet-Tirole (2006) 的统一框架 384
13.5.2 平台定价工具全景 385
13.5.3 Weyl (2010)：绝缘均衡与 Spence 扭曲 385
13.5.4 与经典 IO 的概念对接 386
13.6 市场倾覆与平台包围 387
13.6.1 市场倾覆的条件 387
13.6.2 多平台共存的条件 388
13.6.3 平台包围（Platform Envelopment） 388
13.7 双边市场中的反垄断挑战 390
13.7.1 市场界定：SSNIP 的三重崩溃 390
13.7.2 Ohio v. American Express (2018)：标志性判决 391
13.7.3 零价格市场中的市场势力：SSNDQ 与 SSNIC 392
13.7.4 平台并购分析：再平衡效应 393
13.7.5 掠夺性定价：何时低于成本是正常的？ 394
13.7.6 杀手并购（Killer Acquisitions） 395
13.7.7 全球规制格局比较 396
13.8 平台治理与 B2B 平台 397
13.8.1 B2B 平台的独特性 397
13.8.2 Springer et al. (2025) 的五种 B2B 平台原型 398
13.8.3 平台治理谱系 399
13.8.4 数据飞轮与 B2B 竞争的锁定效应 399
13.9 总结、局限与通往 Ch.14 的桥梁 400
13.9.1 核心要点回顾 400
13.9.2 理论与现实的差距 400
13.9.3 通往 Ch.14 的桥梁 401
练习 402
推荐阅读 404

第十四章 数字时代的竞争政策：算法、数据与 AI **407**
14.1 算法合谋：当定价算法学会”共谋” 407
14.1.1 从 Ch.13 到 Ch.14：竞争过程本身被改变了 407
14.1.2 Ezrachi-Stucke (2016) 四大情景框架【核心】 408
14.1.3 Calvano et al. (2020)：Q-learning 合谋实验【核心】 410

14.1.4 轴辐合谋与共同代理理论【进阶】 . . . 413
14.1.5 算法合谋的检测与竞争法应对【进阶】 . . . 415
14.2 数据作为进入壁垒：当信息成为护城河 . . . 418
14.2.1 数据网络效应（Data Network Effects, DNE）【核心】 . . . 418
14.2.2 数据护城河的经济理论【进阶】 . . . 421
14.2.3 隐私-竞争权衡（Privacy-Competition Tradeoff）【进阶】 . . . 422
14.2.4 数据可携与互操作性【前沿】 . . . 424
14.2.5 中国数据治理：数据作为生产要素的独特路径【前沿】 . . . 424
14.3 生成式 AI 与平台竞争：新范式的冲击 . . . 425
14.3.1 AI 与市场结构：赢者通吃还是促进竞争？【核心】 . . . 425
14.3.2 AI 对搜索与广告市场的冲击【核心】 . . . 427
14.3.3 AI 训练数据的反垄断问题【进阶】 . . . 427
14.3.4 AI 与纵向一体化：AI Stack 的竞争分析【进阶】 . . . 428
14.3.5 中国 AI 竞争格局：从”百模大战”到”五强争霸”【进阶】 . . . 428
14.3.6 Motta & Peitz (2024)：拒绝互操作性的经济学【前沿】 . . . 430
14.3.7 AI 时代的竞争政策：综合评估【前沿】 . . . 432
14.4 全球数字规制：四种制度实验 . . . 432
14.4.1 EU DMA：实施两年后的证据【核心】 . . . 432
14.4.2 UK DMCC Act 2024：量身定制的英国方案【进阶】 . . . 434
14.4.3 美国：诉讼、立法僵局与范式更替【进阶】 . . . 435
14.4.4 中国：从”整改风暴”到”常态化监管”【核心】 . . . 436
14.5 结论：数字时代竞争政策的三个未竟之问【核心】 . . . 439
练习 . . . 441
推荐阅读 . . . 443

第十五章 新布兰代斯学派与反垄断复兴 447
15.1 芝加哥学派的”反垄断遗产” . . . 448
15.1.1 从 Brandeis 到 Bork：反垄断法的范式转换 . . . 448
15.1.2 消费者福利标准（Consumer Welfare Standard, CWS）【核心】 449
15.1.3 错误成本框架（Error Cost Framework）【核心】 . . . 451
15.1.4 芝加哥学派的盲点与批评 . . . 453
15.2 新布兰代斯运动的兴起 . . . 454
15.2.1 布兰代斯的精神源头：为什么”大”本身就是问题？ . . . 454
15.2.2 催化时刻：Khan (2017) 与 Wu (2018) . . . 455
15.2.3 新布兰代斯学派的六大核心纲领 . . . 456
15.2.4 从象牙塔到制度权力：2021–2025 的制度实验 . . . 459
15.3 核心争论点 . . . 462
15.3.1 争论一：反垄断的目标是什么？ . . . 462

15.3.2 争论二：谁的错误更昂贵？ . . . 463
15.3.3 争论三：结构性救济还是行为性救济？ . . . 464
15.3.4 争论四：谁应该承担举证责任？ . . . 465
15.3.5 争论五：经济学在反垄断中应扮演什么角色？ . . . 466
15.4 制度现实与学术回应 . . . 467
15.4.1 法院作为制度壁垒 . . . 467
15.4.2 学术的四类立场 . . . 468
15.4.3 2025–2026 年的制度剧变 . . . 469
15.4.4 国际比较：美国是异类 . . . 469
15.5 中国的位置：第三条道路？ . . . 470
15.5.1 ”社会主义市场经济”的反垄断法 . . . 470
15.5.2 ”防止资本无序扩张”——中国独有的反垄断话语 . . . 471
15.5.3 2020–2022 整改风暴与标志性案件 . . . 472
15.5.4 从”强监管”到”常态化监管” . . . 474
15.5.5 中国、欧盟、美国：三种反垄断哲学 . . . 475
15.5.6 全书收尾：回到 Ch.1 的核心问题 . . . 477
习题 . . . 478
推荐阅读 . . . 480

附录 A 数学附录：核心模型推导与计量方法速查 483
A.1 博弈论基础 . . . 483
A.1.1 策略式博弈与 Nash 均衡 . . . 483
A.1.2 扩展式博弈与子博弈完美均衡 . . . 484
A.1.3 不完全信息博弈与贝叶斯 Nash 均衡 . . . 484
A.1.4 重复博弈与 Folk 定理 . . . 484
A.2 核心模型完整推导 . . . 485
A.2.1 Cournot 寡头：n 个企业的一般解 . . . 485
A.2.2 Bertrand 均衡：差异化产品 . . . 486
A.2.3 Hotelling 线性城市：完整推导 . . . 486
A.2.4 Dixit 产能承诺模型 . . . 487
A.2.5 Laffont-Tirole 激励规制模型 . . . 487
A.3 计量方法速查 . . . 488
A.3.1 广义矩方法（GMM） . . . 488
A.3.2 非线性最小二乘（NLS） . . . 488
A.3.3 收缩映射（Contraction Mapping） . . . 489
A.3.4 Bootstrap 标准误 . . . 489
A.3.5 数值优化方法概述 . . . 489
A.4 BLP 收缩映射的压缩性证明 . . . 490

附录 B 中国产业组织的特殊议题 **491**
B.1 国有企业的竞争与垄断 491
B.1.1 国有企业的双重身份 491
B.1.2 行政垄断 vs. 市场垄断 491
B.1.3 混合所有制改革 492
B.2 中国反垄断法的演变与执法 492
B.2.1 2008-2020：从无到有的制度建设 492
B.2.2 2022 年修订：数字经济的回应 492
B.2.3 标志性执法案例 493
B.2.4 中国反垄断法的哲学基础 493
B.3 中国平台经济的竞争格局 494
B.3.1 市场结构演变 494
B.3.2 ”二选一”——中国版本的排他性行为 494
B.3.3 平台用工：IO 与劳动经济学的交叉 494
B.4 产业政策与竞争政策的张力 495
B.4.1 典型案例 495
B.4.2 ”全国统一大市场” 495
B.5 中国 IO 研究前沿学者与机构 495
B.6 中国数据资源与实证 IO 研究机会 497

第一章　什么是产业组织学？

章首语

2021 年 4 月，中国国家市场监督管理总局对阿里巴巴集团处以 182.28 亿元人民币罚款，原因是该平台滥用市场支配地位，实施” 二选一” 排他性行为。同一年，美国联邦贸易委员会（FTC）对 Facebook（现 Meta）提起反垄断诉讼，指控其通过收购 Instagram 和 WhatsApp 系统性消除竞争威胁。在大西洋彼岸，欧盟委员会对 Google 累计罚款超过 80 亿欧元，涉及购物搜索偏袒、Android 系统捆绑和广告技术反竞争行为。

这三起案件指向同一个核心问题——**当少数企业拥有巨大的市场势力（market power）时，它们如何行使这种权力？这种权力在什么条件下损害消费者和社会福利？政府应该如何应对？**

这些问题，正是产业组织学（Industrial Organization，简称 IO）近两个世纪来不懈追问的。

1.1　核心问题

【核心】

读完本章，你应该能够回答以下问题：

1. **产业组织学研究什么？**它与微观经济学理论、博弈论、计量经济学和反垄断法之间的关系是什么？

2. **为什么不完全竞争是常态？**现实世界中，市场势力从哪些渠道产生？

3. **IO 的核心张力是什么？**为什么” 结构 vs. 行为” 的争论贯穿了整个学科史？

4. **IO 的方法论由哪些部分构成？**理论、实证和政策分析如何相互配合？

5. **本书的组织逻辑是什么？**为什么按历史演进的顺序来学习 IO？

本章是全书的地图和指南针。我们不深入任何具体模型，而是建立一幅关于” 产业组织学是什么、为什么要学习它、以及如何学习它” 的完整认知地图。

1.2　不完全竞争的普遍性

【核心】

1.2.1　教科书世界 vs. 现实世界

任何学过微观经济学的读者都熟悉**完全竞争**（perfect competition）模型：无数原子化的企业生产同质产品，都是价格接受者（price taker），在长期均衡中价格等于边际成本，利润为零。这是一个优雅的基准模型，它证明了竞争市场在理想条件下可以达到 Pareto 效率。

但一旦我们抬头看真实世界，完全竞争的假设几乎全盘崩塌：

- 中国三大电信运营商（中国移动、中国联通、中国电信）瓜分了超过 90% 的市场份额——这是**寡头**（oligopoly），不是完全竞争。
- 据行业估计，苹果公司的 iPhone 在高端手机市场拥有超过 50% 的份额和远超其成本的价格加成（markup）——这来自**产品差异化**（product differentiation）。
- 台积电在先进芯片制造领域独占鳌头——因为它拥有其他企业无法轻易复制的**进入壁垒**（barriers to entry）：数百亿美元的资本投入、数十年的技术积累和专利保护。
- 据 StatCounter 等第三方估计，Google 搜索在全球多数市场拥有超过 90% 的份额，但它对用户**免费**——这种” 零价格” 是否意味着没有市场势力？

不完全竞争（imperfect competition）——即企业拥有某种程度的市场势力，能够将价格设定在边际成本之上而不失去全部顾客——才是市场经济的常态。产业组织学正是为理解这种常态而生的。

1.2.2　市场势力的来源

市场势力从何而来？我们可以识别出五个主要渠道：

1. **规模经济**（economies of scale）：某些行业——如芯片制造、商用飞机制造、铁路——的固定成本如此之高，以至于只有少数几家企业可以在市场上存活。自然垄断（natural monopoly）和自然寡头是技术条件的直接结果。
2. **产品差异化**（product differentiation）：即使是物理上相似的产品，品牌、质量、位置、设计和文化意义也可以创造差异。每一个差异化产品的卖家都面对一条向下倾斜的剩余需求曲线（residual demand curve）——这意味着它可以提高价格而不会失去全部顾客。

3. **进入壁垒**（barriers to entry）：在位企业可能拥有新进入者无法轻易复制的优势：专利和知识产权、对关键投入品的控制、政府许可和特许经营权、品牌忠诚和转换成本（switching cost）、网络效应（network effect）。

4. **策略行为**（strategic behavior）：即使在结构上不存在明显壁垒的市场，在位企业也可以通过策略行为创造或维持市场势力——这包括限制定价、掠夺性定价、排他性合同、捆绑销售和合谋。

5. **数据与算法**（data and algorithms）：近年来变得日益重要的第五个来源。数字平台的个性化推荐、搜索排名和定价算法既可以为消费者创造巨大的价值，也可以成为排斥竞争和维持市场势力的工具。

1.2.3 一个经典的定量快照

让我们用一个简单但有力的数字来感知市场势力的程度。经济学家用 **Lerner 指数**（Lerner Index）来度量一家企业的市场势力：

$$L = \frac{P - MC}{P} \tag{1.1}$$

其中 P 是价格，MC 是边际成本。在完全竞争市场中，$L = 0$（价格等于边际成本）。在纯垄断中，$L = 1/|\varepsilon|$，其中 ε 是需求弹性——需求越缺乏弹性，垄断者的加成越高。

De Loecker、Eeckhout 和 Unger（2020）的研究表明，美国上市公司的平均加成率（markup，定义为 P/MC）从 1980 年的约 1.2 上升到 2016 年的约 1.6——这意味着典型企业的价格超过其边际成本 60%，是四十年前的三倍。换算为 Lerner 指数，$L = (1.6 - 1)/1.6 \approx 0.375$——即每元价格中约 37.5 分钱是超过边际成本的加成。无论你如何解读这一趋势，一个基本事实是清楚的：**市场势力是当代经济的一个核心特征，并且它的重要性在上升**。

1.3 一个永恒的张力：结构 vs. 行为

【核心】

如果说产业组织学有一条贯穿始终的红线，那就是**结构**（structure）与**行为**（conduct）之间的张力。

1.3.1 结构论

结构论的核心直觉是：市场势力的根源在于市场的”解剖学”特征——有多少企业在竞争？进入这个行业有多难？产品有多大差异？

这个思想的古典表达来自 **Joe Bain**（1956），哈佛学派的奠基人之一。Bain 的逻辑链是这样的：

$$\text{高集中度} + \text{高进入壁垒} \longrightarrow \text{在位企业获得持续超额利润} \longrightarrow \text{社会遭受效率损失}$$

如果这是正确的，那么反垄断政策的任务就很简单：关注结构。高集中度本身就是红旗，结构性拆分（structural remedies）是合理的工具。

1.3.2　行为论

行为论的核心直觉是：**同样是三个企业，它们可以激烈竞争，也可以默契合谋。**市场结构只是舞台，企业行为才是戏剧本身。

这个思想在 **Jean Tirole** 的工作中得到了最严格的表达。Tirole（1988）用博弈论证明，市场结果不是由" 企业数量" 这个数字机械决定的——它取决于企业可以使用的策略工具（价格还是产量?）、它们拥有的信息（对手的成本我知道吗?）、和博弈的时间结构（一次性的还是重复的?）。

以著名的**连锁店悖论**（chain-store paradox; Selten, 1978）为例：一个在位者面对 20 个不同市场中的潜在进入者。按照" 结构论" 的直觉，在位者应该在第一个市场中用掠夺性定价打击进入者，以建立" 强硬" 声誉，遏制后续进入。但博弈论告诉我们一个出人意料的结果：用**逆向归纳**（backward induction）——即从最后一个市场的博弈往前推导——来分析这个情境：在最后一个市场中，在位者没有理由掠夺，因为回报不足以弥补成本。知道这一点，进入者就会进入最后一个市场。但在倒数第二个市场中，在位者知道自己在最后一个市场不会掠夺，所以掠夺在倒数第二个市场也无法建立任何声誉优势……以此类推，整个声誉机制从后往前崩溃。**注：这是一个预览——第 ~8~ 章将详细展开掠夺行为、声誉和连锁店悖论的现代解决方案（如 Kreps-Wilson 1982 的" 四人帮" 声誉模型）。**

这个例子（以及大量的后续工作）说明了一个核心教训：**要理解市场势力，你必须打开" 企业行为" 这个黑箱，仔细分析策略互动（strategic interaction）的微观基础。**

1.3.3　为什么这个张力如此重要?

因为不同的诊断导致完全不同的政策处方：

- 如果市场势力主要来自**结构** → 政策应该关注降低集中度、降低进入壁垒、拆分大型企业。
- 如果市场势力主要来自**行为** → 政策应该关注禁止特定的反竞争行为（合谋、排他、掠夺），同时保持对效率抗辩（efficiency defense）的开放性。

本书的 15 章将带你走过这条张力的每一个转折点。你会看到哈佛学派如何试图用结构解释一切，芝加哥学派如何用效率和策略回应，Tirole 如何用博弈论将" 行为" 的分析精确化，NEIO（新实证产业组织）如何用数据区分结构效应和行为效应，以及数字时代如何重新点燃了" 结构 vs. 行为" 的古老争论。

1.4 产业组织学的三根支柱

【核心】

产业组织学由三根互相支撑的支柱组成：**理论**、**实证**和**政策**。

1.4.1 理论（Theory）

IO 理论的核心工具是非合作博弈论（noncooperative game theory）。为什么是博弈论？因为寡头市场的本质特征是**策略依赖**（strategic interdependence）：我的最优选择取决于我认为你会做什么，而你的最优选择又取决于你认为我会做什么。

博弈论为分析这种互相依赖提供了精确的数学语言：

- **策略式博弈**（strategic form）→ Nash 均衡 → 分析 Cournot 竞争
- **扩展式博弈**（extensive form）→ 子博弈完美均衡（SPNE）→ 分析 Stackelberg 领导模型和进入遏制
- **不完全信息博弈** → 贝叶斯 Nash 均衡（BNE）→ 分析拍卖、信号传递和限制定价
- **重复博弈** → Folk 定理 → 分析合谋的可持续性

本书不会预设你是一个博弈论专家。每一个博弈论工具都在它被需要的时刻引入，并在 IO 的具体语境中讲解。附录 A 提供了核心博弈论概念的快速参考。

1.4.2 实证（Empirics）

理论可以告诉我们" 在特定条件下" 市场势力如何产生。但它不能直接回答" 在此时此刻的航空业/手机业/搜索引擎市场，企业到底在行使多大的市场势力？"

回答后一类问题需要**实证产业组织**（empirical IO），特别是 1980 年代后期兴起的新实证产业组织（New Empirical Industrial Organization，NEIO）。

NEIO 的方法论核心是**结构估计**（structural estimation）：从关于消费者效用函数和企业成本函数的具体假设出发，推导出可估计的计量方程，然后用市场数据恢复深层参数（需求弹性、边际成本、竞争行为参数），最后用这些参数进行反事实分析（counterfactual analysis）——例如，" 如果 A 公司和 B 公司合并了，价格会涨多少？"

本书的第 ~10~ 章–第 ~12~ 章将系统介绍 NEIO 的核心方法：Bresnahan 的 conduct parameter 方法、Berry-Levinsohn-Pakes（BLP）的需求估计、以及并购模拟和反垄断损害量化的工具链。

无论你的最终目标是做理论还是做实证，理解 IO 的实证方法在今天已经是必不可少的。正如 MIT 的 Glenn Ellison 教授所观察到的：当今的 IO 博士生在第一年就同时接受理论与实证训练——这一领域的现代研究者几乎不可能只专精其中一者。

1.4.3　政策（Policy）

IO 的终极存在理由，或许就是为竞争政策（competition policy）和规制（regulation）提供分析基础。

在世界主要经济体中，反垄断法（antitrust law）和竞争法（competition law）禁止三类主要行为：

1. **垄断协议**（cartels / collusion）：竞争对手之间的固定价格、划分市场、串通投标——这些在几乎所有法域都是当然违法（per se illegal）

2. **滥用市场支配地位**（abuse of dominance / monopolization）：掠夺性定价、排他性交易、拒绝交易、捆绑销售——这些通常适用合理原则（rule of reason），需要权衡反竞争效应和效率抗辩

3. **反竞争的并购**（anti-competitive mergers）：评估横向并购（竞争者之间）和纵向并购（上下游之间）的单边效应（unilateral effects）和协调效应（coordinated effects）

三大经济体——美国、欧盟、中国——的竞争法体系有着共同的 IO 经济分析内核，但也存在重要差异。这些差异本身也反映了对" 结构 vs. 行为"、" 消费者福利 vs. 多目标" 等核心问题的不同回答。附录 B 专门讨论中国的产业组织与竞争政策议题。

1.5　全书路线图

本书的设计遵循一个核心原则：**按照产业组织学思想的历史演进来组织内容，在每一个历史阶段将理论、实证和政策三个维度有机融合。**

这不是一本" 先讲理论、后讲实证" 的书，也不是一本" 先讲垄断、后讲寡头" 的书。我们的叙事线是：IO 经济学家们是如何思考市场势力的？他们为什么放弃了一些旧方法，为什么发明了一些新方法？每一代学者的突破，都是对前一代范式局限性的回应。

下图展示了本书的叙事结构：

古典奠基 (1838–1933)
Cournot, Bertrand, Hotelling, Chamberlin
"不完全竞争的最早直觉"

哈佛 SCP 范式 (1930s–1970s)
Mason, Bain
"结构决定绩效 → 积极反垄断"

芝加哥革命 (1950s–1980s)
Stigler, Demsetz, Posner, Bork
"效率解释 *vs.* 市场势力 → 收缩干预"

新 IO 理论 (1980s–2000s) ← 全书核心 I
Tirole, Fudenberg, Laffont
"博弈论重塑 *IO* → 策略行为的微观基础"

新实证 IO (1980s–至今) ← 全书核心 II
Bresnahan, Berry, Levinsohn, Pakes
"从 *SCP* 到结构估计 → 让数据说话"

现代前沿 (2000s–至今)
Rochet-Tirole, Calvano, Khan
"平台、算法、新布兰代斯"

注：新 IO 理论与新实证 IO（NEIO）在历史上大致并行发展——Tirole (1988) 与 Bresnahan (1989) 几乎同时面世。本书将理论模块置于实证模块之前，纯粹出于教学考虑：先建立策略互动的分析语言，再用它解读结构估计中的识别策略。

各模块简介

模块 0：序章（本章）你在这里。建立 IO 的认知地图。

模块 1：古典奠基（第 2 章）1838–1933 年，Cournot、Bertrand、Edgeworth、Hotelling、Chamberlin 和 Joan Robinson——在博弈论被发明之前，这些先驱用简单的数理模型捕捉了不完全竞争的本质直觉。这一章也为你提供了从零开始推导寡头模型的练习。

模块 2：哈佛 SCP 范式（第 3–4 章）第 3 章讲述哈佛学派如何将 IO 建立为一个独立的实证研究领域。Mason 的 SCP 框架、Bain 的进入壁垒理论、以及结构主义的反垄断含义。第 4 章讨论 SCP 的实证项目——跨行业回归——为什么最终失败了。Demsetz 的效率批评、计量经济学的方法论缺陷，以及 SCP 留给后世的遗产（HHI、市场界定）。

模块 3：芝加哥革命（第 5–6 章）第 5 章讲述芝加哥学派——Stigler、Demsetz、Posner、Bork——如何以" 效率" 为武器，系统性地颠覆了哈佛的结构主义。他们对进入壁垒

的重新定义、对掠夺性定价的怀疑、以及” 消费者福利标准” 的反垄断革命，至今仍是全球竞争政策的基石。第 6 章聚焦**可竞争市场理论**——如果进入和退出完全自由且无成本，即使高度集中的市场也会产生竞争性结果吗？这个理论推动了全球范围的放松管制（deregulation）运动。

模块 4：新 IO 理论（第 7–9 章） ★ 全书核心第 7 章是全书的理论工具箱——博弈论如何在 Tirole（1988）的框架中重塑了 IO。第 8 章深入到三类核心策略行为：进入遏制（Dixit 产能承诺、Milgrom-Roberts 信号限制定价）、掠夺性定价（Kreps-Wilson 声誉模型、Bolton-Scharfstein 深口袋模型）和合谋（Green-Porter 价格战、Rotemberg-Saloner 繁荣期背叛）。第 9 章进入不对称信息与规制——Laffont-Tirole 的激励规制理论，这是 IO 从” 描述市场” 走向” 设计制度” 的桥梁。

模块 5：新实证 IO（第 10–12 章） ★ 全书核心第 10 章以 Bresnahan（1989）的”NEIO 宣言” 开篇——为什么要放弃跨行业 SCP 回归，转向单一行业的结构估计？你还将学到一个实用的工具：生产函数估计与加成率测算（Olley-Pakes, De Loecker）。第 11 章是 BLP 方法——Berry、Levinsohn 和 Pakes 在 1995 年发表的随机系数需求模型，可能是整个经济学中被引用最多的实证方法论文之一。第 12 章将这些工具链整合为反垄断量化的” 标准操作流程”：并购模拟、SSNIP 市场界定测试、以及卡特尔损害量化的方法。

模块 6：现代前沿（第 13–15 章）第 13 章进入双边市场与平台经济学——当平台服务两组需要互动的用户时（如 Uber 连接司机和乘客、Google 连接搜索用户和广告商），传统的单边市场逻辑在哪些地方失效？第 14 章聚焦数字时代的三个新问题：算法合谋（AI 可以在没有人类指令的情况下学会合谋吗？）、数据作为进入壁垒、以及全球数字规制的兴起（欧盟 DMA、中国 2022 反垄断法修订）。第 15 章是全书的终章——” 新布兰代斯学派”（Lina Khan, Tim Wu）认为，芝加哥消费者福利标准已经无法处理数字时代的经济权力集中。反垄断的目标应该是什么？这场争论还没有答案——这正是你作为新一代研究者的机会。

1.6　方法论预览

【核心】

在正式进入后续章节之前，先简要预登记本书将用到的主要分析工具。

1.6.1　博弈论

博弈论是 IO 理论的语言。本书将在第 7 章系统介绍以下核心概念：

- **Nash 均衡**（Nash Equilibrium）：每个玩家的策略是对其他玩家策略的最优反应。用于分析 Cournot 产量竞争和 Bertrand 价格竞争。

- **子博弈完美均衡**（Subgame Perfect Nash Equilibrium, SPNE）：剔除不可信威胁（incredible threats）的均衡精炼。用于分析 Stackelberg 领导模型、进入遏制和策略投资。

- **贝叶斯 Nash 均衡**（Bayesian Nash Equilibrium, BNE）：玩家拥有不完全信息时的均衡。用于分析拍卖、限制定价的信号模型和规制中的" 菜单合同"。

- **Folk 定理**：在无限重复博弈中，如果参与者足够耐心，任何高于极小极大支付（minimax payoff）的结果都可以作为 SPNE 维持。用于分析合谋的可持续性。

1.6.2 计量经济学

IO 实证方法的核心是**结构估计**，本书将在第 10–11 章系统介绍：

- **工具变量**（Instrumental Variables, IV）：处理价格内生性——价格不仅影响需求，需求也影响价格。

- **广义矩方法**（Generalized Method of Moments, GMM）：BLP 需求估计的核心估计方法。

- **模拟矩方法**（Simulated Method of Moments）：当模型没有解析解时，通过蒙特卡洛模拟近似。

- **数值优化**：非线性目标函数的最小化/最大化——BLP 每天让博士生在服务器前等待的核心原因。

1.6.3 政策分析框架

本书涉及的竞争政策分析工具包括：

- **消费者福利标准**（Consumer Welfare Standard）：反垄断的合法目标是提高消费者福利（通常以价格和产出衡量），还是应追求更广泛的目标（公平、民主、创新）？

- **当然违法 vs. 合理原则**（Per Se vs. Rule of Reason）：某些行为被推定违法（如固定价格的卡特尔），不需要分析其具体效果；其他行为需要逐案权衡。

- **SSNIP 测试**：界定相关市场（relevant market）的标准方法——一个假想垄断者能否有利可图地施加" 小幅但显著且非暂时的价格上涨"（Small but Significant Non-transitory Increase in Price）？

1.7　关键概念预登记

在正式启程之前，以下是本书中反复出现的核心概念。现在只需要对它们有一个初步的印象——每一条都将在后续章节中展开。

表 1.1: 核心概念预登记

概念	简要定义	首次深入讨论
市场势力（market power）	企业将价格设定在边际成本之上的能力	Ch.2
Lerner 指数	$L = (P - MC)/P$，市场势力的标准度量	Ch.2
市场界定（market definition）	识别哪些产品/企业构成一个” 反垄断市场”	Ch.12
进入壁垒（barriers to entry）	使进入者难以与在位者竞争的任何因素	Ch.3, Ch.5
产品差异化（product differentiation）	产品之间不完全替代——水平差异化（口味）或垂直差异化（质量）	Ch.2, Ch.7
价格歧视（price discrimination）	对不同消费者或不同数量收取不同价格	Ch.7
合谋（collusion / cartel）	竞争者之间的协调——明示的（卡特尔）或默契的（tacit）	Ch.8
掠夺性定价（predatory pricing）	以低于成本的价格排挤竞争对手，以在未来收取垄断价格	Ch.8
纵向关系（vertical relations）	上下游企业之间的合同和组织关系	Ch.8
横向并购（horizontal merger）	同一市场中直接竞争者之间的合并	Ch.12
双边市场（two-sided market）	平台服务两组不同的用户，且存在交叉网络外部性	Ch.13
网络效应（network effect）	一个用户使用产品的价值随其他用户数量增加而增加	Ch.13, Ch.6
HHI（Herfindahl-Hirschman Index）	$\sum_{i=1}^{n} s_i^2 \times 10000$，市场集中度的标准度量	Ch.3
消费者福利（consumer welfare）	消费者剩余的总和——反垄断政策的核心标准（有争议）	Ch.5, Ch.15

1.8　本章小结

产业组织学研究**不完全竞争市场中企业的策略行为与市场绩效**。它的核心问题是：市场势力从何而来？企业如何行使它？在什么条件下损害社会福利？我们如何测

量它？政策应该如何回应？

本章建立了四个关键认知：

1. **不完全竞争是常态**。从电信到航空，从搜索引擎到制药，现实世界中的市场几乎都是不完全竞争的。完全竞争是一个有用的基准，但它不是对现实的描述。市场势力有多个来源——规模经济、产品差异化、进入壁垒、策略行为，以及日益重要的数据和算法。

2. **IO 的核心张力是结构 vs. 行为**。市场势力根源于市场的结构性特征（集中度、进入壁垒），还是企业的策略行为（合谋、掠夺、排他）？这个张力从 Bain 到 Tirole，从哈佛到芝加哥，从 SCP 到 NEIO，贯穿了整个学科史——而且今天在数字平台的争论中比以往任何时候都更加鲜活。

3. **IO 由三根支柱构成**：理论（博弈论模型）、实证（结构估计和简化式因果推断）、和政策（反垄断与规制）。这三根支柱不是彼此独立的——今天最好的 IO 研究工作把它们融为一体。

4. **本书的历史演进框架不只是叙事策略，更是学习策略**。要真正理解为什么 Tirole 的博弈论方法能够回应 SCP 范式无法解决的问题，你需要先理解 SCP 是什么、它为什么衰落。要理解 BLP 的结构估计方法为什么是一次突破，你需要先理解 Bresnahan 之前的实证 IO 在做什么。一步一步来，每一章都是下一章的基础。

与后续章节的联系：→ Ch.2 将带你回到 1838 年——Augustin Cournot 如何用一页数学推导，开创了整个不完全竞争经济学的传统。

1.9 延伸阅读

★ 必读（本章核心参考）

- Tirole, J. (1988). *The Theory of Industrial Organization*. MIT Press. **Introduction** —IO 研究生教材的标准开篇。Tirole 在此解释了他为什么用博弈论组织全书，以及新 IO 与传统 SCP 的区别。英文原文 5–10 页。

- Schmalensee, R. (1989). "Inter-Industry Studies of Structure and Performance." In R. Schmalensee & R. Willig (eds.), *Handbook of Industrial Organization*, Vol. 2, Ch.16. North-Holland. —对 SCP 实证传统的权威综述，写于该传统的黄昏时刻。

- Bresnahan, T. (1989). "Empirical Studies of Industries with Market Power." In *Handbook of Industrial Organization*, Vol. 2, Ch.17. North-Holland. —NEIO 的纲领性文献，宣告了从 SCP 到结构估计的范式转换。与 Schmalensee (1989) 对读，可以清晰地看到 IO 实证方法论的断裂线。

★★ 进阶（建构学科史视野）

- De Loecker, J., Eeckhout, J., & Unger, G. (2020). "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics*, 135(2), 561–644. —本章引用的加成率趋势研究的原文。展示了 NEIO 方法如何应用于宏观经济问题。

- Motta, M. (2004). *Competition Policy: Theory and Practice*. Cambridge University Press. **Ch.1** —对竞争政策的经济学基础的精彩概述，补充了本章的政策视角。

- Belleflamme, P., & Peitz, M. (2015). *Industrial Organization: Markets and Strategies*. 2nd ed. Cambridge University Press. **Ch.1–2** —另一本主要 IO 教材的开篇。比较 B&P 和 Tirole 组织方式的不同，本身就很有启发。

★★★ 前沿（IO 思想的当代争论）

- Khan, L. (2017). "Amazon's Antitrust Paradox." *Yale Law Journal*, 126(3), 710–805. —新布兰代斯运动的标志性文献。一个耶鲁法学院学生写的论文，最后改变了美国反垄断的议程。建议在读到 Ch.15 时回头重读。

- Shapiro, C. (2018). "Antitrust in a Time of Populism." *International Journal of Industrial Organization*, 61, 1–25. —Carl Shapiro（Berkeley 教授，前 Obama 政府 DOJ 首席经济学家）对新布兰代斯运动的芝加哥派回应。与 Khan (2017) 对读。

下一章： → Ch.2　古典奠基：Cournot, Bertrand, Hotelling ——不完全竞争的最早直觉

第二章　古典奠基：Cournot, Bertrand, Hotelling

章首语

1838 年，法国数学家 Augustin Cournot 出版了一本在当时完全被忽视的书。他用一个矿泉水双寡头的例子，第一次用数学语言精确地描述了” 当市场上只有两家企业时会发生什么”。他没有使用” 博弈论” 或”Nash 均衡” 这些术语——这些概念要等 100 多年才被发明——但他的推导精确地得到了我们今天称为 Cournot-Nash 均衡的结果。

接下来的 95 年里，一群学者——Bertrand, Edgeworth, Hotelling, Chamberlin, Joan Robinson——以各自的方式回应、修正、综合 Cournot 的开创性工作。本章将带你走过这段历程。六个模型，每一个都是对同一个问题的回答：**在不完全竞争的世界里，价格和数量由什么决定？**

2.1　核心问题

读完本章，你应该能够回答以下问题：

1. **Cournot 模型如何从企业利润最大化推导出寡头均衡？**企业数量的增加如何系统地降低市场势力？

2. **为什么 Bertrand 说” 两家企业就够”？** Cournot 和 Bertrand 的根本分歧在哪里？改变策略变量（产量 vs. 价格）为什么会彻底改变均衡？

3. **产能约束如何化解 Bertrand 悖论？** Edgeworth 和 Kreps-Scheinkman 的答案是什么？

4. **空间竞争如何软化价格竞争？** Hotelling 的” 线性城市” 为什么能产生正利润？最小差异化和最大差异化之争告诉我们什么？

5. **垄断竞争的本质是什么？**大量差异化企业加自由进入，为什么 $P > MC$ 但 $\pi = 0$？

6. **这些古典模型与后续章节的关系是什么？**

第 ~1~ 章建立了 IO 的认知地图，其中" 结构 vs. 行为" 的张力是贯穿全书的红线。本章把这条红线变成可操作的数学模型：在 Cournot 模型中，企业数量 n（结构）直接决定均衡加成（绩效）；但在 Bertrand 模型中，即使 $n = 2$，加成也降为零。这种分歧——结构和行为，哪个更根本？——将从本章开始，贯穿全书。

2.2　历史背景：1838–1933

2.2.1　Cournot 之前的经济学世界

1838 年的经济学还处于古典政治经济学传统之中。Adam Smith 的《国富论》(1776) 已出版 62 年；他本人于 1790 年去世，距 1838 年已 48 年，David Ricardo 已去世 15 年，John Stuart Mill 的《政治经济学原理》(1848) 尚待十年之后。Walras (1874) 和 Jevons (1871) 的边际革命还要等三十多年。

正是在这个没有需求曲线、没有边际分析、没有数学形式化的世界里，Cournot 做了三件划时代的事：

1. 第一次用数学函数定义了**需求曲线**——价格与需求量之间的关系

2. 用**边际分析**（导数）来刻画企业的利润最大化决策

3. 在**一个统一的框架内**分析了垄断、双寡头和完全竞争

2.2.2　Cournot 为什么被忽视了？

答案很简单：19 世纪的经济学家不读数学。法国的学术分工将经济学归入" 道德科学"，数学归入" 自然科学"——两者不应混合。Cournot 的数学化写作对当时的读者而言是不可逾越的障碍。

这本书的命运本身就是一堂关于思想传播的课：Cournot 如此气馁，以至于放弃了经济学整整 25 年。他后来试图在同一主题上写一本没有数学的书（1863），同样未能获得关注。直到 1870 年代边际革命期间，Walras 和 Jevons 才重新发现了他的工作。Nathaniel Bacon 在 1897 年出版了此书的首个英译本，Irving Fisher 为其撰写了长篇导言——这个译本最终将 Cournot 确立为数学经济学的奠基人。

关键人物链：Cournot → Walras → Marshall → Fisher → Samuelson → Tirole。

2.2.3　1883–1933：回应、反驳与综合

Cournot 之后的九十年产生了依次的回应和拓展：

- **Bertrand (1883)**：在一篇对 Cournot 和 Walras 的书评中提出毁灭性批评——如果企业竞争的是**价格**而非产量，则即使是双寡头也产生完全竞争结果。
- **Edgeworth (1897)**：引入产能约束——Bertrand 的结论假设企业可以服务整个市场。现实中，产能有限，结果可能是价格周期而非边际成本定价。
- **Hotelling (1929)**：引入空间/产品差异化——当消费者在意产品的”位置”（字面或隐喻的），他们不会仅仅因为价格差异就切换。差异化软化竞争。
- **Chamberlin & Robinson (1933)**：同年各自出版了关于不完全竞争的理论著作。Chamberlin 开创了垄断竞争理论（产品差异化 + 自由进入），Robinson 则系统化了不完全竞争的分析工具（边际收益曲线、价格歧视、买方垄断）。

为什么是 1933 年？两本关于不完全竞争的著作在同一年分别出自哈佛和剑桥——这不是巧合。Sraffa (1926) 证明递增报酬与完全竞争在逻辑上不相容，摧毁了 Marshall 竞争理论的逻辑根基。大萧条使市场自我修复的假设成为笑话。Chamberlin 和 Robinson 都在为同一个危机提供答案。

2.3 Cournot 模型 (1838)

【核心】

2.3.1 设定

要素	描述
玩家	n 个企业，$i = 1, 2, \ldots, n$
策略	每个企业同时选择产量 $q_i \geq 0$
支付	$\pi_i = P(Q) \cdot q_i - C_i(q_i)$
信息	完全信息——所有企业都知道需求函数和彼此的成本
时机	同时行动，一次性博弈

具体假设：

- 同质产品——企业生产的是完全替代品
- 线性反需求函数：$P(Q) = a - bQ$，其中 $Q = \sum_{i=1}^{n} q_i$，$a > c$，$b > 0$
- 恒定边际成本：$C_i(q_i) = c \cdot q_i$（各企业边际成本相同，无固定成本）
- 每个企业在选择 q_i 时将其他企业的产量 $Q_{-i} = \sum_{j \neq i} q_j$ 视为给定

注记 2.1. *Cournot 的原例是两家矿泉水生产者，开采成本为零（$c = 0$）。这个”矿泉水双寡头”至今仍是教学的标准例子。*

2.3.2　均衡推导：$n = 2$ 情形

步骤 1：写出利润函数。

$$\begin{aligned}\pi_1(q_1, q_2) &= \big[a - b(q_1 + q_2)\big]q_1 - cq_1 \\ &= aq_1 - bq_1^2 - bq_1q_2 - cq_1 \end{aligned} \tag{2.1}$$

同理，$\pi_2(q_1, q_2) = aq_2 - bq_2^2 - bq_1q_2 - cq_2$。

步骤 2：一阶条件。

企业 1 对 q_1 求导，将 q_2 视为给定：

$$\frac{\partial \pi_1}{\partial q_1} = a - 2bq_1 - bq_2 - c = 0 \tag{2.2}$$

二阶条件成立：$\partial^2\pi_1/\partial q_1^2 = -2b < 0$。

步骤 3：反应函数。

从 (2.2) 解出 q_1 作为 q_2 的函数：

$$\boxed{q_1 = R_1(q_2) = \frac{a-c}{2b} - \frac{1}{2}q_2} \tag{2.3}$$

同理，企业 2 的反应函数：$q_2 = R_2(q_1) = \dfrac{a-c}{2b} - \dfrac{1}{2}q_1$。

步骤 4：求解均衡。

将 $R_2(q_1)$ 代入 $R_1(q_2)$（或利用对称性，令 $q_1 = q_2 = q^*$）：

$$q^* = \frac{a-c}{2b} - \frac{1}{2}q^* \implies \frac{3}{2}q^* = \frac{a-c}{2b} \implies \boxed{q_1^* = q_2^* = \frac{a-c}{3b}} \tag{2.4}$$

步骤 5：均衡结果。

$$Q^* = q_1^* + q_2^* = \frac{2(a-c)}{3b} \tag{2.5}$$

$$P^* = a - bQ^* = \frac{a+2c}{3} \tag{2.6}$$

$$\pi_1^* = \pi_2^* = (P^* - c)q^* = \frac{(a-c)^2}{9b} \tag{2.7}$$

2.3.3　一般化到 n 个企业

设 n 个对称企业。在对称均衡中，$q_i = q$，$Q = nq$。企业 i 的一阶条件为：

$$a - 2bq - b(n-1)q - c = 0 \tag{2.8}$$

解得：

$$\boxed{q^* = \frac{a-c}{b(n+1)}, \quad Q^* = \frac{n(a-c)}{b(n+1)}, \quad P^* = \frac{a+nc}{n+1}} \tag{2.9}$$

$$\boxed{\pi_i^* = \frac{(a-c)^2}{b(n+1)^2}}$$

2.3.4 比较静态与 Cournot 极限定理

表 2.1: Cournot 模型中 n 的效应

n	P^*	Q^*	π_i^*	Lerner 指数 L	福利损失占比
1（垄断）	$\frac{a+c}{2}$	$\frac{a-c}{2b}$	$\frac{(a-c)^2}{4b}$	$\frac{a-c}{a+c}$	25.00%
2（双寡头）	$\frac{a+2c}{3}$	$\frac{2(a-c)}{3b}$	$\frac{(a-c)^2}{9b}$	$\frac{a-c}{a+2c}$	11.11%
3	$\frac{a+3c}{4}$	$\frac{3(a-c)}{4b}$	$\frac{(a-c)^2}{16b}$	$\frac{a-c}{a+3c}$	6.25%
5	$\frac{a+5c}{6}$	$\frac{5(a-c)}{6b}$	$\frac{(a-c)^2}{36b}$	$\frac{a-c}{a+5c}$	2.78%
10	$\frac{a+10c}{11}$	$\frac{10(a-c)}{11b}$	$\frac{(a-c)^2}{121b}$	——	0.83%
∞	c	$\frac{a-c}{b}$	0	0	0%

Cournot 极限定理

当 $n \to \infty$ 时：$P^* \to c$，$q_i^* \to 0$，$\pi_i^* \to 0$。Cournot 均衡收敛于完全竞争结果。福利损失占比 $= 1/(n+1)^2$——收敛速度为 $\mathcal{O}(1/n^2)$。仅 5 个企业，福利损失就降至不足 3%。

2.3.5 Lerner 指数与 Cournot

从一阶条件重写：

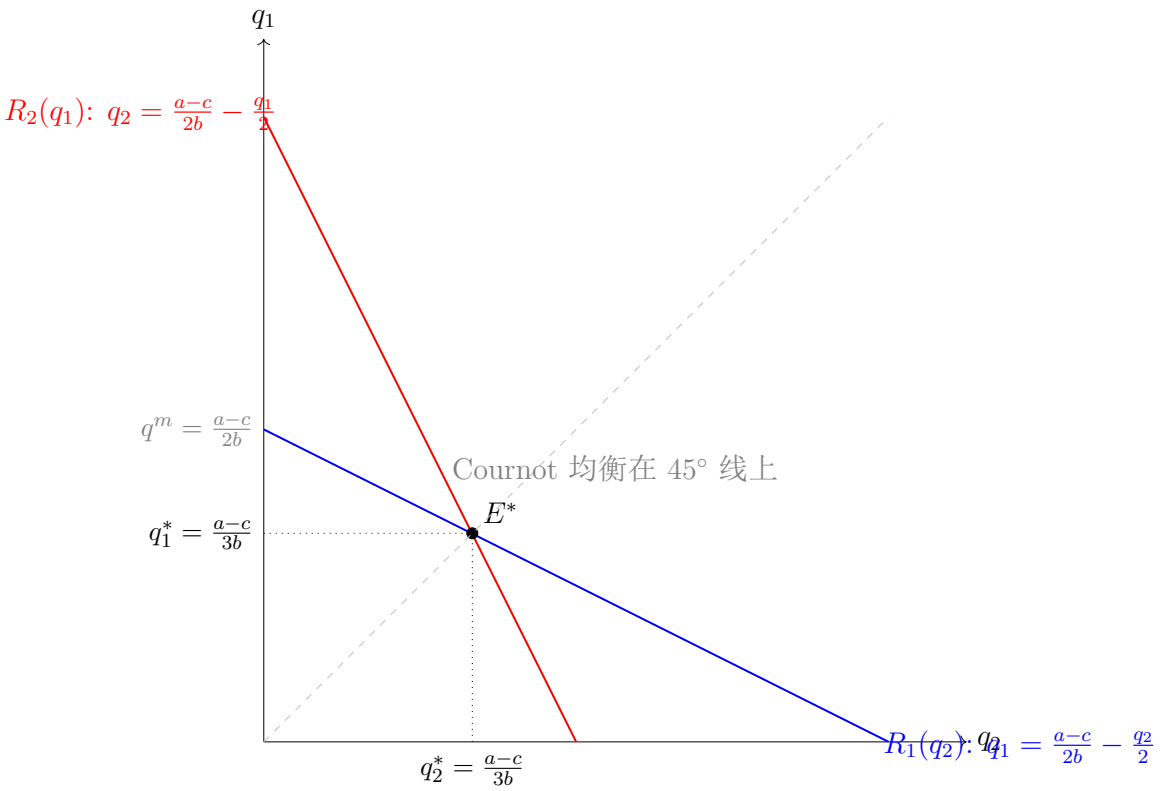
$$P - c = \frac{q_i}{Q} \cdot \frac{P}{|\varepsilon|} \quad \Longrightarrow \quad \boxed{L_i = \frac{P-c}{P} = \frac{s_i}{|\varepsilon|}} \tag{2.10}$$

其中 $s_i = q_i/Q$ 为市场份额。对于对称 Cournot（$s_i = 1/n$）：

$$\boxed{L = \frac{1}{n \cdot |\varepsilon|}} \tag{2.11}$$

企业数量越多、需求弹性越大，市场势力越小。这个公式将 Ch.1 中引入的 Lerner 指数锚定在了具体的市场结构参数上。

2.3.6　图形直觉



反应函数向下倾斜——你增产，我减产。这是**战略性替代**（strategic substitutes）的特征：一个企业更激进的行动（更多产量）降低另一个企业同样行动的边际收益。在 第 ~7~ 章中，Bulow、Geanakoplos 和 Klemperer (1985) 将对此进行系统分类。

Cournot 直觉性地使用了 Nash 均衡概念——在 Nash 定义它的 113 年前。每个企业的产量是其他企业产量的最优反应，均衡是这两个最优反应的固定点。这是经济学史上最卓越的提前发现之一。今天，它被称为 Cournot-Nash 均衡。

2.4　Bertrand 模型 (1883)

【核心】

2.4.1 设定

要素	描述
玩家	2 个企业，同质产品
策略	同时选择价格 $p_i \geq 0$
需求	所有消费者从低价企业购买；若 $p_1 = p_2$，市场均分
成本	相同且恒定的边际成本 c，无产能约束

核心差异：策略变量从**产量**变为**价格**。

2.4.2 均衡推导：削价逻辑

假设企业 2 定价 p_2。企业 1 的选择：

$$D_1(p_1, p_2) = \begin{cases} D(p_1) & \text{若} p_1 < p_2 \\ \frac{1}{2}D(p_1) & \text{若} p_1 = p_2 \\ 0 & \text{若} p_1 > p_2 \end{cases}$$

需求在 $p_1 = p_2$ 处**不连续**——一个无限小的削价，夺取整个市场。
均衡的唯一候选是 $(p_1^* = c,\ p_2^* = c)$：

- 若 $p_1 > p_2 > c$：企业 1 销量得零 $\rightarrow$ 有动机削价
- 若 $p_1 = p_2 > c$：每家都想微幅削价以夺取整个市场
- 若 $p_1 > c = p_2$：企业 1 不生产；企业 2 有动机提价
- 若 $p_1 = p_2 = c$：无动机偏离——削价造成亏损，提价损失全部销量

Bertrand 悖论

$n = 2$ 就足以产生完全竞争结果——$P = MC$，$\pi = 0$。这个结论与直觉严重矛盾：人们预期双寡头应该拥有市场势力和正利润。

2.4.3 Cournot vs. Bertrand：根本分歧

改变策略变量——从产量到价格——改变了交互的全部性质。这告诉我们要非常小心地选择模型假设。

表 2.2: Cournot vs. Bertrand 对比

	Cournot	**Bertrand**
策略变量	产量 q_i	价格 p_i
需求决定方式	$P(Q)$ 自动出清	低价者获取市场
反应函数斜率	负（战略性替代）	正（战略性互补）
双寡头利润	$\pi > 0$	$\pi = 0$
关键直觉	”我多产压低你的价”	”我削价抢走你的客户”

2.4.4　Bertrand 悖论的五种化解方案

Bertrand 的结论依赖强假设：同质产品、无产能约束、一次性互动、完美消费者信息、对称成本。每一个都可以被放松：

1. **产品差异化**（Hotelling, §~2.6）：消费者并不视产品为完全替代品 → 削价夺不到**全部**顾客
2. **产能约束**（Edgeworth, §~2.5）：即使削价，低价企业也产不够 → 高价企业有剩余需求
3. **重复互动**（→ 第 ~8~ 章）：未来惩罚的威胁可以维持合谋定价
4. **搜索成本/信息摩擦**（→ 第 ~13~ 章）：消费者不知道所有价格 → 企业拥有定价权
5. **成本不对称**：若 $c_1 < c_2$，企业 1 可将价格设在略低于 c_2 处，享受 $\pi_1 > 0$

本章的剩余部分展开前两个方案。

2.4.5　不对称成本的 Bertrand

若 $c_1 < c_2$，则无法出现 $p_1 = p_2 = c$ 的均衡（企业 2 卖低于成本就亏损）。均衡变为：

$$p_1^* = c_2 \quad (\text{或略低}), \qquad D_1 = D(c_2), \qquad \pi_1 = (c_2 - c_1)D(c_2) > 0$$

低成本企业获取整个市场并获得正利润——市场势力回来了，但只限于效率更高的那家企业。

2.5　Edgeworth 模型 (1897) 与产能约束

【核心】　【进阶】

2.5.1 核心创新

Bertrand 假设企业总是可以服务整个市场。Edgeworth 指出这通常不成立——任何企业都有产能极限。

设定变化：每家企业的最大产能为 k_i——即使定价最低，最多也只能卖出 k_i。

2.5.2 三个区域

以对称双寡头（$k_1 = k_2 = k$）为例：

- **小产能**（$k \leq q^m$，垄断产量）：纯策略均衡存在——两家都满负荷生产，市价出清
- **大产能**（$k \geq D(c)$）：Bertrand 悖论恢复——$p = c$
- **中间产能**（$q^m < k < D(c)$）：**无纯策略均衡**——结果是价格周期

2.5.3 Edgeworth 价格周期

当两家企业处于中间产能区域时：

1. **削价阶段**：两家相互削价
2. **触底回升**：价格低到一定程度，低价企业产能吃紧，高价企业有利可图地提价
3. **再次削价**：另一家的反应 $\rightarrow$ 周期重来

这不是静态均衡的失败，而是无纯策略均衡时的**混合策略**均衡——企业在价格范围内随机化。现代解释（Dasgupta & Maskin 1986）将 Edgeworth 的价格周期理解为混合策略均衡的现实表现。

2.5.4 Kreps-Scheinkman (1983)：Cournot 与 Bertrand 之间的桥梁

两阶段博弈：

1. 第一阶段：企业选择产能 k_i（不可逆、可观测的投资）
2. 第二阶段：给定产能，企业同时选择价格 p_i

结果：在有效配给（efficient rationing）下，唯一的子博弈完美均衡结果恰好是 **Cournot 均衡**——相同的价格、相同的数量、相同的利润。

为什么？产能是一种承诺装置。通过限制产能，企业向对手发出信号：我无法在价格战中过于激进。预期到这一点，两家企业都在第一阶段选择 Cournot 水平的产能——第二阶段的价格竞争就会产生 Cournot 结果。

Kreps-Scheinkman 为 Cournot 模型提供了微观基础。即使表面上企业在竞争价格，只要：

1. 产能是预先设定且可观测的，
2. 产能调整慢于价格调整，

Cournot 模型就是对均衡结果的良好近似。因此，在产能约束重要的行业（能源、制造、航空），Cournot 仍然是基准分析框架。

2.5.5　现代应用

- **电力市场**：发电厂有额定容量；在电力池中投放量而不是价格
- **航空业**：座位固定（产能），然后是收益管理（价格）
- **半导体制造**：晶圆厂产能约束极紧，定价遵循 Cournot 逻辑

2.6　Hotelling 模型 (1929)：空间竞争

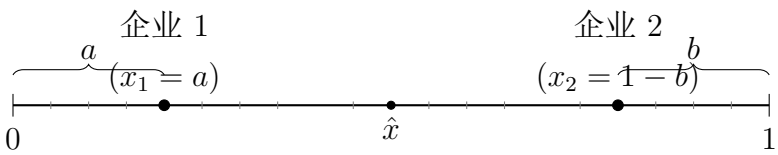
【核心】

2.6.1　为什么需要 Hotelling？

Bertrand 说两家企业就足以产生完全竞争。但现实中的双寡头——可口可乐与百事可乐、Airbus 与 Boeing——赚取正利润。原因就是：**产品差异化**。

Hotelling 用一个巧妙的框架捕捉了这一洞见：把产品差异化为物理距离。消费者既然要花交通成本才能到达不同企业，他们就不会仅仅因为几毛钱的价格差就换供应商。

2.6.2　线性城市的设定



- 消费者在 $[0, 1]$ 上均匀分布（单位密度）

- 企业 1 位于 $x_1 = a$，企业 2 位于 $x_2 = 1 - b$（距左端 a，距右端 b）
- 位于 x 的消费者从企业 i 购买获得效用：$U = v - p_i - t \cdot (\text{距离})^2$
- $t > 0$ 是" 交通成本" 参数——t 越大，产品差异化越大，消费者越不愿意切换

我们用**二次交通成本** $t(x - x_i)^2$（而非原始的线性成本），原因将在下文说明。

2.6.3 基准：固定价格的均衡

先假设价格外生且相等（$p_1 = p_2$）。每个消费者从最近的企业购买。如果企业 1 位于 $a = 0.3$，企业 2 位于 $1 - b = 0.7$，则中点 $\hat{x} = 0.5$ 左边的消费者去企业 1，右边的去企业 2——两家各分一半市场。

Hotelling 的原初洞见：假设企业 1 向右移动到 $a = 0.4$。它现在捕获了原本属于企业 2 的消费者。在固定价格下，两家企业都有动机向中心移动——这就是**最小差异化原则**：企业有动机在产品空间中互相靠近。

2.6.4 完整的两阶段博弈

现在让价格**内生**——在位置给定的情况下，企业选择价格以最大化利润。然后企业再选择位置，预期到后续的价格均衡。用逆向归纳求解。

第二阶段：价格均衡。假设企业位于 $x_1 = a$、$x_2 = 1 - b$（为简化，取对称位置 $a = b$）。无差异消费者满足：

$$v - p_1 - t(\hat{x} - a)^2 = v - p_2 - t((1 - b) - \hat{x})^2$$

解得：

$$\hat{x} = \frac{a + (1 - b)}{2} + \frac{p_2 - p_1}{2t((1 - b) - a)}$$

需求：$D_1(p_1, p_2) = \hat{x}$，$D_2(p_1, p_2) = 1 - \hat{x}$。

（推导略——结果见下方。）若企业对称位于 $a = b$，则均衡价格：

$$\boxed{p^* = c + t(1 - 2a)} \tag{2.12}$$

当企业位于两端时（$a = b = 0$，因此 $1 - 2a = 1$）：

$$\boxed{p^* = c + t, \qquad \pi^* = \frac{t}{2}} \tag{2.13}$$

结论：差异化参数 t 直接决定了加成。$t = 0$（产品同质化）$\to p^* = c$（退化到 Bertrand 悖论）。只要 $t > 0$，企业就拥有市场势力——Hotelling 解决了 Bertrand 悖论。

2.6.5　d'Aspremont 修正：从最小差异化到最大差异化

第一阶段：位置选择。现在问：企业在两端吗？

d'Aspremont、Gabszewicz 和 Thisse (1979) 证明：在二次交通成本下，每家企业的利润随向中心移动而**单调递减**。因此，均衡位置是**最大差异化**——企业位于两端。

两个相反的力量决定位置选择：

1. **市场份额效应（向心力）**：向中心移动 → 捕获更多顾客（固定价格下）
2. **价格竞争效应（离心力）**：向中心移动 → 产品间距变小 → 价格竞争加剧 → 边际降低

在二次交通成本下，**价格竞争效应压倒市场份额效应**。企业理性地选择将自己与对手" 拉开距离"——不是因为他们不喜欢顾客，而是因为他们想**软化价格竞争**。

这修正了 Hotelling 的原始主张。在原始线性成本下，企业确实向中心移动——但当它们走得够近时，纯策略价格均衡根本不存在。d'Aspremont 等人的贡献就是证明了这一点，并用二次成本提供了可行的替代方案。

教学提示：这个故事本身就具有教育意义。Hotelling 在经济直觉上是正确的（差异化确实使产品间不那么容易比较），但他的建模选择（线性成本）存在技术缺陷。直到 1979 年才有人发现这一点——距离原论文整整 50 年。

本章" 最大差异化原则"

企业有动机在产品空间中保持距离以软化价格竞争。在一个维度上，这意味着定位在两端。在现实生活中，这意味着品牌有意与竞争对手区分开来——可口可乐 vs. 百事、BMW vs. 奔驰、iOS vs. Android。

2.6.6　比较静态

- t（**差异化程度**）**增大** → 均衡价格升高 → 利润增加。产品差异化越大，市场势力越大。
- **企业间距缩小** → 竞争加剧 → 价格降低 → 利润减少。
- **福利：**社会最优位置是 $x_1 = 1/4$、$x_2 = 3/4$——最小化总交通成本。

2.6.7　Hotelling 的现代身影

- **政治竞争：**Downs (1957) 中位投票人定理——两党向中间靠拢（Hotelling 逻辑的固定价格版本）

- **平台经济**：两家互相竞争的外卖平台在" 餐厅多样性 vs. 配送速度" 的空间中定位
- **媒体偏见**：两家新闻网站选择意识形态位置以最大化受众
- **产品定位**：任何" 选择特征以吸引消费者同时避开激烈竞争" 的战略

2.7 Chamberlin (1933) 与垄断竞争

【核心】 【进阶】

2.7.1 从寡头到" 众多企业"

Cournot、Bertrand 和 Hotelling 处理的是**寡头**——少数企业，直接相互竞争。但现实世界中还有另一种不完全竞争：许多企业，每家规模很小，但产品有差异——餐馆、服装品牌、理发店、应用程序。

Chamberlin 的《垄断竞争理论》(1933) 就是为这种情形建立的。

2.7.2 设定

- **众多企业**（一个" 产品组"），每家都如此之小以至于可以忽略自己对他人的影响
- **产品差异化**：每家产品是不完全替代品 $\rightarrow$ 面临向下倾斜的需求
- **自由进入**：只要 $\pi > 0$，就有新企业进入
- **非策略行为**：不同于 Cournot/Bertrand，企业不直接与特定对手竞争——它们面对" 群体"

2.7.3 dd 曲线与 DD 曲线——Chamberlin 的创新

Chamberlin 最深刻的洞见是区分两类需求：

- **dd 曲线（感知需求/企业层面）**：一家企业改变价格，**假设其他企业维持原价**时的需求。高度弹性——因为可以夺走大量对手客户。
- **DD 曲线（份额需求/行业层面）**：**所有企业同时调整价格**时的需求。弹性低很多——因为行业齐涨齐跌时，一家企业抢不到对手的客户。

在均衡价格处，dd 和 DD 相交——企业认为的需求与它实际面对的市场份额一致。在此交点之上，dd 高于 DD（你独自提价会失去很多，实际不会那么糟）；在此之下，dd 低于 DD（你独自削价以为能抢很多，实际同行也会跟进）。

2.7.4　长期均衡与过剩产能定理

1. **短期**：给定 n，每家企业在 dd 线上选择 MR = MC $\rightarrow$ $P > AC$ $\rightarrow$ 正利润
2. **进入**：正利润吸引新进入者 $\rightarrow$ 每家企业的 dd 线向左移动
3. **长期**：进入持续到 dd 线与 AC 线**相切**——$\pi = 0$ 但 $P > MC$

过剩产能定理：长期均衡的产出水平低于平均成本曲线的最低点——最小有效规模（MES）。每家企业" 生产得太少"。行业中有" 太多的企业"。

这是 Chamberlin 的核心张力：垄断的一面（$P > MC$，向下倾斜的需求）与竞争的一面（$\pi = 0$，自由进入）相结合，产生了一个矛盾的结果——每家企业的定价高于边际成本，却赚不到经济利润。剩余被低效率" 消耗" 掉了——以过剩产能的形式。

现代经济学对这个定理的态度更温和：消费者重视**多样性**——他们愿意为更多选择支付略高的价格。过剩产能是多样性的成本，不一定是" 浪费"。

2.7.5　Chamberlin vs. Cournot/Bertrand/Hotelling

表 2.3: 竞争模式的对比

	完全竞争	垄断	寡头	垄断竞争
企业数量	∞	1	少	多
产品	同质	同质	同质/差异	差异
进入	自由	封锁	受限	自由
P vs. MC	$P = MC$	$P > MC$	$P > MC$	$P > MC$
长期 π	0	> 0	可变	0
策略互动	无	无	有	无

2.7.6　从遗忘到复兴：Dixit-Stiglitz (1977)

Chamberlin 的框架在 1940–1970 年间几乎停滞。芝加哥学派批评它" 缺少可检验的预测"。但它在不经意间复兴了——通过**国际贸易**和**宏观经济学**。

Dixit 和 Stiglitz (1977) 用一个 CES（常替代弹性）效用函数重新公式化了 Chamberlin：

$$U = \left[\sum_{i=1}^{N} x_i^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}$$

其中 $\sigma > 1$ 是替代弹性。这个公式：

- 使 Chamberlin 的直觉在一般均衡中**数学上可处理**
- 成为**新贸易理论**（Krugman 1979）和**新凯恩斯宏观**的基础
- 为产品差异化产生了精确的福利分析——不像 Chamberlin 的图形方法

现代 IO 更偏爱寡头模型（Cournot, Bertrand, Hotelling + 博弈论），但 Chamberlin 的精神——产品差异化意味着市场势力——弥漫在每一个差异化产品的 IO 模型中。第 ~11~ 章中的 BLP 需求估计在某种意义上就是一个结构化了的高维 Chamberlin 模型。

2.8 Joan Robinson (1933) 与不完全竞争

【进阶】

2.8.1 与 Chamberlin 的分野

同年但有不同焦点。Chamberlin 的核心是**产品差异化**。Robinson 的核心是**垄断权力作为一般概念**——她将 Marshall 的工具体系扩展，以分析任何偏离完全竞争的情形。

表 2.4: Chamberlin vs. Robinson

	Chamberlin	**Robinson**
核心概念	产品差异化	垄断权力的一般化
分析风格	图形（dd/DD 曲线）	边际分析（MR = MC）
关键贡献	垄断竞争模型	边际收益曲线 + 价格歧视 + 买方垄断
对广告的态度	核心特征（”销售成本”）	基本忽略
终极命运	融入贸易理论/宏观经济学	分析工具被微观经济学吸纳

2.8.2 Robinson 的两个主要分析贡献

1. 边际收益曲线。 Robinson 发展了 $MR = MC$ 作为分析不完全竞争的**通用**工具。在完全竞争下，$MR = P$（企业为价格接受者）。在不完全竞争下，$MR < P$（必须降价才能卖出更多）。这在今天看来显而易见，但当时是革命性的。

2. 价格歧视。 Robinson 提供了第三级价格歧视的第一个系统分析——对需求弹性不同的消费者群收取不同的价格。她的形式化条件——垄断者在各市场间使 MR 相等——仍然是现代教科书的范式。

3. 买方垄断（Monopsony）。 Robinson 创造了" 买方垄断" 一词（希腊语 *opsōneîn* = " 购买"）。在买方垄断下，单一买家面对向上倾斜的供给曲线——采购一个额外单位的边际成本超过供给价格。结果：采购量低于完全竞争水平，价格（如工资）低于边际收益产品。

Robinson 将买方垄断应用于劳动力市场——一个矿山作为镇上唯一的雇主——并将买方垄断**剥削**定义为" 工资与工人边际收益产品之间的差距"。她后来 (1938) 用它来解释性别工资差距：如果女性的劳动供给弹性低于男性，即使没有直接的歧视，她们的工资也会更低。

缘何重要：Robinson 创造的" 边际收益曲线" 概念成为微观经济学的标准工具。她关于买方垄断的洞见——近期劳动力市场集中度研究重新点燃——在一个世纪后仍然相关。然而，她后来贬低自己的书为" 学究气" 和" 非原创"——这是高标准的自我批评，而非准确的评估。

2.9　本章小结

这一章覆盖了 95 年的思想史——六位学者，各自以模型回应同一个根本问题。核心收获：

1. **不完全竞争被数学化了。** Cournot 证明了 n 个企业 $\to$ 均衡价格介于垄断价格和边际成本之间。企业越多，市场势力越弱；但收敛是渐进的——仅少数企业就足以产生可观的福利损失（$\approx 1/(n+1)^2$）。

2. **策略变量至关重要。**将变量从产量切换到价格——Cournot 到 Bertrand——改变了关于双寡头利润的每一个预测。这个" 模型不确定性" 不是缺陷，而是特征：它迫使我们去思考，在真实市场中哪种策略变量更合理。

3. **产能是 Bertrand 和 Cournot 之间的桥梁。** Edgeworth 表明产能约束打破了 Bertrand 悖论；Kreps-Scheinkman 证明了两阶段产能 – 价格博弈**就是**简化为了 Cournot 结果——为 Cournot 模型提供了现代微观基础。

4. **差异化意味着市场势力。** Hotelling 的空间模型表明，产品差异化（以 t 量化）软化价格竞争并恢复正利润。d'Aspremont 修正揭示了竞争效应与市场份额效应之间的张力——这对建模者而言是一个警示和教训。

5. **" 大量" 不完全竞争也是有意义的。** Chamberlin 和 Robinson 表明，即使没有直接策略互动——许多小企业、差异化产品、自由进入——市场在 $P > MC$ 的情形下运行，产生了" 多样性成本"。

6. **这些模型不只是历史——它们是现代 IO 的积木。**在 第 ~3~ 章–第 ~12~ 章中，你将反复看到 Cournot、Bertrand 和 Hotelling 被用作理论推导和实证估计的基准框架。

与后续章节的联系：→ 第 ~3~ 章将介绍哈佛 SCP 范式如何将 Cournot 的" 结构 → 绩效" 直觉操作化为一套实证分析工具——以及这项努力为何最终失败。

2.10 习题

基础

2.1 Cournot 比较静态。在一个 n 企业 Cournot 市场（$P = a - bQ$，$MC = c$）中，计算 $n = 1, 2, 3, 5, 10$ 时的 Lerner 指数 L。n 从 1 增到 2 的福利收益大，还是从 9 增到 10 的收益大？这对反垄断政策在高度集中市场（$n = 2 \to 3$）与相对分散市场（$n = 9 \to 10$）上的优先次序有何含义？

2.2 Bertrand 均衡的唯一性。证明：在同质产品 Bertrand 双寡头中，$(p_1 = c, p_2 = c)$ 是唯一的纯策略 Nash 均衡（用偏离逻辑：逐一检查 $(p_1 > c, p_2 > c)$、$(p_1 > c = p_2)$ 和 $(p_1 < c, p_2 < c)$ 的情况）。

中级

2.3 边际成本变化的 Cournot 效应。在 Cournot 双寡头（$P = 100 - Q$）中，假设初始 $c_1 = c_2 = 10$。现在企业 1 的成本降至 $c_1 = 5$。计算均衡产量、价格和利润。企业 2 的利润下降了**因为**对手变得**更有效率**——解释这一表面上的悖论。

2.4 Hotelling 价格推导。在 Hotelling 线性城市中，设 $a = b = 0$（企业位于两端）、二次交通成本 $t(x - a)^2$ 和 $c = 0$。写出无差异消费者的完整推导，推导两家企业的需求函数，并求解均衡价格 p^*。确认 $p^* = t$。

2.5 直接对比 Cournot-Bertrand。设一个行业（$P = 1 - Q$，$c_1 = c_2 = 0$）。

(a) 求解 Cournot 双寡头均衡（q^*、p^*、π^*、CS、DWL）。

(b) 求解 Bertrand 双寡头均衡（同上）。

(c) 哪种竞争模式为消费者产生更高的福利？为生产者呢？

进阶

2.6 垄断竞争均衡。考虑一个有 n 家对称企业的垄断竞争行业。每家面临需求 $q_i = S\left[\frac{1}{n} - b(p_i - \bar{p})\right]$，其中 $\bar{p}$ 是平均行业价格，S 是市场总规模。成本为 $C(q_i) = F + cq_i$。

(a) 给定 n，求解利润最大化价格 p_i^*（用 $\bar{p}$、c、b 表示）。

(b) 施加对称性 $p_i = \bar{p} = p^*$ 并求解 p^* 和 q^*（用 n、S、b、c 表示）。

(c) 施加自由进入零利润条件 $\pi_i = 0$ 并求解 n^* 和 p^*。

(d) 随着 S（市场总规模）增大，n^* 如何变化？提供经济直觉。

2.7 Kreps-Scheinkman 的直觉。（无形式推导。）在 KS (1983) 的两阶段博弈（产能 → 价格）中，解释：**为什么**理性企业在第一阶段选择的产能恰好等于 Cournot 数量？在答案中使用" 承诺" 和" 软化竞争" 两个词。

2.8 建模者的一课。 Hotelling (1929) 在" 稳定性竞争" 中声称得出了最小差异化。d'Aspremont、Gabszewicz 和 Thisse (1979) 证明该模型在原始设定下不存在纯策略均衡。讨论：这个故事对于经济建模中的数学严谨性意味着什么？如果仅凭经济直觉来判断模型是否正确，足以吗？

2.11　延伸阅读

★ 必读（本章核心参考）

- **Cournot, A.** (1838). *Researches into the Mathematical Principles of the Theory of Wealth.* Ch.7 "Of the Competition of Producers." 英文译本（Bacon 1897，Irving Fisher 作序）——读取原文，见证 Cournot 在没有博弈论术语的情况下如何直觉性地写出 Nash 均衡。

- **Tirole, J.** (1988). *The Theory of Industrial Organization.* MIT Press. Ch.5 "Short-Run Price Competition." ——研究生层面 Cournot/Bertrand 的标准处理方法。Tirole 对 Bertrand 悖论的展开是本领域最好的写作之一。

- **Hotelling, H.** (1929). "Stability in Competition." *Economic Journal*, 39, 41–57. ——前 10 页给出了设定和关键直觉。阅读时注意：由于线性成本的存在，该模型在技术上存在缺陷——它是在没有 Nash 均衡概念的情况下写成的。

- **d'Aspremont, C., Gabszewicz, J., & Thisse, J.-F.** (1979). "On Hotelling's 'Stability in Competition'." *Econometrica*, 47, 1145–1150. ——四页，改变了空间竞争理论。证明 Hotelling 的设定存在缺陷，用二次成本修复，结果反转为最大差异化。典范之作：如何用短论文改写一个领域。

★★ 进阶（加深理解）

- **Kreps, D. & Scheinkman, J.** (1983). "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes." *Bell Journal of Economics*, 14, 326–337.

——Cournot 与 Bertrand 之争的关键和解。技术性好读性各半，但洞见深远。

- **Chamberlin, E.** (1933). *The Theory of Monopolistic Competition*. Ch.5. ——垄断竞争的标准一手文献。与 Robinson 对比，注意对产品差异化相较于对分析工具的不同侧重。

- **Corchón, L.** (2008). "Welfare Losses under Cournot Competition." *International Journal of Industrial Organization*, 26, 1120–1131. ——PWL $= 1/(n+1)^2$ 公式的来源。表明在对称线性 Cournot 下，福利损失随 n 极快地平方衰减。

- **Robinson, J.** (1933). *The Economics of Imperfect Competition*. Book V–VI. ——阅读 Robinson 的价格歧视（Book V）和买方垄断（Book VI）以了解她的技术贡献。

★★★ 前沿（该方向的当前工作）

- **Dixit, A. & Stiglitz, J.** (1977). "Monopolistic Competition and Optimum Product Diversity." *American Economic Review*, 67, 297–308. ——CES 形式化让垄断竞争起死回生。是 Krugman 贸易模型和新凯恩斯宏观的数学基础。在第 ~11~ 章（BLP）中也有回响。

- **Thisse, J.-F. & Ushchev, P.** (2018). "Theory of Monopolistic Competition: A Selective Review." ——Dixit-Stiglitz 诞生 40 年以来的最新综述。揭示 CES 框架为何是" 刀刃" 情形，以及超越后的洞见。

- **Colombo, S.** (2025). *Spatial Economic Modelling*. ——最新空间竞争专著，含 50 余幅插图。涵盖政治竞争、环境经济学、市场营销中的 Hotelling 模型应用。

下一章： → 第 ~3~ 章 结构-行为-绩效：一个范式的诞生——哈佛学派如何将古典直觉转化为可操作的实证框架

第三章　结构-行为-绩效：一个范式的诞生

章首语

1950 年代的哈佛经济学家面对一个难堪的处境。Cournot (1838) 已经用数学证明：行业里企业越少，价格越高。但这个结论在真实世界中成立吗？现实中高集中度的行业——汽车、钢铁、卷烟——是否真的在剥削消费者？还是说，那些行业之所以集中，恰恰因为少数高效率企业提供了更好的产品？

Joe Bain 决定用最笨的办法回答这个问题。他选取了 20 个制造业，一个行业一个行业地调查：发问卷、做访谈、估算工厂的最小有效规模、比较利润率。他没有计量软件，没有电子计算机——只有铅笔、纸张，以及一个信念：市场结构的效应是可以被测量的。

他的回答开启了产业组织学第一个系统的经验研究纲领。这个纲领后来被称为**结构-行为-绩效（SCP）范式**。本章讲述它的诞生。

3.1　核心问题

读完本章，你应该能够回答以下问题：

1. **SCP 范式的基本逻辑是什么？**为什么” 结构 → 行为 → 绩效” 这个因果链对反垄断政策如此有吸引力？
2. **Mason 和 Bain 各自贡献了什么？**为什么说 Mason 提供了愿景，Bain 提供了测量工具？
3. **Bain 如何定义和测量” 进入壁垒”？**三个来源——规模经济、产品差异化、绝对成本优势——各自的经济逻辑是什么？
4. **Scherer 的 SCP 框架比 Bain 多了什么？**基本条件、反馈回路、政府政策——这些为什么重要？
5. **SCP 范式的政策含义是什么？**为什么它导致了” 大就是坏”（Big is Bad）的结构主义反垄断？

第 ~2~ 章以 Cournot、Bertrand 和 Hotelling 的数学模型展示了不完全竞争的逻辑可能性：n 个企业 → 价格介于垄断和边际成本之间。但这是一个纯演绎的世界——给定假设，推导结果。SCP 范式走的是相反的路：从真实世界的数据出发，**归纳**出结构与绩效之间的经验规律。这两种知识生产方式——演绎 vs. 归纳——之间的张力，将是 第 ~3~ 章–第 ~10~ 章的核心主题。

3.2　历史背景：为什么是 1930 年代的哈佛？

3.2.1　四种力量的汇聚

产业组织学在哈佛的诞生不是偶然的。四种力量在 1930 年代同时汇聚于剑桥：

力量	关键人物/事件	贡献
理论跳板	Chamberlin (1933)	垄断竞争——在完全竞争和垄断之间建立了理论光谱
制度危机	大萧条 (1929–1939)	市场自我修复假设破产 → 需要理解” 市场为何失灵”
政策需求	Thurman Arnold (DOJ 1938–1943)	新政反垄断高潮 → 律师需要经济学家定义” 什么算垄断”
方法论创新	Mason (1936–1939)	哈佛案例法 + 统计数据的结合

Chamberlin 提供了理论跳板。在 第 ~2~ 章中，我们看到 Chamberlin 的《垄断竞争理论》(1933) 构建了介于完全竞争和完全垄断之间的理论空间——差异化产品、向下倾斜的需求曲线、自由进入导致零利润但 $P > MC$。这个框架暗示：**市场结构（企业数量、产品差异化程度）系统地影响定价和利润**。但 Chamberlin 本人主要做理论推导——他对” 真实世界的市场结构长什么样” 没有给出系统的经验回答。

Mason 认为这个经验问题才是关键。 Edward S. Mason (1899–1992) 是哈佛经济系的支柱人物之一，后来担任哈佛肯尼迪学院院长。他并不满足于 Chamberlin 的理论分类——他认为，经济学家的任务不仅是列出逻辑可能性，更要判断**在真实世界中哪一种逻辑在起作用**。

3.2.2　Mason (1939)：SCP 的创立文献

Mason, E.S. (1939). “Price and Production Policies of Large-Scale Enterprise.” *American Economic Review*, 29(1), 61–74.

这篇 13 页的论文是 SCP 传统的奠基文本。Mason 提出了一个对当时 IO 研究影响深远的诊断和处方：

诊断：新古典价格理论中的核心概念——需求弹性、边际成本、边际收益——在经验上**不可观测**。它们来自公理推导，不是来自数据观察。因此，反垄断政策不能等待经济学家” 测量出需求弹性”——那可能永远等不到。

处方：一种半理论、半经验的混合方法：

1. 理论告诉我们**哪些变量重要**（集中度、进入条件、产品特征、需求增长）
2. 但变量之间的**具体关系**必须通过经验研究来确立——案例研究、行业调查、统计数据
3. 目标不是推导定理，而是**识别经验规律**，为反垄断政策提供可操作的判断标准

Mason 用一句著名的话描述这个领域的状态：**"muddy but not uninteresting"**（泥泞但不乏趣味）。这句话后来被反复引用，既体现了 Mason 的智识诚实——他知道经验 IO 的方法论还远不成熟——也折射出 1930 年代经验研究面对" 没有计量软件、没有大型数据库、没有博弈论" 时的真实处境。

Mason 的" 泥泞"：方法论起点如何决定范式命运。 Mason 选择了一条归纳的、经验驱动的路——从行业研究中归纳出规律，而非从公理中推导定理。这个选择使 SCP 在政策界获得了极大的影响力（它为反垄断提供了" 可操作" 的标准），但也为它后来的方法论困境埋下了伏笔——当芝加哥学派用更精致的价格理论进行演绎攻击时，SCP 的经验基础（小样本、会计数据、无因果识别）显得脆弱不堪。第 ~4~ 章将详细展开这个" 清算" 过程。

注意：Mason 本人**没有**画出任何"S → C → P" 的箭头图。他的框架是松散和归纳的——他强调的是" 应该系统地研究结构与绩效之间的关系"，而不是" 结构机械地决定绩效"。那个著名的箭头图来自 Bain 和 Scherer 的后续形式化。

3.2.3 Mason 1949：SCP 进入反垄断

Mason, E.S. (1949). "The Current Status of the Monopoly Problem in the United States." ***Harvard Law Review*****, 62, 1265–1285.**

这篇发表在《哈佛法律评论》上的论文将 SCP 框架明确引入反垄断政策讨论。Mason 指出：

- 反垄断法需要一个**经济学上可操作的标准**来判断" 什么是垄断"
- 市场结构——企业数量、集中度、进入条件——是可以被观察和测量的
- 因此，反垄断政策应该以**保护和促进" 可行竞争"（workable competition）**为目标——这一概念后来成为哈佛学派的标准政策语言

正是这个论证首次在" 市场结构" 和" 反垄断执法" 之间建立了直接的逻辑链条——这个链条将在 1960 年代的标志性判例中被锻造为法律原则。

3.3　Bain (1956)：进入壁垒的系统理论

【核心】

3.3.1　从 Mason 到 Bain：一个博士生将愿景变为纲领

Joe S. Bain (1912–1991) 在哈佛的博士委员会包括三位 20 世纪经济学殿堂级人物：**Joseph Schumpeter**（创新理论）、**Edward Chamberlin**（垄断竞争）和 **Edward Mason**（导师）。一个学生同时接受这三个人的训练——这是 IO 史上独一无二的” 学术基因” 组合。

Bain 后来去了 UC Berkeley，在那里完成了 IO 历史上第一个系统的经验研究专著。他的贡献可以总结为一句简洁的对比：Mason 说” 我们应该研究结构”；Bain 说” **好的，让我告诉你如何测量它**”。

3.3.2　Bain (1951)：第一次系统检验——42 个制造业

Bain, J.S. (1951). “Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936–1940.” *Quarterly Journal of Economics*, 65(3), 293–324.

这是经济学史上第一次用系统数据检验” 集中度-利润率” 关系的研究：

- **样本**：42 个美国制造业（1936–1940 年数据）
- **集中度指标**：前八家企业集中率（CR8），来自美国制造业普查
- **利润指标**：会计权益回报率（5 年均值），来自 Moody’s Manuals 和 SEC 报告
- **方法**：分组均值比较——**不是回归分析**（1951 年的经济学尚未普及计量回归方法）

核心发现：CR8 > 70% 的行业，其平均利润率**显著高于**更分散的行业。Bain 将此解释为存在一个” 临界集中度”——当少数企业控制了足够多的市场份额时，合谋变得可行，价格得以维持在竞争水平之上。

Bain 的” 临界集中度” 假说

CR8 ≈ 70% 是集中度-利润率关系中的” 拐点”——超过该阈值后，行业利润出现跃升。这一发现为反垄断政策提供了一个简单的经验法则：高度警惕 CR8 > 70% 的行业。

不过，这一” 临界值” 的精确性和普适性后来受到了大量研究的质疑（第 ~4~ 章）。Bain 本人的 42 个行业样本太小，无法精确识别阈值位置。

3.3.3 Bain (1956)：《新竞争者的壁垒》——IO 的" 创世记"

Bain, J.S. (1956). ***Barriers to New Competition: Their Character and Consequences in Manufacturing Industries.*** **Harvard University Press, 329 pp.**（中译本：徐国兴译，人民出版社，2012）

这是 IO 领域第一个系统的经验研究专著。Bain 选取 **20 个美国制造业**（4 位 SIC 代码），采用了一套在今天看来" 手工" 但在当时是开创性的混合方法：

- **行业问卷**：向企业发放半结构化问卷（见原书附录 A——可能是 IO 史上最早的正式企业调查工具）
- **企业访谈**：与行业参与者深度交流
- **工程估算**：由工程师和行业专家估算" 最小有效规模"（MES）——即实现最低单位成本所需的最小工厂产出
- **公开数据**：制造业普查的结构数据、Moody's 和 SEC 的财务数据

这项工作在电子计算机普及**之前**完成——是典型的" 手工经验经济学" 的巅峰之作。

3.3.4 Bain 的进入壁垒三分类

Bain 将进入壁垒定义为：

> "在位销售者相对于潜在进入者的优势，体现为在位者能够持续将价格维持在竞争水平以上，而**不吸引新企业进入**的程度。"

这是**基于后果的定义**——进入壁垒的存在不是通过观察某个具体障碍，而是通过观察结果（持续超额利润且无进入发生）来**推断**。

Bain 识别了**三类进入壁垒**：

3.3.5 Bain 的 20 个行业数据：壁垒与利润

Bain 将每个行业的**整体进入壁垒高度**评定为三个等级：非常高、显著、中低。以下是他的核心发现：

核心发现：壁垒越高，利润越高——**而且是持续性的**（1936–40 和 1947–51 两个子时期高度一致）。这意味着：进入壁垒——而不仅仅是集中度——才是市场势力的根本来源。

表 3.1: Bain 的三类进入壁垒

壁垒类型	经济逻辑	Bain 的测量方法	典型行业举例
大规模经济 (Economies of Large Scale)	MES 占行业产出的比例越大，进入者越难以有效规模进入——要么以亚优规模进入（成本劣势），要么以有效规模进入（大幅增加行业产出，压低价格，可能触发报复）	工程估算 MES ÷ 行业总产出（" 百分比效应"）	汽车、打字机、拖拉机
产品差异化优势 (Product Differentiation Advantages)	在位者的品牌、广告积累、声誉构成需求侧优势——进入者须额外支出以克服消费者对在位品牌的偏好	广告强度、品牌忠诚度评估	卷烟、蒸馏酒、肥皂
绝对成本优势 (Absolute Cost Advantages)	在位者控制专利、优质资源、优越技术或低成本融资渠道——进入者无法复制这些条件	专利持有、投入品价格比较、融资成本	铜、炼油

表 3.2: Bain 的进入壁垒评级与利润率 (20 个制造业)

壁垒等级	行业	平均利润率 (1936–40)	平均利润率 (1947–51)
非常高	汽车 卷烟 蒸馏酒 打字机（办公设备）	$\approx 11.6\%$	$\approx 12.8\%$
显著	钢铁 农用机械 铜 炼油 肥皂 橡胶轮胎 钢笔	$\approx 9.2\%$	$\approx 9.5\%$
中低	水泥 肉类加工 鞋类 罐头蔬果 面粉 乳制品 人造丝 石膏 锡罐	$\approx 7.2\%$	$\approx 7.5\%$

产品差异化是”最强壁垒”。 Bain 的一个重要发现是：在三类壁垒中，**产品差异化**对利润的影响最大。卷烟和蒸馏酒（”非常高”壁垒）的利润远超钢铁和炼油（”显著”壁垒），尽管后两者的资本密集度更高。品牌和广告——而非工厂规模——创造了最持久的市场势力。这一发现在今天看来依然成立：可口可乐的护城河不是它的装瓶厂（完全可以复制），而是它的品牌。

3.3.6　Bain 的方法论遗产与早期局限

遗产：

- 将 Mason 的松散框架**形式化**为可操作的 S → C → P 分析范式
- 撰写了第一本 IO 教材（*Industrial Organization*, 1959, Wiley, 643 pp.），将 IO 确立为独立的经济学分支
- 1982 年 AEA（美国经济学会）授予 Bain “Distinguished Fellow”，称其为”现代产业组织经济学无可争辩之父”

早期局限（此处仅简要标注，第 ~4~ 章详细展开）：

- 样本量小（20–42 个行业）→ 统计推断能力有限
- 依赖会计利润率 → 不等于经济利润（Fisher-McGowan 1983 的批判远在其后）
- 壁垒分类有主观判断成分 → Stigler (1968) 将给予根本性的重新定义
- 将结构视为外生变量 → Demsetz (1973) 的效率假说将翻转因果方向

3.4　SCP 范式的完整表述

【核心】

3.4.1　Bain 的单向模型 (1959)

Bain (1959) 的教科书《产业组织》将 SCP 呈现为一个清晰的线性因果链：

$$\text{市场结构} \longrightarrow \text{市场行为} \longrightarrow \text{市场绩效}$$

- **结构（Structure）**：集中度（CR4/CR8）、进入壁垒、产品差异化。**视为外生变量**——由技术条件（规模经济、需求特征）决定。
- **行为（Conduct）**：定价策略、广告、研发、合谋、并购。**由结构决定**——在集中行业中，合谋更容易；在壁垒高的行业中，进入遏制更可能。

- **绩效（Performance）**：利润率、配置效率、技术进步、公平。**由行为决定**——合谋导致高价低效，竞争推动创新与效率。

这个简单的线性模型有一个**核心假设**：结构是外生的。结构由技术条件给定，不由企业的策略选择决定。这个假设使 SCP 回归变得简单——直接把利润率回归到集中度上，不必担心反向因果——但也使其容易受到 Demsetz 批评的冲击（第 ~4~ 章–第 ~5~ 章）。

3.4.2 Scherer (1970)：SCP 的" 圣经"

Scherer, F.M. (1970). ***Industrial Market Structure and Economic Performance.*** **Rand McNally.**（后续版本：1980；Scherer & Ross 1990, 1996。）

如果说 Bain 提供了 SCP 的骨架，Scherer 则赋予了它血肉和神经网络。Scherer 的教材通常被称为 SCP 的" **圣经**"——它在以下四个方面超越了 Bain 的简化版本。

贡献 1：引入" 基本条件"

Scherer 明确提炼了比市场结构更深层的外生变量——**基本条件**（Basic Conditions）：

供给端基本条件	需求端基本条件
原材料可得性	需求价格弹性
技术特征	市场增长率
产品耐用性	替代品可得性
商业惯例与工会化	购买方式（频率、批量）

基本条件来自技术和偏好，而非企业决策——它们构成整个 SCP 大厦的**地基**。例如：水泥行业的基本条件（产品同质、重运输成本、需求周期性）决定了其市场结构必然以区域性寡头为主。

贡献 2：引入反馈回路

Scherer 最关键的修正：**箭头不仅是单向的**。他加入了一系列反馈回路：

- 行为 $\rightarrow$ 结构：广告改变市场集中度；并购直接改变结构
- 绩效 $\rightarrow$ 结构：高利润吸引新进入者 $\rightarrow$ 结构随时间变化
- 行为 $\rightarrow$ 基本条件：研发改变技术特征；广告影响消费者偏好

这使 SCP 从**静态快照**升级为**动态过程**——结构不再是一成不变的" 给定条件"，而是可以随时间演化的。这一修正在逻辑上更加完整，但在经验操作上增加了复杂性：如果一切都在变化，跨行业横截面回归还能捕捉到什么？

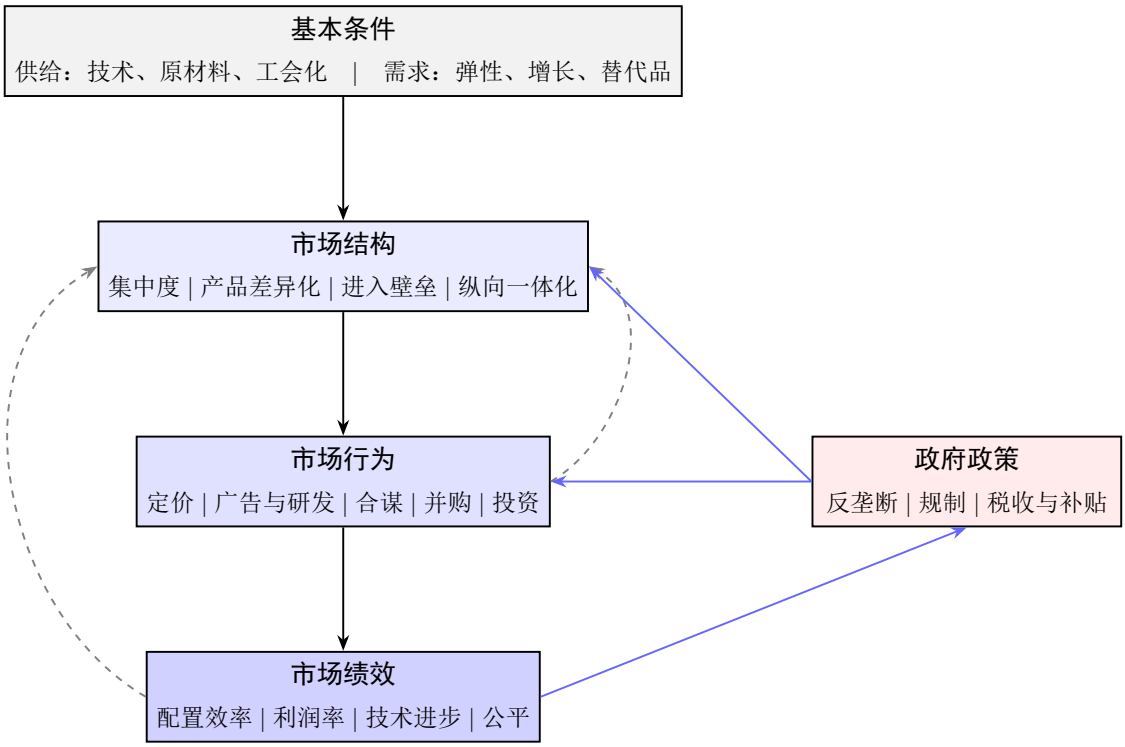
贡献 3：引入政府政策

Scherer 将政府政策（反垄断、规制、税收、贸易政策）明确纳入框架——政策既是影响所有层次的**外部力量**，也受到绩效的**反馈影响**（糟糕的市场绩效触发更积极的执法）。

贡献 4：全面的经验综述

Scherer 综合了几百项经验研究，为 IO 的每个板块（集中度、进入壁垒、定价、广告、创新、规制）提供了”我们目前知道什么”的权威综述。

3.4.3　SCP 流程图



这是 IO 领域最具辨识度的图。每个 IO 研究者都在某个时候见过它——或它的某个变体。让学生从这张图开始建立”认知地图”：SCP 不仅是一个分析框架，更是一种**观察市场的视角**——从结构出发，经由行为，落脚于绩效，同时留意政策对每一个环节的介入和反馈。

SCP 作为一种” 世界观”

SCP 的吸引力不仅在于它是一个分析框架，更在于它是一种**世界观**：

1. 市场不是天然竞争的——**结构预先决定了竞争的可能性**
2. 绩效是结构的产物——要看” 行业长什么样”，才能判断” 行业表现好不好”
3. **政策干预是正当的**——如果结构坏了，就修理结构

这种世界观在 1930–1960 年代是有深厚土壤的：大萧条摧毁了对市场自我修复的信念，新政反垄断提供了政策工具，Chamberlin 的垄断竞争理论提供了理论语言。SCP 是那个时代的产物——理解这一点，就不会简单地把它视为” 错的”，而是看到它在特定历史条件下的合理性。

3.5 测量工具：集中度与进入壁垒的操作化

【核心】 【进阶】

SCP 的核心承诺是**用可观测的结构变量预测市场绩效**。因此，测量的质量直接决定范式的成败。

3.5.1 集中度的测量：从 CR_n 到 HHI

集中率 CR_n 是最简单的市场集中度指标：

$$CR_n = \sum_{i=1}^{n} s_i \times 100 \tag{3.1}$$

其中 s_i 是第 i 家企业的市场份额（如 $0 < s_i \leq 1$）。最常用的是 CR4 和 CR8。

表 3.3: Bain 基于 CR 的市场结构分类

市场类型	CR4	CR8
高度集中型寡头	$\geq 60\%$	$\geq 70\%$
中度集中型寡头	30%–60%	40%–70%
竞争型/原子型	$< 30\%$	$< 40\%$

CR_n 的局限：它完全忽略了前 n 家企业**之外的分布**。两个 CR4 相同的行业可能竞争格局截然不同：

- 市场 A：{25%, 25%, 25%, 25%} → 四家均势，竞争激烈

- 市场 B：$\{70\%, 10\%, 10\%, 10\%\}$ → 一家独大，其余边缘化

赫芬达尔-赫希曼指数（HHI）正是为克服这一缺陷而设计的：

$$HHI = \sum_{i=1}^{N} s_i^2 \tag{3.2}$$

其中 s_i 是百分比形式的市场份额（如 25% 则 $s_i = 25$）。HHI 的范围从接近 0（完全原子化）到 10,000（纯垄断）。

Hirschman 还是 Herfindahl？一个双重独立发明的故事。
HHI 的命名史本身就是一堂学术思想史课：

- **Albert O. Hirschman (1945)** 在《国家权力与对外贸易结构》中首次提出该指数，用于分析**贸易流集中度**。但该书出版于战争末期，在经济学界流通有限。
- **Orris C. Herfindahl (1950)** 在哥伦比亚大学博士论文《美国钢铁业的集中度》中**独立**提出了完全相同的指标，专门应用于**产业集中度分析**。

两个学者在不同大学、不同应用领域、彼此不知道对方工作的情况下，得出了相同的公式。Herfindahl 先将其应用于 IO（1950），Hirschman 先发表（1945）。该指数的正式名称——赫芬达尔-赫希曼指数——反映了经济学界对这一" 双重发明" 的承认。今天，大多数简称为 HHI，但有些文献将其称为 Herfindahl Index 或 Hirschman-Herfindahl Index (HHI)。

HHI 对 CR_n 的优势：

- 平方加权使**大企业获得更大的影响力**——符合" 市场势力主要来自大企业" 的直觉
- 考虑了市场中**所有**企业，而非仅前 n 家
- 能区分不同的份额分布——市场 A（均分）的 $\text{HHI} = 4 \times 25^2 = 2,500$ vs. 市场 B（一家独大）的 $\text{HHI} = 70^2 + 3 \times 10^2 = 5,200$

HHI 的反垄断阈值演进：DOJ/FTC 合并指南中的 HHI 阈值在过去 40 年经历了" 先放宽、再收紧" 的摆动——这本身就是芝加哥学派与新布兰代斯学派交替主导政策议程的反映（第 ~5~ 章、第 ~15~ 章）：

2023 年的变化值得注意：在 2010 年大幅放宽（高度集中阈值从 1,800 升至 2,500）之后，2023 年指南来了一个急转弯——不仅恢复了更低的阈值，甚至取消了" 非集中市场" 的安全港。这是" 新布兰代斯学派"（New Brandeis movement）在反垄断执法中最实质性的制度化胜利之一。第 ~15~ 章将深入讨论这场" 反垄断复兴" 运动。

表 3.4: DOJ/FTC 横向合并指南的 HHI 阈值演进

市场分类	1982	1992	2010（放宽）	2023（收紧）
非集中	$< 1{,}000$	$< 1{,}000$	$< 1{,}500$	取消了安全港
中度集中	1,000–1,800	1,000–1,800	1,500–2,500	——
高度集中	$> 1{,}800$	$> 1{,}800$	$> 2{,}500$	$> 1{,}800$
ΔHHI 推定触发值	> 100	> 100	> 200	> 100

3.5.2 进入壁垒的测量：Bain 之后

Bain 的进入壁垒分类奠定了讨论框架，但操作性定义一直存在争议。主要进展包括：

Stigler (1968) 的重新定义（第 ~5~ 章详细展开）：进入壁垒应被限定为" **进入者必须承担、而在位者无需承担的生产成本**"——一个显著的窄化。按此标准，规模经济不是壁垒（因为进入者也可以达到同等规模）；资本要求不是壁垒（除非存在进入者的融资成本不对称）。

幸存者原则（Stigler 1958；Comanor & Wilson 1967）：不依赖工程估算，而是通过观察存活工厂的规模分布来推断 MES——哪些规模的工厂在长期中存活了下来，哪些消失了？

沉淀成本革命（Baumol, Panzar & Willig 1982）：将分析焦点从" 固定成本" 转移到" **沉淀成本**"（不可回收的投资）。如果进入成本可以完全回收（无沉淀），则即使市场高度集中，" 打了就跑" 的进入威胁也能约束在位者。第 ~6~ 章将详细展开可竞争市场理论。

3.5.3 利润的测量问题（预览 Ch.4 的批判）

SCP 传统使用**会计回报率**（ROA、ROE）作为绩效指标。但会计利润与经济利润之间存在系统性偏差：

$$\text{会计利润} = \text{收入} - \text{显性成本}$$

$$\text{经济利润} = \text{收入} - \text{显性成本} - \text{隐性成本（包括资本的机会成本）}$$

一个典型的例子：如果企业主投入了 200 万元自有资本（市场回报率 10%）并亲自管理企业（市场工资 50 万），会计利润会**高估**真实经济利润 20 万 + 50 万 = 70 万。如果 SCP 回归使用了该企业的会计利润数据，就会将" 企业主的勤劳" 误读为" 垄断势力的回报"。

Fisher & McGowan (1983) 后来证明：**没有任何方法可以从会计回报率推断出相对经济盈利能力的正确排序**——会计惯例（折旧方法、通胀调整、无形资产处理）导致会计回报率与经济内部回报率之间的偏差不仅是数值上的，而且是**排序上的**。这

意味着 SCP 回归中" 利润" 的因变量本身可能存在系统性的测量误差——而如果这种误差与集中度相关，估计结果就是不一致的。

3.6　结构主义的政策含义

【核心】

3.6.1　"大就是坏"：从分析到行动

SCP 范式的政策处方直接且有力。如果：

1. 集中度 + 进入壁垒 → 合谋更容易 → 超额利润 → 消费者受损

2. 结构是**可观测的**（CR4、HHI、壁垒高度），而具体合谋行为**难以直接证明**

那么政策结论就是：**关注结构本身，而非逐一证明行为**。高集中度本身就应引起反垄断警惕。" 大就是坏"（Big is Bad）成为这一时期反垄断的非正式信条。

3.6.2　标志性判例：SCP 逻辑写入法律

Brown Shoe Co. v. United States (1962), 370 U.S. 294.

这是最高法院对 1950 年 Celler-Kefauver 反合并修正案的第一个司法解释。Brown Shoe（全美鞋类生产的约 4%）试图收购 G.R. Kinney（全美鞋类零售的约 1.2%）——合并后份额仅约 **5%**。

法院**阻止了该合并**。首席大法官 Earl Warren 撰写的多数意见使用了典型的 SCP 语言："我们不能忽视国会通过保护可存活的小型地方企业来促进竞争的意愿。" 法院关注的是" **集中度的上升趋势**"，而非已形成的垄断势力。并购带来的效率提升（纵向整合降低价格）被法院视为**威胁竞争对手的武器**——" 有效率" 本身成了罪证。

United States v. Philadelphia National Bank (1963), 374 U.S. 321.

确立了**结构性推定（structural presumption）**：合并后企业拥有" 不当的" 市场份额（本案中约 30%）且在集中市场中，该合并**推定违法**——政府无需逐一证明具体反竞争效果。这一推定成为此后近 20 年合并审查的基石。

United States v. Von's Grocery Co. (1966), 384 U.S. 270.

SCP 结构主义的**高水位线**。洛杉矶的 Von's Grocery 与 Shopping Bag Food Stores 合并，合计份额不到 **8%**。市场本身高度分散，存在大量小型独立杂货商。

法院**仍然阻止了合并**。大法官 Stewart 在异议中写下了著名的批评："在 Section 7 的诉讼中，**政府永远赢**。" 这句话后来被反复引用，作为 SCP 结构主义过度的标志性注脚。

其他重大案例：

- **U.S. v. IBM (1969)**：政府起诉 IBM 垄断通用计算机市场，寻求拆分。诉讼持续 13 年，最终于 1982 年被驳回——标志着结构主义执法的实际极限。

- **U.S. v. AT&T (1974)**:政府起诉 AT&T 纵向垄断。1982 年以同意判决（consent decree）告终——AT&T 被拆分为 7 家区域性运营公司（"Baby Bells"）。这是最后一个重大结构性救济。

3.6.3 1968 年 DOJ 合并指南：结构主义的制度化

DOJ 在 1968 年发布了第一份正式的合并审查指南，将 Brown Shoe 和 Von's Grocery 的判决编纂为量化的执法规则。其核心逻辑完全是 SCP 式的：**以市场份额和集中度作为合并审查的主要标准**，极低的挑战门槛，对横向合并的强烈推定违法。

这份指南的效果是戏剧性的：1960 年代至 1970 年代初，**几乎没有大型横向合并能够获得批准**。

3.6.4 德国平行发展：秩序自由主义（Ordoliberalism）

【进阶】

一个引人入胜的思想史事实是：在大西洋彼岸，一群德国学者从**完全不同的哲学起点**出发，得出了与哈佛 SCP 高度相似的结构主义结论。

表 3.5: 哈佛 SCP 学派 vs. 弗莱堡秩序自由主义

维度	哈佛 SCP 学派	弗莱堡秩序自由主义
诞生时间/地点	1930s–50s, 剑桥 MA	1930s–50s, 弗莱堡
创始人	Mason, Bain	Walter Eucken, Franz Böhm
核心关切	市场势力 → 经济绩效差	市场势力 → **个人自由受威胁**
哲学基础	美国实用经验主义	康德哲学 + 胡塞尔现象学
政策处方	去集中化、结构性救济	" 经济宪法"（Wirtschaftsverfassung）、强力反卡特尔
对竞争法的态度	竞争是**提升效率的手段**	竞争是**保障自由的价值本身**

为什么这很重要？

第一，它表明结构主义不是某个美国学术派别的偏执——它是 1930–50 年代西方世界对” 大企业 + 大萧条 + 极权主义” 的结构性反应。在弗莱堡，学者们在纳粹时期的直接经验（经济权力集中导致政治权力集中）使他们对任何形式的集中都保持了哲学层面的警惕。

第二，秩序自由主义对**欧盟竞争法的影响**是深刻的——特别是 Article 102 TFEU（禁止滥用市场支配地位）。即使在欧盟竞争法越来越” 经济化” 的今天（更多使用效果分析，更少依赖本身违法原则），秩序自由主义的 DNA 仍然嵌入在欧洲竞争法对” 市场结构” 的特殊关注之中。

第三，秩序自由主义提出了一种哈佛 SCP 没有充分发展的论点：**市场势力是自由的问题，而不仅仅是效率的问题**。在国家既是监管者又是市场参与者（通过国企）的中国语境中，这一维度具有特殊的分析价值。

3.7　从古典模型到 SCP：一个认知跃迁

【进阶】

3.7.1　Cournot 的知识生产方式 vs. Bain 的知识生产方式

第 ~2~ 章和 第 ~3~ 章代表着两种截然不同的知识生产方式：

表 3.6: 演绎与归纳：两种 IO 知识生产方式

维度	Cournot (1838)	Bain (1956)
知识类型	演绎——从假设推导均衡	归纳——从数据推断规律
工具	微积分与代数方程	行业问卷 + 利润比较
核心问题	” 给定集中度，均衡价格是多少？”	” 现实中的集中度与利润真的相关吗？”
对政策的意义	提供理论基准（n 企业 $\to P$ 介于垄断和 MC 之间）	提供可执行的行动标准（CR8 $> 70\% \to$ 警惕）
优势	逻辑严密、可推广	贴近现实、可直接指导政策
劣势	假设与现实之间可能存在鸿沟	因果识别脆弱、依赖数据质量

两种方式的张力贯穿全书：演绎的方法（第 ~2~ 章、第 ~7~ 章–第 ~9~ 章）提供了逻辑上严格但假设受限的基准；归纳的方法（第 ~3~ 章–第 ~4~ 章、第 ~10~ 章–第 ~12~ 章）试图用数据回答” 真实世界如何运作”，但面临识别上的根本困难。现代 IO

的精华——从 Bresnahan (1989) 到 BLP (1995)——恰恰在于**将两者结合**：用理论模型指导实证设定，用数据检验和修正理论。

3.7.2 全书的关键知识链

读者应该在此时看到全书的一条核心逻辑链：

$$\text{Ch.1: } L = \frac{P - MC}{P} = \frac{s_i}{|\varepsilon|} \longrightarrow \text{Ch.2: Cournot 证明} L = \frac{1}{n|\varepsilon|}$$
$$\longrightarrow \text{Ch.3: SCP 提供了测量} s_i \text{ 的系统方法（CR}_\text{n}\text{, HHI）}$$

Ch.1 引入 Lerner 指数作为市场势力的理论度量。Ch.2 证明在 Cournot 模型中，Lerner 指数是集中度（$1/n$）和需求弹性（$1/|\varepsilon|$）的简单函数。Ch.3 提供测量集中度的实际工具。Ch.4 将展示第一次系统检验这个理论链的尝试——以及它为何失败。

这是全书最关键的跨章节叙事线索之一：理论 → 操作化 → 经验检验 → 发现局限 → 新理论。

3.7.3 从 Mason 的" 泥泞" 到 Tirole 的" 清晰"

- **Mason (1939)**：" 这个领域 muddy but not uninteresting"——方法论上尚在摸索
- **Bain (1956)**：让我们系统地测量——即使工具粗糙，方向明确
- **Scherer (1970)**：加入反馈和动态——框架更完整，但经验检验更困难
- **Demsetz (1973) & 芝加哥学派**：你们的因果方向反了——颠覆开始（第 ~5~ 章）
- **Tirole (1988)**：一切都可以用博弈论精确建模——新 IO 诞生（第 ~7~ 章）
- **Bresnahan (1989) & BLP (1995)**：结构模型是经验 IO 的未来（第 ~10~ 章–第 ~11~ 章）

本章标记了这个演进弧线的**起点**——SCP 是 IO 从" 纯理论推演" 走向" 系统经验研究" 的第一步。这一步充满了方法论缺陷——但**没有这一步，就没有后来的 NEIO**。

3.8 本章小结

这一章讲述了 IO 第一个系统经验研究纲领的诞生——从 1939 年 Mason 的愿景，到 1956 年 Bain 的操作化，到 1970 年 Scherer 的全面系统化。核心收获：

1. **IO 作为一个经验学科在哈佛诞生是有原因的。** Chamberlin 的理论提供了分类框架，大萧条摧毁了对市场的盲目信任，新政反垄断创造了政策需求，Mason 的方法论提供了经验研究的合法性。四者汇聚，缺一不可。

2. **Mason、Bain、Scherer 各贡献了 SCP 的不同层次。** Mason 提供了**方向**（" 结构与绩效之间的关系值得系统研究"）；Bain 提供了**工具**（进入壁垒的分类和测量）；Scherer 提供了**系统**（基本条件、反馈回路、政府政策）。三十年完成了一个经验研究纲领的基础设施建设。

3. **Bain 的三个进入壁垒至今仍是分析的出发点。**规模经济、产品差异化和绝对成本优势——虽然 Stigler (1968) 和 Baumol–Panzar–Willig (1982) 严格收窄了定义，虽然" 进入壁垒" 的概念如今包含了更丰富的维度（网络效应、转换成本、数据优势），但任何进入分析都无法绕开 Bain 的初始分类。

4. **SCP 是一种世界观，不仅是分析工具。**它的政策含义直接而有力：结构坏了就修结构。" 大就是坏" 的逻辑在 Brown Shoe、Philadelphia National Bank 和 Von's Grocery 中被锻造成法律原则。这套世界观在 1960 年代达到了顶峰——然后在 1970 年代受到了芝加哥学派的根本性挑战。

5. **HHI 是 SCP 最持久的遗产。**从 Herfindahl 1950 年的博士论文到 2023 年的 DOJ/FTC 合并指南，HHI 仍然是全球反垄断执法的核心定量工具。它的双重独立发明故事——Hirschman (1945) 和 Herfindahl (1950)——是学术思想史上的经典篇章。

6. **SCP 的缺陷已在其诞生时埋下。**单向因果假设（结构 $\rightarrow$ 绩效）、会计利润度量（$\neq$ 经济利润）、小样本经验基础（20–42 个行业）、外生性假设（结构不由绩效决定）——这些问题在 1950 年代或许情有可原，但为芝加哥学派在 1970–80 年代的全面攻击提供了靶心。第 ~4~ 章将展示这些缺陷如何被系统地揭露和攻击。

与后续章节的联系：$\rightarrow$ 第 ~4~ 章将展示 SCP 经验传统的" 清算时刻"——跨行业回归的内在局限，Demsetz 效率假说的经验检验，以及为什么一个主导了三十年的范式最终被放弃。

3.9　习题

基础

3.1 SCP 框架应用。阅读以下行业描述，完成 SCP 分析：

”中国乳制品行业：伊利、蒙牛各占约 *25%* 的市场份额，地方品牌合计约占 *30%*，其余为进口品牌。近五年无显著新进入者。广告投入占行业收入约 *12%*。行业平均净利率约 *8%*。”

(a) 识别该行业的市场结构特征（集中度、产品差异化、进入壁垒）

(b) 推测典型企业行为（定价、广告、研发策略）

(c) 评估市场绩效（利润率的含义）

(d) 画一个简化的 S-C-P 因果图

3.2 HHI 的计算与比较。计算以下三个市场的 CR4、CR8 和 HHI：

- 市场 A：{25%, 25%, 25%, 25%}（四家均分）
- 市场 B：{70%, 10%, 10%, 10%}（一家独大）
- 市场 C：{30%, 25%, 20%, 15%, 5%, 5%}（梯度分布）

(a) 市场 A 和 B 的 CR4 均为 100%，但 HHI 分别为多少？这一差异揭示了什么？

(b) 根据 2023 年 DOJ/FTC 合并指南，每个市场分别属于哪个集中度类别？

中级

3.3 Bain 的进入壁垒分类。对以下三个行业案例，识别 Bain 三类壁垒中哪一种（或哪几种）在起作用，并说明理由：

- **制药业：**原研药企业持有 20 年专利保护，仿制药企业只能在专利到期后进入。
- **碳酸饮料业：**可口可乐和百事可乐累计投入了数百亿美元的广告和品牌建设。新品牌需要巨额支出来获得消费者认知。
- **半导体制造：**建设一座先进制程晶圆厂需要 100–200 亿美元资本投入，且需要数年才能投产。

是否存在 Bain 的分类未能充分涵盖的进入障碍？举出一个现代经济中的例子（提示：考虑数字平台）。

3.4 会计利润 vs. 经济利润。某企业年收入 1,000 万元，显性成本（原材料、工资、租金）为 700 万元。企业主投入了 200 万元自有资本（若投资于类似风险资产，预期年回报率为 10%），且企业主本人全职管理该企业（若在别处工作，年市场工资为 50 万元）。

(a) 计算该企业的会计利润。

(b) 计算该企业的经济利润。

(c) 如果 SCP 研究者使用该企业的会计利润率作为" 市场绩效" 指标，会产生什么偏差？这一偏差在跨行业比较中是否系统性地与集中度相关？（提示：大企业更可能由职业经理人而非所有者管理。）

3.5 Scherer 的反馈回路。选择一个你熟悉的行业（如智能手机、网约车、快递），描述其市场结构的演变历程，并说明：

(a) 初始结构 → 行为 → 绩效的因果链

(b) 至少两个反馈回路（绩效或行为**反馈**到结构）的具体例子

进阶

3.6 SCP vs. 效率假说：设计一个检验。 SCP 范式认为" 集中度 → 高利润" 反映了市场势力。Demsetz (1973) 的效率假说认为这一相关性反映了高效企业同时获得高市场份额和高利润（效率 → 集中度和利润）。在阅读 第 ~4~ 章之前：

(a) 设计一个经验检验策略来区分这两种解释。你需要什么样的数据结构（行业层面、企业层面、面板还是横截面）？

(b) 你的策略面临什么实施困难？

(c) 这两种解释对反垄断政策的意义有何不同？

3.7 秩序自由主义 vs. SCP：两种结构主义的比较。阅读以下关于 EU 竞争法（Article 102 TFEU）与美国反垄断法（Sherman Act Section 2）在" 滥用市场支配地位" 上的简要描述：

> *EU* 竞争法不仅禁止剥削性滥用（如超额定价），还禁止排他性滥用（如独家交易、搭售、拒绝交易），且对支配地位企业施加了 " 特殊责任 "（*special responsibility*）以不扭曲竞争——即使某些行为在不具支配地位时是完全合法的。美国反垄断法主要关注排他性行为，且通常要求证明对消费者的具体伤害。

(a) EU 竞争法的" 特殊责任" 概念反映了哪种思想传统——哈佛 SCP 还是弗莱堡秩序自由主义？为什么？

(b) 在数字平台（如 Google、Amazon、TikTok）的反垄断中，哪种传统更有优势？论证你的立场。

3.10 延伸阅读

★ 必读（本章核心参考）

- **Chamberlin, E.H.** (1933). *The Theory of Monopolistic Competition*. Harvard University Press. ——SCP 范式的理论基础。Chamberlin 提供了" 不完全竞争" 这一概念空间——Mason 和 Bain 的工作就是将它填满真实世界的数据。重点阅读第 5 章（产品差异化与 dd/DD 曲线）。

- **Mason, E.S.** (1939). "Price and Production Policies of Large-Scale Enterprise." *American Economic Review*, 29(1), 61–74. ——SCP 的创立文献。仅 13 页，短小精悍。阅读时注意 Mason 的谨慎语气——"muddy but not uninteresting"——与后人" 强结构主义" 之间的对照。

- **Bain, J.S.** (1956). *Barriers to New Competition*. Harvard University Press. Ch.1–3. ——阅读 Bain 对三个进入壁垒的原始阐述。中译本：徐国兴译（2012），《新竞争者的壁垒》，人民出版社。Bain 的写作清晰而谨慎——留意他在每个结论后附加的保留和限定。

- **Scherer, F.M. & Ross, D.** (1990). *Industrial Market Structure and Economic Performance*. 3rd ed. Houghton Mifflin. Ch.1–2. ——SCP 的" 圣经"。重点阅读第一章的流程图（Fig. 1.1）和 Scherer 对框架的整体阐述。比较 1970 第一版、1980 第二版和 1990 第三版在结构上的变化——你能看到博弈论革命渗透进来的痕迹。

- **Panhans, M.T.** (2023). "The Rise, Fall, and Legacy of the Structure-Conduct-Performance Paradigm." *Journal of the History of Economic Thought*, 46(3), 337–357. ——对 SCP 范式最好的当代学术史评估。Panhans 的叙述平衡了"SCP 被芝加哥学派击败" 的常见叙事——他指出 SCP 的很多洞见以变换的形式存活在 NEIO 中。

★★ 进阶（加深理解）

- **Bain, J.S.** (1951). "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936–1940." *Quarterly Journal of Economics*, 65(3), 293–324. ——第一次系统检验。观察一个顶级经验经济学家在计量方法成熟之前如何进行" 手工" 数据分析。Bain 的 42 个行业的分组均值比较——今天看起来简单到几乎天真——但它是 IO 经验传统的起点。

- **Schmalensee, R.** (1989). "Inter-Industry Studies of Structure and Performance." In *Handbook of Industrial Organization*, Vol. 2, Ch.16. North-Holland.

——对 SCP 经验文献的权威综述。为 第 ~4~ 章的阅读做热身。Schmalensee 的核心判断：跨行业 SCP 回归" 能识别经验规律，但不能识别因果关系"。

- **Comanor, W.S. & Wilson, T.A.** (1967). "Advertising, Market Structure and Performance." *Review of Economics and Statistics*, 49(4), 423–440. ——广告作为进入壁垒的经典实证研究。发现高广告-销售比行业的利润率平均高出约 4 个百分点——这是 SCP 经验传统中最有力也最具争议的发现之一。

- **Kwoka, J.** (2014). "Structure, Conduct, Policy, and Performance: The 'Simple Model' of Frederic M. Scherer." *Economia e Politica Industriale*. ——追溯 Scherer SCP 图的演化史。Kwoka 展示了 Scherer 图示如何在不同的版本中被修改、简化和重新诠释。

★★★ 前沿（现代再评价与延续）

- **Bougette, P. & Marty, F.** (2024). "The SCP Paradigm Revisited: What Structuralism Really Contributed to U.S. Antitrust." *Working Paper*, hal-04577182. ——对反哈佛" 黑色传说" 的修正主义重评。Bougette 和 Marty 认为，哈佛结构主义的真实贡献被芝加哥学派的胜利叙事系统地低估了——许多被归功于芝加哥学派的改革实际上在结构主义时代已有萌芽。

- **Bernstein, D.** (2023). "The Empirical SCP Tradition: A Reassessment." ——在 NEIO 主导的今天，重新评估传统 SCP 经验研究的当代意义。Bernstein 的一个核心论点：在" 大数据" 时代，某些 SCP 的初衷——用丰富的跨行业变化来识别市场势力的结构性来源——可能以新的形式复兴。

- **Hovenkamp, H.** (2024). "The SCP Paradigm and the Rise of Neo-Brandeisian Antitrust." *Yale Law Journal Forum*. ——分析 2023 年 DOJ/FTC 合并指南对 HHI 阈值的收紧（$> 1{,}800$ 回归为高度集中；Δ HHI > 100 推定触发；取消安全港）是否标志着 SCP 结构主义的暗中回归。Hovenkamp 的答案：是，也不是——新布兰代斯学派重拾了结构主义的手段，但目标已从" 经济效率" 转向了更广泛的" 权力分散"。

下一章： → 第 ~4~ 章　SCP 的实证遗产与衰落——跨行业回归告诉我们什么，以及它为什么最终被放弃

第四章　SCP 的实证遗产与衰落

章首语

1974 年，Leonard Weiss 发表了一篇著名的综述。他审视了 54 篇检验集中度与利润率关系的经验研究，得出结论：46 篇发现了正相关。这个结果——超过 85% 的研究指向同一个方向——似乎是 SCP 范式经验成功的最高光时刻。如果科学进步是证据的累积，那么 SCP 纲领似乎正在稳步逼近真理。

但就在同一时期，Harold Demsetz 在 *Journal of Law and Economics* 上发表了一篇仅三页的论文。他没有增加任何新的数据，没有跑任何新的回归。他只是问了一个令整个领域不安的问题：你确定是集中度导致了高利润？如果高效率的企业恰好同时获得了高市场份额和高利润呢？

这个简单的问题撬动了 IO 历史上最深刻的一次方法论清算。在随后的二十年里，计量经济学家证明了跨行业 SCP 回归无法区分" 集中度 $\rightarrow$ 市场势力 $\rightarrow$ 利润" 与" 效率 $\rightarrow$ 集中度 + 利润" 这两种因果叙事。博弈论理论家证明，同样的市场结构可以支撑完全不同的竞争行为。会计学家证明，SCP 使用的利润率数据——会计回报率——根本不能反映经济盈利能力。

到 1990 年代，跨行业 SCP 回归在顶级期刊上几乎绝迹。统治 IO 长达 30 年的经验研究纲领就这样谢幕了。

但这不是一个简单的" 英雄打败反派" 的故事。SCP 虽然作为**活跃的研究纲领**衰落了，但它的**概念 DNA** 以出人意料的方式存续下来：HHI 至今仍是全球反垄断执法的核心工具；进入壁垒分析仍是任何合并审查的自然起点；" 市场界定" 的逻辑可以一路追溯到 Bain 对" 行业" 的执念。而 2023 年 DOJ/FTC 新合并指南中 HHI 阈值的大幅回调——某些评论者称之为" 结构性推定的回归"——在某种程度上是 Mason 和 Bain 的幽灵重新出现在政策舞台上。

本章讲述这个充满戏剧性和认知张力的兴衰故事。它的核心教训超越 IO 本身：一个经验研究纲领可以提出清晰、可检验的假说，产生丰富的数据证据，被政策采纳为行动指南——而**同时**在方法论上是不可靠的。认识到这一点，不仅是理解 IO 历史的关键，也是理解经验经济学本身的科学地位的关键。

4.1　核心问题

读完本章，你应该能够回答以下问题：

1. **跨行业 SCP 回归的经典设定是什么？**为什么 $\text{Profit}_i = \alpha + \beta \cdot \text{CR}_i + \gamma \cdot \text{Barriers}_i + \varepsilon_i$ 这个方程在 1960–70 年代被视为" 科学地检验市场势力" 的黄金标准？
2. **SCP 经验文献发现了什么？**集中度与利润率的关系有多强？有多稳健？存在" 临界集中度" 吗？
3. **Demsetz 效率假说如何挑战 SCP 的因果解释？**为什么" 集中度与利润正相关" 不必然意味着" 集中度导致了市场势力"？
4. **SCP 经验纲领在方法论上有哪些致命缺陷？**内生性、测量误差、市场界定问题、会计利润 vs. 经济利润——这些批判各自致命在哪里？
5. **SCP 的衰落是单一原因还是多重原因？**芝加哥理论批判、博弈论革命、计量可信性革命、反垄断实践变革——它们如何共同推动了一个范式的终结？
6. **SCP 留下了什么遗产？**哪些概念和工具被后续范式继承并继续塑造当代 IO 和政策？

4.2　SCP 经验纲领的黄金时代（1960s–1970s）

4.2.1　从 Bain 的手工分析到跨行业回归

第 ~3~ 章讲述了 Joe Bain 如何在 1950 年代用手工方式——问卷、访谈、工程估算——构建了 SCP 的经验基础。Bain (1951) 的方法论从今天的角度看极为朴素：42 个制造业行业，按 CR8 是否超过 70% 分为两组，比较两组的平均利润率。这本质上是一个**均值比较**（comparison of means），而非回归分析。

但它的核心逻辑——" 高集中度行业应该有更高利润"——被后续研究完整继承。转变发生在 1960 年代，驱动力量是两项技术变革：

1. **电子计算机的普及**：使研究者能够对数百个行业同时跑回归，而无需像 Bain 那样手工计算
2. **Census of Manufactures 数据的数字化**：提供了 4-digit SIC 层级上数百个制造业行业的标准化数据（集中度、利润率、投入产出结构）

经典 SCP 跨行业回归设定：

$$\pi_i = \alpha + \beta_1 \cdot \text{CR}_i + \beta_2 \cdot \text{Barriers}_i + \beta_3 \cdot \text{Controls}_i + \varepsilon_i \tag{4.1}$$

其中：

- π_i：行业 i 的利润率度量——通常是会计利润率（ROA、ROE）或价格-成本边际（PCM $= \frac{\text{Value Added}-\text{Payroll}}{\text{Value Added}+\text{Material Cost}}$）
- CR_i：行业 i 的卖方集中度——CR4、CR8 或（后期）HHI
- Barriers_i：进入壁垒的代理变量——广告-销售比（产品差异化）、MES/行业产出比（规模经济）、资本-产出比（资本需求）等
- Controls_i：控制变量——行业增长率、进口渗透率、购买者集中度、工会化程度等
- ε_i：误差项

SCP 回归的核心假设

式 (4.1) 看似简单，但它隐含着三个关键假设：

1. **结构外生**：集中度和进入壁垒被视为**由技术条件外生决定**——它们不是企业策略选择的结果
2. **行为同质**：给定相同的结构，所有行业中的企业行为方式相同——即”行为（C）”不需要单独建模，它由结构机械决定
3. **利润反映市场势力**：利润率差异主要反映了市场势力的差异，而非效率差异或会计方法的差异

这三个假设在 1970–80 年代将逐一被击破。但它们在 1960 年代构成了一个内部一致的逻辑闭环，使 SCP 回归成为当时”科学地研究市场势力”最先进的工具。

4.2.2 关键经验发现

集中度-利润关系的存在性与量级

SCP 经验纲领在 1960–70 年代积累了数十篇跨行业回归研究。表 4.1 汇总了其中最关键的几篇。

Weiss (1974) 综述：经验证据的巅峰时刻。 Weiss 的综述在 IO 历史上具有里程碑意义。他对 54 篇经验研究的系统梳理可以说是 SCP 经验纲领的”官方状态报告”。关键数字：

- 54 篇研究中，**46 篇**（$\approx 85\%$）发现了集中度与利润率的正相关
- 其中 **36 篇**在统计上显著（给定当时的显著性标准）
- 典型弹性：CR4 上升 10 个百分点 $\to$ 利润率上升 **1–2 个百分点**

表 4.1: 关键 SCP 经验研究汇总

研究	样本	集中度	利润率	核心发现
Bain (1951)	42 行业 (1936–40)	CR8	权益回报率	CR8 $>$ 70% 的行业利润显著更高（均值 11.6% vs. 7.2%）
Collins & Preston (1969)	∼100 行业 (1963)	CR4	价格-成本边际	正相关，但 CR4 仅解释利润变异的 5–10%
Weiss (1974) 综述	54 篇研究	各种	各种	46/54 发现正相关；典型弹性 $\approx$ 0.1–0.2
Domowitz, Hubbard & Petersen (1986)	284 行业 (1958–81)	CR4	PCM	正相关存在，但在 1970 年代后显著减弱
Schmalensee (1985)	456 企业, 242 行业 (1975)	—	ROA	行业效应 $\approx$ 20%，市场份额效应仅 $\approx$ 0.6%
Salinger (1990)	再分析	CR4	PCM	控制资本存量后，关系大幅弱化

- 集中度能解释的利润跨行业变异比例（R^2）：通常 **低于 10%**

" 显著" 但不" 大"：一个被忽视的区分

Weiss (1974) 的 46/54 这个数字经常被引为 SCP 经验成功的证据。但注意两个关键限定：

第一，**统计显著 $\neq$ 经济显著**。集中度上升 10 个百分点只伴随 1–2 个百分点的利润上升——这个效应在统计上可以检测到（因为样本量足够大），但在经济上是否" 大" 则取决于判断。

第二，$R^2 < 10\%$ **意味着什么？** 超过 90% 的行业间利润变异是由集中度**之外**的因素驱动的。SCP 发现了一个真实但微弱的经验规律——它远非" 决定性的"。这在方法论上是一个重要的教训：不要把" 存在" 与" 重要" 混为一谈。

" 临界集中度" 假说

Bain (1951) 的原始发现包含一个微妙但重要的非线性：CR8 $>$ 70% 时利润才" 跳升"。这意味着集中度-利润关系**不是线性的**——存在一个阈值，超过它之后合谋变得容易，利润显著上升。

后续研究（White 1976; Bradburd & Over 1982）使用更复杂的方法（spline regression、threshold regression）大体确认了这一模式，但争议持续存在：

- **统计问题**：临界值的选择是数据驱动的（data-driven）——Bain 选择 70% 是因为他的数据在这个点出现了跳跃，而非基于理论预测
- **非线性转换的敏感性**：任何变量（即使与利润无关）对其自身进行平方或分段转换后，也可能" 发现" 一个临界值——这是多重比较（multiple comparisons）问题的一种形式
- **理论解释的模糊性**：临界集中度反映的可能是合谋的" 启动阈值"（coordination threshold），但也可能反映了不同 SIC 分类对" 市场" 定义的差异——高集中度行业恰好是有清晰市场边界、SIC 分类恰好映射到经济市场的行业

政策含义：如果临界集中度假说正确，那么反垄断只需要关注那些超过阈值的行业——大多数行业是" 安全的"。这使得 SCP 的政策处方更加精准，但也意味着跨行业回归的平均效应（β_1）可能低估了高集中度行业中的真实效应。

进入壁垒的作用

Comanor & Wilson (1967) 是 SCP 传统中最重要的壁垒研究之一。他们发现：

- **广告强度**（广告支出/销售收入）与利润率正相关——即使在控制集中度之后

- 高广告-销售比行业的利润率平均高出约 4 个百分点
- 解释：广告创造了**产品差异化的进入壁垒**——新进入者不仅需要建工厂，还需要投入大量广告来克服在位者的品牌忠诚度

这意味着**集中度 × 壁垒**的交互作用可能是理解市场势力的关键——单独的集中度不足以产生持续超额利润，集中度**加上**壁垒才是完整的图景。Bain (1956) 的原始直觉——三个壁垒类型的分类——在跨行业回归中获得了经验支持。

但这里同样存在一个" 鸡生蛋还是蛋生鸡" 的问题：高广告行业的高利润是因为广告壁垒保护了市场势力，还是因为高利润企业**有更多资源**投入广告？这个问题在 SCP 框架内部是无法回答的。

4.2.3　PIMS 数据库：SCP 的咨询业变体

SCP 经验纲领在 1970 年代产生了一个出人意料的衍生产品——PIMS（Profit Impact of Market Strategy）数据库。PIMS 源自 GE 的内部管理数据库，1970 年代由哈佛商学院和 Strategic Planning Institute (SPI) 推广。它包含了数百个" 业务单元"（strategic business units）的详细数据——不是 Census 意义上的" 行业"，而是企业内部的分部。

PIMS 的核心发现——" 市场份额是盈利能力的最强预测因子"——被管理咨询业（BCG 矩阵、McKinsey 的增长-份额矩阵等）广泛引用。但这个发现受到了严厉的方法论批评：

- **样本选择偏差**：只有" 成功的" 公司才会留在 PIMS 数据库中（参与是自愿的，失败的企业没有动机报告数据）——这种选择偏差会人为地强化" 份额 → 利润" 的相关性
- **反向因果**：高利润企业有更多资源投资于增长 → 获得更高的市场份额——与效率假说一样的问题
- **会计数据的固有问题**：PIMS 使用与 SCP 相同的会计利润数据，面临相同的度量问题（详见后文 §4.4 的讨论）

SCP 的双重生命

SCP 有一个有趣的现象：它同时存在于**两个世界**中。在经济学界，它到 1990 年代基本被放弃。但在管理学、战略咨询和商学院课程中，它的核心直觉——" 行业结构决定企业绩效"——通过 Porter (1980) 的五力模型（Five Forces）获得了第二生命，至今仍在被广泛教授。

这两个世界的"SCP" 共享了一个祖先（Mason 和 Bain），但走上了完全不同的演化路径。经济学中的 SCP 因为缺乏因果识别而被淘汰；管理学中的 SCP 因为提供了" 可操作的分析框架" 而繁荣。这个分歧本身值得深思。

4.2.4 1968 年 DOJ 合并指南：SCP 的政策巅峰

第 ~3~ 章已描述了 SCP 如何影响反垄断判例（Brown Shoe 1962、Philadelphia National Bank 1963、Von's Grocery 1966）。1968 年 DOJ 合并指南将这些判例中的结构主义逻辑编纂为**量化的执法规则**：

- **高度集中市场**：CR4 $\geq$ 75% 且合并方之一 $\geq$ 15% $\rightarrow$ 推定违法
- **低门槛**：合并方之一仅需 4% 市场份额即可触发审查
- **横向合并**的强烈推定违法——几乎任何增加集中度的合并都可能被挑战

SCP 的政策逻辑链

如果跨行业回归显示" **集中度** $\uparrow\rightarrow$ **利润** $\uparrow\rightarrow$ **消费者受损**" 成立，那么政策处方是直接的：**事前阻止集中度上升**。不需要逐一证明具体行为（如合谋或掠夺性定价）——结构本身就是充分证据。

这个逻辑在实践中极其强大，因为它为反垄断执法提供了一个" 简易的操作标准"：

$$\text{高集中度} \Rightarrow \text{推定存在竞争问题} \Rightarrow \text{干预}$$

但正如本章后续各节将展示的，这个逻辑链的每一个环节——从集中度的度量到因果方向的推断——都将在 1970–80 年代受到系统的拆解。

4.3 Demsetz 效率假说：芝加哥学派的第一枪

4.3.1 Demsetz (1973) 的核心论证

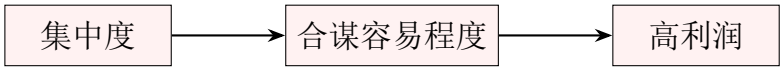
Demsetz, H. (1973). "Industry Structure, Market Rivalry, and Public Policy." ***Journal of Law and Economics*****, 16(1), 1–9.**

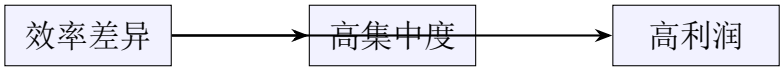
这三页论文可能是 IO 历史上影响力最大的一篇三页文章。Demsetz 没有增加新数据，没有跑新回归。他只是改变了因果叙事的**方向**——而方向决定了一切。

Demsetz 的核心论证可以用以下逻辑链来表述：

1. **效率异质性**：同一行业内的企业效率不同——有些企业生产成本更低、产品质量更好、或管理更高效
2. **效率 $\rightarrow$ 市场份额**：高效率企业增长更快，获得更大的市场份额 $\rightarrow$ **行业集中度上升**
3. **效率 $\rightarrow$ 利润**：高效率企业同时也获得更高利润（成本优势或质量溢价）

SCP 假说：集中度 → 合谋 → 利润





Demsetz 效率假说：效率 → 集中度 + 利润

图 4.1: 两种因果叙事：SCP 假说 vs. Demsetz 效率假说。两个故事产生**完全相同的可观测相关性**（集中度 ↑ 与利润 ↑ 正相关），但因果方向和福利含义完全不同。

4. **行业层面的相关性是成分效应（composition effect）**：CR ↑ 与利润 ↑ 不是因果关系——它们是同一个底层原因（效率差异）的**两个后果**

Demsetz 的核心洞察

”” 有效竞争不需要所有企业规模相同……集中度与盈利能力之间的经验关联可能反映的是大企业优越的效率，而非市场势力的行使。””
——Demsetz (1973, p. 3)

4.3.2　一个数值例子

为了让 Demsetz 的逻辑更具体，考虑以下两个行业：

两个行业的 CR4 都是 100%，但行业 B 的” 平均利润” 显著更高。SCP 的跨行业回归会将此归因为” 高集中度导致了高利润”。但实际原因是行业 B 有更显著的**效率异质性**——一家特别高效的企业同时获得了高份额和高利润。行业 B 中 10% 份额的小企业的利润（5%）与行业 A 中的企业（5%）完全相同——这暗示” 高行业利润” 并未” 均摊” 到所有企业。

表 4.2: 数值例子：效率假说 vs. 市场势力假说

	行业 A（低效率同质）		**行业 B（效率异质）**	
	份额	利润率	份额	利润率
企业 1	25%	5%	70%	20%
企业 2	25%	5%	10%	5%
企业 3	25%	5%	10%	5%
企业 4	25%	5%	10%	5%
CR4	**100%**		**100%**	
行业平均利润		**5%**		**15.5%***

* $(0.7 \times 20\%) + (0.1 \times 5\%) + (0.1 \times 5\%) + (0.1 \times 5\%) = 14\% + 1.5\% = 15.5\%$ ——远超行业 A 的 5%，但这完全是因为高效率企业拉高了均值。

4.3.3 Demsetz 的经验检验策略

Demsetz 提出了一个巧妙的方法来区分效率假说和市场势力假说：

市场势力假说预测 ：在集中行业中，**所有企业**都应该受益于高价格（" 伞形定价" umbrella pricing）→ 小企业在集中行业中的利润也应该比低集中行业中的企业更高

效率假说预测 ：只有**大企业**（高效率）获得高利润 → 集中行业中的小企业利润与低集中行业中的企业利润**无显著差异**

Demsetz 使用 IRS 1963 年的数据，按资产规模将企业分为五组，比较不同集中度行业中各组别的利润率。他发现：

- 在集中行业中，**大企业**的利润确实更高
- 但集中行业中的**小企业**利润与低集中行业中的小企业利润**没有显著差异**
- 这支持了效率假说：高行业利润是由高效率大企业拉动的，而非所有企业共享的" 伞形" 定价

后续验证与批评：

- **Smirlock (1985)** 使用 FTC LB 数据重新检验了 Demsetz 假说，发现支持证据——市场份额（而非集中度）是利润的主要驱动因子
- **Martin (1983)** 指出 Demsetz 的检验存在统计效力问题——如果集中行业中的小企业数量很少，它们的" 平均利润" 估计就非常不精确，难以检测到伞形效应的存在

- **Clarke, Davies & Waterson (1984)** 提供了一个理论整合——在 Cournot 模型的框架下，集中度、市场份额和利润之间存在一个**结构关系**，而不仅是二选一的竞争假说

4.3.4　Peltzman (1977) 的价格-成本分解

Peltzman, S. (1977). "The Gains and Losses from Industrial Concentration." *Journal of Law and Economics*, 20(2), 229–263.

Peltzman 提出了一个更精细的检验框架：将集中度变化对价格-成本边际的效应分解为**价格效应**和**成本效应**。

令 $M = \frac{P-AC}{P}$ 为价格-成本边际（PCM）。集中度 ΔC 的变化对 PCM 的效应为：

$$\frac{dM}{dC} = \underbrace{\frac{\partial M}{\partial P} \cdot \frac{dP}{dC}}_{\text{价格效应}} + \underbrace{\frac{\partial M}{\partial AC} \cdot \frac{dAC}{dC}}_{\text{成本效应}} \tag{4.2}$$

两种假说对这两个效应给出不同的预测：

表 4.3: Peltzman 分解：两种假说的预测对比

假说	价格效应 dP/dC	成本效应 dAC/dC
市场势力假说（SCP）	$+$（集中 $\uparrow \to$ 合谋 $\uparrow \to$ 价格 $\uparrow$）	≈ 0（成本不受集中度影响）
效率假说（Demsetz）	≈ 0（竞争仍然有效）	$-$（集中 $\uparrow \to$ 效率 $\uparrow \to$ 成本 $\downarrow$）

Peltzman 的经验发现：历史上的集中度上升**主要伴随成本下降而非价格上升**。换言之，$\frac{dM}{dC} > 0$ 的主要驱动力来自 AC/P 项的缩小（平均成本相对价格下降），而非 P 的上升——这与效率假说一致。

同一现象，两种福利含义

Peltzman 的分解框架是经济学思维的典范。同一个可观测结果——PCM 上升——可以对应两种完全不同的福利解释：

- 如果 PCM 上升来自**价格上升**（市场势力）：消费者受损，社会福利下降
- 如果 PCM 上升来自**成本下降**（效率改善）：消费者可能不受影响或受益（如果成本节省部分传递给价格），社会福利上升

这个区分不仅仅是一个方法论细节——它决定了反垄断政策应该怎么做。阻止一个"由效率驱动的集中度上升"会损害社会福利；阻止一个"由合谋驱动的集中度上升"会增进社会福利。SCP 回归只看 PCM 与 CR 的相关性符号——它无法做这个区分。

4.3.5 其他芝加哥学派的早期批评

- **Brozen (1970, 1971)**：Bain (1951) 发现的高利润行业在后续时期利润**回归均值**（regression to the mean）——超额利润不是" 持续的"（persistent），而是" 暂时的"（transitory）。如果市场势力真如 SCP 所说那样持久，利润应持续偏高；如果利润是暂时的，更可能是短暂的效率租（efficiency rents）或需求冲击。

- **Mancke (1974) " 运气之块"（lump-of-luck）**：即使所有企业完全相同，由于随机因素——一次成功的 R&D 突破、一个偶然的需求波动——某些企业会在特定时期同时获得高市场份额和高利润。SCP 回归会错误地将这种**随机现象**解释为系统性关系。Mancke 的论证本质上是一个" 虚假回归"（spurious regression）的警告——在纯粹的随机游走过程中，两个独立的时间序列也可能表现出高度相关性。

- **Stigler (1968) 对进入壁垒的重新定义**：Bain 将规模经济列为进入壁垒。Stigler 反驳：如果进入者可以建设与在位者相同规模的工厂，那么规模经济对在位者和进入者的影响是对称的——它不是" 壁垒"。这种逻辑大大缩小了" 进入壁垒"的概念范围，进而缩小了反垄断的干预空间。这个问题将在 第 ~5~ 章详细展开。

SCP 的" 阿喀琉斯之踵"

Demsetz 在 SCP 的逻辑链条上找到了一个致命脆弱点：**集中度是内生的**。在 Bain 的框架中，集中度被视为由技术条件（规模经济、MES）外生给定——这是从结构推断绩效的逻辑前提。Demsetz 指出，集中度本身可能是企业效率差异的**结果**——高效率企业通过竞争**赢得**了高市场份额。如果集中度是内生的，那么 SCP 回归的 OLS 系数就不具有因果解释——它只是一个相关性，可能反映的是效率差异而非市场势力。

这个问题——内生性——不仅是 SCP 的阿喀琉斯之踵。它也是整个经验经济学的核心方法论挑战。SCP 的经验失败是计量经济学" 可信性革命"（credibility revolution）的早期和富有教育意义的案例之一。

4.4 测量问题：会计利润、市场界定与变量误差

4.4.1 Fisher-McGowan (1983)：会计利润的致命缺陷

Fisher, F.M. & McGowan, J.J. (1983). "On the Misuse of Accounting Rates of Return to Infer Monopoly Profits." *American Economic Review*, 73(1), 82–97.

这是对 SCP 经验基础最彻底、最令人不安的攻击。Fisher 和 McGowan 不是在做通常的" 测量误差" 批评——他们在论证一个**不可能定理**（impossibility theorem）：**没有任何方法可以从会计回报率推断相对经济盈利能力。**

他们的论证基于以下观察：

1. **折旧方法的任意性：**考虑两个企业，经济上完全相同——相同的投资、相同的现金流、相同的经济寿命。但一家使用直线折旧，另一家使用加速折旧（如双倍余额递减法）。在资产寿命的早期，加速折旧企业的账面资产更低（更多折旧已计提），账面利润也更低。两家企业的经济盈利能力完全相同，但会计 ROA 可能相差数个百分点。

   令 I 为投资额，n 为资产经济寿命，r 为经济回报率。经济回报率是 r——一个数字。会计回报率取决于折旧方法的选择——一个本质上**任意的会计惯例**。

2. **无形资产的费用化处理：**在标准会计准则下，R&D 和广告支出在当期被**费用化**（expensed）——即直接从当期利润中扣除。但它们的经济效应是**跨期**的——今天的 R&D 创造的是未来的收入流。

   这意味着：R&D 密集型企业的**资产被系统性低估**（自创的无形资产不在资产负债表上），它们的**会计 ROA 被系统性高估**（分母缩小而分子包含了无形资产的回报）。

   回忆 Comanor & Wilson (1967) 的发现：广告强度与利润率正相关。Fisher-McGowan 指出这个" 发现" 可能**完全**是会计处理的产物——广告支出高的企业，费用化处理导致其资产被低估，ROA 被高估，与市场势力毫无关系。

3. **通胀效应：**历史成本会计下，折旧费用基于资产的**原始购买价格**而非重置成本。在通胀时期，历史成本低于重置成本 $\rightarrow$ 折旧费用被低估 $\rightarrow$ 利润被高估。而且这个偏差在资本密集型行业中更为严重——恰好也是 Bain 分类中" 高壁垒" 的行业。

4. **并购会计的扭曲：**并购中的商誉（goodwill）处理和购买法（purchase accounting）会彻底改变账面资产和利润，而完全不影响经济盈利能力。

Fisher-McGowan 的核心结论

"" 没有任何方法可以查看会计回报率并从中推断有关相对经济盈利能力的任何信息……那些据称基于会计数据将集中度与经济盈利能力联系起来的文献实际上没有做到任何此类事情。""

——Fisher & McGowan (1983, p. 82)

如果 Fisher 和 McGowan 是正确的，那么整个 SCP 经验文献的基础——数百篇论文，数千个回归——建立在一个不可靠的度量基础之上。这不是一个" 测量误差" 问题（可以通过更好的数据或计量方法缓解）——这是一个**概念上不可能**的问题。

但也要注意反对方的论证。Martin (1983) 等学者认为 Fisher-McGowan 夸大了问题：

- 如果会计扭曲在行业间足够相似——如大多数制造业使用相似的折旧方法——跨行业回归仍然可以检测到经济利润的**差异**（即使不能检测其**绝对水平**）
- 在差分中，行业间的会计惯例差异如果与集中度不相关，它只是一个" 噪声"——它降低估计精度但不导致系统性偏差
- 但 Martin 也承认：如果会计扭曲**系统性地**与集中度相关（如资本密集度 $\uparrow \rightarrow$ 通胀扭曲 $\uparrow$，且资本密集度 $\uparrow \rightarrow$ 进入壁垒 $\uparrow$），那么偏差的方向和量级都无法确定

教学意义：Fisher-McGowan 的论文是" 度量问题可以致命" 的经典例证。在经验经济学中，我们经常花大量时间讨论识别策略——工具变量、DID、RD——但对**度量**的关注往往不够。Fisher-McGowan 提醒我们：如果度量无效（measurement validity）不成立，再精巧的识别策略也无济于事。

4.4.2 市场界定问题

SCP 经验研究使用的分析单位是 Census 的 SIC（Standard Industrial Classification）分类。但 **SIC 分类 $\neq$ 经济市场**：

- **过宽**：SIC 可能将地理上分隔的市场聚合为一个" 全国行业"。水泥行业是最经典的例子——运输成本高，有效市场半径通常不超过 200–300 公里。把美国 50 个地方水泥市场聚合为一个" 行业"，CR4 可能只有 10–20%——但每个地方市场的 CR4 可能是 80–100%。SCP 回归使用的是错误的集中度度量。
- **过窄**：SIC 可能忽略来自相邻行业的竞争。铝和钢材在某些用途（如汽车车身）中竞争，但它们属于不同的 SIC。如果铝价上升，钢材的需求会上升——但这在 SIC 分类中不可见。

- **进口竞争**：国内集中度（CR4 基于国内生产）忽略进口——而进口渗透率可能比国内集中度对企业定价能力的约束更大。一个 CR4 = 80% 的行业如果面临 30% 的进口渗透率，有效集中度可能远低于 80%。

- **买方势力**：SCP 传统关注卖方集中度，但**买方集中度**（buyer concentration）同样影响企业定价能力。一个卖方集中的行业如果面对同样集中的买方（如汽车零部件供应商面对少数汽车制造商），其" 市场势力" 可能被买方势力抵消——利润率可能低于卖方集中度预测的水平。

4.4.3　变量误差：集中度作为" 市场势力" 的代理变量

SCP 的逻辑是：集中度 → 合谋容易程度 → 市场势力 → 高利润。但集中度只是一个**代理变量**（proxy variable）——它不直接度量市场势力：

- 一个 CR4 = 100% 的两企业行业，如果企业激烈竞争（Bertrand 行为），P = MC

- 一个 CR4 = 60% 的四企业行业，如果企业暗中协调，P ≫ MC

集中度度量的是结构，不是行为——而 SCP 跳过了行为（C），直接从结构（S）推断绩效（P）。这个简化在政策上很方便，但在分析上是致命的——它假设了" 结构 → 行为" 的映射是**确定性**的，但现实中这个映射依赖于模型中不可见的因素（企业目标函数、信息结构、博弈的时间维度等）。

Salinger (1990) 进一步指出：如果集中度存在**经典测量误差**（classical measurement error）——即 SIC 集中度 = 真实经济市场集中度 + 噪声——那么 OLS 系数存在衰减偏差（attenuation bias），即倾向于低估真实关系。但 Salinger 论证，测量误差可能**不是经典的**——它与真实集中度的关系是系统性的。例如，SIC 分类对高集中度行业的" 真实市场" 可能捕捉得更准确——因为高集中度行业恰好有清晰的市场边界。这种情况下，偏差方向不确定。

4.5　内生性问题：因果方向的根本不确定性

4.5.1　为什么相关性 ≠ 因果关系在这里特别严重

在大多数经验经济学中，" 相关不等于因果" 是一个标准警告。但在 SCP 回归的语境中，这个问题尤其严重——因为**可替代的因果叙事不仅在理论上是可能的，而且有系统的经验证据支持**。

4.5.2 反向因果（Reverse Causality）

SCP 的因果方向是：集中度 ↑ → 利润 ↑。

但反向同样合理：

- 高利润行业吸引进入 → 进入可能采取**并购**形式 → 并购提高集中度
- 高效率企业获得高利润 → 用利润投资扩大产能 → 通过内部增长提高集中度
- 高利润企业收购退出企业 → 集中度上升（即使行业总利润未变）

在单一横截面回归中，这两种因果方向**完全无法区分**。数据集只有一个时间点（或一个时期均值），你只能看到一个正相关。它告诉你" 这些变量一起变动"——但**不能**告诉你" 谁在驱动谁"。

4.5.3 遗漏变量（Omitted Variables）

SCP 回归的误差项 ε_i 包含大量不可观测的行业特征，其中许多与集中度相关：

表 4.4: SCP 回归中的关键遗漏变量

遗漏变量	定义	与 CR 的关系	对利润的效应方向
技术机会	行业中的 R&D 和创新潜力	可能 +（高机会 → 创新 → 赢者通吃）或 −（高机会 → 频繁进入）	+（创新租）
需求增长	行业需求的趋势性增长	−（增长吸引进入，降低集中度）或 +（增长使领先企业扩大份额）	+（需求拉力）
工会力量	工会化程度和集体谈判能力	?	−（工会提取一部分租）
进口竞争	进口渗透率	对国内 CR 的测量无影响，但降低有效集中度	−（压低国内价格）
规制环境	价格规制、进入许可等	+（规制常保护在位者）	取决于规制类型

如果这些遗漏变量与集中度相关（它们确实相关），那么 OLS 估计的 $\hat{\beta}_1$ 就是**有偏且不一致**的。偏差方向取决于遗漏变量与集中度和利润的协方差方向——在 SCP

的案例中，这个方向**在理论上模糊**（有些遗漏变量产生正偏差，有些产生负偏差），净偏差方向未知。

4.5.4　联立性（Simultaneity）

利润和集中度可能不仅相关——它们可能是**联合决定**的（jointly determined）。Martin (1983) 提出了一个简单的联立方程系统：

$$\pi_i = \alpha_1 + \beta_1 \cdot \mathrm{CR}_i + \gamma_1 \cdot Z_{1i} + \varepsilon_{1i} \tag{4.3}$$

$$\mathrm{CR}_i = \alpha_2 + \beta_2 \cdot \pi_i + \gamma_2 \cdot Z_{2i} + \varepsilon_{2i} \tag{4.4}$$

如果 $\beta_2 \neq 0$（利润影响集中度——如 Brozen 和 Demsetz 论证的），那么方程 (4.3) 中 CR_i 与 ε_{1i} 相关（因为 CR_i 同时出现在方程 (4.4) 的左边和方程 (4.3) 的右边，而 ε_{1i} 通过 π_i 间接影响 CR_i）。这导致 OLS 的 $\hat{\beta}_1$ 不一致。

工具变量解法及其困难：解决联立性偏误的经典方法是用 2SLS——找一个与集中度相关但不直接影响利润的变量作为工具变量。但这里存在一个根本难题：

- **什么变量影响集中度但不影响利润？**几乎任何影响集中度的变量——技术特征、需求条件、制度环境——都可能通过某些渠道同时影响利润
- **排他性约束（exclusion restriction）无法检验：**我们无法用数据检验工具变量是否” 只通过集中度影响利润”——这是先验的经济学理论判断
- Salinger (1990) 尝试使用 2SLS，但承认” 不存在完全令人满意的工具变量”——这是困扰 SCP 直到其终结的问题

4.5.5　Schmalensee (1985, 1989)：权威的” 验尸报告”

Schmalensee, R. (1985). “Do Markets Differ Much?” *American Economic Review*, 75(3), 341–352.

Schmalensee 使用 FTC Line of Business 数据（1975 年，456 个企业，跨 242 个 4-digit 制造业行业），进行了 IO 历史上影响最深远的方差分解。他的模型是：

$$\pi_{ij} = \mu + \alpha_i + \beta_j + \gamma \cdot s_{ij} + \varepsilon_{ij} \tag{4.5}$$

其中：

- π_{ij}：企业 j 在行业 i 中的利润率
- μ：总体均值
- α_i：**行业效应**（industry effect）——特定于行业 i 的不可观测因素

- β_j：**企业效应**（firm effect）——特定于企业的不可观测因素
- s_{ij}：企业 j 在行业 i 中的**市场份额**
- γ：市场份额对利润的效应

方差分解的核心发现

Schmalensee (1985) 的核心发现——这可能是 SCP 经验传统中最具破坏性的单一结果：

- **行业效应**（α_i）：解释利润变异的 **≈20%**——行业确实不同，但**不是决定性的**
- **市场份额效应**（s_{ij}）：解释 **≈0.6%**——**几乎可以忽略！**虽然统计显著
- **企业效应**（β_j）：**可以忽略**——永久性的企业”效率”差异在跨行业中不显著
- **残差：≈80%——我们不了解的比了解的更多**

集中度本身只是”行业效应”的一部分——而行业效应总共只解释 20%。如果集中度只占这 20% 的一部分，那么它直接解释的利润变异比例可能低至 1–3%。

Schmalensee (1989) 综述：在 Handbook of Industrial Organization 的综述中，Schmalensee 做出了可能是 SCP 经验传统的**官方讣告：**

> ”” 跨行业截面中卖方集中度与会计盈利能力之间的相关性，虽然常常在统计上显著，但其**如此微弱**，以至于对结构主义反垄断政策几乎无法提供任何支持。””
>
> ——Schmalensee (1989, Ch.16, p. 976)

同时也应注意到 Schmalensee 的谨慎：

> ”” 这并不意味着市场结构不重要。它只意味着我们使用的工具——跨行业简约式回归——太粗糙了，无法捕捉到市场结构的效应。””

这个区分很重要：Schmalensee 不是在说”集中度对绩效没有影响”——他说的是”跨行业回归**作为检测这种效应的工具**失败了”。

4.5.6 集中度-利润关系的时变：DHP (1986–87)

Domowitz, Hubbard & Petersen (1986, 1987) 发现了另一个令人不安的模式：集中度-利润率关系**正在随时间减弱**。

- 1958–1970 年：集中度与 PCM 之间存在清晰的正相关
- 1970–1981 年：这种关系显著弱化，在某些设定中甚至消失

可能的解释：

1. **进口竞争加剧**：1970 年代国际贸易扩张——国内集中度对定价的约束力被进口削弱
2. **反垄断执法效果**：1960–70 年代的积极反垄断可能已经降低了最严重市场势力的发生率
3. **经济周期效应**：1970 年代的石油冲击和滞胀可能改变了利润和集中度之间的关系
4. **更根本的解释**：关系一直都不强——早期的正相关可能是**出版偏差**（publication bias）的结果（只有正且有显著的研究被发表）

DHP 的结果——在 SCP 经验纲领的全盛时期，其核心经验规律正在实时消失——进一步削弱了 SCP 作为政策指南的可信度。

SCP 的”移动靶”问题

DHP 的结果指出了一个更广泛的问题：即使我们相信 SCP 回归在 1950–60 年代发现了真实的集中度-利润关系，这种关系本身可能不是**跨时间稳定**的（time-invariant）。随着全球化、技术进步和制度变革，同样的集中度水平在不同时期可能具有完全不同的竞争含义。

政策含义：基于 1950 年代数据制定的合并指南，可能在 1980 年代已经过时——即使指南本身在制定时是”科学地”基于最佳可用证据的。

4.6　SCP 衰落的多重原因

SCP 的衰落不是一夜之间发生的，也不是单一原因造成的。它是多条”战线”同时溃败的结果。理解这个多维度性很重要——它告诉我们：一个研究纲领的终结通常需要**多个条件同时满足**。

4.6.1　五线并进的解释框架

4.6.2　各战线的相互作用

这五条战线不是独立的——它们相互强化：

表 4.5: SCP 衰落的五条战线

战线	核心批判	关键人物/文献	后果	→ 后续章节
经验	内生性、测量误差、弱相关性、方差分解	Schmalensee (1985, 1989), Fisher-McGowan (1983), DHP (1986–87)	跨行业回归丧失学术合法性	—
理论（右翼）	效率 → 集中度 + 利润（因果倒置）；进入壁垒定义过宽	Demsetz (1973), Peltzman (1977), Stigler (1968), Brozen (1970–71)	SCP 的核心因果链被颠倒	Ch.5
方法论（博弈论）	同样的结构可产生完全不同的行为——跳过了” 行为” 黑箱	Tirole (1988), Shapiro (1989)	新-IO 用博弈论提供行为的微观基础	Ch.7
计量（可信性革命）	相关性 $\neq$ 因果；缺乏外生变异来源	Angrist & Pischke (2010), 自然实验转向	跨行业 OLS 成为” 糟糕经验研究” 的代名词	Ch.10
政策	从结构推定转向效果分析；消费者福利标准取代结构主义	Bork (1978), Posner, Easterbrook, 1982/2010 合并指南	合并指南大幅放宽；结构性救济式微	Ch.5, Ch.15

- **芝加哥理论批判**为**政策变革**提供了智识合法性（”我们不是在放纵垄断者——我们是在防止过度干预损害效率”）

- **博弈论革命**提供了**替代的理论框架**——它解释了为什么同样的结构可以产生不同的绩效，而芝加哥学派解释了为什么高效率企业在同样的结构下胜出

- **计量可信性革命**为**经验批判**提供了更精确的工具——它不只是说”SCP 回归有问题”，而是系统地展示了问题的深度

- **政策变革**创造了**对新研究的需求**——如果合并指南不再使用结构推定，反垄断机构就需要新的、不依赖结构推定的分析框架（如 SSNIP 测试、合并模拟、UPP）

关键时间节点：

- **1974**：Weiss 综述（SCP 经验巅峰）

- **1973–1977**：Demsetz (1973), Peltzman (1977)（芝加哥理论反攻）

- **1982**：DOJ 合并指南重大修订——首次引入 HHI 代替 CR4，引入 SSNIP 和”安全港”

- **1983**：Fisher-McGowan（度量基础攻击）

- **1985–1989**：Schmalensee (1985) 方差分解 + Schmalensee (1989) Handbook 综述 + Bresnahan (1989) Handbook——**新旧纲领的”交接仪式”**

- **1990s**：跨行业 SCP 回归在顶刊几近消失；NEIO 结构估计和新一代准实验研究兴起

为什么 SCP 的衰落是”好的科学失败”

SCP 的衰落不是一个”坏理论被推翻”的故事——它是科学进步的标准模式：

1. SCP 提出了一个**清晰、可检验的假说**：集中度与利润相关，且这种相关反映了市场势力
2. 它用**当时最好的可用方法**（跨行业 OLS、Census 数据）系统检验了这个假说
3. 最初的证据似乎支持这个假说（46/54 正相关）
4. 批评者提出了**替代假说**（Demsetz 效率假说）和**方法论批判**（Fisher-McGowan、内生性）
5. **更精巧的经验分析**（方差分解、时变分析）表明原假说获得的经验支持比最初认为的弱得多
6. 这个领域**承认了失败**并转向了更强大的工具——博弈论 + 结构估计 + 准实验方法

这与 Lakatos (1970) 对”退化的研究纲领”的描述高度一致：当辅助假说（如”我们需要更好的集中度度量”、”我们需要控制更多变量”）的增殖不再产生新的经验发现时，纲领进入退化阶段。但 Lakatos 也强调——退化的研究纲领可以持续很长时间，而 SCP 在政策界的持续影响力就是例证。

4.7 SCP 的遗产：哪些东西被保留？

SCP 作为**活跃的研究纲领**衰落了。但它的**概念 DNA**——它所创造的术语、分类法、分析直觉和制度基础设施——在更大的 IO 生态系统中存续下来，并在许多情况下成为了”理所当然”的背景知识。

4.7.1 遗产 1：HHI——从学术公式到全球监管标准

HHI 的故事（已在 第 ~3~ 章 §~3.5 中详细讲述）是 SCP 最持久和最具影响力的遗产。它的轨迹完美地说明了 SCP 的悖论性命运：

- **1945/1950**：Hirschman（贸易分析）和 Herfindahl（美国钢铁业集中度的博士论文）独立发明 HHI——纯粹的学术兴趣
- **1982**：DOJ 合并指南首次引入 HHI 阈值——取代 CR4，成为反垄断执法的核心标准

- **2010**：HHI 阈值大幅放宽（安全港从 1,000 升至 1,500；高度集中从 1,800 升至 2,500）——芝加哥学派主导的政策环境

- **2023**：新 DOJ/FTC 合并指南将阈值大幅回调——HHI $>$ 1,800 回归为" 高度集中"，Δ HHI $>$ 100 推定触发审查——某种程度上回到了 1982 年的标准

HHI 的演变反映了更深层的思想潮流：**同样的数字，不同的政策含义**。HHI 本身只是一个加总市场份额平方和的公式——赋予它什么阈值和什么政策含义，完全取决于时代的主导范式。

4.7.2　遗产 2：进入壁垒分析

Bain (1956) 的三个壁垒来源——规模经济、产品差异化、绝对成本优势——至今仍是任何进入分析的**自然分类起点**。

但从 Bain 的" 有什么壁垒？" 到现代的" 这个壁垒有多大？它的来源是什么？它是否可以由策略行为解释？"——分析框架已经深刻演化：

- **Stigler (1968)** 收窄了定义：只有" 在位者与进入者之间的不对称成本" 才算壁垒（$\rightarrow$ 第 ~5~ 章）

- **Baumol, Panzar & Willig (1982)** 提供了极端基准：如果进入和退出完全自由（无沉没成本），甚至垄断也会按边际成本定价——" 可竞争市场" 将壁垒的概念逻辑推向极限（$\rightarrow$ 第 ~6~ 章）

- **Sutton (1991)** 提供了新的理论语言：区分**外生沉没成本**（exogenous sunk costs——如建厂成本，与市场规模无关）和**内生沉没成本**（endogenous sunk costs——如 R&D 和广告，随市场规模上升）。后者会产生比前者更持久的集中度

Bain 的遗产：虽然他的具体分类法在今天看来过于静态和描述性，但他问的那个问题——" 为什么高利润行业的在位者不被进入者取代？"——仍然是 IO 的**元问题**之一。

4.7.3　遗产 3：市场界定的核心地位

SCP 传统的一个副产品是将" **市场界定**"（market definition）推向反垄断分析的中心——虽然是以一种有问题的方式。

- **SCP 时代**：市场 = SIC 分类（行政便利驱动的）

- **1982 年以后**：市场 = SSNIP 测试（经济学驱动的）

SSNIP（Small but Significant Non-transitory Increase in Price）测试的核心理念——”哪些产品和企业构成了彼此的有效竞争约束？”——可以从 Bain 和 Mason 的核心洞见回溯：”你不能在不先界定市场的情况下谈论集中度。”

现代演进：

- **UPP（Upward Pricing Pressure）**：Farrell & Shapiro (2010)——不先界定市场，直接测试合并是否会创造单边价格上升的激励
- **合并模拟（Merger Simulation）**：Nevo (2000)——用结构需求模型模拟合并后的价格效应，不需要先验的市场界定

这些现代工具都是在对 SCP ”先界定市场，再计算集中度”的范式做出回应和改进——它们试图**绕过** SCP 的市场界定问题，直接估算合并的价格效应。

4.7.4 遗产 4：结构性推定的政策直觉

2023 年 DOJ/FTC 新合并指南中，HHI 阈值的大幅回调（与 2010 年指南相比）引发了关于”结构性推定回归”的讨论。这是否意味着 Mason 和 Bain 的幽灵回到了政策舞台？

2023 年合并指南的关键变化

- **HHI 高度集中阈值**：从 2,500（2010）降至 1,800（2023）——回到 1982 年水平
- **Δ HHI 推定触发**：从 > 200（2010）降至 > 100（2023）——”推定”门槛大幅降低
- **取消”安全港”**：不再有 HHI 低于特定阈值即可自动通过的安全性规定
- **新增对纵向合并、连续收购（roll-up acquisitions）和平台市场中多归属（multi-homing）的关注**

但对这一”回归”的解释需要谨慎：

- 2023 年指南仍然保留了**效果分析**（effects-based analysis）作为最终标准——HHI 阈值只是筛选工具（screening tool），不是决定性标准
- 新指南增加了很多**非结构性**考量——劳动力市场效应、创新效应、平台网络效应等
- Hovenkamp (2024) 认为，新布兰代斯学派重拾了结构主义的手段，但目标已从”经济效率”（芝加哥标准）转向了更广泛的”权力分散”（dispersion of power）——这是 Mason 和 Brandeis 的结合，而非纯粹的 Harvard SCP 回归

更准确的描述：2023 年指南不是 SCP 的” 复辟”——它是结构性推定和效果分析的**新的混合体**。结构变量（HHI）重新获得了更大的权重，但决策的最终依据仍然是经济分析（效率抗辩、进入可能、买方势力等）。详见 第 ~15~ 章。

4.7.5　遗产 5：SCP 在邻域的存续

SCP 的基本框架在经济学核心之外找到了持久的位置：

- **战略管理**：Porter (1980) 的五力模型（Five Forces）本质上是 SCP 的管理学衍生——行业结构（供方议价力、买方议价力、新进入者威胁、替代品威胁、现有竞争者间的竞争强度）决定企业战略和行业” 吸引力”。学术研究在 1990 年代初期从五力框架转向了 RBV（Resource-Based View），但 MBA 课程至今仍在教授 Porter。

- **发展经济学**：跨国/跨行业回归用于研究市场结构、竞争与增长的关系——概念框架直接继承自 SCP

- **国际贸易**：集中度与出口绩效、进口竞争的影响——使用的变量（CR4、HHI、PCM）来自 SCP 工具箱

- **金融经济学**：银行业集中度与竞争——银行并购审查中使用 HHI 的方式直接源自 SCP 传统

SCP 的” 学术谱系”

SCP 在经济学核心领域的” 死亡” 和在管理/政策领域的” 存活” 构成了经济学思想传播的有趣案例。一个可能的解释是：

- 经济学对**因果识别**的要求在过去 40 年急剧提高——SCP 回归无法通过这个标准
- 管理学对分析**可用性**（actionability）的要求高于因果精确性——企业战略家需要的是一个思考框架，而非一个点估计
- 政策制定需要**可操作的规则**（administrable rules）——HHI 阈值提供了这样的规则，即使其因果基础存在争议

这也解释了为什么 SCP 在 MBA 课程和竞争政策指南中持续存在——那里看重的是 SCP 的**框架功能**（frame function：帮助组织思维），而非其**因果推断功能**（causal inference function：精确估计效应大小）。

表 4.6: SCP 遗产清单：什么存活，什么被取代

概念/工具	SCP 中的原始形式	当代演化	存续状态	关键差异
HHI	CR4/CR8 的替代集中度度量	全球反垄断执法的核心筛选标准	**完全存续**	阈值政策含义随范式变化而变
进入壁垒	Bain (1956) 三类壁垒	沉没成本模型 (Sutton 1991); 策略壁垒 vs. 结构壁垒	**重新定义后存续**	从静态分类到动态策略分析
市场界定	SIC 分类 = 市场	SSNIP 测试 (1982–); UPP (2010); 合并模拟	**完全存续并深刻演化**	从行政便利到经济学驱动
结构性推定	” 大就是坏”——高集中度本身推定违法	2023 指南中的 HHI 阈值回调	**部分回归**	从确定性推定到可反驳推定
跨行业分析	OLS 简约式回归	跨国面板回归 (非 IO 主流的但活跃)	**在 IO 核心被放弃**	在邻域（发展、金融、管理）存续
S → C → P 框架	单向因果链	结构 ↔ 行为 ↔ 绩效——反馈回路（已由 Scherer 纳入）	**作为启发式框架存续**	不再被视为因果模型，而是组织工具

4.7.6　遗产清单

4.8　从 SCP 到 NEIO：一个研究纲领的嬗变

4.8.1　Bresnahan (1989) 的诊断与处方

Bresnahan, T.F. (1989). "Empirical Studies of Industries with Market Power." ***Handbook of Industrial Organization*****, Vol. 2, Ch. 17.**

Bresnahan 的 Handbook 章节与 Schmalensee (1989) 的综述同期发表——它们共同构成了 SCP 经验传统的" 讣告" 和 NEIO 新纲领的" 出生证"。

Bresnahan 将 SCP 的失败诊断为**三个方法论问题**：

1. **跨行业回归不能识别行为**：集中度 → 利润的相关性对于" 合谋" 和" 效率" 两种叙事是**观察等价**的（observationally equivalent）。同一组数据，不同的因果故事——回归无法区分它们

2. **SCP 回归缺乏微观基础**：它们没有从企业的最优化行为推导出来——它们是 ad hoc 的关联。这与 Lucas 批判的逻辑相似：简约式参数（β_1）不是" 深层参数"——它们随制度和博弈规则而变化

3. **市场界定问题**（§4.4）——SIC $\neq$ 市场

Bresnahan 的 NEIO 处方：

- 研究**单个行业**（single-industry studies）而非跨行业——消除市场界定问题
- 从**理论模型**（博弈论）推导**结构方程**——供给关系（supply relation）和需求系统（demand system）
- **估计行为参数**（conduct parameter）θ——直接度量行为，而非从结构推断行为

4.8.2　行为参数 θ：度量市场势力的新方法

NEIO 的核心创新之一是" 行为参数"（conduct parameter）的概念。在标准的寡头供给关系中：

$$P + \theta \cdot P'(Q) \cdot q_i = \text{MC}_i \tag{4.6}$$

其中 θ 是行为参数：

- $\theta = 0 \to P = \text{MC}$：完全竞争 / Bertrand 结果
- $\theta = 1 \to P + P'(Q)q_i = \text{MC}$：Cournot 结果或完全合谋（单一企业）

- $\theta \in (0,1)$：部分合谋或不完全竞争——企业在一定程度上" 内化" 了产出扩张对市场价格的影响

关键创新：θ 不再是理论假设（如在 Cournot 模型中 $\theta \equiv 1$），而是**从数据中估计**的对象。需求弹性可以从数据中估计（通过需求系统），成本可以从供给关系中联合估计。然后，θ 的估计值告诉研究者这个行业的竞争强度——这是直接度量" 市场势力" 的行为表现，而非从结构推断行为。

与前文的知识链连接：

- 第 ~1~ 章定义了 Lerner 指数 $L = \frac{P-\text{MC}}{P} = \frac{s_i}{|\varepsilon|}$
- 第 ~2~ 章证明在 Cournot 模型中 $L = \frac{s_i}{|\varepsilon|}$——结构直接影响绩效
- 第 ~3~ 章提供了测量 s_i 的工具（CR4, HHI）
- **本章**展示了直接检验这个关系的第一步（SCP 跨行业回归）——以及它的失败
- 第 ~10~ 章将展示 Bresnahan 如何通过直接估计 θ 来**绕过** SCP 回归的所有方法论陷阱

4.8.3 SCP 跨行业回归 vs. NEIO 结构估计：全面的方法论对比

为什么 NEIO 是 SCP 的" 继承者" 而非完全决裂

NEIO 虽然全面拒绝了 SCP 的方法论（跨行业回归 + 结构外生假设），但它继承了 SCP 的**核心问题**：

如何用数据来检测和度量市场势力？

SCP 用" 集中度" 来回答这个问题（代理变量策略）。NEIO 用" 行为参数" 来回答这个问题（结构估计策略）。问题相同，工具有别。
这是科学进步的常态模式——不是" 否定老问题"，而是" 用更好的工具重新回答老问题"。

4.8.4 总结性时间线：IO 经验研究的范式转换

4.9 本章小结

本章讲述了 IO 历史上最重要的方法论转折。以下是七个关键收获：

1. **SCP 经验纲领是 IO 第一个系统的大规模经验研究纲领。**从 Bain (1951) 的 42 个行业均值比较到 1970 年代的数百个行业回归，SCP 产生了庞大的经验文献。

表 4.7: SCP 跨行业回归 vs. NEIO 结构估计

维度	**SCP 跨行业回归**	**NEIO 结构估计**
分析单位	行业横截面（50–400 个行业）	单个行业（或紧密相关的一组行业）
理论基础	S → C → P 启发式 + Bain 壁垒分类	博弈论纳什均衡（Cournot、Bertrand 或一般行为参数模型）
计量方法	OLS / 2SLS	GMM / 极大似然 / 模拟矩方法（SMM）
识别策略	无明确识别——依赖结构外生假设	从需求和供给一阶条件联合估计——需求弹性从价格-数量协变中识别
度量目标	集中度-利润相关性（R^2, $\hat{\beta}_1$）	行为参数 θ——直接度量市场势力
政策含义	” 集中度太高了——干预”	” 合并将使 θ 从 0.3 上升至 0.5——价格预期上升 7%”
市场界定	SIC 分类（行政便利）	模型内生决定——哪些产品属于同一需求系统
数据需求	中等（公开 Census 数据）	高（产品级价格、数量、特征、成本工具变量）
反事实模拟	不可能（简约式参数不是” 深层参数”）	可能——模型结构允许预测合并、税收、规制的福利效应

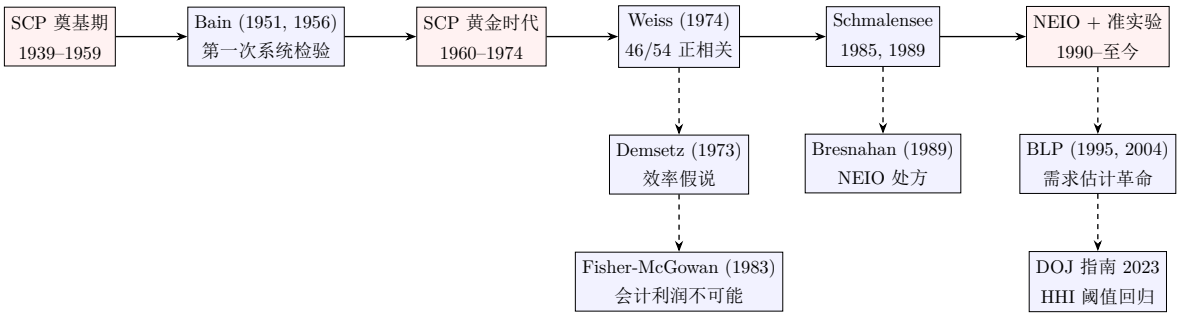


图 4.2: IO 经验研究的范式转换：从 Bain (1951) 到 2023 年合并指南。实线箭头为主流发展路径，虚线箭头为批判/替代路径。

Weiss (1974) 的综述——46/54 研究发现集中度-利润正相关——标志着这个纲领的经验巅峰。

2. **但核心发现是脆弱的。**虽然大多数研究发现正相关，但量级很小（$R^2 < 10\%$），且对设定选择高度敏感。更关键的是——这个" 正相关" 对因果解释几乎没有约束力。

3. **Demsetz (1973) 提出了一个至今无解的挑战。**效率 $\rightarrow$ 集中度 + 利润 vs. 集中度 $\rightarrow$ 合谋 $\rightarrow$ 利润——这两个故事在行业横截面数据中是**观察等价**的。跨行业回归无法区分它们。Demsetz 的论证不仅适用于 IO——它是所有经验经济学中" 内生性" 问题的经典教科书案例。

4. **Fisher-McGowan (1983) 动摇了度量基础。**如果会计利润率不能反映经济盈利能力——而三个具体原因（折旧任意性、无形资产费用化、通胀效应）使这成为一个**不可能定理**而非仅仅是" 噪声"——那么数百篇 SCP 回归建立在不可靠的度量基础之上。

5. **SCP 的衰落是多维度的。**经验战线的内生性问题、芝加哥学派的理论批判、博弈论革命、计量可信性革命、反垄断实践变革——这些力量共同而非单独地将 SCP 推向了衰落。单一原因的解释（如"SCP 因为内生性而失败"）过于简化——内生性提供了一个诊断，但范式转换需要替代范式的存在。

6. **SCP 的 DNA 比 SCP 本身更持久。**HHI 仍是全球反垄断的核心筛选工具。进入壁垒分析仍是任何竞争评估的自然起点。市场界定的核心地位可追溯到 Bain 对" 行业" 的执念。2023 年合并指南中 HHI 阈值的回调引发了" 结构性推定回归" 的讨论。SCP 作为分析框架（而非因果模型）的持久性是其影响力最深刻的证明。

7. **SCP 的失败是" 好的科学失败"。**它提出了清晰、可检验的假说；用当时最好的方法去检验它们；当证据和方法论批判积累到无法忽视时，领域转向了更强大的工具（博弈论 + 结构估计）。这不是一个" 错误的理论" 的故事——这是**知识积累的正常模式**。

4.10 习题

基础（3 题）

4.1 SCP 回归的解读。一个典型的跨行业 SCP 回归结果如下：

$$\hat{\pi}_i = \underset{}{0.05} + \underset{(0.02)}{0.12} \cdot \text{CR4}_i + \underset{(0.04)}{0.08} \cdot \text{Barriers}_i + \underset{(0.01)}{0.03} \cdot \text{Growth}_i, \quad R^2 = 0.15,\ N = 200$$

括号中为标准误。

(a) 解释 CR4 的系数：集中度上升 10 个百分点伴随着利润率上升多少？

(b) $R^2 = 0.15$ 告诉你什么？集中度在多大程度上" 解释" 了跨行业利润变异？

(c) 基于本章的讨论，列举至少三个原因说明为何不能将 CR4 的系数解释为" 集中度对利润率的因果效应"。

4.2 效率假说 vs. 市场势力假说。

(a) 用你自己的语言（不逐字引用原文）阐述：如果 Demsetz (1973) 的效率假说正确，那么 SCP 回归中 CR4 的显著正系数实际上度量的是什么？

(b) Demsetz 建议用什么经验检验来区分效率假说和市场势力假说？（提示：比较集中行业和分散行业中**小企业**的利润率。）

(c) 这个检验的一个局限性是什么？（提示：如果高效率企业的低成本的来源是**规模经济**而非外生的" 管理质量"，那么" 企业层面 vs. 行业层面" 的区分是否还有说服力？）

4.3 会计利润 vs. 经济利润。

(a) 列出 Fisher & McGowan (1983) 认为会计 ROA 不能度量经济利润的三个具体原因。

(b) 为什么 R&D 密集型行业中，会计 ROA 尤其可能高估经济盈利能力？

(c) 如果所有行业都使用相同的会计规则，且会计扭曲在不同行业间具有相似的方向和量级，那么跨行业 SCP 回归是否仍然可以检测到经济利润的**差异**？为什么能或为什么不能？

中级（3 题）

4.4 内生性诊断。考虑以下联立方程系统：

$$\pi_i = \alpha_1 + \beta_1 \cdot \mathrm{CR}_i + \gamma_1 \cdot \mathrm{Growth}_i + \varepsilon_{1i} \tag{4.7}$$

$$\mathrm{CR}_i = \alpha_2 + \beta_2 \cdot \pi_i + \gamma_2 \cdot \mathrm{MES}_i + \varepsilon_{2i} \tag{4.8}$$

其中 MES_i 是最小有效规模（假设由技术条件外生决定），π_i 是行业 i 的利润率。

(a) 证明：如果 $\beta_2 \neq 0$（利润影响集中度），那么第一式的 OLS 估计量 $\hat{\beta}_1$ 不一致。请说明偏差的方向与 β_2 符号的关系。

(b) MES_i 是第一式有效的工具变量吗？请从**相关性**（第一阶段 F 统计量）和**排他性**（MES 是否只通过集中度影响利润?）两个条件进行论证。

(c) 如果 $\beta_2 < 0$（高利润吸引进入，降低集中度），它会对第一式中 $\hat{\beta}_1$ 的 OLS 估计产生什么方向的偏差？

4.5 临界集中度假说。 Bain (1951) 发现 CR8 $> 70\%$ 的行业利润" 跳升"，而非线性增加。

(a) 如果临界集中度假说正确，为何这很重要？对反垄断政策有何含义？

(b) 建议一种经验策略来检验临界集中度假说。讨论它可能遇到的计量问题：临界值的内生选择、结果对临界值的敏感性、以及" 数据窥探"（data snooping）的风险。

4.6 Schmalensee 方差分解。阅读 Schmalensee (1985) 的方差分解——行业效应 $\approx 20\%$，市场份额效应 $\approx 0.6\%$，残差 $\approx 80\%$。

(a) 为什么 0.6% 的市场份额效应是" 对 SCP 的沉重打击"——集中度不就是由市场份额构建的吗？这两个结果（行业效应显著、市场份额效应接近零）如何同时成立？

(b) " 残差 $\approx 80\%$" 告诉我们关于 SCP 经验纲领的什么局限性？这 80% 的残差可能包含哪些未观测因素？

(c) Schmalensee 的结果是否意味着市场结构" 不重要"？论证你的立场。（提示：区分" 是否存在真实效应" 与" 这个效应是否可以通过跨行业回归可靠地度量"。）

进阶（2 题）

4.7 SCP 的兴衰作为" 科学研究纲领" 的变迁。用 Lakatos (1970) 的" 科学研究纲领方法论"（Methodology of Scientific Research Programmes）框架分析 SCP 的兴衰：

- **硬核**（hard core）：S $\to$ C $\to$ P 单向因果链；结构由技术和偏好外生决定；集中度可以衡量市场势力
- **保护带**（protective belt）：具体度量方法（CR4 vs. HHI）、壁垒分类（Bain vs. Stigler）、计量设定（线性 vs. 含临界值）

- **正面启发法**（positive heuristic）：找更多行业、更好度量、控制更多变量

(a) SCP 纲领在 1950–1970 年代是" 进步的研究纲领" 还是" 退化的研究纲领"？在 1980 年代之后呢？根据 Lakatos 的区分标准——进步纲领的每次修正都产生" 新颖的事实"（novel facts）——来论证你的判断。

(b) NEIO 是否构成了一个" 新的研究纲领"？如果是，它的硬核是什么？它与 SCP 纲领的关系是" 继承"、" 替代" 还是" 更高层次的综合"？

4.8 中国的集中度-利润关系。阅读一篇关于中国制造业集中度与利润率关系的实证研究（如使用中国工业企业数据库 1998–2013 的研究），并讨论：

(a) 中国制造业的集中度-利润率关系是否类似美国 SCP 文献的发现？量级和稳健性如何？

(b) 在中国制度背景下——国有企业（SOE）、地区保护主义（regional protectionism）、产业政策（industrial policy）——集中度-利润关系的解释面临哪些美国文献中不存在的额外挑战？

(c) 如果 Joe Bain 今天用他的方法（行业访谈 + 利润率比较）研究中国制造业，你认为他会发现什么？

4.11　延伸阅读

★ 必读（本章核心参考）

- **Weiss, L.W. (1974).** "The Concentration-Profits Relationship and Antitrust." In H.J. Goldschmid, H.M. Mann & J.F. Weston (eds.), *Industrial Concentration: The New Learning*, pp. 184–233. Little, Brown. ——46/54 的经典综述。阅读时需注意：它发表于 SCP 经验的巅峰时期——Weiss 本人对这个证据的解读是"SCP 获得了经验支持"。从今天的角度看，它是一个方法论的" 时间胶囊"。

- **Demsetz, H. (1973).** "Industry Structure, Market Rivalry, and Public Policy." *Journal of Law and Economics*, 16(1), 1–9. ——三页改变了一个领域的因果叙事。学习如何用最简单的逻辑和最小的数据来推翻一个庞大的经验文献。这是一篇" 少即是多" 的经典示范。

- **Schmalensee, R. (1989).** "Inter-Industry Studies of Structure and Performance." In *Handbook of Industrial Organization*, Vol. 2, Ch. 16, pp. 951–1009. North-Holland. ——SCP 经验传统的权威验尸报告。如果你的时间只够读一篇本章参考文献，读这篇。Schmalensee 的综述不仅是总结——它是" 如何系统评估一个经验文献" 的方法论教科书。

- **Fisher, F.M. & McGowan, J.J. (1983).** "On the Misuse of Accounting Rates of Return to Infer Monopoly Profits." *American Economic Review*, 73(1), 82–97. ——对 SCP 度量基础的彻底批判。读完之后问自己：在我的经验研究中，我使用的度量变量是否经得起类似的检验？

★★ 进阶（加深理解）

- **Peltzman, S. (1977).** "The Gains and Losses from Industrial Concentration." *Journal of Law and Economics*, 20(2), 229–263. ——价格-成本分解的精妙框架。学习如何在不直接观测成本和价格的情况下，区分" 成本驱动的利润上升" 和" 市场势力驱动的利润上升"。

- **Schmalensee, R. (1985).** "Do Markets Differ Much?" *American Economic Review*, 75(3), 341–352. ——方差分解的经典。行业效应 vs. 企业效应 vs. 市场份额效应——这是一个" 元分析"（meta-analysis）思维的早期体现。

- **Domowitz, I., Hubbard, R.G. & Petersen, B.C. (1986).** "Business Cycles and the Relationship between Concentration and Price-Cost Margins." *RAND Journal of Economics*, 17(1), 1–17. ——集中度-利润率关系的时间维度。一个经验规律在实时削弱——对这个现象的观察本身就是对 SCP 经验纲领的重击。

- **Salinger, M.A. (1990).** "The Concentration-Margins Relationship Reconsidered." *Brookings Papers on Economic Activity: Microeconomics*, 287–335. ——方法论上最严谨的 SCP 再检验之一。控制资本存量和测量误差后，关系显著弱化。Salinger 包含讨论者的评论（其中多位是 SCP 辩论的关键人物），阅读整篇对话式评论可以让你感受 1990 年前后 IO 领域内部的争论状态。

- **Bresnahan, T.F. (1989).** "Empirical Studies of Industries with Market Power." In *Handbook of Industrial Organization*, Vol. 2, Ch. 17, pp. 1011–1057. North-Holland. ——NEIO 的宣言。与 Schmalensee (1989) 同时阅读——这两章共同构成了 IO 经验研究的" 岔路口"。

★★★ 前沿（当代重新评价与发展）

- **Berry, S., Gaynor, M. & Scott Morton, F. (2019).** "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization." *Journal of Economic Perspectives*, 33(3), 44–68. ——当代加价（markup）上升的讨论，可以视为 SCP 核心问题——" 市场势力在上升吗？"——在现代方法下的重访。三位作者分别来自 NEIO 结构估计传统（Berry）、医疗经济学（Gaynor）和反垄断经济学（Scott Morton）。

- **De Loecker, J., Eeckhout, J. & Unger, G. (2020).** "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics*, 135(2), 561–644. ——使用企业层面数据（而非行业数据！）重新审视市场势力问题。方法论上完全不同于 SCP（成本份额法 + 生产函数估计），但回答的核心问题正是 Mason 和 Bain 在一个世纪前提出的。

- **Hovenkamp, H. (2024).** "The SCP Paradigm and the Rise of Neo-Brandeisian Antitrust." *Yale Law Journal Forum*. ——2023 年 DOJ/FTC 合并指南是否标志着 SCP 结构主义的回归？Hovenkamp 分析了新布兰代斯学派与哈佛结构主义之间的连续与断裂。

- **Berry, S. & Haile, P. (2024).** "Foundations of Demand Estimation." In *Handbook of Industrial Organization*, Vol. 5. North-Holland. ——BLP 之后 30 年的需求估计：方法论综述。追溯从 BLP (1995) 到 Berry-Levinsohn-Pakes (2004) 到现代微观数据时代的需求估计演化。（为 第 ~11~ 章的阅读做热身。）

下一章： → 第 ~5~ 章　效率 vs. 市场势力：芝加哥学派的颠覆——Demsetz 的效率假说只是芝加哥革命的第一枪。Stigler 如何重新定义进入壁垒？Posner 和 Bork 如何将消费者福利标准推上神坛？芝加哥学派如何在一个十年里改变了一个领域的世界观？

第五章　效率 vs. 市场势力：芝加哥学派的颠覆

章首语

1946 年，一位名叫 Aaron Director 的经济学家加入了芝加哥大学法学院。他没有博士学位，在接下来的 20 年里几乎没有发表任何署名论文。但他的周四晚反垄断研讨课——围坐在一张长桌旁，由 Friedman 主导讨论，Stigler 做笔记，年轻的 Posner 和 Bork 洗耳恭听——成为了 20 世纪经济学最具影响力的" 口述传统"。George Stigler 后来说：**"Aaron 的大部分论文都以他同事的名义发表了。"**

从这张桌子旁诞生了一套完整的替代世界观：市场是自我修正的；垄断是暂时的；政府通常是问题而非解决方案；效率——而非对小企业的保护——应该是反垄断的唯一目标。这套世界观用一代人的时间征服了美国反垄断政策，并在全球范围内重塑了竞争法的智识基础。

第 ~4~ 章讲述了 SCP 经验纲领的衰落——它被内生性、测量误差、博弈论和计量可信性革命一一击破。但经验失败本身并不足以杀死一个研究纲领。Lakatos 教我们：退化的研究纲领可以持续很久。SCP 的终结需要的不仅是" 你的回归有问题"——还需要一个**完整的替代方案**：一套能解释同样现象（集中度、利润、进入、纵向约束）但给出截然不同政策处方的理论体系。

芝加哥学派提供了这个替代方案。而且，它不仅仅是一个替代的理论——它是一种完全不同的**世界观**。哈佛问：" 市场哪里失灵了？" 芝加哥问：" 你有什么证据证明市场失灵了？" 哈佛看集中度看到了市场势力；芝加哥看集中度看到了效率差异。哈佛认为进入壁垒无处不在；芝加哥认为真正的壁垒只有政府创造的那一种。

本章讲述这个颠覆的故事。核心张力在于：**芝加哥学派几乎在每一个具体问题上都提供了与 SCP 相反的答案——而这些答案在 1980 年代成为了美国反垄断政策的操作标准**。理解芝加哥学派不是要" 选边站"——而是理解为什么聪明的、善意的经济学家可以从完全相同的可观测事实中得出完全相反的结论。

5.1　核心问题

读完本章，你应该能够回答以下问题：

1. **芝加哥学派是什么？**它与哈佛 SCP 学派的分歧不仅是经验问题（相关性 vs. 因果性），更是世界观问题——关于市场如何运作、政府应该做什么、以及经济学应该如何做的根本分歧。

2. **Aaron Director 为什么是芝加哥 IO 的" 隐形创始人"？**他的口头传统如何比大多数人的论文产生了更大的影响？" 一个垄断利润" 定理如何仅用价格理论逻辑就推翻了搭售的杠杆化理论？

3. **Stigler 如何——以及为什么——重新定义进入壁垒？** Bain 的三个壁垒（规模经济、产品差异化、绝对成本优势）在 Stigler 的定义下还剩下什么？为什么这一定义转变具有巨大的政策后果？

4. **Posner 和 Bork 如何改变反垄断法？**消费者福利标准、" 合理原则" 的扩张、以及" 错误成本" 框架如何重塑了执法？

5. **芝加哥学派的政策胜利有多完整？** 1982 年合并指南、里根政府的反垄断革命、放松管制运动——什么成功了，什么没有？

6. **芝加哥的遗产是什么？**后芝加哥综合如何接受了芝加哥的一些洞察而拒绝了另一些？新布兰代斯学派的挑战在动摇什么？

5.2　"桌子旁的经济学"：芝加哥口述传统与 Aaron Director

5.2.1　Aaron Director：没有论文的创始人

【核心】

Aaron Director (1901–2004)：关键事实

- 生于乌克兰，1913 年移民美国，耶鲁本科，芝加哥大学师从 Frank Knight 和 Jacob Viner
- 1946 年加入芝加哥大学法学院——这标志着” 法律与经济学” 的制度化诞生
- 与 Edward Levi（后来的法学院院长、芝加哥大学校长、美国司法部长）共同教授反垄断课程
- Milton Friedman 的姐夫——Rose Director 在 1938 年与 Friedman 结婚
- 在 20 年学术生涯中几乎未发表署名论文
- Ronald Coase 将自己比作圣保罗，将 Director 比作基督：” 他创立了教义，我的工作是将它带给外邦人”

Director 的 seminar 风格是传奇性的。Kenneth Scott 将 Director 的课堂评论比作”Thurber 漫画中的那个击剑手——一剑封喉，只留下一句 ‘Touché’”。Director 会提出一个简单的问题，然后整条论证线就瓦解了。

芝加哥的” 价格理论帝国” 在 Director 的 seminar 桌旁锻造，而非在期刊论文中诞生。这一事实本身就意味深长：**芝加哥 IO 的智识 DNA 中刻着口头论证和” 纯逻辑”——而非计量识别或结构估计**。这解释了为什么芝加哥学派在 1950–60 年代几乎没有产生 SCP 式的大规模经验研究，却能够从根本上挑战 SCP 的逻辑基础。

5.2.2 “一个垄断利润” 定理

【核心】

Director 对搭售（tying）的分析是他最著名的贡献——也是芝加哥” 纯价格理论” 方法的最纯粹展示。

传统智慧——被法院接受——是：在 X 产品上拥有垄断地位的企业可以通过搭售将垄断” 杠杆化” 到 Y 产品市场。逻辑看似直观：” 如果有一个垄断是好的，有两个更好。”

Director 用纯粹的价格理论逻辑推翻了这一点。**“一个垄断利润” 定理**（single monopoly profit theorem）：

> 垄断者可以从其在一个市场上的垄断地位中榨取**所有**可获得的垄断利润——不需要第二个垄断。被搭售产品面临更高价格的消费者会减少对主产品的需求。两个市场的联合利润不可能超过对垄断产品单独最优定价的利润。

直觉：假设你在 widget 上拥有垄断地位——消费者**必须**购买你的 widget 才能使用它。通过 widget 的垄断价格，你已经从每个消费者那里榨取了最大可能的剩余。试图通过搭售一个竞争性配件来" 再次" 榨取剩余是徒劳的——消费者只会减少 widget 的购买。垄断者只有一个垄断利润可以榨取。

为什么这个论证如此有力

Director 没有使用任何数据。他没有做任何回归。他只是画了一条需求曲线，然后问了正确的问题：" 如果垄断者在 widget 的价格中已经榨取了所有的垄断利润，搭售如何能增加总利润？"

这就是芝加哥方法论的缩影：**价格理论 + 逻辑一致性 = 一个世纪的判例法被推翻**。

对" 搭售悖论" 的解决：如果搭售不能增加垄断利润，为什么真实世界中的企业会这么做？Director 的答案：**搭售是一种计量/价格歧视工具**。当 (a) 消费者使用主产品的强度不同（高强度和低强度使用者），(b) 显式价格歧视由于套利或无法预先识别消费者类型而不可行时，企业可以通过低价出售固定组件（如机器）并高价出售可变搭售组件（如卡片、墨盒、盐）来**隐含计量使用强度**——重度使用者支付更多。

唯一的署名发表：Director & Levi (1956), "Law and the Future: Trade Regulation," *Northwestern University Law Review*, 51, 281——这篇论文制定了以经济效率为核心的反垄断研究议程。

5.2.3　反垄断项目 (The Antitrust Project, 1953–)

【核心】

1953 年，Director 被任命为**反垄断项目**的负责人。这是一个从自由市场第一原理系统地重构美国反垄断法的研究计划。其核心论点是激进的：

私人垄断只能来自政府行为——任其自行，市场会自我修正。

该项目最著名的经验产品是 **G. Warren Nutter 的博士论文**，*The Extent of Enterprise Monopoly in the United States, 1899–1939*（芝加哥大学出版社，1951 年）。虽然完成于项目正式启动之前，Nutter 的论文体现了该项目的核心精神——在 Director、O.H. Brownlee 和 Friedman 的指导下，Nutter 的定量研究显示垄断并未在美国经济中显著增加——垄断行业仅占国民收入的约五分之一。

口述传统 vs. 发表记录

芝加哥 IO 的一个显著特征是：其最具影响力的思想往往**首先**以口头形式传播——在 Director 的研讨课、Friedman 的 workshop、或 Coase 的晚宴对话中——**然后**才出现在出版物中。这与哈佛 SCP 传统形成对比，后者从 Mason 和 Bain 开始就是高度经验性的、基于出版的。

这种差异不仅是风格问题——它反映了对" 什么构成了有效的经济学论证" 的根本不同理解。芝加哥相信**逻辑**（价格理论的一阶含义）优先于**证据**（计量估计）。

5.2.4　芝加哥反垄断的" 神圣三位一体"

【核心】

Director 的学生形成了芝加哥反垄断的核心：

- **Robert Bork**：称 Director 为" 芝加哥法律与经济学的真正教父"（《华尔街日报》，2004 年），撰写了《反垄断悖论》（1978 年）
- **Richard Posner**：第七巡回法院法官，撰写了《法律的经济分析》（1973 年），将芝加哥经济学扩展到法学的每一个角落
- **Frank Easterbrook**：第七巡回法院首席法官，发表了" 反垄断的限度"（1984 年，《德克萨斯法律评论》）

2018 年，助理司法部长 Makan Delrahim 在芝加哥演讲时宣称：" 我感觉自己站在了神圣的土地上，这里是世界上一些领先的反垄断学者发展现代教义的地方——Ed Levi、Aaron Director、Robert Bork……Frank Easterbrook、Richard Posner……"

5.2.5　芝加哥学派的两代人

这一代际转变对 IO 至关重要。第一代芝加哥经济学家（尤其是 Henry Simons）对垄断持高度批评态度，并支持积极的反垄断执法。第二代——在 Director 和 Friedman 的影响下——将经济论证推向了一个更为彻底的**亲市场**方向。第二代的 IO 立场可以概括为：" 如果市场是集中的，那是因为有效率的企业赢得了竞争。干预只会惩罚成功者。"

5.3　Stigler 的进入壁垒重定义：一场定义之战

5.3.1　Bain 的原定义（简要回顾）

【核心】

表 5.1: 芝加哥学派的两代人

维度	第一代（战前）	第二代（战后）
核心人物	Frank Knight, Jacob Viner, Henry Simons	Milton Friedman, George Stigler, Gary Becker
价格理论的角色	在不同制度安排下研究市场适应的工具	识别唯一配置解的工具
方法论	Marshall 的局部均衡，制度敏感性	硬理性选择，扩展到所有社会行为
核心关切	自由社会的制度基础	市场在任何环境下的效率属性

学生已在 第 ~3~ 章学习了 Bain (1956) 的三个壁垒来源：规模经济、产品差异化和绝对成本优势。这里简要回顾 Bain 的**基于效果的定义**：

> 进入壁垒是" 在位者能够持续将价格提高到竞争水平之上而不吸引新企业进入的任何在位者优势"。

这个定义不问**为什么**进入被阻止——只问进入是否事实上被阻止。如果观察到持续的超额利润且没有进入发生 → 存在壁垒。这是**结果导向**的定义。

5.3.2 Stigler (1968) 的精确措辞与逻辑

【核心】

Stigler, G.J. (1968). *The Organization of Industry*. Ch.6 "Barriers to Entry, Economies of Scale, and Firm Size." Irwin.

Stigler 提出了一个截然不同的定义：

> 进入壁垒是" 一种生产成本（在某个或每个产出率下），寻求进入行业的企业必须承担，而行业中已有的企业不承担。"

这是**基于不对称成本的定义**。核心不是" 进入是否发生了"——而是" 进入者是否面临在位者不需要支付的差异成本"。

5.3.3 为什么规模经济不是壁垒

【核心】

Stigler 的论证直接且有力：

1. 如果进入者可以建设与在位者" 恰好相同规模" 的工厂达到 MES，那么它面临与在位者**完全相同**的长期平均成本曲线。

表 5.2: Bain (1956) vs. Stigler (1968) 进入壁垒定义的对比

维度	**Bain (1956)**	**Stigler (1968)**
定义标准	基于效果：能否持续获得超额利润而不吸引进入？	基于不对称：进入者是否面临在位者没有的成本？
检验问题	”进入发生了吗？”	”进入者支付了在位者不需要支付的成本吗？”
规模经济	是壁垒——如果 MES 相对于市场大	**不是壁垒**——进入者可以建同样规模的工厂
资本需求	是壁垒——如果绝对资本需求大	**不是壁垒**——资本市场为所有正 NPV 项目提供资金
产品差异化	是壁垒——在位者品牌忠诚度	**不是壁垒**——进入者可以投资品牌建设
专利	是壁垒	是壁垒（满足不对称条件）
政府创造的排他性	未单独分类	**是壁垒**——真正的壁垒
政策含义	广泛的干预范围	狭窄的干预范围——主要是政府壁垒

2. 两个企业为在有效规模上运营支付相同的成本。

3. 市场可能无法支撑两个 MES 工厂（需求相对于技术太小）这一事实——是**需求和技术的特征**，而非进入者单独承担的差异成本。

4. 关键点：在位者**也曾**必须在 MES 上建设和承担该成本。**没有不对称**。

定义之战的本体论含义

Stigler 的重新定义不仅仅是术语上的修正——它是一种**本体论上的减少**。一度是”壁垒”的东西（规模经济、广告、资本需求）突然”不是壁垒”了。一个纯粹分类学上的、非经验性的转变，戏剧性地缩小了反垄断干预的合法范围。

这是本章最深层的方法论教训之一：**定义很重要**。在经济学的政策应用中，你对一个关键术语的**定义**——在你跑任何回归或引用任何数据之前——已经预先决定了你的政策结论的大部分。

5.3.4　为什么资本需求不是壁垒——Modigliani-Miller 连接

【核心】

教学中必须强调的知识链：Stigler 诉诸 Modigliani & Miller (1958) 的逻辑：

> 在有效资本市场中，资本成本取决于项目的**风险和预期回报**，而非借款人的**身份**。一个提议建设正 NPV 有效规模工厂的进入者面临的资本成本，与一个提议相同扩张的在位者面临的相同。

所需资本的绝对数量仅仅是" 入场券的价格"——任何进入者平等支付。它**不是差异性的**。

局限性：M-M 假设要求苛刻——无税收、无破产成本、无信息不对称。在现实中，由于信息不对称（Akerlof 1970），新的、未经证实的进入者确实比有业绩记录的在位者面临更高的外部融资成本。Stigler 和芝加哥学派的立场是：**M-M 定理对政策目的而言" 足够正确"**——证明资本需求的差异性是那些声称存在壁垒者的举证责任。

芝加哥的举证责任分配

Stigler 的定义之战中隐含着一项深刻的制度选择：**举证责任在谁？**在 Bain 的框架下，证明*没有*壁垒的举证责任在合并方或被告。在 Stigler 的框架下，证明壁垒*存在*的举证责任在执法者或原告。

这不仅仅是证据规则——这是世界观。芝加哥将市场效率作为默认假设，要求干预者提供令人信服的证据。哈佛将市场失灵作为默认关切，要求市场证明其良性。

5.3.5　幸存者原则（1958 年）

【核心】

Stigler, G.J. (1958). "The Economies of Scale." ***Journal of Law and Economics*****, 1, 54–71.**

Stigler 提出了一个深具芝加哥特色的替代 MES 测量方法：

1. 将行业中的所有企业按规模类别分类

2. 追踪每个类别在行业总产出中的份额随时间的变化

3. **份额增长的规模类别 = 有效率的**；份额下降的类别 = 规模太大或太小

与 Bain 的工程方法的对比：Bain 试图通过调查工程师和工厂经理来估计 MES → 询问他们在不同规模下的成本结构。Stigler 认为该方法本质上不可靠——工程师不知道真实的成本；只有市场知道。

幸存者原则的深层芝加哥信念

幸存者原则体现了芝加哥的深层信念：**市场竞争是一个达尔文选择机制。只有最有效规模的企业才能生存和增长。无需询问工程师——让市场告诉你什么有效。**

这种”市场知道得最多”的认识论在芝加哥的几乎所有领域中反复出现：从 Stigler 的幸存者原则到 Fama 的有效市场假说，到 Posner 的”普通法的效率理论”。

5.3.6　在 Stigler 的定义下什么仍算作壁垒？

只有真正差异性的成本符合 Stigler 的标准：

- **专利**：进入者必须获得许可或发明替代方案；在位者已经拥有专利。这是差异成本。
- **稀缺资源控制**：如果在位者拥有唯一的铝土矿矿床，进入者面临的资源获取成本是差异性的。
- **政府创造的排他性**：许可证、特许经营权、保护在位者的关税——这些是 Stigler 眼中**唯一的真正壁垒**。
- **声誉资本/干中学**：如果在位者通过在位生产积累了进入者不生产就无法复制的效率——一个 Catch-22。这在 Stigler 的框架中处于灰色地带。

5.3.7　关键歧义：McAfee、Mialon 和 Williams (2004)

【进阶】

McAfee、Mialon 和 Williams 指出了 Stigler 措辞中的一个关键歧义：

> ”不承担”——这是指**当前**成本，还是包括**已经沉没的**成本？如果在位者已经承担了一项沉没成本，进入者仍然必须现在承担它。这是**差异性**的吗？

这个歧义是后续大部分争论的枢纽。如果沉没成本算作不对称成本，那么规模经济**确实**构成壁垒——因为在位者在进入者必须建设工厂的时候已经在运营了。如果沉没成本不算（因为在位者也曾经历过它们），那么 Stigler 的排除成立。

5.3.8　Demsetz (1982)：Nirvana 谬误应用于壁垒

【进阶】

Demsetz, H. (1982). “Barriers to Entry.” *American Economic Review*, 72(1), 47–57.

Demsetz 将 Nirvana 谬误的论证（已在 第 ~4~ 章 §4.3 中介绍）扩展到壁垒：

表 5.3: 进入壁垒的七个定义（基于 Gilbert 1989 扩展）

作者	年份	定义	关键贡献
Bain	1956	在位者能够持续提价而不吸引进入的任何优势	基于效果的定义；三种壁垒来源
Stigler	1968	进入者必须承担而在位者不承担的差异生产成本	基于不对称的定义；几乎清空壁垒清单
Demsetz	1982	政府创造的进入限制才是真正的壁垒；其他是信息/偏好的合法特征	Nirvana 谬误应用于壁垒
von Weizsäcker	1980	从社会角度看，进入壁垒是**对社会含意的**进入的限制	引入福利标准——不是所有进入都含意
Gilbert	1989	进入壁垒是" 在位者赚取的租金，而具有相同生产技术的潜在进入者无法复制"	综合 Bain 和 Stigler ——聚焦于**租金的不可复制性**
Carlton & Perloff	2005	进入壁垒是" 阻止企业家立即创建一个新的、有利可图的企业长期因素"	强调**长期**——短期障碍不是壁垒
McAfee, Mialon & Williams	2004	区分" 经济壁垒"（从社会角度看不应存在）和" 反垄断壁垒"（从竞争者角度看不应存在）	分类学清理；双重分类

> 将广告和品牌忠诚度称为" 进入壁垒"，隐含地假设正确的基准是一个信息无成本、产品同质的完美竞争世界——这个世界不存在，也不可能存在。

如果消费者真正重视品牌——因为品牌传递关于质量的信息、降低搜索成本、或直接提供消费效用——那么品牌忠诚度是合法的消费者偏好，而非市场失灵。投资建立品牌的在位者承担了真实成本；进入者**在原则上**没有面临成本劣势，因为他们也可以投资于品牌建设。

Demsetz 确实接受为真正壁垒的：政府创造的壁垒（许可证、关税、排他性特许经营权）。这些满足 Stigler 的不对称要求，且无法基于效率进行辩护。

5.4 Demsetz 效率假说：从批判到世界观

5.4.1 在 Ch.4 基础上构建

【核心】

第 ~4~ 章 §4.3 将 Demsetz (1973) 呈现为对 SCP 的方法论批判。此处将其展开为芝加哥世界观的一个正面组成部分。

核心思想回顾：

> 效率异质性 → 高效率企业获得高市场份额 + 高利润 → 可观测的集中度-利润相关性。

集中度与利润之间的关系不是因果的——它们是效率这一共同原因的两个结果。本章不再重复技术细节（Demsetz 两组对比、数值例子等），而是聚焦于**这个论证对政策的含义**。

如果 Demsetz 是对的：

- 将高集中度行业拆分为较小企业 → 摧毁效率优势 → 消费者价格上升，而非下降
- 针对" 垄断利润" 的反垄断干预 → 惩罚成功者 → 降低企业投资于效率创新的激励
- 观察到的集中度 → 不是问题的信号，而是效率差异的信号

5.4.2 "Nirvana 谬误"：芝加哥的方法论签名

【进阶】

Demsetz, H. (1969). "Information and Efficiency: Another Viewpoint." ***Journal of Law and Economics*****, 12(1), 1–22.**

Demsetz 直接回应了 Kenneth Arrow 对信息市场的分析，其中 Arrow 论证了固有的市场失灵问题（可占有性、不可分割性、不确定性）为政府干预提供了理由。Demsetz 识别出了 **“Nirvana 谬误”**：

> 将现实世界的市场与一个不可能的完美政府政策理想进行比较，而非与其他现实世界的制度安排进行比较。

三个子谬误：

1. **”邻家草地总是更绿”谬误**：假设政府行动会解决一个感知到的市场问题，而不证明政府替代方案在实践中会更好
2. **”免费午餐”谬误**：忽视实施和管理提议替代方案的真实资源成本
3. **”人非其人”谬误**：假设限制市场表现的行为现实（道德风险、自利、风险规避）在政府安排下会奇迹般地消失

Nirvana 谬误的方法论处方

Demsetz 的核心方法论处方是：**比较制度分析**——将现实世界市场与现实世界替代方案进行比较，而非与不可达到的涅槃进行比较。
这不仅是 IO 的方法论——它是制度经济学的核心原则。当你看到一篇论文说”市场在 X 方面失灵了”，问：

1. 与什么相比？
2. 政府替代方案在实践中的表现如何？
3. 提议的方案是否经过了与市场相同的严格审视？

5.4.3　Demsetz 论产权与效率

【进阶】

Demsetz, H. (1967). “Toward a Theory of Property Rights.” *American Economic Review*, 57(2), 347–359.

产权的发展是为了**内部化外部性**——当内部化的收益变得大于内部化的成本时。由技术和相对价格的变化驱动。

著名经验例证——魁北克的蒙塔格奈印第安人：

- 在欧洲毛皮贸易之前：共有土地上的公共狩猎——海狸相对于人口是丰富的
- 欧洲需求急剧提高了海狸毛皮的价值后：公地的外部性（过度狩猎）变得严重

- 蒙塔格奈人的回应：发展私有家庭狩猎领地——包括对 trespass 的强制执行、轮换狩猎和继承
- **解释**：私有财产出现，因为内部化的收益（保护、养护）现在超过了界定和执行边界的成本

与 IO 的连接：同样的产权逻辑支撑着芝加哥对" 壁垒" 的处理——在产权清晰、交易成本低的情况下，市场会自我修正。当存在明显的" 市场势力" 问题时，芝加哥经济学家会问：(1) 是否存在未解决的产权问题？(2) 是否存在政府制造的进入限制？

5.5　Posner 与 Bork：反垄断的重塑

5.5.1　Posner (1976)：反垄断的经济学重构

【核心】

Posner, R.A. (1976). *Antitrust Law: An Economic Perspective*. University of Chicago Press.

Posner 书名的副标题——*An Economic Perspective*——是经过深思熟虑的挑衅：暗示存在其他" 视角" 需要被取代。核心论点：

A. 经济效率作为反垄断的唯一目标。反垄断不应保护小企业、不应分散经济权力、不应追求公平——它应仅最大化经济效率（以消费者福利度量）。

B. 激进简化。 Posner 提议用" 一个简单的对不合理地限制竞争的明示或默示协议的禁止" 来取代反垄断复杂的、多种罪名的教义架构。

C. 三重优先序。只有三个问题值得认真的执法关注：

(a) 明示横向价格固定卡特尔

(b) 对垄断的大规模横向合并

(c) 政府创造的进入壁垒

其他一切——纵向约束、搭售、排他性交易、掠夺性定价主张、价格歧视、小型横向合并、以及 Robinson-Patman 法案——应被普遍视为合法和有益的。

为什么掠夺性定价几乎不存在：Posner 的逻辑链——

1. 掠夺者遭受比猎物更大的损失（掠夺者更大，以低于成本的价格销售更多单位）
2. 即使猎物被驱逐，回收需要阻止新竞争者进入的壁垒——壁垒是罕见的（在芝加哥的定义下）

3. 大多数” 掠夺性定价” 投诉是由无法匹敌更高效竞争对手价格的低效率竞争者提起的

4. 结论：掠夺性定价是一种非理性的策略，理性企业不会采用

5.5.2 Posner (2001) 第二版：什么被改变、什么没有

【进阶】

- 副标题”An Economic Perspective” 被删除——到 2001 年，不再有” 替代视角” 的任何暗示

- 新第 8 章：” 排他性实践（II）：新经济”——将芝加哥框架应用于软件、互联网服务提供商和通信设备

- 25 年的判例法整合：*Brown Shoe*、*Von's Grocery*、*Schwinn*、*Utah Pie* 被斥为” 路边警示性的尸骸”

- 博弈论被承认但被打发——” 产生的含义与 Stigler 时代的分析并无不同”

5.5.3 Posner 作为第七巡回法院法官：关键判例

【核心】

- ***Blue Cross & Blue Shield v. Marshfield Clinic* (1995)**：驳回了针对一家大型医生所有诊所的几乎所有垄断化主张

- ***In re Text Messaging Antitrust Litigation* (2015)**：” 跟随领导者” 定价（有意识平行行为）不违反《谢尔曼法》——只有明示合谋才违反

- ***Joe Sanfelippo Cabs v. City of Milwaukee* (2016)**：驳回了出租车牌照持有者对 Uber/Lyft 进入的挑战。Posner 的著名判词：

> ” 出租车牌照仅赋予运营出租车的权利……它不赋予成为寡头垄断者的权利。”
>
> ” 如果旧的经济主体被认为享有阻止新的经济主体进入旧的市场——经济进步可能会戛然而止。替代出租车的可能是马和马车；替代电话的可能是电报；替代计算机的可能是算尺。”

Posner 与芝加哥的”创造性破坏”精神

Posner 在 *Sanfelippo* 案中的判词捕捉到了芝加哥学派最深刻的本能：**对”创造性破坏”的信念——新的、更好的企业取代旧的企业的过程不仅是可接受的，而且是含意的**。保护在位者免受竞争——即使是来自颠覆性新进入者的竞争——正是反垄断法被设计来**阻止**的事情。

5.5.4　Bork (1978)：《反垄断悖论》

【核心】

Bork, R.H. (1978). ***The Antitrust Paradox: A Policy at War with Itself.*** **Basic Books.**

核心论点：反垄断执法已经变得反竞争——通过保护低效率的竞争对手免受更高效企业的竞争，人为地提高了价格。”悖论”在于：一部表面上旨在促进竞争的法律，正被用来压制竞争。

Bork 的立法意图论证：

”国会意图法院只实施今天我们将称之为消费者福利的那种价值。”

对小商人的保护被赋予”仅具有补充性而非冲突性的作用。”

关键——现代学术共识认为 Bork 的解读是错误的：

- ”消费者福利”这一短语在 1890 年不存在，现代福利经济学也不存在
- Robert Lande (1982) 证明了主要关切是防止不公正的财富从消费者转移到垄断者——而非配置效率
- Bork 自己也用”消费者福利”来指代**配置效率（总剩余）**，而非消费者剩余——这是一个至关重要的滑移

Bork 对消费者福利的模棱两可的使用：

- Bork 的”消费者福利” = **总福利**（消费者剩余 + 生产者剩余）= 配置效率
- 实际消费者福利标准 = 仅消费者剩余 → 任何对消费者的伤害超过生产者收益
- Hovenkamp：”Bork 将他的标准称为‘消费者福利’标准，但随后将其等同于一般福利”
- 在实践中，法院已经应用了更接近真正消费者福利（价格）检验的方法

术语陷阱：Bork 的” 消费者福利”

这是反垄断史上最重要的术语滑移之一。Bork **说**” 消费者福利”，但他**指的是**总福利（配置效率）。在一个配置效率标准下，如果生产者收益超过消费者损失，一项价格上涨的合并可能是合法的。在一个消费者福利标准下则不会。当阅读 Bork 时——以及当法院引用” 消费者福利标准” 时——总是问：他们是指**总剩余**还是**消费者剩余**？答案往往决定了合并是被通过还是被阻止。

Bork 的具体政策建议：

表 5.4: Bork (1978) 对各类反垄断实践的政策建议

实践	Bork 的建议
横向价格固定/市场分割（卡特尔）	保留当然违法
对垄断的大规模横向合并	保留禁止
纵向合并	应该合法
综合合并	应该合法
RPM（纵向价格固定）	应该合法
纵向地域/客户限制	应该合法
排他性交易/需求合同	应该合法
掠夺性定价	应该合法
搭售安排	应该合法
垄断化（Sherman 法案第 2 条）	应该被取消或严重缩减
Robinson-Patman 法案	应该被废除——” 无法容忍的草案起草与完全错误的经济理论的畸形后代”

Donald Dewey (1984) 总结道：” 在反垄断法现在包含的许多禁令中，Bork 只想保留两条——对卡特尔和大规模横向合并的禁令。”

5.5.5　关键芝加哥反垄断概念

【核心】

消费者福利标准 (Consumer Welfare Standard)

Bork 将该标准推向了前台。最高法院在 *Reiter v. Sonotone* (1979) 中采纳了他的原话：Sherman 法案是一部” 消费者福利处方”。反垄断从一项多目标事业（分散经济权力、保护小企业、维护民主价值、保护消费者）变为一个单一指标：消费者价格和产出。

合理原则的扩张

表 5.5: 从 *Dr. Miles* 到 *Leegin*：纵向约束的司法采纳时间线

年份	案例	裁决	芝加哥连接
1911	*Dr. Miles Medical Co.*	最低 RPM 当然违法	前芝加哥时代
1960	——	Telser 论文发表	免费搭车者效率理论进入文献
1977	*Continental T.V. v. GTE Sylvania*	非价格纵向约束 → 合理原则（推翻 *Schwinn* 1967）	Posner 的影响；法院开始引用芝加哥论证
1986	*Matsushita v. Zenith*	掠夺性定价主张需"更具说服力的证据"	McGee (1958) + Posner (1976) 的掠夺性定价怀疑论
1993	*Brooke Group v. Brown & Williamson*	两步检验：低于成本 + 回收的危险概率	自此后没有原告在联邦法院赢得掠夺性定价案件
1997	*State Oil v. Khan*	最高 RPM → 合理原则	Posner 撰写第七巡回法院意见
2007	*Leegin v. PSKS*	最低 RPM → 合理原则（以 5-4 推翻 *Dr. Miles*）	Telser → Posner → Bork 智识链的最终胜利

一个重要的细微差别：芝加哥学者并不一致赞成逐案的合理原则。他们倾向于**事前将经济学纳入规则、推定和结构化测试**——而非事后逐案平衡。这是由对法律确定性和可管理性的关切驱动的——形式法治的理想。

Edmund W. Kitch 描述了芝加哥学者偏好的轨迹：" 从当然违法到合理原则下的推定合法，再到接近当然合法。"

错误成本框架——Easterbrook (1984)

【核心】

Easterbrook, F.H. (1984). "The Limits of Antitrust." ***Texas Law Review*****, 63, 1–40.**

Easterbrook 的论证是芝加哥反垄断的智识巅峰。核心主张：

假阳性（错误地谴责有益实践）比假阴性（允许有害实践通过）更可怕。

三个机制：

1. **市场自我修正假阴性**：卡特尔崩溃（成员有激励背叛）；垄断利润吸引进入；即使法院做不了什么，竞争也会侵蚀市场势力
2. **假阳性通过 stare decisis 持久存在**：一旦法院宣告一种实践非法，该实践被永久放弃——即使后来证明它是促进竞争的。法院的先前判例不会被市场过程"修正"
3. **效率收益影响所有产出，垄断损失只影响被削减的产出**：一项效率创新惠及所有购买该产品的消费者。垄断的福利三角仅涉及价格提高所排斥的消费者。效率收益的范围可能比垄断损失更广

Easterbrook 的五重筛选：在法院谴责一项实践为反竞争之前，它应通过以下筛选：

1. 被告是否拥有市场势力？
2. 该实践是否削减了产出？
3. 是否存在看似合理的效率合理性？
4. 排除意图证据（意图是模棱两可的——意图"击败竞争对手"可能是竞争的，也可能是掠夺性的）
5. 消费者是否受到了损害？

错误成本框架的持久影响

Easterbrook 的文章可能是现代反垄断中最具影响力的单篇法律评论文章。它的框架——假阳性 vs. 假阴性的相对成本——不仅塑造了反垄断，还塑造了更广泛的规制分析。当 FTC 或 DOJ 评估一项合并时，他们隐式地或显式地在问：**阻止一项促进竞争的合并（假阳性）与允许一项反竞争的合并（假阴性），哪种错误更昂贵？**

芝加哥的答案是明确的：假阳性更昂贵。但后芝加哥和新布兰代斯学派挑战了这种不对称性——见 §~5.8。

5.6　芝加哥方法论武器：Friedman、Coase 与 Telser

5.6.1　Friedman (1953) 与"好像"方法论

【核心】

Friedman, M. (1953). "The Methodology of Positive Economics." In *Essays in Positive Economics*. University of Chicago Press.

核心论点：科学假说应仅由其**预测力**来判断，而非其假设的描述性现实主义。

> ” 真正重要和显著的假说将被发现具有 ‘假设’——这些假设是对现实的极其不准确描述性表征，并且，一般来说，理论越重要，假设（在这个意义上）就越不现实。”

” 好像” 教条：企业不需要实际计算边际成本和边际收益——如果它们的行动**好像**正在最大化利润，那就是相关的检验。Friedman 将询问商人如何做决策比作” 通过询问八旬老人如何解释长寿来检验关于长寿的理论。”

芝加哥 IO 中的应用：完全竞争的假设（许多小企业、同质产品、完全信息）是不现实的——但这不重要。重要的是这些模型是否比更” 现实” 的寡头模型预测得更好。

Friedman 的方法论与芝加哥对 SCP 的批判之间的联系

Friedman 的预测主义方法论与芝加哥对 SCP 的批判之间存在一个未被充分认识的联系。SCP 的” 结构 → 行为 → 绩效” 链条——以其对行业集中度、进入壁垒和其他结构变量的详细分类——恰恰是试图让假设**更加**描述现实。芝加哥的回应：我们不需要” 现实的” 结构模型。具有完全竞争假设的简单价格理论预测得更好。

更一般地说：博弈论革命（第 ~7~ 章）削弱了 Friedman 方法论中预测主义的部分——博弈论中的均衡对假设极其敏感，因此假设的” 现实性” 确实重要。这就是为什么芝加哥首先拒绝了博弈论——它不仅是不必要的复杂性，而且它的非稳健性与预测主义方法论相矛盾。

5.6.2　Coase 的连接

【核心】

Coase (1937)：“企业的性质”

为什么企业存在于市场经济中？**交易成本**——使用市场的成本超出商品本身的价格：搜索和信息成本、讨价还价和谈判成本、监督和执行合同的成本。企业存在是因为它们通过行政指导在内部协调生产来节省交易成本。企业的边界是内部组织的边际成本等于使用市场的边际成本之处。

与 IO 的连接：Coase 对交易成本的分析暗示，纵向整合可能具有效率合理性——而非市场势力驱动的。当企业” 制造而非购买” 时，这可能不是排斥竞争——而是节省交易成本。

Coase (1960)：“关于社会成本的问题” 与 Coase 定理

Journal of Law and Economics, 3, 1–44——历史上被引用最多的法律评论文章。

著名的晚宴故事：1959 年，Coase 在" 联邦通信委员会" 一文中论证广播频谱应被拍卖而非政府分配。芝加哥经济学家（由 Friedman 和 Stigler 领导）认为他犯了一个错误。Aaron Director 在他家举办了一场晚宴。约二十位资深经济学家出席。

Stigler 后来叙述道：

> "Milton Friedman 承担了大部分的发言，如常。他也承担了大部分的思考，如常。在长达两个小时的争论过程中，投票从 21 票对 Coase 不利、1 票对 Coase 有利，变为 21 票对 Coase 有利。多么令人振奋的事件！我后来哀叹我们没有那个远见来录下它。"

Coase 自己说：" 如果不是芝加哥大学这些经济学家认为我在 '联邦通信委员会' 的文章中犯了一个错误，'关于社会成本的问题' 很可能永远不会被写出来。"

由 Stigler 命名的 Coase 定理：在交易成本为零的世界中，无论最初的权利分配如何，私人谈判会导致有效配置。

与 IO 的连接：Coase 定理暗示许多感知到的" 市场势力" 问题实际上是产权界定或交易成本的问题。如果频谱的产权被清晰界定（而非由政府任意分配），Coase 的晚宴论证意味着市场会将其交易至最高价值的使用。芝加哥的处方：在呼吁反垄断干预之前，先问——**产权是否清晰？交易成本是否低？**

Coase (1972)：“耐久性与垄断”——Coase 猜想

Coase, R.H. (1972). “Durability and Monopoly.” *Journal of Law and Economics*, 15(1), 143–149.

惊人的猜想：如果无法承诺未来价格，一个销售耐用品（如汽车、机器）的垄断者没有任何垄断势力。

逻辑链：

1. 耐用品垄断者最初收取高价以捕捉高价值消费者

2. 那些消费者购买后，垄断者有激励降价以捕捉低价值消费者

3. 理性的、向前看的消费者预期到未来的降价并延迟购买

4. 预期到这一点，垄断者被迫立即降价

5. 最终，价格崩溃到边际成本——垄断者与其**未来的自己**竞争

后来由 Gul、Sonnenschein 和 Wilson (1986) 正式证明。

Coase 猜想与芝加哥世界观

Coase 猜想是芝加哥" 垄断是暂时的" 这一信念的极端版本。它说：**即使一个企业是市场中唯一的生产者——即使没有进入——如果它不能承诺不降价，它的垄断势力也是零。**

芝加哥学派中有一条共同的线索：市场势力比表面看起来更难维持。卡特尔成员背叛。进入侵蚀利润。Coase 猜想表明，即使没有进入，承诺问题也可能剥夺垄断者所有的市场势力。

5.6.3　Telser (1960) 与纵向约束的免费搭车者逻辑

【核心】

Telser, L.G. (1960). "Why Should Manufacturers Want Fair Trade?" *Journal of Law and Economics*, 3, 86–105.

谜题：为什么制造商会想要**阻止**零售商对他们的产品打折？乍一看，较低的零售价格应当增加销售数量并使制造商受益。然而许多制造商激烈地执行" 公平贸易"（最低转售价格）协议。

免费搭车者解释：

1. 某些零售商提供昂贵的" 特殊服务"——广告、知识渊博的销售人员、维修服务、引人入胜的展厅、售前演示

2. 消费者可以访问提供全面服务的零售商以获取这些服务，然后从价格更低但不提供服务的折扣零售商处购买

3. 折扣零售商免费搭车了提供服务零售商的投入

4. 最终，提供服务的零售商无法回收成本并停止提供服务

5. 结果：次优服务水平和制造商产品总需求降低

RPM 解决了这个问题：通过执行最低转售价格，制造商保证了一个足够高的利润以补偿零售商提供服务，阻止折扣免费搭车者削价。

Telser → Leegin：一个知识链的解剖

Telser (1960) 的免费搭车者论证是现代反垄断中最具影响力的智识谱系之一：

Telser (1960) → Posner (1976) → Bork (1978) → *Sylvania* (1977) → *Leegin* (2007)

这一链条跨越了 47 年——从一篇期刊文章到一个最高法院判决。很少有经济学论证能够追溯到一个如此直接、不间断的谱系。学生应反思：**为什么这个特定的论证如此有说服力？**部分原因在于它符合芝加哥更广泛的”理性效率”叙事——纵向约束是高效的，除非证明并非如此。

5.7　政策的芝加哥化：里根革命与放松管制

5.7.1　里根政府的反垄断革命

【核心】

1981 年里根上台，带来了反垄断执法史上最急剧的方向转变。

William Baxter (DOJ 反垄断局，1981–1984)：

- **撤销 IBM 案（1982 年 1 月）**：结束了长达 13 年的垄断化诉讼。Baxter 的理由：IBM 在计算领域的地位已经被削弱——案件没有意义。这在 DOJ 内部产生了激烈的争议。

- **AT&T 解体（1982 年同意判决）**：矛盾的是，Baxter 对 AT&T 诉讼采取了激进的推进态度（”打到眼球里”）。修改后的最终判决将 AT&T 的长途业务与七家地区贝尔运营公司（Baby Bells）分离。长途费率从每分钟约 50 美分降至 1994 年的低于 20 美分。

James Miller (FTC，1981–1985)：

第一个担任 FTC 主席的经济学博士。回忆录：《*作为改革者的经济学家：重振 FTC，1981–1985*》。

- FTC 预算被削减三分之一；人员减半
- 将执法限制于硬核违规行为（横向价格固定、串通投标）
- 几乎停止审查消费者广告
- 在 1980 年代的并购浪潮中，只阻止了数千起公司合并中的极少数

5.7.2　1982 年合并指南

【核心】

1982 年合并指南是芝加哥最大的制度性胜利。它从根本上重新定义了合并审查的分析框架。

表 5.6: 1968 年 vs. 1982 年合并指南对比

维度	1968 年指南	1982 年指南
集中度指标	CR4	HHI
推定门槛	CR4 $\geq$ 75%（在许多情况下）	HHI $>$ 1,800（合并后）；$\Delta > 100$
进入壁垒	宽泛地解释（Bain 式）	狭义地解释（Stigler 式）
效率抗辩	未获承认	获承认（在 1984 年修订中明确化）
纵向合并	推定可疑	普遍不予挑战
市场界定	相对僵硬（基于产品/地理界限）	SSNIP 检验（基于弹性的）
”安全港”	非常狭窄	HHI $<$ 1,000 非集中（安全）
分析目标	市场结构	实际价格效应
经济学角色	辅助性（法律主导）	核心（经济学主导）

HHI 取代 CR4 的制度逻辑

为什么芝加哥推动 HHI 取代 CR4？表面上是技术性的：HHI 是更好的集中度度量——它既反映企业数量也反映不平等程度。但存在更深层的原因：

CR4 只关注前四大企业——忽略了行业内其他企业的分布。**HHI** 对大型和小型企业之间的不对称给予权重，因此更不容易仅仅因为行业中存在几家大企业就标记问题。在芝加哥框架下，如果市场中有大量小竞争者，少数大企业的集中可能完全良性。

从 CR4 到 HHI 的转变不是纯粹的技术改进——它是对”只关注最大企业”的结构主义方法的一个微妙但重要的纠正。

5.7.3　放松管制运动

【核心】

航空放松管制的标志性意义：Fred Kahn 的 CAB 在 1985 年被废除——他解散了自己的机构。这是放松管制智识的象征性巅峰：一个监管者得出结论，他的机构正在伤害而非帮助消费者。

表 5.7: 主要放松管制立法

立法	年份	受影响的机构	芝加哥连接与结果
航空放松管制法案	1978	CAB 被废除（1985）	Alfred Kahn（Cornell）；Stigler 的俘获理论作为智识正当性。实际票价下降约三分之一；乘客数量翻倍
汽车承运人法案	1980	ICC 放松管制	到 1985 年，放松管制每年为托运人节省 $78 亿
电信法案	1996	FCC 重组	部分/被扭曲的芝加哥应用（不是真正的放松管制——是重新管制）

5.7.4　四代合并指南：一个世纪的演变

5.8　胜利的限度：后芝加哥综合与当前的挑战

5.8.1　芝加哥的胜利有多完整？

【核心】

在联邦反垄断执法和最高法院判例中——近乎全面：

到 1990 年，消费者福利标准成为了最高法院判例中的" 统一主题"。合理原则急剧扩张。效率抗辩成为合并审查的标准。联邦执法预算和人员被大幅削减。

在学术 IO 中——部分的：

博弈论革命（第 ~7~ 章）**并非**芝加哥——它接受了芝加哥对 SCP 的批判，但拒绝了芝加哥对简单价格理论的依赖。NEIO（第 ~10~ 章）发展了结构估计方法，可以直接估计行为参数——而非采纳芝加哥的" 市场效率推定"。

这意味着 1990 年以后的 IO 处于一个奇特的位置：**芝加哥在政策中占据霸权，但在前沿学术中已不再如此。**

5.8.2　后芝加哥综合的关键思想

【进阶】

后芝加哥综合始于一个简单的前提：芝加哥对竞争过程的乐观在大多数市场中是合理的，但在某些市场中——尤其是那些具有不完全信息、战略互补、或强转换成本的市场——可能不成立。

核心人物和贡献：

- **Steven Salop**（乔治城法学院）：提高竞争对手成本的理论

表 5.8: 四代合并指南对比：1968 → 1982 → 2010 → 2023

维度	**1968（哈佛时代）**	**1982（芝加哥革命）**	**2010（后芝加哥）**	**2023（钟摆回摆）**
集中度指标	CR4	HHI	HHI	HHI + 市场份额
高集中度门槛	CR4 $\geq 75\%$	HHI $> 1{,}800$	HHI $> 2{,}500$	HHI $> 1{,}800$
Δ 触发	未明确	$\Delta > 100$	$\Delta > 200$	$\Delta > 100$
市场份额推定	CR4 高即非法	不存在	不存在	合并份额 $> 30\%$ 推定非法
效率抗辩	否	是（1984 明确化）	是（增强）	是（但更严格）
进入分析	Bain 式宽泛	Stigler 式狭窄	及时、可能、充分	增加了网络效应/数据
安全港	极窄	HHI $< 1{,}000$	HHI $< 1{,}500$	取消
纵向合并	推定可疑	不予挑战	新指南（2020）	强化审查
新生竞争	未涉及	未涉及	初步关注	核心关注（杀区收购）
劳动力市场	未涉及	未涉及	未涉及	首次纳入
分析模式	结构性推定	效果分析	效果分析	结构性推定 + 效果分析混合

- **Salop & Scheffman (1983, 1987)**：“Raising Rivals’ Costs”——主导企业可以通过提高竞争对手的**成本**（而非削减自己的价格）来排斥它们。不要求短期利润牺牲。不要求竞争对手退出——只要求它们被迫削减产出
- **Krattenmaker & Salop (1986)**：“Anticompetitive Exclusion: Raising Rivals’ Costs to Achieve Power over Price”——将 RRC 框架从经济学转化为法律教义
- **Jonathan Baker (1999)**：“Policy Watch: Developments in Antitrust Economics” (*Journal of Economic Perspectives*)——后芝加哥共识的权威陈述

后芝加哥综合的五个前沿领域（Baker 1999）：

1. **合并效率**：效率是真实的，但可能达不到芝加哥声称的规模
2. **进入条件**：进入比芝加哥宣称的更困难、更慢、更不完整
3. **便利协调的实践**：某些纵向约束（信息交换、最惠国待遇条款）可以便利合谋——即使没有明示协议
4. **排他性实践**：在某些条件下可以提高竞争对手成本（RRC）——**条件很重要**
5. **单边竞争效应**：差异化产品市场中的合并可以在没有合谋的情况下提高价格——仅仅通过消除竞争约束

5.8.3　三校对比：一个分析框架

【进阶】

5.8.4　错误成本辩论的继续

【进阶】

- **Easterbrook (1984)**：假阳性 > 假阴性（经典芝加哥立场）
- **Kaplow (2011–2013)**：为证明负担分析提供了形式化的决策理论框架——表明最优证明标准取决于两类错误的相对成本
- **Mungan & Wright (2021–2022)**：在 Easterbrook 的实质假设下，反垄断中的最优证明标准**高于**优势证据——而非更低。这是对芝加哥立场的一个令人惊讶的正式确认
- **新布兰代斯学派的反驳**：在数字平台市场中，假阴性可能**更为**致命——杀区收购（killer acquisitions）、网络效应、数据护城河使市场势力远比芝加哥模型暗示的更持久。如果假阴性创造了自我强化的垄断，市场的” 自我修正” 机制可能被关闭

表 5.9: 哈佛 SCP vs. 芝加哥 vs. 后芝加哥 vs. 新布兰代斯

维度	哈佛 SCP	芝加哥	后芝加哥	新布兰代斯
反垄断目标	多目标：竞争 + 公平 + 分散 + 小企业	消费者福利（总剩余/配置效率）	消费者福利（扩展至交易伙伴福利）	多目标：民主、自由、分散权力、消费者价格
市场看法	结构决定一切；失灵常见	自我修正；干预通常有害	自我修正有时失灵；博弈论识别反竞争场景	市场势力持久且自我强化；结构性集中是问题
进入壁垒	广泛（Bain 三类）	狭窄（仅政府创造）	有条件地广泛（取决于信息和策略条件）	广泛（包括网络效应、数据护城河）
政策处方	结构性推定；积极干预	默认合法；仅干预卡特尔和大规模合并	具有经济推定的结构化合理原则	明线规则；结构性推定；重振执法
关键人物	Mason, Bain, Weiss	Director, Stigler, Demsetz, Posner, Bork, Easterbrook	Salop, Baker, Shapiro, Krattenmaker	Khan, Wu, Kanter, Stoller
代表文本	Bain (1956, 1959)	Bork (1978), Posner (1976)	Baker (1999), Salop & Scheffman (1983)	Khan (2017), 2023 Merger Guidelines
认识论	经验归纳主义	价格理论推演	博弈论 + 有限经验	经验观察 + 结构主义

5.8.5　2023 年合并指南：钟摆的回摆

【前沿】

2023 年 DOJ/FTC 合并指南通过以下方式部分地回归了结构主义推定：

- HHI 高度集中阈值从 2,500 降至 1,800
- ΔHHI 推定触发从 > 200 降至 > 100
- 新的 30% 合并市场份额推定
- 取消” 安全港”（HHI $< 1{,}500$ 不再自动安全）
- 对” 巩固或延伸主导地位” 的收购的新挑战
- 包括劳动力市场效应（反垄断应保护工人，而不仅仅是消费者）
- 对连续收购（roll-up acquisitions）/ 平台市场 / 新生竞争威胁的更强关注

2023 指南不是 SCP 的” 复辟”

2023 年指南不是简单地回到 1968 年。准确描述是：**它是结构性推定和效果分析的一种新的混合体**。结构变量（HHI、市场份额）重新获得了更大的权重——但经济分析（效率、进入、价格效应）仍保留为决策的最终依据。
这是三个传统的融合：

- **哈佛**：结构变量作为初始推定（”HHI 高 $\rightarrow$ 需要解释”）
- **芝加哥**：经济分析作为审查工具（” 效率是否可以证成集中？ ”）
- **后芝加哥/新布兰代斯**：对新生竞争、劳动力效应、平台动态的扩展关切

详见 第 ~15~ 章。

5.8.6　欧洲的平行道路：“更加经济学的方法”

【进阶】

欧洲的反垄断演变与美国平行但不同步：

- **与芝加哥的相似之处**：转向基于效果的分析；引入了首席竞争经济学家；提高了经济分析在决策中的作用
- **与芝加哥的不同之处**：
  - 欧盟拥有更广泛的目标（市场一体化、公平性、对经济自由的保护）——而非单一的消费者福利

  - 芝加哥的消费者福利焦点 vs. 弗莱堡学派的" 竞争作为过程" 传统
  - 欧盟法院保留了比委员会更强的秩序自由主义印记——关注竞争过程而不仅是结果

三个传统的一行总结：

- **哈佛 SCP**：结构决定一切——干预以修复结构
- **芝加哥**：效率决定一切——除非政府造成了它，否则市场不会失灵
- **弗莱堡/秩序自由主义**：国家设定了竞争的宪法规则——然后让位

5.9 本章小结

1. **芝加哥学派不仅仅是一系列对 SCP 的批判——它是一种完整的替代世界观。**哈佛问：" 市场在哪里失灵了？" 芝加哥问：" 你有什么证据？" 举证责任在主张市场失灵者，而非为市场辩护者。

2. **Aaron Director 是芝加哥 IO 的" 隐形创始人"。**他的口述传统——周四晚的晚宴研讨课——培养了一代学者（Stigler、Posner、Bork、Easterbrook），他们彻底重塑了反垄断。Ronald Coase 称自己为圣保罗之于 Director 之基督。

3. **Stigler 的进入壁垒重定义是一场定义之战，具有巨大的政策后果。**通过将定义从 Bain 的" 任何阻止进入以获取持续超额利润的东西" 改为" 仅限进入者与在位者之间的成本不对称"，Stigler 几乎清空了壁垒的概念空间——并清空了反垄断干预的范围。规模经济、资本需求和产品差异化一度不再是" 壁垒"。

4. **Posner 和 Bork 将反垄断简化为一个单一问题：这种行为是增加了还是减少了消费者福利？** Bork 的《反垄断悖论》（1978 年）是将芝加哥经济学转化为法律教义的关键文本。最高法院在 1977 年至 2007 年间通过一系列案件中的引用和采纳，将芝加哥框架奉为圭臬。

5. **芝加哥的政策胜利是迅速的且近乎全面的。** 1982 年合并指南牢固地确立了基于经济学的（而非基于结构的）合并审查。放松管制运动（航空、卡车运输、电信）拆除了整个监管机构。到 1990 年，消费者福利标准在美国反垄断中占据霸权地位。

6. **但芝加哥的学术胜利是不完整的。**博弈论革命（第 ~7~ 章）既不是哈佛也不是芝加哥。NEIO（第 ~10~ 章）接受了芝加哥对更好经验方法的需求，但提供了结构估计而非简单价格理论。后芝加哥综合（Salop、Baker、Shapiro）表明，**在某些条件下**——不完全信息、战略互补性、提高竞争对手成本——芝加哥关于效率的简单故事可能会崩溃。

7. **芝加哥遗产的矛盾核心：**它同时是 IO 历史上最成功和最被错误理解的思想运动。它赢得了政策战争——但在此过程中，它的立场的极端性（”没有掠夺性定价”、”没有进入壁垒”、”政府是唯一的垄断来源”）被后来的学术研究所削弱。2023 年合并指南和当前新布兰代斯学派的挑战代表了钟摆远离芝加哥方向的最显著摆动。这个摆动可能有多大、可能持续多久——详见 第 ~15~ 章。

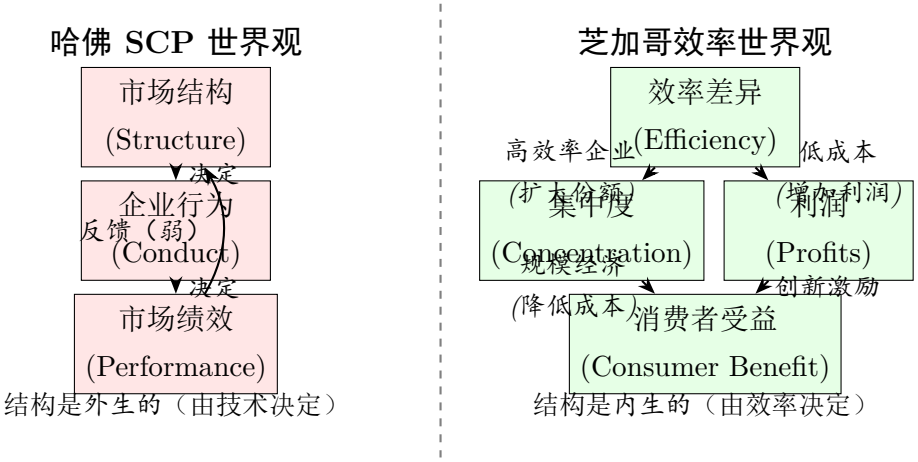


图 5.1: 哈佛 SCP 与芝加哥效率的世界观对比。在哈佛框架中，结构外生地决定行为和绩效——因果箭头从左到右。在芝加哥框架中，效率是深层原因——集中度和利润都是其结果。结构是内生的。

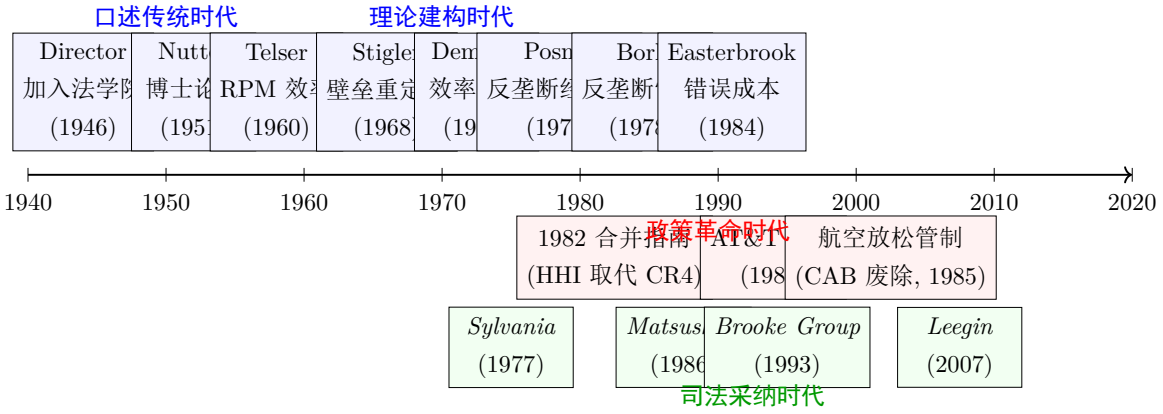


图 5.2: 芝加哥反垄断的智识谱系：从 Director 的口述传统到最高法院的采纳。注意理论建构与政策实施之间的时间差：Stigler (1968) → 1982 年指南（14 年）；Bork (1978) → *Leegin* (2007)（29 年）。

5.10 习题

基础

5.1. Bain vs. Stigler 关于进入壁垒的定义。

(a) 分别陈述 Bain (1956) 和 Stigler (1968) 对进入壁垒的定义。

(b) 在 Stigler 的定义下，Bain 三类壁垒（规模经济、产品差异化、绝对成本优势）中哪些仍算作壁垒？哪些不算？为什么？

(c) 举出一个在 Bain 的定义下算作壁垒但在 Stigler 的定义下不算的具体例子（一个真实行业或一个构造的数值例子），并解释差异。

5.2. 消费者福利标准。

(a) Bork 所说的" 消费者福利" 是什么意思？它与你作为消费者的直观理解有何不同？

(b) 假设一项合并导致价格上升 5%（消费者损害），但产生足以使合并后成本降低 10% 的效率收益。在 Bork 的总剩余标准下，这项合并是否应该被批准？在一个只考虑消费者价格的标准下呢？

(c) 为什么这种区分（总剩余 vs. 消费者剩余）在实践中很重要？

5.3. 一个垄断利润定理。

(a) 用你自己的语言（不用数学！）解释为什么一个在 widget 上拥有垄断地位的企业不能通过将 widget 的销售搭售一个竞争性供应的配件来增加其总利润。

(b) 如果搭售不能增加垄断利润，Director 给出了什么解释来说明为什么真实世界中的企业会这样做？

(c) 想出一个 Director 的解释成立的真实世界搭售例子，以及一个可能需要不同解释的例子。

中级

5.4. Easterbrook 的错误成本框架。

(a) Easterbrook (1984) 为什么论证假阳性（错误地谴责有益实践）比假阴性（允许有害实践通过）更昂贵？给出他提供的两个机制。

(b) 在什么条件下——如果有的话——Easterbrook 的不对称性可能会被逆转？考虑一个具有强网络效应和高转换成本的市场。

(c) 将 Easterbrook 的框架应用于一项提议的政策：所有 HHI > 1,800 的合并应被推定非法。基于错误成本的理由——赞成和反对该政策——进行论证。

5.5. Stigler 的幸存者原则。下表给出了三个规模类别的企业在 10 年间的市场份额数据（占行业总产出的百分比）：

表 5.10: 按规模类别划分的市场份额演变

规模类别（员工数）	1970	1975	1980
小型（1–99）	30%	22%	15%
中型（100–499）	45%	50%	55%
大型（500+）	25%	28%	30%

(a) 根据幸存者原则，确定 MES 范围。

(b) 解释为什么幸存者原则体现了芝加哥方法论（相对于 Bain 的工程方法）。

(c) 讨论该技术的局限性——在什么条件下最小有效规模可能**不会**通过幸存者原则揭示？

5.6. Coase 猜想与芝加哥世界观。

(a) Coase 猜想指出，如果无法承诺未来价格，一个耐用品垄断者被迫按边际成本定价。用你自己的语言解释逻辑。

(b) Coase 猜想与芝加哥学派更广泛的" 垄断是暂时的" 论题有何关联？它是一个正式化、还是一个反例、还是无关的？

(c) 现实世界中的企业使用什么策略来逃避 Coase 猜想？为每种策略给出一个真实世界的例子。

进阶

5.7. 芝加哥 vs. 后芝加哥关于纵向约束的观点。 Telser (1960) 提供了 RPM 的免费搭车者效率论证。后芝加哥经济学家（如 Salop）同意免费搭车是真实的，但论证在某些条件下——尤其是当 RPM 由具有市场势力的制造商或零售商联盟强加时——它可能反竞争。

(a) 重构 Telser 的效率论证。

(b) 在什么市场结构下，RPM 可能从促进竞争转变为反竞争？

(c) 在设计 RPM 的法律规则时，法院应如何在错误地谴责效率性 RPM 的风险与允许反竞争 RPM 的风险之间取得平衡？在判决中引用至少一个最高法院案例（*Dr. Miles*、*Sylvania* 或 *Leegin*）。

5.8. 平台市场中的进入壁垒。考虑一个具有强网络效应的数字平台（如拼车或社交媒体）。

(a) 在 Stigler (1968) 的定义下，网络效应是否构成进入壁垒？（提示：在位者的用户基础是进入者面临的差异成本吗？在位者是否也曾不得不从头建立用户基础?）

(b) 你的答案是否取决于市场是" 倾覆的"（winner-take-all）还是可以支撑多个平台？

(c) 2023 年合并指南将网络效应和" 数据优势" 列为潜在进入壁垒。这在 Bain-Stigler 争论的谱系中处于什么位置？为你的立场辩护。

5.9. 芝加哥学派的兴起与（可能的）衰落。从 1960 年到 2023 年追踪反垄断政策的主导范式：

(a) 1960 年代的结构主义共识是什么？什么证据和论证导致了它在 1970 年代的瓦解？(第 ~3~ 章–第 ~4~ 章)

(b) 芝加哥共识在 1980 年代取代了它。芝加哥的理论和经验论证中哪一个最有力地推动了这一转变？

(c) 2023 年合并指南是否标志着芝加哥共识的终结？论证" 是" 和" 否"——然后采取一个立场。

5.11 延伸阅读

★ 必读（本章核心参考）

- **Stigler, G.J. (1968).** *The Organization of Industry.* Ch.6 "Barriers to Entry, Economies of Scale, and Firm Size." Irwin. ——定义了进入壁垒争论的文本。注意 Stigler 是如何从 Bain 的具体分类过渡到基于不对称成本的一般定义的。这是一位价格理论大师在工作——每个定义都服务于一个论证目的。

- **Demsetz, H. (1973).** "Industry Structure, Market Rivalry, and Public Policy." *Journal of Law and Economics*, 16(1), 1–9. ——效率假说。如果你在 第 ~4~ 章只读了总结，在这里读原文。三页纸，永久性地改变了一个领域的因果叙事。

- **Posner, R.A. (1976).** *Antitrust Law: An Economic Perspective.* University of Chicago Press. Ch.1–2, 7–8. ——芝加哥关于反垄断应该是什么样子的愿景。Posner 的散文风格清晰得残酷——论证是逐步收紧的，不给异议留空间。

- **Bork, R.H. (1978).** *The Antitrust Paradox: A Policy at War with Itself.* Basic Books. Ch.1–3, 21. ——改变了反垄断法律教义的书。读它提出的问题和它回避的问题。中译本：《反垄断悖论》，商务印书馆。

★★ 进阶（加深理解）

- **Demsetz, H. (1969).** "Information and Efficiency: Another Viewpoint." *Journal of Law and Economics*, 12(1), 1–22. ——Nirvana 谬误。这是 Demsetz 对比较制度分析的最全面阐述——这不是微调，这是他整个研究纲领的哲学基础。

- **Easterbrook, F.H. (1984).** "The Limits of Antitrust." *Texas Law Review*, 63, 1–40. ——错误成本框架。也许是现代反垄断中最具影响力的单篇法律评论文章。五个筛选为反垄断分析提供了一个可操作的结构化框架。

- **Telser, L.G. (1960).** "Why Should Manufacturers Want Fair Trade?" *Journal of Law and Economics*, 3, 86–105. ——免费搭车者逻辑的奠基性陈述。在 *Leegin* (2007) 中仍被引用。一篇期刊文章在 47 年后成为最高法院判决的核心——这是经济学影响法律的范本。

- **McGee, J.S. (1958).** "Predatory Price Cutting: The Standard Oil (N.J.) Case." *Journal of Law and Economics*, 1, 137–169. ——表明 Standard Oil 并未从事系统性掠夺性定价。被最高法院在 *Matsushita* (1986) 和 *Brooke Group* (1993) 中引用。（注：Dalton & Esposito (2007) 后来挑战了 McGee 的结论，发现了" 相当多的掠夺性定价证据"——这种辩论本身就是一个教训。）

- **Coase, R.H. (1960).** "The Problem of Social Cost." *Journal of Law and Economics*, 3, 1–44. ——历史上被引用最多的法律评论文章。著名的晚宴故事。Coase 定理。还有 Coase **实际上**感兴趣的东西——正交易成本的世界。

★★★ 前沿（当代重新评价与当前争论）

- **Baker, J.B. (1999).** "Policy Watch: Developments in Antitrust Economics." *Journal of Economic Perspectives*, 13(1), 181–194. ——后芝加哥共识的权威陈述。在 5 个前沿领域中发生了什么：合并效率、进入条件、便利协调的实践、排他性实践、单边竞争效应。

- **Khan, L.M. (2017).** "Amazon's Antitrust Paradox." *Yale Law Journal*, 126, 564–907. ——标题是对 Bork 的有意回应。论证消费者福利标准在结构上无法认知平台市场势力。为 第 ~15~ 章的阅读做铺垫。

- **Salop, S.C. (2024).** "The 2023 Merger Guidelines: A Post-Chicago and Neo-Brandeisian Integration." *Review of Industrial Organization*, forthcoming. ——从一位后芝加哥经济学家（他自身也是芝加哥传统的一部分）的视角解读 2023 年指南。

- **Stigler Center Report (2019).** *Final Report of the Committee on Digital Platforms*. University of Chicago Booth School of Business. ——来自芝加哥自身的数字平台委员会。具有讽刺意味的是，以狭义定义的经济学家命名——该报告的分析基本上是 Bain 式的：它处理网络效应、数据优势、消费者行为偏差和" 瓶颈权力" 作为有意义的壁垒。

下一章： → 第 ~6~ 章　可竞争市场理论：芝加哥逻辑的极端版本——如果进入完全自由且退出完全无成本，垄断也会产生竞争性结果。Baumol、Panzar 和 Willig 将芝加哥的世界观推向了它的逻辑极限——而这一极限恰好暴露了芝加哥框架中隐含的紧张关系。

第六章　可竞争市场与进入威胁

章首语

1982 年，William Baumol 在 AEA 主席致辞中提出了一个惊人的命题：如果进入和退出完全自由，即使市场中**只有一家企业**——即使存在自然垄断——结果仍然是竞争性的。潜在的" 打了就跑" 进入者不需要实际进入；仅仅是进入的**威胁**就约束了在位者的定价。

这个被称为" 可竞争市场理论" 的框架，代表了 20 世纪产业组织学中最极端的亲市场命题。它将芝加哥学派的逻辑——市场通常自我修正，进入壁垒被夸大，政府干预通常是问题——推向了它的逻辑极限：**如果进入壁垒真的是零，那么垄断结构根本不需要规制。**

可竞争市场理论在 1980 年代取得了惊人的政策成功。它为航空、卡车、铁路和电信放松管制提供了关键的智识信心。然而，正是它的示范行业——航空业——随后提供了拒绝该理论最强预测的最有力证据。Borenstein (1989, 1992) 证明，在航空枢纽市场（理论预测竞争定价的地方），主导航空公司系统性地收取**更高**而非更低的票价。

本章讲述这个" 成功与失败并存" 的故事。可竞争市场理论不是一个简单的" 对" 或" 错" 的故事——它是一个关于理论如何能够既成功（作为政策催化剂和概念创新）又失败（作为经验描述）的故事。而且，这个故事完美地设立了 第 ~7~ 章的主题：为什么策略行为——在位者的承诺、威胁和阻挠——在塑造市场结果中如此重要。

6.1　核心问题

读完本章，你应该能够回答以下问题：

1. **可竞争市场理论的核心主张是什么？**如何从" 进入和退出完全自由" 这两个简单条件推导出" 即使垄断也产生竞争性结果" 的结论？" 打了就跑" 进入的机制是什么？

2. **完全可竞争性与完全竞争有什么不同？**为什么前者可以在完全竞争失效的行业（如自然垄断）中成立？

3. **沉没成本为什么是关键？**固定成本与沉没成本的区别如何成为可竞争性理论的核心概念创新——同时又是它被经验推翻的基础？

4. **可竞争性理论如何影响政策？**为什么它成为 1970–80 年代全球放松管制运动的智识基础？1982 年合并指南如何反映了可竞争性思维？

5. **经验证据如何评价可竞争性？** Borenstein 的”枢纽溢价”发现是什么？为什么航空业同时提供了放松管制的最大成功和可竞争性理论的最大失败？

6. **可竞争性理论留下了什么？**哪些概念工具被保留？理论如何奠定博弈论革命（第 ~7~ 章）的基础？

6.2　可竞争市场理论的核心框架

6.2.1　完全可竞争市场的定义

【核心】

可竞争市场理论由 **Baumol, Panzar 和 Willig (1982)** 在 *Contestable Markets and the Theory of Industry Structure* 中系统阐述，并在 **Baumol (1982)** 的 AEA 主席致辞中得到最简洁的表述。

定义 6.1 (完全可竞争市场). 完全可竞争市场（*perfectly contestable market*）是进入**绝对自由**且退出**绝对无成本**的市场。

这个简短的定义包含两个紧密相连的条件：

1. **进入绝对自由**（absolutely free entry）：潜在进入者面临与在位者完全相同的成本函数和技术，没有 Stigler 意义上的不对称壁垒。进入者可以以与在位者相同的条件获得所有投入品。

2. **退出绝对无成本**（absolutely costless exit）：零沉没成本。所有在进入时投入的资本在退出时可以按原始价值完全回收——无论是通过转售还是重新部署到其他用途。

固定成本 $\neq$ 沉没成本：可竞争性理论的核心概念贡献

可竞争性理论最关键的区分是：**固定成本不构成进入壁垒**（如果它们在退出时可以回收），但**沉没成本构成进入壁垒**。高固定成本本身不是问题——只要这些成本在退出时可以回收。构成壁垒的是成本的**沉没程度**（degree of sunkness），而非成本的大小。

例子：一架飞机是固定成本（不随乘客数量变化），但它是否构成沉没成本取决于退出时能否转售。一个活跃的二手飞机市场意味着飞机的沉没程度较低；高度专业化的设备（如专门的冶炼炉）几乎没有转售价值，因此是高度沉没的。

6.2.2 完全可竞争 $\neq$ 完全竞争

【核心】

这是学生必须掌握的关键区分：

表 6.1: 完全可竞争 vs. 完全竞争

维度	完全竞争	完全可竞争
企业数量	众多（原子式）	可以只有一家（垄断）
规模报酬	不变或递减	可以递增（自然垄断）
定价规则	P = MC = AC	P = AC（自然垄断）；P = MC = AC（多家企业）
利润	零经济利润	零经济利润
关键机制	**实际**竞争	**潜在进入的威胁**
沉没成本	无关（假设自由进入退出）	**定义性条件**（必须为零）
福利属性	Pareto 最优	拉姆齐最优（受利润约束的次优）

可竞争性理论最激进的贡献

可竞争性理论实现了竞争性结果，但**不要求竞争性市场结构**。”效率”与”原子式结构”的脱钩是这一理论最激进的智识贡献。在完全竞争的世界中，Pareto 最优要求许多小企业。在完全可竞争的世界中，**一个企业就够了**——只要进入和退出是自由的。

这一洞察直接挑战了自 Adam Smith 以来将”竞争”等同于”众多竞争者”的智识传统。

6.2.3 “打了就跑”进入：可竞争性的约束引擎

【核心】

”打了就跑”（hit-and-run）进入是可竞争性理论最生动的概念。其逻辑如下：

1. 在位者设定价格 $P > \text{AC}$（获得正经济利润）
2. 潜在进入者观察到此利润机会，以与在位者相同的技术进入，并设定 $P^e = P - \varepsilon$（略低于在位者价格）
3. 进入者捕获**整个市场**——消费者完全知情、产品同质、所有人瞬时切换到最低价卖家（Bertrand 假设）
4. **在在位者能够降低价格之前**，进入者已经赚取了短期利润并**退出**市场

5. 由于退出无成本（零沉没成本），进入者不会被” 困在” 市场中，也不面临任何风险

最精致的扭曲：进入者甚至不需要实际进入。仅仅是**可信的进入威胁**就约束了在位者。如果在位者将价格设定在平均成本之上，进入威胁立即出现。如果在位者将价格设定为等于平均成本，进入威胁消失。因此，在位者的唯一可维持策略是 $P = \text{AC}$——零经济利润。

这一机制需要**三个必要条件**同时成立：

1. **零沉没成本**：所有资本在退出时可完全回收。如果存在任何不可回收的投资，进入者会因在位者报复而被” 困住”。

2. **价格粘性滞后**：进入者设定价格与在位者做出价格反应之间存在一个时间差。如果不存在此滞后——如果在位者可以在进入者公布价格后**瞬间**匹配价格——进入者赚取零利润，永远不会进入。这是**最脆弱的假设**（Schwartz & Reynolds 1983）。

3. **Bertrand 式需求**：产品同质，消费者完全知情，所有需求瞬时切换到最低价卖家。进入者以仅 ε 的价格优势捕获整个市场。

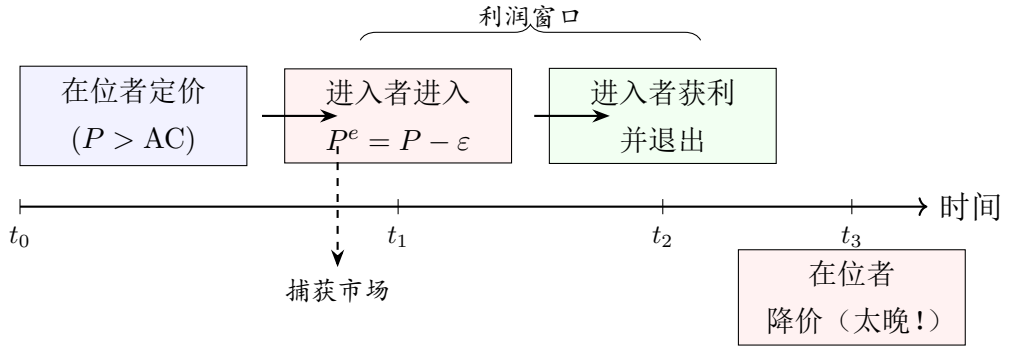


图 6.1: ” 打了就跑” 进入的时序。进入者在 t_1 以略低于在位者价格的价格进入，捕获整个市场，在 t_2 获利并退出——**在**在位者在 t_3 做出降价反应之前。关键假设：t_3 显著晚于 t_2（价格粘性滞后）。

6.2.4　三个核心定理

【核心】

BPW (1982) 建立了三个互相支撑的定理：

命题 6.1 (弱无形之手定理，Part I). *在完全可竞争市场中，任何可维持的行业配置必须*__最小化行业总成本__*。无效率的企业无法生存——它会被成本更低的进入者以” 打了就跑” 方式驱逐。*

直觉：如果一个行业以总成本 C_{actual} 生产总产出 Q，而存在任何替代配置（不同企业数量、不同产出分配）可以按更低的总成本 $C_{\min} < C_{\text{actual}}$ 生产相同的 Q，那么潜在进入者可以复制该成本最小化配置、以低于在位者的价格提供服务、获得利润并退出。因此，原始配置不是可维持的。唯一免受进入的配置是成本最小化的配置。

命题 6.2 (可维持配置定理). *在完全可竞争市场中，可维持价格向量存在当且仅当市场完全可竞争。可维持价格向量与**拉姆齐最优价格**一致——即在总收入等于总成本（零经济利润）的约束下最大化总剩余的价格。*

对于多产品企业，拉姆齐最优加价满足著名的" 逆弹性规则"（inverse elasticity rule）：

$$\frac{p_i - \text{MC}_i}{p_i} = \frac{k}{|\varepsilon_i|} \tag{6.1}$$

其中 $k \in [0, 1]$ 取决于盈亏约束的松紧程度，ε_i 是产品 i 的需求弹性（绝对值）。

经济学含义：

- 需求弹性越低的产品，加价越高——因为对这种产品提价造成的需求扭曲较小
- $k = 0$：一阶最优边际成本定价（仅当规模报酬不变时才可能——MC = AC）
- $k = 1$：无约束垄断定价
- 在可竞争的自然垄断中，k 被设定为恰好使 $\sum p_i y_i = C(y)$

命题 6.3 (自然垄断结果). *在完全可竞争的自然垄断中（成本函数在相关范围内是次可加的），均衡价格为：*

$$P = AC \quad (\text{而非} P = MC)$$

*垄断者获得**零经济利润**。任何 $P > AC$ 都会吸引" 打了就跑" 进入。*

与传统观点的关键区别：

- 传统自然垄断观：$P = \text{MC} < \text{AC} \to$ 企业亏损 $\to$ 需要政府补贴或规制
- 可竞争性观：进入威胁迫使 $P = \text{AC} \to$ 企业盈亏平衡 $\to$ **不需要规制**
- $P = \text{AC}$ 不是不受规制的垄断价格（后者是 MR = MC，导致 $P \gg \text{AC}$ 和高利润）——进入威胁将价格一路压缩至平均成本

6.2.5　自然垄断的重新思考：次可加性

【核心】

可竞争性理论为" 什么是自然垄断" 提供了一个更精确的定义。

定义 6.2 (成本次可加性). *成本函数 $C(y)$ 在产出 y 处是**次可加的***（*subadditive*），*如果对于 y 的任何分割 $y^1, y^2, \ldots, y^K$（满足 $\sum y^k = y$）：*

$$C(y) < \sum_{k=1}^{K} C(y^k)$$

即，由一个企业生产整个产出比由两个或更多企业分割生产更便宜。

次可加性，而非规模经济，是自然垄断的正确标准。在单一产品情形中，规模经济（平均成本下降）是次可加性的充分但非必要条件。在多产品情形中，**范围经济**（economies of scope）与规模经济共同决定次可加性：

定义 6.3 (范围经济).

$$C(y_1, y_2) < C(y_1, 0) + C(0, y_2)$$

由一家企业联合生产两种产品比由两家专业化企业分别生产更便宜。

实际例子：一条同时承载客运和货运的铁路网络。轨道是共享投入。C(客运 + 货运) $< C$(客运, 0) $+ C$(0, 货运)——这是范围经济。次可加性的问题是：一家铁路公司能否比两家拥有重复轨道的竞争铁路公司更便宜地服务整个市场？

Panzar & Willig (1977, 1981) 的这项工作将范围经济引入经济学词汇，现在已成为 IO 分析多产品企业的标准工具。

6.2.6　Baumol 的智识轨迹：AT&T 联系

【进阶】

一个引人深思的制度细节：Baumol 不是芝加哥经济学家——他在 LSE 师从 Lionel Robbins，根植于 Pigou 的福利经济学传统。他早期的工作（*Welfare Economics and the Theory of the State*, 1952）是正统的福利经济学，分析外部性和政府干预的理由。

然而 Baumol 却生产了 IO 史上最极端的亲市场理论。为什么会这样？答案的一部分在于他的制度位置：

- 从 1966 年起，Baumol 在 AT&T 的 Bell 实验室担任顾问和专家证人——跨越约三十年
- 可竞争性理论是在 1970 年代与 Bell 实验室的经济学家团队（Elizabeth Bailey、John Panzar、Robert Willig）共同开发的

- 该理论被” 部署为证明放松管制正当性的政策工具”（Colton 2022），直接服务于 AT&T 在放松管制辩论中的利益

一位福利经济学家如何成为亲市场理论的旗手

Baumol 的智识轨迹说明了经济学中一个更深层的模式：制度位置会影响理论方向。一个为自然垄断辩护的经济学家——证明即使只有一家企业，市场也可以产生竞争性结果——恰好服务于一个面临分拆威胁的受规制垄断企业（AT&T）的智识需求。

但这并不意味着可竞争性理论是” 腐败的” 或” 无效的”。好的经济学可以服务于客户利益而同时是真正的智识贡献。问题在于：理论是否经得起**独立**的检验——而本章的后半部分正是关于检验的结果。

6.3 政策的可竞争性化：放松管制运动

6.3.1 来自 Demsetz (1968) 的谱系

【核心】

可竞争性理论的直接前身是 **Demsetz, H. (1968). “Why Regulate Utilities?” *Journal of Law and Economics*, 11(1), 55–65.**

Demsetz 挑战了传统的自然垄断三段论：

1. 规模经济 → 一个企业最有效（自然垄断）
2. 一个企业 → 垄断定价（市场失灵）
3. 垄断定价 → 需要规制

Demsetz 攻击了第二步：即使最终只有一家企业服务市场，**为市场而竞争**（competition *for* the market）可以替代**市场内竞争**（competition *in* the market）。如果有足够多的合格竞标者拥有获得相同投入的途径，特许经营权的竞标过程会将中标价格压缩至单位成本。

BPW (1982) 推广了这一洞察：甚至不需要显式拍卖。潜在” 打了就跑” 进入的威胁本身就构成了一种**隐式的、持续的拍卖**。潜在进入者通过不断评估能否以低于在位者的价格获利来持续” 竞标”。

从 Demsetz 到 Baumol 的智识链：Director/Stigler/Demsetz（芝加哥学派）→ 潜在竞争原则 → Baumol/Panzar/Willig（可竞争性理论）。这一链条构成了从芝加哥价格理论到放松管制政策的最直接的知识传递。

6.3.2 航空放松管制：可竞争性的伟大实验

【核心】

管制前时代

从 1938 年到 1978 年，美国民用航空委员会（CAB）控制着航空业的三个方面：

- **航线**：航空公司未经 CAB 批准（通常耗时 2–3 年）不得进入新市场。1938 年以来没有新的干线航空获得许可
- **票价**：CAB 通过成本加成公式设定票价。航空公司围绕非价格维度竞争——餐食服务、座位间距、航班频次
- **进入**：新企业进入本质上是不可能的。CAB 将" 过度竞争" 视为不稳定因素

1977 年，卡特总统任命了 Cornell 经济学家 **Alfred E. Kahn** 为 CAB 主席。Kahn 不是正式意义上的可竞争性理论家，但他吸收并操作化了其核心洞察：进入威胁可以约束在位者。

Kahn 的名言捕捉了他的先验信念：

> "I really don't know one plane from the other. To me they are all marginal costs with wings."

Kahn 在国会行动前就行政性地放松了票价管制：允许深度折扣票价，允许自由进入选定的市场。这些行政实验表明更低的票价显著扩大了需求，从而建立了对放松管制的政治支持。

1978 年航空放松管制法案

1978 年 10 月，两党支持通过了《航空放松管制法案》：

- 到 1983 年逐步取消 CAB 对票价的权力
- 到 1981 年逐步取消 CAB 对航线的权力
- 于 1985 年 1 月 1 日**完全废除 CAB**
- 明确排除各州对航空公司的经济规制

关键的遗漏：该法案未涉及机场时刻分配、登机口租赁、或计算机订座系统（CRS）的反垄断含义。这些遗漏后来被证明是**决定性的**。

初步结果：第一个辉煌的十年

深度折扣票价战产生了戏剧性的例子：纽约至洛杉矶 $99、纽约至佛罗里达 $39、纽瓦克至水牛城 $19（名义价格）。

这些总体数字似乎验证了可竞争性命题：放松管制、允许自由进入和退出、市场会——即使集中度可能很高——实现竞争性结果。

表 6.2: 航空放松管制前后关键指标对比

指标	管制前（1978）	放松管制后（1980s 末）	变化
每客英里实际收益	——	——	**−30%** (Kahn, 1976–1990)
年客运量	∼2.4 亿	∼4.47 亿 (1987)	∼+86%
曾乘机美国人比例	∼50%	∼70% (1988)	+20 ppt
客座率	∼55–60%	∼65–70%	+10–15 ppt
年消费者节省	——	∼$100 亿/年	Kahn 估计首十年累计 $1000 亿

6.3.3 1982 年合并指南：进入分析框架的制度化

【核心】

1982 年 DOJ 合并指南首次将**进入分析**作为合并审查的核心要素——这是可竞争性推理的直接制度遗产。核心分析转向：不再仅看当前的市场结构和集中度，而是问：**如果合并后的企业试图提高价格，新的进入能否及时、可能、充分地约束它？**

”及时、可能且充分”（timely, likely, and sufficient）的进入测试在 1992 年横向合并指南中被正式化，成为此后合并审查的标准分析框架。详见 第 ∼15∼ 章。

6.4 枢纽-辐射式批判：可竞争性在经验上的失败

可竞争性理论的伟大讽刺是：它为放松管制提供了智识基础，而放松管制的经验观察随后证明了该理论的最强预测是错误的。

6.4.1 Borenstein (1989)：枢纽与高票价

【核心】

Borenstein, S. (1989). “Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry.” *RAND Journal of Economics*, 20(3), 344–365.

Severin Borenstein 的工作是产业组织学中最具影响力的经验发现之一。他的核心发现与可竞争性理论的预测直接矛盾：

1. **机场支配力 ≫ 航线集中度**：一家航空公司在其端点机场总登机量中的份额是**比航线层面的集中度强得多的票价预测因子**。支配两个端点机场的航空公司可以收取显著高于非支配航线上同等距离的票价。

2. **枢纽溢价**：在支配性枢纽机场，票价在控制成本因素后比非枢纽航线高 **10–20%**。这不是微小的统计异常——它是经济上显著且稳健的。

3. **没有"保护伞效应"**：较小的航空公司**不能**搭主导航空公司高票价的便车。即使主导航空公司提价，竞争对手的票价维持低位。这意味着进入确实发生了——但进入者没有约束在位者的定价。这与可竞争性预测恰恰相反。

表 6.3: Borenstein (1989) 的核心发现：可竞争性预测 vs. 经验现实

预测	完全可竞争性	经验现实
集中度与票价关系	无关系——只有可竞争性条件重要	强正相关——枢纽支配力推高票价
空中服务频次	仅效率差异	频次优势构成产品差异化壁垒
枢纽票价 vs. 非枢纽票价	应该相同（控制数量）	枢纽显著更高（10–20%）
进入对票价的影响	纯粹威胁即足矣	只有**实际**进入显著压低票价（Southwest 效应）

6.4.2　为什么可竞争性在航空枢纽失灵？

【核心】

Borenstein 和后续研究者识别了使航空枢纽不可竞争的多个具体机制：

可竞争性失败的系统性特征

航空枢纽的可竞争性失败不是孤立的异常——它是**系统性的**。几乎每一个机制都涉及企业**策略性地**创造的沉没成本或转换成本。常旅客计划不是外在的——它们是航空公司的策略性营销工具。CRS 偏见不是偶然的——它是供应商航空公司有意设计的。登机口租赁是航空公司游说获得的合同条款。

这正是 Sutton (1991) 的观点：沉没成本往往是**内生的**——由在位者和进入者策略性地选择和塑造——而非外在给定的技术参数。可竞争性理论将沉没成本视为外在的、要么存在要么不存在的二元变量——而这恰恰是与现实脱节的地方。

6.4.3　Morrison & Winston (1987)：正式检验

【核心】

Morrison, S.A. & Winston, C. (1987). "Empirical Implications and Tests of the Contestability Hypothesis." ***Journal of Law and Economics*****, 30(1), 53–66.**

表 6.4: 破坏航空市场可竞争性的机制

壁垒	机制	可竞争性含义
常旅客计划 (FFP)	商务旅客（不自付机票）依据里程累积而非价格选择航空公司。GAO (1991)：商务旅客超 50% 时间基于 FFP 选择。阶梯式奖励结构（累计里程阈值后的奖金里程）创造了锁入效应	需求方转换成本——减少进入者可获得的剩余需求
CRS 屏幕偏见	最大的两个 CRS（Sabre/American, Apollo/United）控制 70–74% 的旅行社预订。屏幕显示偏见有利于供应商航空公司。对竞争对手航空公司的订座费超过竞争水平	在消费者搜索层面创造在位者优势
登机口长期租赁	排他性使用登机口租约长达 20–30 年。航空公司囤积不用登机口。多数利益条款赋予主导航空公司对机场扩建的否决权	在位者对稀缺机场设施的控制——经典沉没成本壁垒
机场时刻控制	FAA 在华盛顿国家、O'Hare、JFK 和 LaGuardia 的时刻控制使在位者于 1969 年以零成本获得时刻。后续买卖规则允许交易但时刻价格使得进入代价高昂	在位者以低于市场价值获得瓶颈设施
旅行社佣金加成 (TACO)	非线性佣金激励旅行社将业务引导至主导航空公司——在旅客和旅行社之间创造了委托-代理问题	扭曲了需求引导，有利于在位者
密度经济	枢纽运营从更高客座率、更大飞机和更密集航班安排带来真实成本优势	真实的效率优势——但被保留为在位者租金，而非传递给消费者

这是对可竞争性最直接的经验检验。Morrison 和 Winston 使用多项式 logit 需求模型估计了放松管制后美国航空市场中的补偿变异福利度量，区分了：

1. **完全可竞争性**（BPW 表述）：进入/退出的福利变化为零（价格始终处于竞争水平）

2. **不完全可竞争性**（Bain 式）：存在一些壁垒，但潜在竞争发挥缓和效应

发现：

- 航空市场**不是**完全可竞争的——福利变化度量在统计和经济上显著不为零
- 但潜在竞争**确实**具有一定约束效应——市场是" 不完全可竞争的"
- 政策含义：继续鼓励实际**和**潜在竞争，但**不要仅依赖**潜在竞争来约束在位者

6.4.4 Southwest 效应：实际 > 潜在

【核心】

Morrison, S.A. (2001). "Actual, Adjacent, and Potential Competition: Estimating the Full Effect of Southwest Airlines." ***Journal of Transport Economics and Policy*****, 35(2), 239–256.**

Morrison 的后续工作将 Southwest Airlines 的竞争效应分解为三个渠道：

表 6.5: Southwest 效应的分解（1998 年数据）

渠道	描述	年消费者节省
实际竞争	Southwest 直接在该航线上运营	$34 亿
邻近竞争	Southwest 服务于一个端点机场，使多航段旅行的部分路段存在竞争	包含在 $95 亿之中
潜在竞争	Southwest 在端点机场有存在，使进入物理上可行	包含在 $95 亿之中

总节省：$129 亿（约为美国国内定期客运收入的 20%）。

对可竞争性的关键解读：即使是潜在竞争效应，其机制也更类似于**限制性定价**（limit pricing）——在位者因担心 Southwest **实际**进入而降价——而非可竞争性的" 打了就跑" 机制。Southwest **确实有**沉没成本（飞机、机组基地、登机口租赁），并且一旦进入就可以可信地承诺**留下**。威胁之所以有效，**正是因为**进入是可信的——而非因为它是无成本可逆的。

6.5 理论批判：为什么完全可竞争性在逻辑上失效

6.5.1 Shepherd (1984)：同义反复批判

【核心】

Shepherd, W.G. (1984). "'Contestability' vs. Competition." *American Economic Review*, 74(4), 572–587.

William Shepherd 发动了对可竞争性理论最有影响力的攻击。他的核心论证——三个字——是：

可竞争性理论是同义反复。

论证的逻辑：

1. 理论**定义**" 完全可竞争市场" 为进入绝对自由、退出绝对无成本的市场——即，没有任何壁垒的市场
2. " 完全自由进入" 意味着：无沉没成本、进入者与在位者技术相同、在位者无价格反应优势、无品牌忠诚、无转换成本——简而言之，**任何壁垒都没有**
3. 在假设所有壁垒都不存在之后，理论随后**得出结论**：市场产生竞争性结果，不存在市场势力问题
4. **结论已经包含在前提之中**。理论没有产生一个可检验的经验预测——它只是将假设重述为结论

Shepherd 在 1995 年的回访（*Land Economics*）中进一步论证该理论要求两个**相互不一致的假设**同时成立：

- **假设 1**：进入者必须能**完全取代**在位者（在位者做出反应前捕获整个市场）——否则" 打了就跑" 利润太小，进入无利可图
- **假设 2**：进入者必须被视为**太渺小**以至于在位者不会费力做出反应——否则在位者降价，进入者一无所获

Shepherd 称这一矛盾为理论的" 致命缺陷"：如果进入者大到能捕获整个市场，它就大到足以引起在位者的即时反应——消除利润窗口。

6.5.2 Schwartz & Reynolds (1983)：微小沉没成本问题

【核心】

Schwartz, M. & Reynolds, R.J. (1983). "Contestable Markets: An Uprising in the Theory of Industry Structure: Comment." *American Economic Review*, 73(3), 488–490.

这是对可竞争性最精辟的理论批判。核心主张：

即使微小的沉没成本也打破完全可竞争性结果。如果进入需要任何不可回收的成本——哪怕只有 $1——在位者可以通过足够快速地降价来消除利润窗口，使" 打了就跑" 无利可图。

同时，Schwartz & Reynolds 识别了一个微妙的悖论：

一些退出摩擦是" 打了就跑" 威胁具有约束力所必需的。

如果退出真正瞬时而零成本，在位者也没有退出成本——所以如果在位者降价匹配进入者，在位者也零损失。战斗对在位者同样是" 免费的"。但如果在位者可以无成本地战斗，进入者的" 打了就跑" 利润窗口消失——因此进入永远不会发生。

Schwartz (1986) 的提炼：承诺问题双向运行。为让可竞争性起作用：

- 进入者必须能承诺快速退出（" 打了就跑"）——仅在零沉没成本下可能
- 在位者必须**无法**承诺战斗——但在零沉没成本下，在位者没有战斗成本，因此战斗是**可信的**

因此，使进入者进行" 打了就跑" 成为可能的同一条件（零沉没成本），也使在位者的战斗承诺可信——而这又使" 打了就跑" 无利可图。

6.5.3　其他理论批判

【进阶】

- **Brock (1983, JPE 书评)：**结果对假设极度敏感。零沉没成本假设的小偏离产生质的变化。理论缺乏稳健性。
- **Weitzman (1983, AER 评论)：**价格刚性假设至关重要且不现实。理论隐含地假定进入者在调整速度上的不合理不对称——进入者被假定为无限快，在位者反应迟缓。
- **Spence (1983, JEL 书评)：**最精辟的批判——" 该理论在市场竞争**不是**完全可竞争时没有提供任何指导。" 既然没有真实市场是完全可竞争的，这意味着该理论对**任何**真实市场都没有提供指导。它是一个二进制框架：要么完全可竞争（竞争性结果），要么不是（未知结果）。
- **Stiglitz (1981)：**" 潜在竞争可能降低福利。" 当存在固定成本时，过度或重复进入会耗散资源而不改善配置效率。
- **Dasgupta & Stiglitz (1988)：**当存在沉没成本时，" 潜在竞争是实际竞争的劣质替代品。" 市场结构必须被视为**内生的**——由技术、沉没成本和策略行为共同决定。

表 6.6: 可竞争性理论的主要理论批判

作者	年份	核心论点	目标假设
Shepherd	1984, 1995	同义反复：假设零壁垒，结论零问题；两个相互不一致的假设	整个框架
Schwartz & Reynolds	1983	微小沉没成本打破结果；无摩擦退出反而使战斗可信	零沉没成本 + 价格粘性
Brock	1983	结果对假设极度敏感；缺乏稳健性	零沉没成本
Weitzman	1983	进入者/在位者的调整速度不对称不合理	价格粘性滞后
Spence	1983	理论只在完全可竞争时给出预测——而这永不为真	外部有效性
Stiglitz	1981	潜在竞争在某些条件下可能降低福利	福利单调性
Dasgupta & Stiglitz	1988	潜在竞争是实际竞争的劣质替代品；市场结构是内生的	潜在/实际竞争等价

BPW (1983) 的回复与”基准”辩护

Baumol、Panzar 和 Willig 在同一期 AER 中发表了回复。他们的核心辩护：完全可竞争性旨在作为**理论基准**——完全类似于完全竞争。没有人相信任何真实市场是完全竞争的（原子式企业、同质产品、完全信息），但该概念作为福利基准服务了数个世纪。同理，完全可竞争性为具有规模经济的行业提供了一个**更为一般**的基准。

这一辩护有部分道理——但有一个关键区别：完全竞争至少在一部分真实市场中（农产品、大宗商品）是不错的近似。完全可竞争性在任何真实市场中是否是不错的近似——这就是经验证据介入之处。而经验证据给出了否定的回答。

6.6 博弈论批判：策略行为为什么重要

6.6.1 可竞争性的核心盲点：在位者是被动的

【进阶】

可竞争性理论**刻意忽略了策略行为**。理论假定在位者对进入的价格反应存在一个滞后——在进入者捕获市场并退出之前，在位者不会降价。这一假定并非无心的简

化——它是可竞争性结果**能够成立的前提**。

博弈论革命（第 ~7~ 章将详细展开）揭示了为什么这一假定是致命的：

1. 如果进入在当前价格下有利可图，在位者有激励在进入者退出前立即降价
2. 如果进入者知道在位者会降价，进入者不会进入
3. 如果在位者无法可信地承诺不降价，进入威胁消失
4. 可竞争性要求：在位者 = 非策略性的。博弈论证明：在位者在子博弈精炼均衡中不能可信地承诺非策略性行为

6.6.2 Dixit (1980)：产能作为承诺工具

【进阶】

Dixit, A. (1980). "The Role of Investment in Entry-Deterrence." *Economic Journal*, 90(357), 95–106.

Dixit 的模型是可竞争性理论的博弈论替代方案。考虑一个三阶段博弈：

1. **在位者选择产能** k_1（沉没投资）
2. **潜在进入者观察** k_1 并决定是否进入
3. 如果进入发生，**Cournot 竞争**（产能约束产出）

关键机制：一旦产能安装，它是**沉没的**——无法转售。这使在位者的产能选择成为一个**可信承诺**（credible commitment），承诺在进入后生产高产出。进入者观察到此承诺并——如果产能足够高——决定不进入。

巨大的讽刺：可竞争性理论假定沉没成本为零以确保进入。Dixit 表明**在位者策略性地创造沉没成本以阻挠进入**。沉没成本不是可以" 移除" 的外在壁垒——它们是竞争过程本身的一部分。

6.6.3 Fudenberg & Tirole (1984)：策略行为的分类学

【进阶】

Fudenberg, D. & Tirole, J. (1984). "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look." *American Economic Review*, 74(3), 361–366.

这一经典论文为策略性进入阻挠提供了一个统一的分类学——" 肥猫效应"、" 小狗策略"、" 饿虎相"、" 头犬"——基于两个维度：

- 投资使在位者变得**更强硬**（tough）还是**更温和**（soft）?

表 6.7: 可竞争性理论 vs. Dixit (1980) 进入阻挠

维度	**可竞争性理论 (BPW 1982)**	**承诺模型 (Dixit 1980)**
沉没成本的角色	消除之——它们是" 壁垒"	**策略的核心**——在位者通过沉没投资承诺激进行为
在位者行为	被动（不能承诺战斗）	**主动**（策略性地过度投资以阻挠进入）
进入阻挠	假设为不可能	可能、理性且经验可观察
均衡概念	可维持性（静态）	子博弈精炼均衡（动态）
政策含义	取消规制壁垒，市场自我约束	策略性壁垒存在；需要基于效果的分析
进入者观察	进入者在位者价格中观察到的利润机会	进入者观察在位者的**沉没产能**后做进入决策

- 进入后竞争涉及**策略替代**（strategic substitutes，如 Cournot 数量）还是**策略互补**（strategic complements，如 Bertrand 价格）?

Tirole (1988, Ch.8) 将这一分类学作为 IO 教范的核心部分。对可竞争性的含义：在位者有**系统性的策略工具**来影响进入者的盈利性——即使在没有传统 Bain 式壁垒的情况下。

6.6.4 可竞争性 → 博弈论的完美过渡

可竞争性理论和博弈论革命之间的关系是教学上极为有力的**辩证结构**：

1. **正题**（可竞争性）：如果没有任何承诺工具，进入威胁足够约束垄断者。P = AC，零利润。不需要规制。

2. **反题**（博弈论）：在位者可以**策略性地创造**承诺工具（沉没产能、品牌投资、长期合同）。这些工具使进入威胁失效。

3. **合题**（经验检验）：问题不是理论上的" 是或否"，而是经验上的" **在什么条件下**进入威胁约束在位者，在什么条件下在位者的策略性承诺阻挠进入？"

为什么可竞争性的失败是博弈论革命的最佳引言

可竞争性理论展示了当不存在策略行为时会发生什么。博弈论展示了为什么策略行为重要——以及承诺（通过沉没成本）如何使进入阻挠即使在”无壁垒”市场中也成为可能。

学生应该从这个故事中获得的核心教训是：**IO 的核心问题不是”市场集中吗？”而是”什么使承诺可信或不可信？”**可竞争性的世界是承诺不可能的世界。博弈论的世界是承诺可能且被策略性地使用的世界。真实世界位于两者之间——而 第 ~7~ 章的工具帮助我们分析具体的案例。

6.7　Sutton 与内生沉没成本

6.7.1　外生 vs. 内生沉没成本

【进阶】

Sutton, J. (1991). *Sunk Costs and Market Structure*. MIT Press.

John Sutton 为沉没成本的分析引入了一个关键区分：

定义 6.4 (外生沉没成本). 由**技术**决定的设立成本——建立最小有效规模的工厂、购买设备、获得基本监管批准。所有进入者均等地承担这些成本。

定义 6.5 (内生沉没成本). 企业**策略性地选择**的支出——研发、广告、品牌建设——以提高消费者的支付意愿。企业可以选择在这些维度上支出多少，且这一水平直接塑造了市场结构。

对可竞争性的含义：

1. 可竞争性理论将沉没成本视为**外生壁垒**——要么存在，要么不存在；政策可以”移除”它们
2. Sutton 表明企业在**策略性地选择**它们的沉没成本水平——在位者可以通过增加广告或研发支出升高沉没成本壁垒
3. 结果是，市场可能随时间变得**越来越不可竞争**——在位者投资于创造壁垒
4. 在内生沉没成本的市场中，增加的市场规模**不会**降低集中度（与外生沉没成本的情形相反），因为对沉没成本的竞争性升级维持了——甚至提高了——均衡集中度水平

对航空业的应用：航空枢纽壁垒的大部分——FFP、CRS 投资、登机口租赁、旅行社合同——不是外在的”给定”特征。它们是航空公司**策略性地**创造和维护的。可竞争性理论无法把握这一事实，因为它将壁垒视为需要被”移除”的——但实际上，它们是竞争过程本身的产物。

6.8 可竞争性的遗产：什么留下了，什么没有

6.8.1 持久的概念贡献

【核心】

尽管经验上失败，可竞争性理论留下了多项被保留至今的贡献：

1. **沉没成本 vs. 固定成本的区别**：或许是最持久的贡献。高固定成本本身并不构成进入壁垒——仅当成本是沉没的（不可回收的）才构成。这一区分重构了数十年的反垄断和规制分析。

2. **范围经济**：BPW 提供了多产品生产最全面的分析处理，引入了现已为 IO 基础的概念。

3. **可维持性概念**：一个统一的框架，用于分析从完全竞争到自然垄断的所有市场结构——在企业和产品空间的两个维度上。

4. **拉姆齐最优定价作为福利基准**：为规模报酬递增的行业提供了一个逻辑一致的福利基准，替代了不适用的 Pareto 最优。

5. **"可竞争程度"作为谱系**：政策制定者现在问"这个市场有多可竞争？"而非"它是完全可竞争的吗？"——一个更微妙的、经验可处理的框架。

6.8.2 "基准"辩护：Baumol & Willig (1986)

【进阶】

Baumol, W.J. & Willig, R.D. (1986). "Contestability: Developments Since the Book." ***Oxford Economic Papers*****, 38(Supplement), 9–36.**

Baumol 和 Willig 对批评者的核心回应是：

> **完全可竞争市场作为基准，类似于完全竞争。**没有人相信任何真实市场是完全竞争的（原子式企业、同质产品、完全信息），但完全竞争作为福利分析的基准服务了数个世纪。同理，完全可竞争性提供了一个**更为一般**的基准——一个在规模经济和小数量企业的条件下仍然产生 Pareto 类比性质的基准。

这一辩护值得严肃对待：

- **支持的论证**：从逻辑严谨性而言，完全可竞争市场确实可以作为完全竞争市场的替代理性基准。它比完全竞争更一般，因为它不要求原子式企业数量。

- **反对的论证：**完全竞争至少在一些真实市场（农产品、大宗商品）中是不错的近似。可竞争性似乎不在任何真实市场中成立。如果基准与现实差距太大，它就是一个糟糕的基准。

- **中间立场：**一个基准的价值不在于它是否" 真实"——而在于它是否提出了正确的比较问题。在可竞争性之前，标准比较是" 垄断 vs. 完全竞争"——这忽略了对规制和反垄断至关重要的整个制度维度（进入条件、沉没成本、承诺能力）。可竞争性迫使经济学家问更好的问题。

6.8.3 "可竞争性"一词的制度化

【进阶】

" 可竞争性"（contestability）一词现已完全融入全球竞争经济学的标准词汇：

- **EU 竞争法：**数字经济中的" 可竞争性" 是一个核心概念
- **UK CMA 市场调查：**常规性地评估" 可竞争程度"
- **OECD 竞争政策建议：**将可竞争性作为市场运作良好的标准之一
- **2023 年美国合并指南：**虽然从可竞争性为基础的推理转向，但该概念仍旧出现在分析框架中

这个词本身是 Baumol 持久的修辞贡献：它重新定义了政策问题，从" 这个市场有多少企业？" 转变为" 企业进出这个市场有多容易？"

6.8.4 当代相关性与前言

【前沿】

可竞争性概念在现代政策辩论中持续出现——虽然常常以 Baumol 不会预料到的方式：

- **EU 数字市场法案 (DMA)** 的明确目标是" 通过减少进入壁垒，提高 EU 数字市场的**可竞争性**"（De Streel, 2024）。讽刺的是，DMA 的核心工具是**规制**（互操作性要求、数据可携带性、自我优待禁止）——这与可竞争性理论" 不需要规制" 的精神恰恰相反。Baumol 的思想遗产已经从放松管制的理论基础转变为再规制的工具。

- **2023 年美国合并指南**代表了从可竞争性推理的**部分撤回**。Structural presumption 被恢复——但以一种新型混合体的方式，而非简单回到 1968 年。Guideline 4（潜在竞争）和 Guideline 6（新生竞争）是两个方向的背离：指南对**合并方**诉诸潜在竞争（" 进入会约束合并后的企业"）持怀疑态度，而对**机构**诉诸潜在竞争（" 合并消除了一个潜在进入者"）持更开放态度。

- **AI 与可竞争性**：GenAI 同时降低和升高进入壁垒——对下游应用降低（小团队可以更高效），对基础设施层升高（训练成本指数级上升，芯片获取限制）。

从放松管制到再规制：可竞争性遗产的讽刺

可竞争性理论的最大讽刺可能是：它的核心概念——" 进入壁垒决定市场绩效" ——现在被用来**证明**管制干预的正当性（如 DMA、2023 年合并指南），而非质疑它。Baumol 构建了让政府撤出的理论框架；他的继承者使用相同的概念来证明让政府重新介入的正当性。

这种从" 可竞争性 → 不需要规制" 到" 可竞争性不足 → 需要规制" 的转变，是思想史中最引人入胜的辩证法之一。

6.9 本章小结

1. **可竞争市场理论将竞争政策的目光从" 有多少企业" 转向" 企业进出有多容易"**。其核心洞察是：进入和退出的自由程度——而非在位者的数量——决定了市场绩效。一个只有一家企业的市场，如果进入和退出完全自由，可以产生竞争性结果。

2. **" 打了就跑" 进入是可竞争性的约束引擎**。核心机制：潜在进入者可以进入、瞬时捕获整个市场、赚取利润、并在在位者能做出反应之前退出。关键是：进入者甚至不需要实际进入——仅仅是可信的进入威胁就约束了在位者。

3. **可竞争性理论为 1980 年代的放松管制运动提供了关键的智识信心**。美国航空（1978）、卡车（1980）、铁路（1980）和电信（1982/1984）放松管制均受到可竞争性推理的影响。航空业展示了惊人的总体成功——票价下降约 30%，客运量近乎翻倍。

4. **然而，经验检验拒绝了完全可竞争性——正是在其示范行业中**。 Borenstein (1989, 1992) 证明航空枢纽市场产生**更高**而非更低的票价。枢纽溢价在控制成本因素后约为 10–20%。几乎每一个破坏可竞争性的机制——FFP、CRS、登机口租赁、时刻控制——都是航空公司**策略性地**创造和维护的。

5. **理论批判揭示了可竞争性失败的根本原因：忽略了策略行为**。 Shepherd (1984) 指出理论是同义反复。Schwartz & Reynolds (1983) 证明微小沉没成本也打破结果。Dixit (1980) 和 Fudenberg & Tirole (1984) 表明在位者通过沉没投资**策略性地承诺**激进行为——使进入阻挠即使在" 无壁垒" 市场中也成为可能。

6. **可竞争性理论留下了持久的概念工具，尽管它在经验上失败了**。沉没成本 vs. 固定成本的区别、范围经济、成本次可加性、" 可竞争程度" 作为谱系——这些被

保留。Baumol & Willig (1986) 的” 基准” 辩护——完全可竞争性是” 有规模经济的完全竞争”——有部分道理，但经验差距比完全竞争的更大。

7. **可竞争性的失败是博弈论革命的最佳引言。**故事的辩证结构完美地设立了 第 ~7~ 章：可竞争性 = 承诺不可能的世界 → 博弈论 = 承诺可能且被策略性使用的世界 → 经验检验 = ” 在什么条件下？ ”

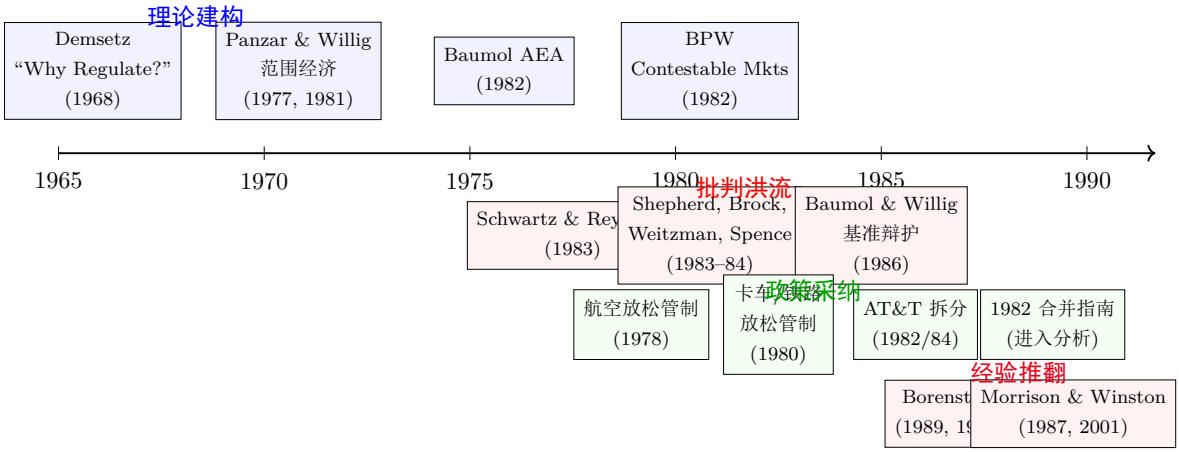


图 6.2: 可竞争性理论在 IO 思想史中的位置。注意时间线的紧凑性：理论建构（1977–1982）、批判洪流（1983–1986）、政策采纳（1978–1984）和经验推翻（1987–1992）几乎同时发生。IO 史上几乎没有另一个理论在如此短的时间内经历了如此完整的兴衰周期。

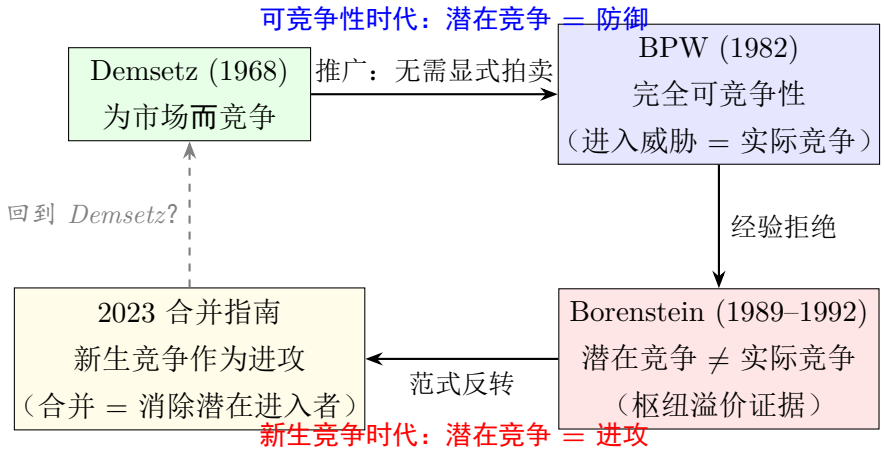


图 6.3: 潜在竞争概念的演变：从 Demsetz 到 2023 年合并指南。可竞争性时代将潜在竞争用作合并的**防御**：” 即使合并增加集中度，潜在进入者仍将约束在位者。”2023 年指南反转了这一逻辑，将潜在竞争用作**进攻**：” 合并消除了本将成为实际竞争者的潜在进入者。

6.10 习题

基础

6.1. 完全可竞争 vs. 完全竞争。

(a) 分别陈述完全可竞争市场和完全竞争市场的定义条件。

(b) 为什么一个完全可竞争的自然垄断者定价在 $P = \text{AC}$ 而非 $P = \text{MC}$？在这两种价格下利润分别是多少？

(c) 举出一个可能比" 完全竞争" 更接近" 完全可竞争" 的现实世界行业——以及一个两者都不接近的行业。解释你的理由。

6.2. 沉没成本 vs. 固定成本。

(a) 定义沉没成本和固定成本。给出一个同时为两种的成本的例子，以及一个是固定但不是沉没的成本的例子。

(b) 为什么在可竞争性理论中，区分固定成本和沉没成本很重要？

(c) 在航空业中，一架飞机是固定成本、沉没成本、还是两者都是？你的答案对航空市场的可竞争性意味着什么？

6.3. " 打了就跑" 进入。

(a) 用你自己的语言解释" 打了就跑" 进入的逻辑。为什么即使它从未实际发生也能约束在位者？

(b) " 打了就跑" 进入要有利可图，必须满足哪三个条件？

(c) 对于每个条件，给出一个它可能在现实世界中被违反的具体理由。

中级

6.4. Borenstein 的枢纽溢价。下表给出了两条距离相同的航线 A-B 和 C-D 的虚构（但典型）数据：

表 6.8: 两条可比航线的票价与集中度数据

航线	距离	枢纽支配？	HHI	平均票价
A–B	1,000 英里	是（一家在两端支配）	4,200	\$289
C–D	1,000 英里	否	3,800	\$198

(a) 根据可竞争性理论，两条航线的票价应该是怎样的关系？根据 Borenstein (1989) 的经验发现呢？哪个预测与表格一致？

(b) 给出三个具体的、可检验的机制来解释为什么 A-B 的票价更高。

(c) 如果 FAA 将登机口租约从长期排他性合同改为年度拍卖，票价差距会发生什么变化？

6.5. Shepherd 的同义反复批判。

(a) 重构 Shepherd (1984) 的论证：为什么可竞争性理论是" 同义反复"？

(b) Baumol & Willig (1986) 的回应是：完全可竞争性是一个" 基准"——类似于完全竞争。评价这一辩护。一个从未在任何真实市场中被观察到的基准有用吗？

(c) 完全竞争和完全可竞争性作为基准的区别是什么？对规制政策来说，哪个基准设立的门槛更高？

6.6. Dixit vs. BPW 关于进入。

(a) 在可竞争性理论中，为什么在位者不能阻挠进入？在 Dixit (1980) 中，为什么在位者**能够**阻挠进入？这两种情形之间的关键区别是什么？

(b) 沉没成本在这两个模型中扮演什么角色？为什么这个角色的不同导致了完全相反的结论？

(c) 哪个模型更好地描述了航空枢纽市场的行为？在你的答案中引用至少一个 Borenstein 的发现。

进阶

6.7. 平台市场中的可竞争性。考虑拼车平台（如 Uber/Lyft）：

(a) 评估拼车市场在 BPW 每个条件上的可竞争程度。有实质性的沉没成本吗？在位者能多快地调整价格？进入者能瞬时捕获市场吗？

(b) " 多归属"（用户和司机同时使用多个平台）如何影响市场的可竞争性？

(c) 2023 年合并指南明确提到" 网络效应" 和" 数据优势" 是潜在的进入壁垒。在 Bain-Stigler-Baumol 壁垒争论的谱系中——Bain 的" 任何在位者优势"、Stigler 的" 不对称成本"、Baumol 的" 沉没成本"——将网络效应视为壁垒的论点处于什么位置？为你的答案辩护。

6.8. 新生竞争 vs. 潜在竞争。

(a) " 潜在竞争" 和" 新生竞争" 之间的区别是什么？为什么这一区别在反垄断中重要？

(b) 为什么 2023 年合并指南降低了机构证明潜在进入被消除的负担，而提高了合并方证明潜在进入将约束合并后企业的负担？

(c) 考虑 Facebook/Instagram (2012) 收购。使用可竞争性逻辑为收购辩护。使用新生竞争逻辑挑战收购。哪种论证更契合当今的数字市场现实？

6.9. **可竞争性理论的兴衰。**从 1968 年到 2023 年追踪可竞争性概念在政策辩论中的角色：

(a) Demsetz (1968) 的原创论证是什么？它与 BPW (1982) 有何不同？

(b) 1980 年代放松管制运动在哪些方面依赖可竞争性推理？这些推理在哪些行业中最有道理，在哪些行业中最没道理？使用航空和卡车的证据。

(c) 为什么 2023 年合并指南代表了对可竞争性逻辑的部分拒绝？什么经验教训导致了这一转变？

6.11 延伸阅读

★ 必读（本章核心参考）

- **Baumol, W.J. (1982).** "Contestable Markets: An Uprising in the Theory of Industry Structure." *American Economic Review*, 72(1), 1–15. ——AEA 主席致辞。可竞争性理论最易读、最有力的单篇文章陈述。注意 Baumol 如何首先陈述三个核心定理，然后引出政策含义。14 页涵盖整个研究纲领。

- **Baumol, W.J., Panzar, J.C., & Willig, R.D. (1982).** *Contestable Markets and the Theory of Industry Structure*. Harcourt Brace. Ch.1–3. ——核心文本。特别注意固定成本与沉没成本的区别（Ch.1）、可维持性和弱无形之手定理（Ch.2）、以及多产品成本概念（Ch.3）。

- **Borenstein, S. (1992).** "The Evolution of U.S. Airline Competition." *Journal of Economic Perspectives*, 6(2), 45–73. ——对航空放松管制后竞争的权威综合。清晰阐述了为什么枢纽市场的结果与可竞争性预测背道而驰。可读性极强——JEP 风格。

★★ 进阶（加深理解）

- **Demsetz, H. (1968).** "Why Regulate Utilities?" *Journal of Law and Economics*, 11(1), 55–65. ——可竞争性理论的直接前身。"为市场竞争" vs. "市场内竞争" 的经典表述。注意 Demsetz 关注的是显式特许经营权拍卖，而 BPW 将洞察推广至隐式持续进入威胁。

- **Colton, R. (2022).** AT&T 制度背景与可竞争性理论的政策应用。关于该理论如何被部署为放松管制正当性的制度分析。

- **Shepherd, W.G. (1984).** "'Contestability' vs. Competition." *American Economic Review*, 74(4), 572–587. ——最有影响力的批判。即使你最终不同意 Shepherd，" 同义反复" 论证是必须认真对待的。

- **Schwartz, M. & Reynolds, R.J. (1983).** "Contestable Markets: An Uprising in the Theory of Industry Structure: Comment." *American Economic Review*, 73(3), 488–490. ——微小沉没成本批判。建议在同一期 AER 中读 Schwartz & Reynolds、Weitzman 的评论和 BPW 的回复，以了解辩论的完整结构。

- **Dixit, A. (1980).** "The Role of Investment in Entry-Deterrence." *Economic Journal*, 90(357), 95–106. ——作为承诺工具的产能的博弈论模型。可竞争性理论说" 进入威胁约束在位者"，Dixit 说" 在位者通过沉没投资约束进入威胁"。哪个正确？这取决于承诺是否可能。

- **Morrison, S.A. & Winston, C. (1987).** "Empirical Implications and Tests of the Contestability Hypothesis." *Journal of Law and Economics*, 30(1), 53–66. ——对可竞争性的最直接经验检验。区分完全与不完全可竞争性。方法比结论更值得关注——如何检验一个市场是否可竞争？

★★★ 前沿（当代重新评价与当前争论）

- **Sutton, J. (1991).** *Sunk Costs and Market Structure.* MIT Press. Ch.1–2. ——外生 vs. 内生沉没成本。为" 沉没成本在什么程度上构成进入壁垒" 这个程度问题提供了最系统的处理。

- **Morrison, S.A. (2001).** "Actual, Adjacent, and Potential Competition: Estimating the Full Effect of Southwest Airlines." *Journal of Transport Economics and Policy*, 35(2), 239–256. ——将竞争效应分解为实际、邻近和潜在渠道。发现潜在竞争确实重要——但机制是限制性定价而非" 打了就跑" 进入。参见 第 ~7~ 章的进入阻挠模型。

- **Salop, S.C. (2024).** "The 2023 Merger Guidelines: A Post-Chicago and Neo-Brandeisian Integration." *Review of Industrial Organization*, 65(1), 79–128. ——2023 年指南如何从可竞争性逻辑转向新生竞争和结构性推定。参见 第 ~15~ 章。

- **De Streel, A. (2024).** "The Digital Markets Act and Contestability." *Journal of Antitrust Enforcement.* ——DMA 如何试图通过规制使数字市场具有可竞争性——Baumol 的智识遗产如何被翻转。

下一章：→ 第 ~7~ 章　博弈论重塑产业组织：Tirole 的统一框架——可竞争性理论假定在位者是被动的：面对进入时，他们坐视不管，让进入者捕获市场。但如果他们**不**被动呢？如果——通过策略性地沉没产能、建立品牌、或锁定客户——在位者可以**承诺**在进入发生时采取激进行为呢？进入阻挠变成了一个博弈——而博弈论是分析它的语言。

第七章　博弈论重塑产业组织：Tirole 的统一框架

章首语

1988 年，Jean Tirole 出版了 *The Theory of Industrial Organization*。这本书的第一句话不是关于市场或企业的——而是关于博弈论的。在整整 400 页的分析之后，他将最后一章命名为” 非合作博弈论：用户手册”——一本应该” 放在书的开头” 的参考工具，却被他故意放在最后。

这个看似怪异的决定体现了一个深刻的教学洞见：博弈论是在解决具体 IO 问题的过程中被需要的，而不是先学一堆抽象概念再去找应用。学生在求解 Cournot 模型时已经在使用 Nash 均衡——只是不知道它的名字。在分析 Stackelberg 先发优势时已经做了逆向归纳——只是不知道它叫子博弈完美均衡。

30 年后，Tirole 因” 对市场势力和规制的分析” 获得诺贝尔经济学奖。瑞典皇家科学院的颁奖词说，他” 驯服了那些拥有市场势力的大公司”。但更准确地说，他驯服的是*关于市场势力的思考*——他证明了非合作博弈论可以为一门原本碎片化的学科提供统一的理论语言。

本章讲述这个故事。它既是一本为 IO 量身定做的” 博弈论用户手册”，也是一部关于 IO 本身如何在 1980 年代被重新发明的思想史。前六章讲述了 IO 的” 前史”——从 Cournot 的直觉天才到可竞争市场理论的智识极限。本章展示博弈论如何将这一切重塑为一门严谨的分析科学。

7.1　核心问题

读完本章，你应该能够回答以下问题：

1. **为什么 IO 需要博弈论？** SCP 范式和芝加哥价格理论分别遗漏了什么？博弈论填补了哪些根本性空白？

2. **核心博弈概念如何在 IO 中起作用？** Nash 均衡 → Cournot 和 Bertrand；子博弈完美均衡（SPE）→ Stackelberg 和进入承诺；Bayesian Nash 均衡（BNE）→

拍卖和限价信号；重复博弈 → 合谋；完美 Bayesian 均衡（PBE）→ 声誉和掠夺。

3. **Tirole (1988) 如何组织现代 IO？**为什么”垄断力量的行使”（Part I）在”策略互动”（Part II）之前？博弈论用户手册为什么在第 11 章（末尾）？

4. **什么是一个好的 IO 博弈模型？**”Tirole 七步法”——玩家、策略、时序、信息、支付、均衡概念、比较静态——作为思考任何 IO 问题的系统框架。

5. **博弈论革命如何改变了 IO 的方法论？**从”描述和分类”到”模型化并求解”——旧 IO 与新 IO 在核心问题、分析方法、理论地位、政策含义上的系统性差异。

6. **博弈论 IO 有什么局限？**多重均衡（”任何事都可以发生”）、经验相关性的丧失、对制度细节的敏感——以及 NEIO 和 Sutton 边界方法如何回应这些批评。

7.2　博弈论以前的 IO：困境与缺口

7.2.1　SCP 的”没有理论的实证”

【核心】

第 ~3~ 章–第 ~4~ 章呈现了哈佛 SCP 传统的核心回归方程：

$$\text{PCM}_i = \beta_0 + \beta_1 \cdot \text{CR4}_i + \beta_2 \cdot \text{MES}_i + \varepsilon_i \tag{7.1}$$

其中 PCM 为价格-成本边际，CR4 为四企业集中度，MES 为最小有效规模。

这个方程的两个根本性缺陷无法在 SCP 范式内部解决：

1. **没有策略互动的概念**。回归隐含假设市场结构（CR4）是外生的——由技术条件（MES）决定——企业只能被动接受，根据给定的结构选择行为。但在现实中，企业通过定价、投资、进入遏制*主动塑造*市场结构。SCP 无法处理这个反馈回路：结构影响行为，但行为（通过进入遏制、合并、产能投资）也改变结构。

2. **没有区分行为机制的理论基础**。即使 $\beta_1 > 0$ 显著（集中度与利润正相关），方程无法告诉我们*为什么*。两种完全不同的机制产生同样的相关性：

   - **合谋假说**：高集中度 → 企业更容易协调合谋 → 提高加成率 → 应加强反垄断
   - **效率假说**（Demsetz 1973）：高效率企业增长 → 市场份额增加 → 集中度上升 + 利润上升 → 不应干预

   回归本身无法区分这两种机制——它们给出了完全相反的政策处方。

7.2.2 芝加哥的”没有策略互动的价格理论”

【核心】

第 ~5~ 章展示了芝加哥学派如何有效地批判了 SCP——Demsetz 的效率假说、Stigler 的壁垒重定义、Posner-Bork 的反垄断革命。但其自身的价格理论工具箱存在一个根本的盲区。

Marshall 的供需分析框架只能处理两种极端情形：

- **完美竞争**：原子式企业，价格接受者
- **垄断**：一家企业，价格制定者

而中间地带——2 到 8 家企业的寡头市场——恰恰是大多数 IO 问题的所在地——在这个框架中缺乏分析语言。芝加哥的核心论断” 市场是自我修正的”（第 ~5~ 章 §5.2）依赖于一个隐含假设：进入威胁足够约束在位者。

第 ~6~ 章将这一假设推向了逻辑极限：如果进入壁垒真的为零（包括沉没成本），那么即使垄断也产生竞争性结果——这就是可竞争市场理论。但正如我们在第 ~6~ 章 §6.4–§6.6 所见，经验证据（航空枢纽溢价）和理论批判（Dixit 产能承诺）一致显示：在位者可以通过策略性行动来改变进入条件——而这正是芝加哥价格理论——没有博弈论工具箱——无力分析的行为。

7.2.3 博弈论填补的三个空白

【核心】

表 7.1: 博弈论填补的三个根本空白

空白	博弈论的解答	关键概念
战略相互依赖	我的最优选择取决于你的选择 → 不动点条件：每个人的策略是其他每个人策略的最优反应	反应函数、Nash 均衡
动态与承诺	当企业序贯行动时，什么威胁是可信的？→ 要求每个子博弈中的策略都是最优的	逆向归纳、可置信威胁、沉没成本作为承诺工具
信息不对称	当企业不知道对手的成本/需求/意图时，如何形成最优策略？→ 类型依存的策略 + Bayes 信念更新	类型空间、信念、信号传递

你已经在使用博弈论——只是不知道它的名字

回顾你在前六章中已经做了什么：

- 在 Cournot 模型中（第 ~2~ 章 §2.1），你计算了反应函数的交点——这就是 **Nash 均衡**。
- 在 Stackelberg 模型中（第 ~2~ 章 §2.2），你用逆向归纳求解了两阶段博弈——这就是**子博弈完美均衡**。
- 在可竞争性批判中（第 ~6~ 章 §6.6），你看到没有沉没成本承诺工具的进入威胁是不可信的——这正是 **SPE 排除不可置信威胁**的逻辑。

本章的教学任务不是向你介绍全新的概念——而是将你已经积累的" 模糊的直觉" 转化为精确的博弈论语言，并展示这些工具的系统应用范围远超你迄今为止看到的。

7.3　博弈论工具箱：核心概念与 IO 应用

在展开具体概念之前，先引入一个将贯穿全章的**结构性装置**，也是本章对读者最重要的方法论贡献：

Tirole 七步建模法——本章的"北星"

任何 IO 问题都可以通过以下七步转化为一个定义明确的博弈：

1. **玩家（Players）**：谁做决策？企业（利润最大化）、消费者（效用最大化）、规制者（社会福利最大化）。
2. **策略空间（Strategy spaces）**：每个玩家能做什么？价格、数量、产能投资、广告、研发、进入/不进入……
3. **时序（Timing）**：谁在何时行动？同时（静态博弈）还是序贯（动态博弈）？一期还是重复？
4. **信息结构（Information）**：谁在行动时知道什么？完全信息还是不完全信息？完美还是不完美？
5. **支付函数（Payoffs）**：作为策略的函数，每个玩家的期望收益是什么？
6. **均衡概念（Equilibrium concept）**：根据 [时序]+[信息] 选择合适的解概念。
7. **比较静态与福利（Comparative statics & welfare）**：参数变化如何改变均衡？均衡是否有效率？

这七步构成了一个**决策流程图**（图7.1）。本章及后续所有章节中，每当引入一个新模型，我们都会以这七步格式呈现。

7.3.1 Nash 均衡 → Cournot 与 Bertrand

【核心】

七步设定：

- **玩家：** n 家企业，$i = 1, \ldots, n$
- **策略：** $q_i \in [0, \infty)$（Cournot）或 $p_i \in [0, \infty)$（Bertrand）
- **时序：** 同时行动
- **信息：** 完全信息（所有支付为共同知识）
- **支付：** $\pi_i = (P - c) \cdot q_i$，其中 $P = a - b\sum_j q_j$（Cournot）或由需求系统决定（Bertrand）
- **均衡概念：** Nash 均衡

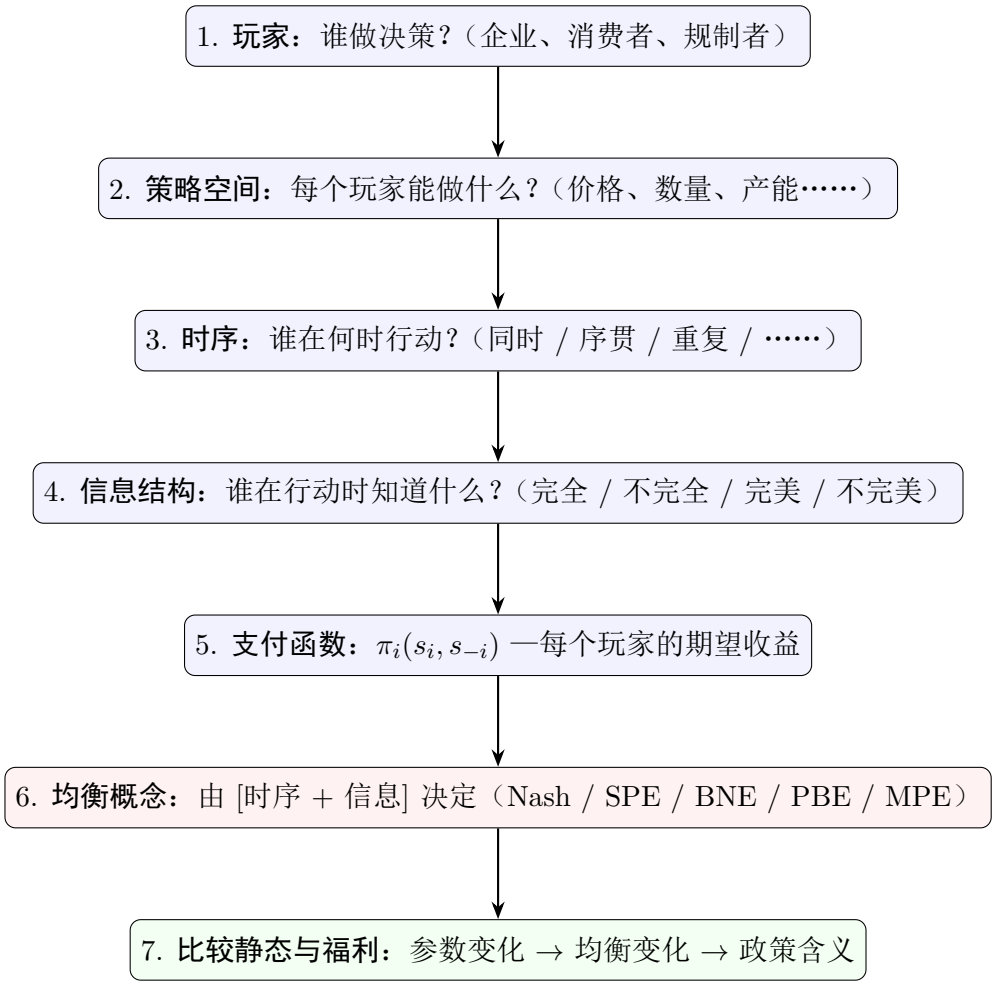


图 7.1: Tirole 七步建模法——任何 IO 问题的系统分析框架

定义 7.1 (Nash 均衡（Nash Equilibrium）). 策略组合 $(s_1^*, \ldots, s_n^*)$ 是一个 *Nash* 均衡，当且仅当对每个玩家 i 和每个替代策略 s_i：

$$\pi_i(s_i^*, s_{-i}^*) \geq \pi_i(s_i, s_{-i}^*) \tag{7.2}$$

即：给定所有其他玩家的均衡策略，没有玩家可以通过单方面偏离获利。

Cournot-Nash 均衡（回顾第 ~2~ 章 §2.1）。企业 i 的利润：$\pi_i = (a - b\sum_j q_j - c) \cdot q_i$。一阶条件（对 q_i 求导并令其为零）：

$$\frac{\partial \pi_i}{\partial q_i} = a - bQ_{-i} - 2bq_i - c = 0 \quad \Rightarrow \quad q_i^*(Q_{-i}) = \frac{a - c - bQ_{-i}}{2b} \tag{7.3}$$

这就是企业 i 的**最优反应函数**（best response function）——给定对手的总产出 Q_{-i}，企业 i 的最优产出。对称均衡条件（$q_i = q^*, \forall i$）给出：

$$q^* = \frac{a - c}{b(n+1)}, \quad Q^* = \frac{n(a-c)}{b(n+1)}, \quad P^* = \frac{a + nc}{n+1} \tag{7.4}$$

Bertrand-Nash 均衡。同质产品、价格竞争 $\rightarrow$ 关键的不连续性：如果 $p_i < p_j$，企业 i 获得整个市场；如果 $p_i = p_j$，各得一半。任何 $p_i = p_j = p > c$ 都不是 Nash 均衡——每家企业都有动机设定 $p_i = p - \varepsilon$ 来捕获整个市场。这种" 降价螺旋" 一直持续到 $p_i = p_j = c$——此时进一步降价意味着亏损。唯一纯策略 Nash 均衡：$p_1^* = p_2^* = c$，$\pi_1^* = \pi_2^* = 0$。

这就是 **Bertrand 悖论**：即使只有两家企业，价格竞争也足以产生完全竞争结果。

两个模型的根本对比：

表 7.2: Cournot vs. Bertrand ——同一博弈形式，不同策略空间

	Cournot（数量）	**Bertrand（价格）**
策略空间	$q_i \in [0, \infty)$	$p_i \in [0, \infty)$
反应函数	向下倾斜（**战略替代品**）	向上倾斜（**战略互补品**，差异化）
$n = 2$ 时的均衡	$P > c$（正利润）	$P = c$（零利润，同质）
$n \to \infty$	渐进趋近 $P = c$	已收敛至 $P = c$（同质）
核心直觉	数量承诺缓和竞争	价格竞争激烈直到无利可图

定义 7.2 (战略替代品 vs. 战略互补品（Bulow, Geanakoplos & Klemperer 1985）). 企业 j 的策略 s_j 增加时，对企业 i 的最优反应的影响：

- **战略替代品**（*strategic substitutes*）：$\frac{\partial^2 \pi_i}{\partial s_i \partial s_j} < 0$ —反应函数向下倾斜。我的最优策略随你的策略增加而减少。*Cournot* 数量竞争是典型例子。

- **战略互补品**（*strategic complements*）：$\frac{\partial^2 \pi_i}{\partial s_i \partial s_j} > 0$ —反应函数向上倾斜。我的最优策略随你的策略增加而增加。*Bertrand* 价格竞争（差异化产品）是典型例子。

这一区分不仅影响均衡的比较静态——它还决定了企业最优的"强硬"或"温和"投资策略（见 §~7.3.3）。

为什么 Kreps-Scheinkman（1983）是一个关键结果

Kreps 和 Scheinkman（1983）证明了：如果企业先选择产能（K_i），然后进行 Bertrand 价格竞争，子博弈完美均衡就是 *Cournot* 结果。直觉：产能投资作为承诺机制——如果两家企业都选择了"Cournot 产能"，每家企业都无法独自服务整个市场，因此没有动机大幅降价（因为不能服务激增的需求）。产能承诺有效地将 Bertrand 竞争转化为 Cournot 竞争。

这一结果的深层含义：**选择什么模型（Cournot 还是 Bertrand）取决于制度细节——最重要的是是否存在产能约束。**

7.3.2　子博弈完美均衡 → Stackelberg 与 Dixit 承诺

【核心】

七步设定（Stackelberg）：

- **玩家**：企业 1（领导者），企业 2（追随者）
- **策略**：q_1（先），$q_2(q_1)$（后，观察到 q_1 之后选择）
- **时序**：序贯——企业 1 先动，企业 2 后动
- **信息**：完全信息——企业 2 观察到 q_1
- **支付**：$\pi_i = (a - b(q_1 + q_2) - c) \cdot q_i$
- **均衡概念**：子博弈完美均衡（SPE）

定义 7.3 (子博弈完美均衡（Subgame Perfect Equilibrium, SPE）). 一个策略组合是一个 *SPE*，当且仅当在每个子博弈中它都构成 *Nash* 均衡。这意味着策略不仅在均衡路径上是最优的——在每一个可能到达的决策节点上也都是最优的。*SPE* 的核心功能是消除**不可置信的威胁**（*non-credible threats*）。

Stackelberg 均衡的逆向归纳。

- **阶段 2（追随者）**：企业 2 观察到 q_1 后，最大化 $\pi_2 = (a - b(q_1 + q_2) - c)q_2$。

$$q_2^*(q_1) = \frac{a - c - bq_1}{2b}$$

- **阶段 1（领导者）**：企业 1 预见 $q_2^*(q_1)$，最大化 $\pi_1 = (a - b(q_1 + q_2^*(q_1)) - c)q_1$。

$$q_1^* = \frac{a-c}{2b}, \quad q_2^* = \frac{a-c}{4b}$$

先发优势：$\pi_1^{\text{Stack}} = (a-c)^2/(8b) > \pi_1^{\text{Cournot}} = (a-c)^2/(9b)$。领导者通过承诺一个产量水平获得更高利润——因为它知道追随者将最优反应（retreat），而非匹配。

Dixit (1980) 承诺模型（与第 ~6~ 章 §6.6 衔接）：

- **阶段 1（投资）**：在位者选择产能 K_1（每单位成本 $c_K = 5$）。K_1 是沉没成本——投入后不可回收。
- **阶段 2（竞争）**：进入者决定是否进入（进入成本 F）→ 如进入，双方 Cournot 竞争。
- **关键机制**：因为 K_1 是沉没的，在位者的事后边际成本在 $q \le K_1$ 区间仅为可变成本（低），超过 K_1 才产生额外成本。进入者知道：一旦进入，在位者有动机利用已投资的产能——使进入者的剩余需求更小、进入后利润更低。

Dixit (1980) vs. Spence (1977) ——SPE 如何改变结论

Spence (1977) 直观地认为在位者可以通过持有过剩产能来阻挠进入。Dixit 用 SPE 证明这是不可置信的：一旦进入实际发生，在位者的利润最大化选择是充分利用（而非闲置）产能。因此，只有在产能投资改变了在位者第二阶段激励（即降低了进入后的竞争边际成本）时，它才能阻挠进入。

这个微妙的逻辑修正——从" 过剩产能威胁" 到" 沉没投资改变激励"——是博弈论如何将 IO 从直觉提升到分析的完美范例。

可竞争性批判的重申（延续第 ~6~ 章 §6.6）。可竞争市场理论（第 ~6~ 章 §6.2）假定在位者是被动的——进入者捕获市场后，在位者才调整价格。Dixit 模型展示了为什么这是一个不可置信的假设：在位者可以通过沉没投资事先改变事后激励，从而策略性地阻止进入。**可竞争性只有在零沉没成本——即没有任何承诺工具——的世界中才成立**。一旦存在沉没成本，在位者有承诺工具——可竞争性崩溃。

7.3.3 Fudenberg-Tirole 策略动物分类学

【进阶】

Fudenberg 和 Tirole (1984) 提供了统一分析在位者第一阶段投资策略的框架。分类依赖于两个维度：

1. **投资使在位者" 更强硬" 还是" 更温和"？**
   - **强硬**（tough）：投资提高了在位者的边际利润，使其在第二阶段更激进（如降低成本的投资）

- **温和**（soft）：投资降低了在位者的边际利润，使其在第二阶段更温和（如高质量广告提升品牌价值但不降成本）

2. **第二阶段的竞争是战略替代品还是战略互补品？**

- 战略替代品（Cournot）：对手收缩 → 我扩张（反应函数向下倾斜）
- 战略互补品（Bertrand）：对手降价 → 我也降价（反应函数向上倾斜）

表 7.3: Fudenberg-Tirole (1984) 策略动物分类学

		第二阶段竞争类型	
		战略替代品（Cournot）	**战略互补品（Bertrand）**
投资使在位者…	**强硬**	**领头狗（Top Dog）** 过度投资。显得激进，阻挠进入。 例：*Dixit (1980)* 产能投资	**小狗（Puppy Dog）** 投资不足。避免显得激进，防止价格战。 例：避免成本降低以避免触发对手降价
	温和	**饥肠辘辘（Lean and Hungry）** 投资不足。保持精干和激进。 例：不投资品牌广告以保持低价威胁	**肥猫（Fat Cat）** 过度投资。显得温和，让进入者也温和定价。 例：投资高质量广告 → 高成本 → 高价格 → 对手也高价格

为什么"Dixit 产能投资

Dixit 的产能投资降低了在位者的第二阶段边际成本——使其更强硬。面对的是 Cournot 竞争——战略替代品。当在位者变得更强硬时，进入者理性地收缩（反应函数向下倾斜）。因此，在位者有动机过度投资产能，锁定一个更具攻击性的进入后姿态。

7.3.4　Bayesian Nash 均衡 → 拍卖与不完全信息寡头

【核心】

Harsanyi 转换（1967-68）。将不完全信息博弈——每个玩家知道自己类型的成本/估值，但不知道他人的——转化为不完美信息博弈：自然（Nature）先抽取每个

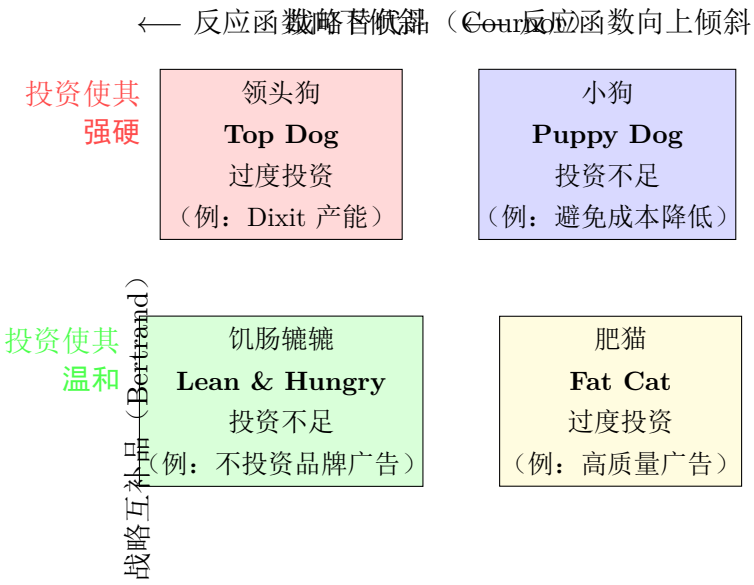


图 7.2: Fudenberg-Tirole 策略动物分类学的结构

玩家的" 类型"$t_i \in T_i$，每个玩家知道自己的类型但只知道他人类型的分布。策略是类型依存的（type-contingent）：$s_i(t_i)$。

定义 7.4 (Bayesian Nash 均衡（Bayesian Nash Equilibrium, BNE）). *一个策略组合 $(s_1^*(\cdot), \ldots, s_n^*(\cdot))$ 是一个 BNE，当且仅当对每个玩家 i、每种类型 t_i、每个替代策略 s_i：*

$$\mathbb{E}_{t_{-i}}\left[\pi_i(s_i^*(t_i), s_{-i}^*(t_{-i}); t_i, t_{-i}) \mid t_i\right] \geq \mathbb{E}_{t_{-i}}\left[\pi_i(s_i, s_{-i}^*(t_{-i}); t_i, t_{-i}) \mid t_i\right] \tag{7.5}$$

即：给定信念（关于他人类型的分布），每种类型的策略必须是期望最优。

IO 应用一：拍卖。 n 个竞标者参与一级价格拍卖。每个竞标者的私有价值 $v_i \sim$ i.i.d. F（独立同分布）。出价函数 $b(v_i)$ 是类型依存的策略。对称 BNE 出价函数（在均匀分布 $F(v) = v$ 下）：

$$b(v) = \frac{n-1}{n} \cdot v \tag{7.6}$$

直觉：每个竞标者策略性地将出价低于真实估值（bid shading），且 n 越大，折扣越小（竞争越激烈 $\to$ 出价越接近真实估值）。

IO 应用二：限价作为信号（Milgrom-Roberts 1982，预览 Ch.8）。在位者的成本类型是私人信息（高 c_H 或低 c_L）。第一阶段，在位者选择 p_1；进入者观察到 p_1 但不观察成本类型，然后决定进入/不进入。这就是一个不完全信息动态博弈——需要在 §7.3.5 引入 PBE 后才能完整分析。

7.3.5　重复博弈与无名氏定理 $\to$ 合谋

【核心】

七步设定（重复 Cournot 合谋）：

- **玩家**：n 家企业
- **策略**：每期选择 q_{it}
- **时序**：无限期重复，贴现因子 $\delta = 1/(1+r)$
- **信息**：完全信息（每期结果可完美观察到——本节约定）
- **支付**：折现利润和 $V_i = \sum_{t=0}^{\infty} \delta^t \pi_{it}$

囚徒困境作为合谋结构。在一次性 Cournot 中，每家企业有动机比合谋水平生产更多（因为给定对手限制产出，增加自己的产出有利于短期利润）。但如果博弈无限重复 + δ 足够大，**背叛的短期收益可能被未来惩罚的现值所抵消**。

触发策略（Grim Trigger）。开始时合作（生产合谋产量 q^c）。如果任何企业曾在任何期背叛，从下一期开始永远回到 Nash 均衡 q^n。

合谋可持续性条件。

$$\text{合谋的现值：} \quad V^{\text{coll}} = \frac{\pi^{\text{coll}}}{1-\delta} \tag{7.7}$$

$$\text{背叛 + 惩罚：} \quad V^{\text{dev}} = \pi^{\text{dev}} + \frac{\delta \cdot \pi^{\text{Nash}}}{1-\delta} \tag{7.8}$$

合谋可持续当且仅当 $V^{\text{coll}} \geq V^{\text{dev}}$：

$$\delta \geq \delta^* \equiv \frac{\pi^{\text{dev}} - \pi^{\text{coll}}}{\pi^{\text{dev}} - \pi^{\text{Nash}}} \tag{7.9}$$

对于线性需求 Cournot 双寡头（$a - c = 1$ 表示）：

- $\pi^{\text{coll}} = \pi^m/2 = 1/8$（各生产垄断产量一半）
- $\pi^{\text{dev}} = 9/64$（对手合作，我背叛的最优产量）
- $\pi^{\text{Nash}} = 1/9$（Cournot-Nash）
- $\delta^* = \frac{9/64 - 1/8}{9/64 - 1/9} = \frac{1/64}{17/576} = \frac{9}{17} \approx 0.529$

这意味着：如果年利率 $r \leq 89\%$，合谋在离散时间（以年为单位）下是可以维持的。

命题 7.1 (无名氏定理（Folk Theorem）——直觉). *对于 δ 足够接近 1，在无限重复博弈中任何满足个人理性的支付向量都可以作为 SPE 被支持。对 IO 的核心含义：给定足够的耐心，企业可以在任何水平上维持合谋——从 Cournot 利润到完全垄断利润。博弈论没有告诉我们哪个均衡会发生——它只告诉我们哪些结果是可能的。*

两个关键拓展：

1. **Green-Porter (1984) ——不完美监控**。当需求受随机冲击影响且企业无法直接观察对手的产量时，价格战在均衡路径上发生。低于阈值的市场价格触发惩罚期——即使实际上没有企业背叛（价格低可能仅因需求冲击）。均衡价格战不是错误——它们是维持威慑可信性的必要代价。

2. **Rotemberg-Saloner (1986) ——繁荣期的价格战**。当需求高时，背叛的诱惑更大（当前需求高）但未来惩罚的期望不变（需求回归均值）。为维持合谋，企业在繁荣期必须降低合谋价格。这意味着加成率可能是反周期的——繁荣期价格低，衰退期价格高——与直觉相反，但来自博弈论逻辑。

7.3.6 完美 Bayesian 均衡 → 信号传递与声誉

【核心】

当不完全信息遇到序贯行动时，博弈同时涉及策略精炼（如 SPE）和信念更新（如 BNE）。完美 Bayesian 均衡（PBE）将两者结合起来。

定义 7.5 (完美 Bayesian 均衡（Perfect Bayesian Equilibrium, PBE）). 一个 *PBE* 包括：*(1)* **策略**：每个信息集上每个玩家的策略；*(2)* **信念**：每个玩家在每个信息集上关于过去未观察事件的概率分布。这些必须满足：

- **序贯理性**：给定信念，在每个信息集上的策略都是最优的
- **信念一致性**：
  - 在均衡路径上：信念通过 *Bayes* 规则更新
  - 在非均衡路径上：信念必须"合理"（由具体的精炼概念——如 *Cho-Kreps* 直观准则——来界定）

Milgrom-Roberts (1982) 限价信号（概要，详细推导见第 ~8~ 章 §8.1.3）。

- 在位者的成本类型 $\theta \in \{c_L, c_H\}$ 是私人信息——进入者只知道先验概率 $\mu_0 = \Pr(c = c_L)$
- 第一阶段：在位者选择 p_1（可能低于短期垄断价格——限价）
- 进入者观察到 p_1，更新信念 $\mu(p_1) = \Pr(c = c_L \mid p_1)$，决定进入/不进入
- **分离均衡**：低成本在位者设置 p_1 低于高成本在位者可以模仿的水平 → 进入者推断 p_1 低 ⇒ 低成本 ⇒ 不进入
- **混同均衡**：两类在位者设置相同价格 → 没有信息被传递 → 进入者基于先验信念做决策

- Cho-Kreps (1987) **直观准则**（Intuitive Criterion）：消除" 不合理" 的混同均衡——如果低成本类型永远可以从某个偏离中获益（在高成本类型不会选择该偏离的价格下），那么观察到该偏离的进入者应将信念集中在低成本类型上

Kreps-Wilson (1982) / " 四人帮" ——连锁店悖论的解决。

- Selten (1978) **连锁店悖论**：一个连锁店同时面对 N 个市场的潜在进入者（依次进入）。完全信息下，逆向归纳（SPE）预测：在位者在最后一个市场中总是容纳（因为没有后续市场需要威慑）$\to$ 倒数第二个市场也容纳（因为最后一个已确定容纳）$\to$ ……$\to$ 第一个市场就容纳。掠夺从不发生。

- **悖论**：这与直觉和商业现实完全相悖。

- **Kreps-Wilson 的解决**：引入不完全信息——在位者以极小概率 $\varepsilon > 0$ 是" 强硬型"（总是掠夺）。结果是：即使 ε 非常小，理性的" 软弱型" 在位者在早期市场也有动机模仿强硬型——通过掠夺来建立声誉——以威慑后期市场的进入者。

- **核心洞见**：芝加哥学派（第 ~5~ 章 §5.4）说" 掠夺从不理性"，但这是完全信息下的结论。引入不完全信息后——即使是微小的不完全信息——掠夺可以是完全理性的声誉投资。

声誉均衡的经济学

在重复互动中，当前的行动不仅影响当前的利润——它还通过改变他人对未来的信念来影响未来的利润。这就是声誉效应的核心：

当前掠夺的损失 $<$ 未来各期因" 强硬" 声誉而获得的更高利润的现值

当 δ 足够大时，" 声誉租金" 超过掠夺的短期成本——即使对弱在位者也如此。这为掠夺性定价、限价和其他看似" 非理性" 的策略行为提供了博弈论基础。

7.3.7　Markov 完美均衡 $\to$ 动态寡头

【前沿】

定义 7.6 (Markov 完美均衡（Markov Perfect Equilibrium, MPE）). *一个 MPE 是随机博弈中策略仅依赖于支付相关的状态变量（payoff-relevant state）——而非完整博弈历史——的 SPE。MPE 极大地简化了动态寡头分析：企业不需要跟踪完整历史的每一个细节，只跟踪那些直接影响未来支付的变量。*

Maskin-Tirole (1988a, 1988b) ——交替行动动态寡头。

- 企业 1 在奇数期选择 q_1 并承诺一期；企业 2 在偶数期选择 q_2

- 状态变量：对方当前约束的产量
- 值函数满足 Bellman 方程：

$$V_1(q_2) = \max_{q_1} \left[\pi_1(q_1, q_2) + \delta \cdot V_2(q_1)\right], \quad V_2(q_1) = \max_{q_2} \left[\pi_2(q_2, q_1) + \delta \cdot V_1(q_2)\right] \tag{7.10}$$

- 核心结果：对大固定成本，MPE 表现为自然垄断——但活跃企业定价低于短期垄断价格（限价效应），以阻挠进入
- 随着承诺期缩短，MPE 结果趋近可竞争性（Bertrand）——为" 可竞争性何时成立" 提供了博弈论微基础

Ericson-Pakes (1995) 将 MPE 扩展到企业通过 R&D 投资改善" 质量" 状态，且进入和退出内生发生的行业演化模型——这是当代 NEIO 结构性估计的基石（见第 ~10~ 章–第 ~12~ 章）。

7.4 Tirole (1988) 的体系结构

7.4.1 总体结构

【核心】

为什么垄断先于策略互动？

Tirole 的组织逻辑包含三个层次：

1. 垄断是寡头的基本构件。在博弈论寡头模型中，每家企业的利润函数首先需要被定义——而垄断模型提供了这个" 原子"：给定一个残余需求，企业如何最优定价？质量选择、价格歧视、纵向约束——这些都是寡头环境中每家企业面临决策的微观基础。

2. 许多 IO 问题本身不涉及企业间的策略依赖。垄断定价的时间不一致性（Coase 猜想）、激励相容的价格歧视菜单、纵向外部性——这些问题的分析已经足够丰富和复杂，不需要同时引入企业间的策略互动。

3. 教学递进。让学生在" 简单" 环境（非策略性定价问题）中熟悉 IO 的语言和工具，再引入策略互动的复杂性——类似于先学单一变量微积分再学博弈论。

7.4.2 为什么" 博弈论用户手册" 在最后一章？

【核心】

这是一个著名的教学决定。Tirole 的哲学是" **先用它，后命名它**"：

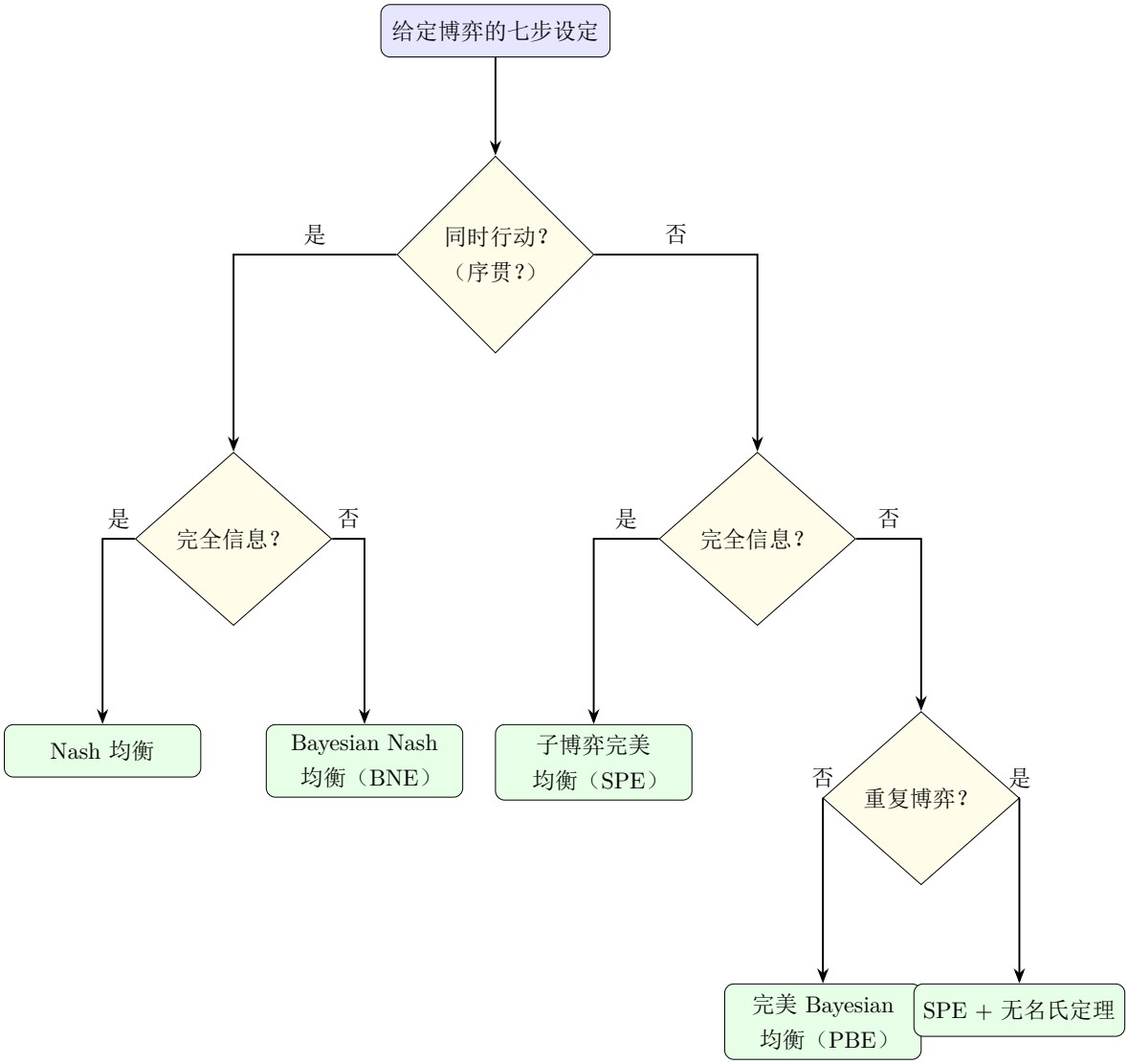


图 7.3: 博弈论均衡概念决策树：根据 [时序] + [信息] 选择正确的均衡概念

表 7.4: Tirole (1988) 的体系结构

部分	章节	内容	核心博弈概念
绪论	—	企业理论（企业为什么存在?）	—
第一篇：垄断力量的行使	Ch.1	垄断定价（单一产品、多产品、耐用品）	Coase 猜想、时间不一致性
	Ch.2	产品质量与产品选择	信息不对称
	Ch.3	价格歧视	激励相容
	Ch.4	纵向控制（RPM、排他性区域）	纵向外部性
第二篇：策略互动	Ch.5	短期价格竞争	**Nash 均衡**
	Ch.6	动态价格竞争与默契合谋	**重复博弈、无名氏定理**
	Ch.7	产品差异化	两阶段定位-价格博弈
	Ch.8	进入、容纳与退出	**SPE、Dixit、Fudenberg-Tirole**
	Ch.9	信息与策略行为	**PBE、信号、声誉**
	Ch.10	研发与技术采纳	专利竞赛、网络外部性
附录	Ch.11	非合作博弈论：用户手册	**所有概念的正式定义**

- 读者在 Ch.5 中求解 Cournot 模型——实际上已经使用了 Nash 均衡，只是不知道这个名称
- 读者在 Ch.6 中分析触发策略——实际上已经使用了 SPE 的逻辑，只是没有被形式化告知
- 读者在 Ch.9 中遇到信号传递——实际上已经使用了 PBE，只是在 Ch.11 中才获得精确的数学定义

Ch.11 的功能是**事后形式化**：将学生通过反复应用积累的直觉转化为精确的博弈论概念。

本章的策略（改良的混合方法）：采用 Belleflamme & Peitz (2015) 的改进方案——在 §~7.3中提前提供压缩版博弈论工具箱（每个概念立即应用于 IO），但保留 Tirole 的" 问题驱动" 精神——博弈论是工具，不是主角。

从" 工具箱" 到" 用户手册"

本章的 §~7.3可以看作 Tirole Ch.11 的" 前置压缩版"——但两者有本质区别：

- Tirole 的 Ch.11：系统的博弈论定义和定理——适用于已经完成了全书学习的读者用于巩固
- 我们的 §~7.3：IO 用途导向的博弈概念——每一概念以" 为什么 IO 需要这个工具" 为动机引入，立即应用于具体 IO 问题

推荐学习路径：先读本章 §~7.3获得 IO 博弈论的工具直觉 → 学完本章 →（可选）阅读 Fudenberg & Tirole (1991) Ch.3-5 获得更系统的博弈论处理。

7.5　深层统一性：不同的 IO 问题，相同的博弈结构

7.5.1　统一性对比表

【核心】

7.5.2　三条统一原理

【核心】

1. **Cournot vs. Bertrand = 同一博弈形式，不同策略空间**。q_i vs. p_i → 反应函数斜率反转（战略替代 vs. 战略互补）→ 完全不同的均衡结果。这说明**策略空间的选择不是建模技巧——它是关于制度现实的实质假设**。Kreps-Scheinkman (1983) 的深刻之处在于：通过两阶段博弈，它证明了" 选择什么模型" 取决于" 是否存在产能约束"——制度细节决定策略空间。

表 7.5: 博弈论统一的 IO 问题——共享的深层结构

IO 问题	玩家	策略	时序	信息	均衡	核心洞见
Cournot	n 企业	q_i	同时	完全	Nash	数量承诺缓和竞争
Bertrand	n 企业	p_i	同时	完全	Nash	价格竞争更激烈
Stackelberg	领导者 + 追随者	q_L, q_F	序贯	完全	SPE	先发优势，承诺价值
Dixit 进入阻挠	在位者 + 进入者	K, E, q	两阶段	完全	SPE	沉没成本 $\rightarrow$ 可信承诺
限价信号	在位者 + 进入者	p_1, E	两阶段	不完全	PBE	价格 $\rightarrow$ 信号 $\rightarrow$ 进入
合谋	n 企业	q_i	无限重复	完全	SPE(trigger)	δ 大 $\rightarrow$ 合谋可持续
Green-Porter	n 企业	q_i	无限重复	不完全	SPE(trigger)	价格战是均衡一部分
声誉	连锁店 +N 进入者	E,F/A	N 期序贯	不完全	PBE	ε 不确定性 $\rightarrow$ 声誉

2. **限价 vs. 产能承诺 = 同一两阶段结构，不同承诺工具。** Milgrom-Roberts 限价（信息作为承诺工具）和 Dixit 产能投资（沉没成本作为承诺工具）都遵循相同的 SPE/PBE 结构：**第一阶段不可逆行动 → 改变第二阶段博弈的激励 → 阻挠/容纳进入**。不同的是承诺工具——前者通过信息不对称（成本类型的信号传递），后者通过改变事后成本结构。

3. **合谋 vs. 进入阻挠 = 同一重复博弈逻辑，不同玩家。**两者都依赖于" 如果你背叛/进入，我将在未来惩罚你" 的重复博弈结构。差别仅在于：合谋发生在现有企业之间；进入阻挠涉及在位者和潜在进入者。深层数学结构完全相同：临界贴现因子、触发策略、无名氏定理。

" 啊哈" 时刻：从记住模型到掌握框架

当学生意识到 Stackelberg、Dixit 产能承诺和可竞争性批判都在说同一个故事——承诺为什么重要——他们的理解就从" 记住几个模型" 跃升到" 掌握了分析任何 IO 问题的通用框架"。这就是 Tirole 七步建模法的力量——它将所有的 IO 问题还原为相同的七个基本元素。

7.6　方法论对比：旧 IO vs. 新 IO

一个关键例子。在旧 IO 框架中，一个市场有 3 家企业且价格高于平均成本 → " 集中导致市场势力" → 应拆分。在新 IO 框架中，完全相同的可观测事实有三种可能的解释：

1. **合谋**（→ 反垄断介入，拆分可能有效）
2. **效率差异**（→ 不应介入，芝加哥解释）
3. **策略性进入阻挠**（→ 应禁止某些策略行为，但不必分拆结构——因为即使拆分，新企业仍可能面临同样的策略性阻挠）

博弈论 IO 的核心信息是：**不能从市场结构直接跳跃到政策结论。**必须先分析策略行为的具体机制——这正是 Ch.8 将要做的事。

7.7　博弈论 IO 的批评与回应

【进阶】

7.7.1　Fisher (1989)：" 任何事都可以发生" 批判

Franklin Fisher 在 *RAND Journal of Economics* 上发表了尖锐的方法论批评。核心指责：

表 7.6: 旧 IO vs. 新 IO ——范式转变的系统对比

维度	**旧 IO（哈佛 SCP）**	**新 IO（博弈论/Tirole）**
基本方法	描述性分类学——” 这个行业集中吗？ ”	分析性均衡——” 给定基本条件，均衡是什么？ ”
核心问题	利润与集中度相关吗？	在什么条件下产生什么策略行为？
理论地位	微弱——回归缺乏微观基础	强——理性行为的严格假设推导均衡
因果关系	单向：结构 → 行为 → 绩效	双向：行为可内生改变结构
进入壁垒	结构性（规模经济、绝对成本优势）	结构性 + 策略性（限价、过度产能）
时间维度	静态	动态（承诺、声誉、学习、演化）
信息	假设完美信息	信息不对称是核心分析变量
均衡概念	没有均衡概念	多层次均衡（Nash → SPE → BNE → PBE → MPE）
实证方法	跨行业会计数据 → R^2 和系数	单一行业详细数据 → 结构估计参数
政策含义	粗放：高集中度 → 自动关注	精细：需分析策略行为的理性基础

- **无名氏定理**意味着：在无限重复博弈中，几乎任何结果——从完全合谋到价格战到随机波动——都可以作为某个均衡被支持
- **均衡选择**意味着：通过选择不同的精炼概念和信念设定，研究者可以”解释”任何观察到的行为
- **后果**：一个理论可以解释一切 → 它什么都解释不了。博弈论没有排除任何行为——因此它不能做出可检验的预测

这是一个严重的智识挑战。连 Kreps 本人后来都承认：”如果模型可以通过选择精炼来产生任何结果，那么模型就不是在做出预测——它是在提供事后的合理化。”

7.7.2 Sutton (1991) 的边界方法：一个建设性回应

John Sutton 的 *Sunk Costs and Market Structure* (1991) 提供了最有创意的建设性回应。

核心思想。与其问”这个市场的确切均衡是什么？”（对博弈形式的细节——时序、信息结构、先验信念——高度敏感），不如问”**什么预测在所有合理的博弈形式中都是稳健的？**”

Sutton 区分两类行业：

1. **外生沉没成本行业**（厂房、设备）：沉没成本固定于技术条件。当市场扩大时，集中度下限趋于零——碎片化不可避免。
2. **内生沉没成本行业**（R&D、广告）：企业内生选择沉没成本投资水平。当市场扩大时，企业在这些成本上展开竞相升级——集中度有正的下限。无论市场多大，集中度不会趋于零。

Sutton 的边界方法避开了”任何事都可以发生”问题——因为它只寻找跨模型稳健的预测（集中度的下限），而不需要选择任何特定的博弈形式。

7.7.3 NEIO 的经验回应（预览第 ~10~ 章）

Bresnahan (1989) 的新实证 IO 纲领提供了第三种回应：**用数据约束模型**。

- 不选择”哪个均衡”，而是从数据中估计行为参数 θ（conduct parameter）
- $\theta = 0$：Bertrand；$\theta = 1/N$：Cournot；$\theta = 1$：完全合谋
- 博弈论模型被用于规范估计方程——数据的拟合度决定了哪个行为假设最可信

现代共识：这三种回应构成了现代 IO 的方法论三角——边界方法（Sutton）+ 结构估计（NEIO）+ 特定行业的制度知识。

7.8 博弈论 IO 的中国脉络

【进阶】

7.8.1 核心中文教材与翻译

表 7.7: 中国博弈论 IO 的核心学习资源

教材	作者/译者	出版社	定位
《产业组织理论》	梯若尔（Tirole），张维迎总译校	中国人民大学出版社 (2015, 2024 新版)	研究生核心教材——Tirole 原著的中文权威译本
《博弈论与信息经济学》	张维迎	上海人民出版社/格致出版社 (1996)	中国最经典的中文博弈论教材——IO 相关章节直接衔接本章
《博弈论基础》	Gibbons（吉本斯），高峰译	中译本	博弈论入门标准教材——中国研究生普遍使用
《产业组织理论》(第 2 版)	李明志、柯旭清、罗金峰	清华大学出版社 (2014)	中国 IO 教材——遵循 Tirole 体系
《产业组织理论》	吴汉洪	中国人民大学出版社	本科 IO 教材——博弈论处理较浅，适合初学者
《博弈论教程：理论 • 应用》	黄涛	北京大学出版社	系统的中文博弈论教材——覆盖 IO 应用

7.8.2 张维迎与博弈论 IO 的中国引入

张维迎是将博弈论系统引入中国产业组织研究的关键人物。他在牛津大学的博士论文运用了博弈论和信息经济学方法分析企业理论。其《博弈论与信息经济学》(1996) 成为中国被引次数最高的经济学著作之一——覆盖 Nash 均衡、SPE、BNE、PBE 等核心概念，并在进入阻挠、信号传递、拍卖等章节直接应用于产业组织问题。

推荐学习路径：本章 §~7.3（IO 博弈论工具）→ 张维迎《博弈论与信息经济学》相关章节（系统博弈论）→ Tirole (1988) 的对应章节（原著应用）。

7.9　本章小结

核心概念（8 个）：

1. **Nash 均衡：**策略组合中每个玩家的策略是对手策略的最优反应——Cournot 和 Bertrand 的博弈论基础
2. **子博弈完美均衡（SPE）：**在每个子博弈中都是 Nash 均衡——消除不可置信的威胁——Stackelberg 和 Dixit 承诺的基础
3. **战略替代品 vs. 战略互补品：**反应函数斜率——Fudenberg-Tirole 策略分类学的基石
4. **Bayesian Nash 均衡（BNE）：**类型依存的策略——不完全信息下的最优行为——拍卖和限价信号的基础
5. **重复博弈与无名氏定理：**δ 足够大 $\rightarrow$ 任何个人理性支付可维持——合谋可持续性的博弈论基础
6. **完美 Bayesian 均衡（PBE）：**序贯理性 + Bayes 信念更新——信号传递和声誉模型的基础
7. **Tirole 七步建模法：**任何 IO 问题的统一分析框架——玩家、策略、时序、信息、支付、均衡概念、比较静态
8. **旧 IO $\rightarrow$ 新 IO 的范式转变：**从描述性分类学到分析性均衡——从” 结构决定绩效” 到” 策略互动决定结果”

核心公式（3 个）：

1. Cournot-Nash n 企业均衡：$q^* = \dfrac{a-c}{b(n+1)}, \quad P^* = \dfrac{a+nc}{n+1}$
2. 合谋临界贴现因子：$\delta^* = \dfrac{\pi^{\text{dev}} - \pi^{\text{coll}}}{\pi^{\text{dev}} - \pi^{\text{Nash}}}$
3. Stackelberg 均衡：$q_L^* = \dfrac{a-c}{2b}, \quad q_F^* = \dfrac{a-c}{4b}$

本章位置：

- $\leftarrow$ 第 ~6~ 章以” 可竞争性 $\rightarrow$ 博弈论的完美过渡” 结尾——本章正是这个过渡的系统展开
- $\rightarrow$ 第 ~8~ 章将本章的通用博弈论工具箱应用于进入阻挠、掠夺和合谋的具体策略行为分析
- $\rightarrow$ 第 ~10~ 章的 NEIO 方法是博弈论模型的经验实现

- → 本书所有后续章节（第 ~8~ 章–第 ~15~ 章）都以 Tirole 七步框架为方法论骨架

7.10 习题

基础（概念理解）

习题 7.1 为什么 Cournot 均衡是一个 Nash 均衡？写出 Cournot 双寡头的策略式博弈表示（玩家、策略空间、支付函数），并验证均衡的一阶条件等价于 Nash 均衡的定义——每个企业的产量是对手产量的最优反应。

习题 7.2 Nash 均衡和子博弈完美均衡有什么不同？考虑以下进入博弈：进入者选择进入（Enter）或不进入（Stay Out）；如果进入，在位者选择容纳（Accommodate）或战斗（Fight）。收益为：(Enter, Accommodate) = (2, 4)，(Enter, Fight) = (-1, 1)，(Stay Out) = (0, 6)。

(a) 写出这个博弈的所有 Nash 均衡（提示：不要忘记考虑在位者的非均衡路径策略）。

(b) 哪些 Nash 均衡是子博弈完美的？哪些不是？为什么不是？

习题 7.3 解释为什么在 Bertrand 双寡头中（同质产品、相同边际成本、完全信息），Nash 均衡是 $P = MC$。然后解释：为什么在现实中我们很少看到 $P = MC$？（给出至少三种解释——每种都对应 Bertrand 假设的放松）

习题 7.4 ”战略替代品”和”战略互补品”的区别是什么？各给出一个 IO 中的例子。为什么在战略互补品的环境中，”更强硬”可能使自己处境*更差*？

中级（模型推导）

习题 7.5 [**Stackelberg vs. Cournot**] 考虑 Stackelberg 双寡头。反需求为 $P = 100 - Q$，两企业的边际成本均为 0。

(a) 用逆向归纳求 SPE。（提示：先求追随者的反应函数，然后领导者代入。）

(b) 计算消费者剩余（$\text{CS} = Q^2/2$）和总福利（$\text{CS} + \pi_1 + \pi_2$）。与 Cournot 均衡的结果比较。谁从先发优势中获益？

(c) 现假设两家企业都可以在博弈开始前宣布一个产量承诺，然后同时执行。求这个两阶段博弈的 SPE，并解释为什么结果与原始 Stackelberg 不同。

习题 7.6 [**合谋的可持续性**] 两家企业在无限重复的 Cournot 竞争中互动。需求为 $P = 120 - Q$，边际成本均为 $c = 0$。

(a) 单期 Cournot-Nash 均衡中每家企业各生产多少？利润是多少？

(b) 如果两家企业完美合谋（联合垄断，平分产出），每家企业每期生产多少？利润是多少？

(c) 如果一家企业继续生产合谋产量而另一家背叛，背叛企业的最优产量和利润是多少？

(d) 如果两家企业都使用" 冷酷触发策略"（grim trigger），维持合谋所需的最小贴现因子 δ^* 是多少？如果年利率为 10%，合谋可持续吗？

习题 7.7 [Dixit 进入阻挠] 考虑以下两阶段进入博弈（改编自 Dixit 1980）：

- 阶段 1：在位者选择产能 K（每单位成本 $c_0 = 5$）。
- 阶段 2：进入者决定是否进入（进入有固定成本 $F = 100$）。如果进入，两家企业进行 Cournot 竞争。需求为 $P = 100 - Q$，可变成本均为 $c = 5$。
- 在位者的产能使其以可变成本 $c = 5$ 生产最多 K 单位；超过 K 的生产另有边际成本 10（加班/外包）。

(a) 如果进入者进入，在位者的事后边际成本函数是什么？这如何影响在位者进入后产量的" 攻击性"？

(b) 分析 K 如何影响进入者的事后利润。

(c) 求进入被阻挠所需的最小 K。

(d) 如果所有成本都是可变的（无沉没成本），K 的选择可以阻挠进入吗？解释你的答案与可竞争性理论（第 ~6~ 章）的关系。

进阶（概念延伸）

习题 7.8 [可竞争性批判：从 Ch.6 到 Ch.7] 可竞争性理论（第 ~6~ 章 §6.2）假定了在位者是被动的——不能对进入威胁做出策略性反应。用 Tirole 七步框架来形式化" 打了就跑" 进入：

(a) 写出博弈的玩家、策略空间、时序和信息结构（按七步格式）。

(b) 在可竞争性的假设下（零沉没成本、在位者被动），博弈的 SPE 是什么？

(c) 现在修改模型：在位者可以在进入者行动之前沉没地投资产能 $K \geq 0$。新的 SPE 是什么？为什么结果变了？

(d) 解释你的结果如何体现了" 可竞争性的失败 → 博弈论革命的必要性" 这一辩证结构（第 ~6~ 章 §6.6 的过渡）。

习题 7.9 [模型建构：将七步框架应用于一个真实 IO 谜题] 选择一个真实的 IO 谜题（建议：航空业价格战、超市降价竞赛、平台补贴竞争、制药专利申请竞赛……）。用 Tirole 七步框架构建一个简单的博弈模型：

(a) 按七步格式明确：玩家、策略、时序、信息、支付函数。

(b) 选择均衡概念并解释这个选择为什么是合适的。

(c) 讨论：模型的预测在什么条件下会被推翻？即——有什么经验证据可以证伪你的模型？

习题 7.10 [连锁店悖论与声誉] 考虑一个 3 市场依次进入的连锁店博弈。每个市场中：（1）进入者选择进入/不进入；（2）如进入，在位者选择容纳（收益 2,2）或掠夺（收益-1,-1）。不进入收益为（0,5）。

(a) 对于 $N = 1$（一个市场），求 SPE。

(b) 对于 $N = 2$（两个市场），求 SPE。逆向归纳如何从最后一个市场" 拆解" 到第一个市场？

(c) 现在引入不完全信息：在位者有 5% 的先验概率是" 强硬型"（总是掠夺）。论证为什么即使" 软弱型" 在位者在早期市场也可能掠夺。芝加哥学派（第 ~5~ 章 §5.4）关于" 掠夺从不理性" 的论断在引入不完全信息后还成立吗？

7.11 延伸阅读

★ 必读（核心）

Tirole, J. (1988). *The Theory of Industrial Organization*. MIT Press. Ch.11.

非合作博弈论用户手册——Nash → SPE → BNE → PBE。读完本章后回头再读可获得" 事后形式化" 的理解提升。

Fudenberg, D. & Tirole, J. (1991). *Game Theory*. MIT Press. Ch.3–5.

更系统的博弈论处理。如果本章的博弈论工具箱让你想要更系统的理解，这是标准参考。

Shapiro, C. (1989). "The Theory of Business Strategy." *RAND Journal of Economics*, 20(1), 125–137.

博弈论概念如何应用于商业策略——博弈论 IO 与战略管理的桥梁。

Dixit, A. (1980). "The Role of Investment in Entry-Deterrence." ***Economic Journal*****, 90, 95–106.**
SPE + 承诺价值的经典示范——从" 可竞争性批判" 到" 博弈论革命" 的关键论文。

★★ 进阶（深化理解）

Fudenberg, D. & Tirole, J. (1984). "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look." ***AER P&P*****, 74, 361–366.**
策略动物分类学的原始论文——本章 §7.3.3 的完整展开（也是第 ~8~ 章的基础）。

Kreps, D. & Wilson, R. (1982). "Reputation and Imperfect Information." ***Journal of Economic Theory*****, 27, 253–279.**
声誉博弈——PBE 用于解决连锁店悖论的经典论证。

Green, E. & Porter, R. (1984). "Noncooperative Collusion under Imperfect Price Information." ***Econometrica*****, 52(1), 87–100.**
重复博弈 + 不完美监控 = 价格战作为均衡的一部分。

Sutton, J. (1991). ***Sunk Costs and Market Structure*****. MIT Press. Ch.1–2.**
边界方法——对博弈论 IO" 任何事都可以发生" 批评的建设性回应。

张维迎 (1996).《博弈论与信息经济学》. 上海人民出版社/格致出版社.
中国最经典的中文博弈论教材——信号传递、拍卖和进入阻挠章节直接衔接本章。

★★★ 前沿（研究前沿）

Fisher, F. (1989). "Games Economists Play: A Noncooperative View." ***RAND Journal of Economics*****, 20(1), 113–124.**
对博弈论 IO 的经典方法论批评——与 Shapiro (1989) 在同期刊出，必读的思想辩论。

Maskin, E. & Tirole, J. (1988a, 1988b). "A Theory of Dynamic Oligopoly, I & II." ***Econometrica*****, 56(3), 549–600.**
MPE 应用于动态寡头的完整处理——现代 NEIO 的理论基础之一。

Ericson, R. & Pakes, A. (1995). "Markov-Perfect Industry Dynamics." ***Review of Economic Studies*****, 62, 53–82.**
包含进入/退出/投资的动态行业演化——NEIO 中结构性估计的基石。

Bresnahan, T. (1989). "Empirical Studies of Industries with Market Power." In *Handbook of IO*, Vol. 2, Ch.17.

NEIO 纲领——博弈论模型如何与数据结合——第 ~10~ 章的预览。

→ 第 ~8~ 章：策略行为的微观基础

博弈论工具箱已经备好。在第 ~8~ 章中，我们将用本章的语言——Nash 均衡、SPE、BNE、PBE——逐一分析 IO 的核心策略行为：

- 进入遏制（entry deterrence）：Dixit 产能承诺、Milgrom-Roberts 限价信号、Fudenberg-Tirole 策略分类学的完整展开
- 掠夺性定价（predatory pricing）：连锁店悖论的 Kreps-Wilson 声誉模型、深口袋模型
- 合谋（collusion）：Green-Porter 不完美监控、Rotemberg-Saloner 繁荣期价格战、Porter (1983) 联合执行委员会的实证检验

第 ~7~ 章给出了语言和框架；第 ~8~ 章讲述 IO 的核心故事。

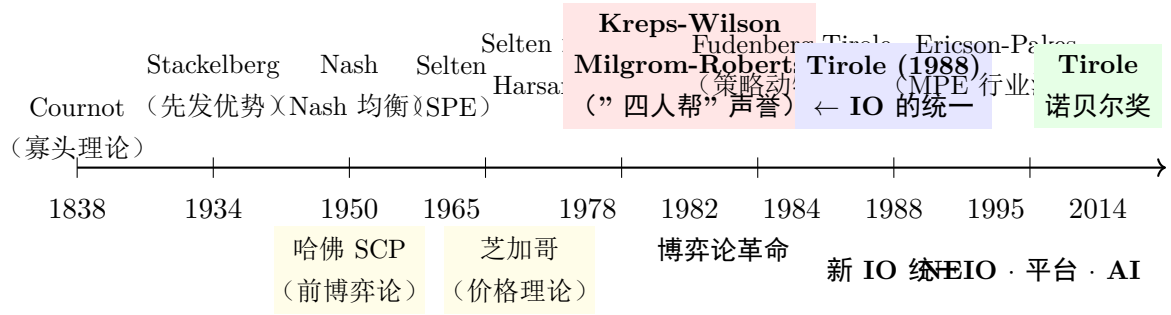


图 7.4: IO 思想演化中的博弈论革命——关键节点时间线

第八章　策略行为的微观基础：进入遏制、掠夺与合谋

章首语

1982 年，三位斯坦福经济学家——David Kreps、Paul Milgrom、John Roberts——以及哈佛的 Robert Wilson——发表了两篇论文，证明了在位企业可以通过在一个市场中掠夺来在所有市场中建立” 强硬” 声誉。在此之前，博弈论似乎**支持**了芝加哥学派的论证：在 Selten (1978) 的连锁店悖论中，掠夺在完全信息下的逆向归纳逻辑中从不发生。但” 四人帮” 证明了，只需引入一丝丝不完全信息——即使进入者认为在位者有 0.1% 的概率是” 疯狂” 的——掠夺就从” 从不理性” 变成了” 完全理性”。

这个结果捕捉了本章的深层主题：**策略行为是否存在，不是一个” 是或否” 的问题，而是一个” 在什么条件下” 的问题**。Dixit 产能承诺——条件：沉没成本 + 产能先于定价。Milgrom-Roberts 限价信号——条件：在位者成本是私人信息。Kreps-Wilson 掠夺声誉——条件：进入者不确定在位者的类型。合谋可持续——条件：贴现因子足够大 + 监测足够可靠。

本章用第 ~7~ 章建立的博弈论工具箱逐一分析这些条件。它不是宣称” 策略行为总是有效的”——而是提供” 分析策略行为在什么条件下有效的精确框架”。第 ~7~ 章给了我们博弈论语言；本章用这种语言讲述 IO 的三个核心故事。

8.1　核心问题

读完本章，你应该能够回答以下问题：

1. **进入遏制如何运作？**在位者如何通过沉没投资（产能）、信息传递（限价）和结构改变（品牌增殖、排他合同）在进入发生**前**使进入无利可图？承诺（commitment）为什么是一切进入遏制策略的核心？

2. **掠夺性定价何时是理性的？**芝加哥学派说” 掠夺从不理性”——在什么条件下这个命题成立？在什么条件下它不成立？声誉机制、深口袋机制和信号机制如何为掠夺提供理性基础？

3. **合谋如何克服囚徒困境？**在一次性博弈中背叛是占优策略——什么使企业在重复互动中维持合作？贴现因子、监测精度和惩罚可信性如何决定合谋的可持续性？

4. **这三种策略行为共享什么深层结构？**进入遏制、掠夺和合谋是否只是同一个"囚徒困境 → SPE/重复博弈/PBE"逻辑在不同情境下的应用？Fudenberg-Tirole (1984) 分类学如何统一进入遏制和进入容纳？

5. **博弈论如何改变了掠夺和合谋的法律分析？**从 Areeda-Turner (1975) 的 AVC 测试到 Brooke Group (1993) 的 recoupment 标准，再到 EU 的 AKZO 测试——法律如何回应博弈论关于掠夺何时理性的洞见？宽大计划如何基于博弈论逻辑瓦解合谋？

6. **中国的策略行为实践是什么？**中国《反垄断法》（2022 修订）第 22 条如何界定"市场支配地位的滥用"？Alibaba"二选一"（~182.28 亿）、360 v. Tencent、Didi-Kuaidi 补贴战中体现了哪些策略行为逻辑？

决策树：本章的导航图

在位企业面对进入威胁时的策略选择遵循一个清晰的决策树，这也是本章的结构性装置：

$$\text{进入威胁} \rightarrow \begin{cases} \text{能事前遏制？} \rightarrow \textbf{进入遏制}\ (\S8.2\text{–}8.3) \\ \text{不能遏制，能事后消灭？} \rightarrow \textbf{掠夺性定价}\ (\S8.4\text{–}8.5) \\ \text{消灭不了，能和平共处？} \rightarrow \textbf{合谋}\ (\S8.6\text{–}8.7) \end{cases}$$

8.2　进入遏制 I：产能承诺与限价信号

【核心】

8.2.1　进入遏制的博弈论骨架：一个统一的七步框架

所有进入遏制模型共享同一个基本结构。在展开具体模型之前，建立这个骨架有助于读者看到 Dixit 和 Milgrom-Roberts——表面上处理不同问题（产能 vs. 价格）——实际上共享完全相同的深层博弈逻辑。

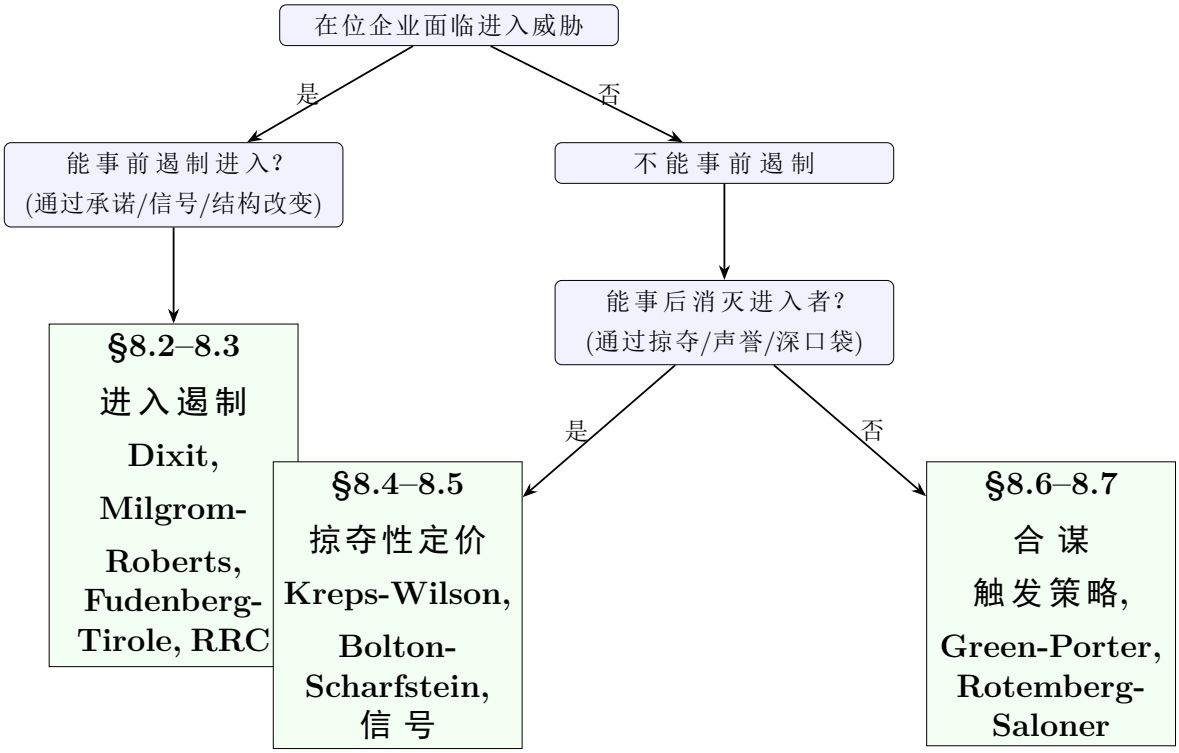


图 8.1: 在位企业策略选择决策树——本章的结构性装置

Tirole 七步框架：进入遏制的通用结构

1. **玩家**：在位者（I）+ 潜在进入者（E）
2. **策略**：在位者第一阶段选择某种不可逆行动 A_I（产能、价格、产品品种、合同……）；进入者第二阶段选择进入/不进入
3. **时序**：序贯——在位者先行动，进入者后行动
4. **信息**：完全信息版本——所有支付为共同知识；不完全信息版本——在位者有私人信息（成本类型、需求强度……）
5. **支付**：$\pi_I(A_I, \text{进入决策})$，$\pi_E(A_I, \text{进入决策})$
6. **均衡概念**：完全信息 $\to$ SPE；不完全信息 $\to$ PBE
7. **比较静态**：沉没成本的大小、信息不对称程度、需求弹性如何影响进入遏制是否可行？

定义 8.1 (承诺). 一个行动构成**承诺**（*commitment*）当且仅当：

1. 它是**不可逆的**（沉没成本 $\to$ 在位者事后不能"撤回"），且
2. 它**改变了在位者的事后最优策略**（使在位者在进入后更激进或更温和地竞争）。

只满足 (1) 不满足 (2) 是空洞的承诺；只满足 (2) 不满足 (1) 是不可信的威胁。

承诺为什么重要？因为**在没有承诺工具的世界中，在位者无法策略性地影响进入决策**——这正是第 ~6~ 章可竞争市场理论的世界。一旦存在承诺工具（沉没成本），在位者可以主动塑造进入条件——可竞争性崩溃，策略行为登场。

8.2.2 Dixit (1980)：产能承诺模型

【核心】

背景。Spence (1977) 提出在位者可以投资过剩产能来阻挠进入——如果进入者进入，在位者使用过剩产能增加产量、压低价格、使进入无利可图。但 Spence 的论证有一个漏洞：**如果在位者在进入者进入后仍有动机使用过剩产能，威胁是可信的；但如果进入后再增产是无利可图的，威胁就是空洞的。**Spence 假定了在位者会增产，但没有证明增产是进入后的最优反应。

Dixit (1980) 的贡献是用子博弈完美均衡严格化了这个论证。

模型设定（七步框架）

- **玩家**：在位者（I）+ 潜在进入者（E）
- **策略**：阶段 1——在位者选择产能 K（单位成本 c_0，一旦投资即沉没）；阶段 2——进入者决定进入（支付固定进入成本 F）或不进入；阶段 3（如进入）——两家企业进行 Cournot 数量竞争
- **时序**：三阶段序贯博弈
- **信息**：完全信息——所有成本和需求参数为共同知识
- **支付**：在位者的成本函数是分段线性的：

$$MC_I(q_I) = \begin{cases} w, & q_I \le K \quad \text{（仅可变成本）} \\ w + r, & q_I > K \quad \text{（超额产量需加班/外包，成本更高）} \end{cases}$$

  进入者的边际成本 $= w + r$（假设进入者没有在位者的产能优势）；需求 $P = a - b(q_I + q_E)$
- **均衡概念**：子博弈完美均衡（SPE）——逆向归纳

逆向归纳求解

阶段 3（给定 K 且进入已发生）。在位者和进入者进行 Cournot 竞争。在位者的反应函数在 $q_I = K$ 处有一个**拐折**（kink）——因为边际成本在此处从 w 跳升至 $w+r$。进入者的反应函数是标准的：

$$q_E^*(q_I) = \frac{a - w - r - bq_I}{2b}$$

在拐折点，在位者既可以向左（利用低成本产能，边际成本 w）也可以向右（使用昂贵加班产能，边际成本 $w+r$）。Cournot-Nash 均衡出现在反应函数的交点——K 的位置决定了交点的位置。

阶段 2（给定 K）。进入者进入当且仅当在阶段 3 Cournot 均衡中 $\pi_E(K) \geq F$。

阶段 1。在位者选择 K 以最大化：

$$\pi_I(K) = \begin{cases} \pi_I^M(K), & \text{如果 } \pi_E(K) < F \text{（进入被遏制）} \\ \pi_I^A(K), & \text{如果 } \pi_E(K) \geq F \text{（进入被容纳）} \end{cases}$$

核心结果

1. **产能投资降低在位者的事后边际成本**（对于 $q_I \leq K$）$\rightarrow$ 在位者在进入后 Cournot 竞争中的最优产量更高 $\rightarrow$ 进入者的剩余需求更小、利润更低
2. **存在临界产能** K^* 使进入者的事后利润恰好等于 F——在位者选择 $K \geq K^*$ 即可阻挠进入
3. **在位者可能选择"过剩产能"**：K^* 可能超过如果没有进入威胁时在位者的最优产量
4. **关键的"可信性"机制**：因为 K 是沉没的，在位者在阶段 3 **确实**有动机充分利用产能——威胁是**可信的**

Dixit & Spence：关键区别

Spence (1977) 的论证是**前博弈论的**：在位者"选择"产能，进入者"反应"——未证明威胁的可信性。Dixit (1980) 用 SPE 证明了产能投资改变了在位者的事后边际成本，因此威胁是可信的。**区别在于：Dixit 要求——并满足——子博弈完美性；Spence 没有。**

对可竞争性的回应（衔接 第 ~6~ 章 §6.6）。可竞争性理论 = 零沉没成本 $\rightarrow$ 在位者没有任何承诺工具 $\rightarrow$ 市场是可竞争的。Dixit 模型 = 沉没成本 > 0 $\rightarrow$ 在位者有了承诺工具 $\rightarrow$ 进入可以被阻挠。**可竞争性是 Dixit 世界在 $K = 0$ 处的退化解——而 $K = 0$ 在现实世界中几乎从不成立。**

8.2.3 战略替代品与战略互补品

在展开 Fudenberg-Tirole 分类学之前，需要引入 Bulow, Geanakoplos & Klemperer (1985) 的关键区分——因为它决定了在位者应该"过度投资"还是"投资不足"。

定义 8.2 (战略替代品与战略互补品).

- **战略替代品**（*strategic substitutes*）：我的最优反应随你的策略增加而**减少** $\rightarrow$ 反应函数向下倾斜 $\rightarrow$ *Cournot* 数量竞争

- **战略互补品**（*strategic complements*）：我的最优反应随你的策略增加而**增加** → 反应函数向上倾斜 → *Bertrand* 价格竞争（差异化产品）

IO 中的深远意义：

- 在战略替代品中，" 更强硬"（更大量） → 对手退缩（减少产量） → 可以有利
- 在战略互补品中，" 更强硬"（更低价格） → 对手也降价 → 双方都受损

这一区分是 Fudenberg-Tirole (1984) 策略动物分类学的数学基础。

8.2.4 Fudenberg-Tirole (1984)：策略动物分类学

【核心】

Fudenberg & Tirole (1984) 的"The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look"（肥猫效应、小狗策略、饿虎相和领头狗）为所有进入遏制/容纳策略提供了统一的数学框架。这是本章——乃至全书——最优雅的理论统一之一。

两阶段结构

- **阶段 1**：在位者选择投资水平 K（产能、广告、研发……）
- **阶段 2**：两家企业进行产品市场竞争（Cournot 或 Bertrand）
- 第一阶段投资影响在位者的事后边际成本/需求——从而影响其在阶段 2 中的" 强硬程度"

核心的二阶导数：$\partial^2\pi_I/\partial K\partial q_E$ 的符号决定投资使在位者变得更强硬（+）还是更温和（−）。

四种策略原型

取决于 (a) 投资使在位者变得更强硬（tough）还是更温和（soft），以及 (b) 产品市场竞争是战略替代品还是战略互补品：

四个具体案例：

1. **领头狗（Top Dog）——Dixit (1980) 产能投资。**投资（产能） → 降低事后边际成本 → 在位者更强硬（生产更多）。战略替代品（Cournot） → 对手退缩（生产更少）。→ 在位者**过度投资**产能以阻挠进入。

2. **小狗（Puppy Dog）。**投资（产能） → 降低事后边际成本 → 在位者更强硬（降价）。战略互补品（Bertrand 差异化产品） → 对手也降价 → 价格战。→ 在位者**投资不足**以避免显得激进，防止破坏性价格战。例子：在位者故意不投资于降低成本的技术，保持高成本 → 高价格 → 进入者也可定高价格 → 双方获利。

表 8.1: Fudenberg-Tirole (1984) 策略动物分类学

	投资使在位者更强硬	投资使在位者更温和
战略替代品 （Cournot，数量）	**领头狗（Top Dog）** 过度投资，显得激进，阻挠进入	**饥肠辘辘（Lean and Hungry）** 投资不足，保持精干和激进
战略互补品 （Bertrand，价格）	**小狗（Puppy Dog）** 投资不足，避免显得激进，防止价格战	**肥猫（Fat Cat）** 过度投资，显得温和，容纳进入

3. **饥肠辘辘（Lean and Hungry）。**不投资（广告/质量）→ 保持低边际成本 → 在位者更强硬。战略替代品（Cournot） → 对手退缩。→ 在位者**不投资**于质量/广告以保持成本优势和激进性。

4. **肥猫（Fat Cat）。**投资（质量/广告） → 提升边际利润 → 在位者更温和（定更高价格）。战略互补品（Bertrand） → 对手也定更高价格 → 双方获利。→ 在位者**过度投资**于质量/广告以显得温和，容纳进入。

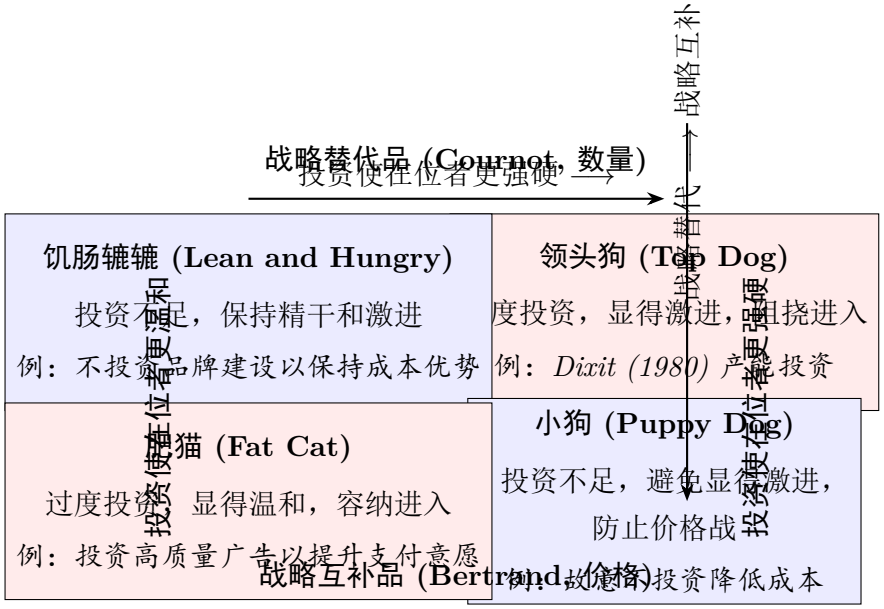


图 8.2: Fudenberg-Tirole (1984) 策略动物分类学——2×2 矩阵

不存在一刀切规则

Fudenberg-Tirole 分类学的核心洞见是：**不存在"总是应该过度投资"或"总是应该投资不足"的一刀切规则**。最优策略取决于 (a) 投资如何影响在位者的强硬程度，以及 (b) 产品市场竞争的性质（战略替代 vs. 战略互补）。这使得分类学同时成为教学工具和实践性的管理指南。

8.2.5　Milgrom-Roberts (1982)：限价作为信号

【核心】/【进阶】

核心问题。Dixit 模型假设进入者知道在位者的成本——完全信息。但如果进入者**不知道**在位者的成本类型呢？那么在位者的进入前价格就可能起作用——不是通过改变事后博弈结构（像 Dixit 产能那样），而是通过**传递关于在位者类型的信息**。

芝加哥批判回顾（第 ~5~ 章）。芝加哥学派已经指出了限制定价概念的问题：如果进入发生了，在位者的最优反应是最大化进入后利润，而**不是**继续执行压低进入者利润的价格。限制定价的威胁是不可信的——除非在位者能在进入前做出某种不可逆的承诺。

Milgrom-Roberts (1982) 的贡献是证明了：即使没有沉没投资作为"硬"承诺工具，不完全信息也能使限价在均衡中发生。机制不是承诺——是**信号传递**。

模型设定（七步框架）

- **玩家**：在位者（I）+ 进入者（E）
- **策略**：阶段 1——在位者选择价格 p_1；阶段 2——进入者观察到 p_1（但不知道在位者成本类型），选择进入/不进入；如进入，阶段 2 两家企业竞争
- **时序**：两阶段序贯博弈
- **信息**：在位者的边际成本是私人信息——$c \in \{c_L, c_H\}$，先验概率 $\Pr(c = c_L) = \mu_0$；进入者的成本为共同知识
- **支付**：$\pi_I(p_1, c) + \delta \cdot \pi_I^{\text{阶段 }2}(\text{进入决策}, c)$，$\pi_E(\text{进入决策}, c)$
- **均衡概念**：完美 Bayesian 均衡（PBE）

关键机制：单交叉条件

在位者的阶段 1 价格有两个效应：

1. **直接利润效应**：p_1 影响阶段 1 的利润

2. **信号效应**：p_1 影响进入者对在位者类型的信念 → 影响进入决策 → 影响阶段 2 的利润

低成本在位者有动机将自己与高成本在位者区分开来——如果进入者知道在位者是低成本，它就不会进入（进入后竞争会很惨烈）。高成本在位者想模仿低成本在位者——因为它想阻挠进入。关键是：**设定低价格对于低成本在位者来说"更便宜"**（因为它的利润边际更大，产量更大 → 价格下降同等幅度时利润损失相对更小）。这就是**单交叉条件**（single-crossing property）——它使分离成为可能。

两种均衡

1. **分离均衡**（separating equilibrium）：

   - 低成本在位者设定 $p_1 = p^L < p^M(c_L)$（低于短期垄断价格）
   - 高成本在位者设定 $p_1 = p^M(c_H)$（自己的短期垄断价格——不模仿）
   - 进入者观察到 p^L → 推断 $c = c_L$ → 不进入
   - 进入者观察到 $p^M(c_H)$ → 推断 $c = c_H$ → 进入
   - **限价发生了**：低成本在位者为了阻挠进入而定价低于短期最优

2. **混同均衡**（pooling equilibrium）：

   - 两种类型都设定相同的 p_1
   - 进入者观察到 p_1 → 信念不更新 → 基于先验 μ_0 做进入决策
   - 如果 μ_0 足够低（进入者先验认为低成本的概率足够小），进入者进入；反之不进入

表 8.2: Dixit (1980) vs. Milgrom-Roberts (1982) 进入遏制机制对比

	Dixit (1980)	**Milgrom-Roberts (1982)**
机制	沉没产能改变事后激励	价格传递事前私人信息
信息假设	完全信息	在位者成本为私人信息
均衡概念	SPE	PBE
承诺工具	实物投资（产能）	价格（信号）
可信性来源	沉没成本改变事后最优行为	均衡信念更新使信号价格最优

两种机制的互补性

这两个模型不是”竞争”的——它们捕捉了同一现象的不同机制。在位者可以**同时**使用两种工具：投资产能（像 Dixit）**和**设定限价（像 Milgrom-Roberts）。完整的进入遏制策略是产能承诺 + 价格信号的组合。

8.3　进入遏制 II：产品增殖、提高对手成本与排他性合同

【进阶】

Dixit 和 Milgrom-Roberts 分别展示了在位者如何通过**改变自身成本**（产能）或**传递自身类型信息**（限价）来遏制进入。本节呈现三种替代策略——它们共同的特征是**改变市场结构本身**。

8.3.1　Schmalensee (1978)：品牌增殖（Brand Proliferation）

核心直觉：如果在位者在进入者进入之前用多个品牌”填满”产品空间，进入者的剩余需求可能太小而无法覆盖固定进入成本。

机制（圆形城市 Salop 1979 框架）：

- 消费者均匀分布在圆周上，每个品牌需要固定成本 f
- 在位者可以在进入者行动之前选择多个品牌的位置
- 在位者事先选择 n 个品牌并均匀分布 $\to$ 相邻品牌之间的最大间距 $= 1/n$
- 进入者可以进入在位者品牌之间的”缝隙”——但如果缝隙不够大，进入后的利润不足以覆盖 f
- 在位者可能选择**超过短期利润最大化的品牌数量**来”填满”空间 $\to$ 阻挠进入

关键条件：品牌增殖必须是**不可逆的**（存在品牌设立的沉没成本——产品开发、市场调研、初始广告）——否则在位者在进入发生后会”撤回”多余的品牌。这与 Dixit 逻辑一致：承诺需要沉没成本。

现实案例：早餐麦片行业（Schmalensee 的原始动机）——Kellogg's、General Mills、Post 等企业在 1950–1970 年代推出了大量品牌变体；中国方便面市场——康师傅和统一的多品牌策略。

8.3.2　Salop & Scheffman (1983)：提高对手成本（Raising Rivals' Costs, RRC）

核心直觉：进入阻挠不一定需要降低进入者的需求——也可以**提高进入者的成本**。

三种机制：

1. **通过要素市场：**在位者大量购买稀缺投入品 → 抬高投入品价格 → 进入者的成本上升。或与供应商签订排他性合同 → 进入者只能使用效率更低的替代品。条件：在位者自己的要素购买量足够大——或法律的帮助。

2. **通过规制：**在位者游说政府制定更严格的标准（环保、安全、质量……）。在位者具有合规成本优势（已安装合规设备） → 新标准对进入者的成本增加大于在位者。**祖父条款**（grandfathering）——现有企业免于新规——是最极端的 RRC。

3. **通过工会/工资**（Williamson 1968）：在位者与工会签订全行业工资协议 → 劳动密集型进入者受损更多。

RRC vs. 掠夺性定价

掠夺性定价	RRC
在位者承担低价的损失	在位者也面临成本上升 ——但**上升幅度更小**
可行性：在位者比进入者财务更强	可行性：在位者和进入者的成本结构不对称
可能被视为掠夺（反垄断）	通常更难被反垄断法触及 （要素购买通常合法）

8.3.3 Aghion & Bolton (1987)：排他性合同（Exclusive Contracts）

核心问题：在位者可以通过与下游买家或供应商签订排他性合同来阻挠进入——但芝加哥学派论证说买家不会自愿签订对自己有害的合同。如果真的”有害”（进入会降低价格），买家应该拒绝签订排他性合同。

Aghion-Bolton 的回应：买家和在位者之间存在**外部性**。单个买家签订排他性合同不会改变在位者的总市场覆盖——但**所有**买家都签订时，进入者被排除在市场之外，这损害了**所有买家**的利益。

模型机制：

- 在位者向买家提供合同：如果你承诺只从我这里购买，我给你折扣价格
- 如果有足够多的买家签订合同 → 进入者的剩余市场太小 → 进入不盈利 → 进入不发生
- 即使每个买家都**知道**如果所有人都不签合同进入就会发生（价格会更低），单个买家仍然可能签合同——因为一个买家不签不足以使进入盈利

- 存在一个**临界签约比例**：超过此比例 → 即使有损于买家整体，所有买家也可能签约（囚徒困境）

关键洞察：排他性合同的损害不在于" 合同本身强制买家只从在位者购买"，而在于**合同的外部性**——单个买家的决定影响所有其他买家面临的竞争环境。这为反垄断干预提供了 Aghion-Bolton 式的经济学基础。

8.3.4 Gilbert & Newbery (1982)：先占专利（Preemptive Patenting）

在位者是否有比潜在进入者更强的动机去获取新专利？答案是肯定的。如果在位者获得专利 → 不进入 → 垄断利润继续。如果进入者获得专利 → 进入 → 双寡头利润。因为**垄断利润 > 双寡头利润之和**（标准 Cournot 结果——竞争者不能完全内化进入对在位者利润的负外部性），在位者为了保护垄断地位的支付意愿（垄断利润 − 双寡头在位者利润）**大于**进入者为了获得双寡头地位的支付意愿（双寡头进入者利润 − 0）。→ 在位者有更强的动机获取专利——即使该专利本身不创造新价值。

效率含义：先占专利可能导致" 沉睡专利"（sleeping patents）——在位者获得专利仅为了防止别人使用，本身不商业化。

8.4 掠夺性定价 I：芝加哥批判与连锁店悖论

【核心】

8.4.1 什么是掠夺性定价？

定义 8.3 (掠夺性定价). *在位者将价格设定在成本以下（或足够低），以迫使竞争对手退出市场，然后在退出发生后将价格提高到垄断水平以回收前期损失。*

三个关键阶段：

1. **牺牲期**（sacrifice phase）：在位者承担亏损——以低于成本的价格销售

2. **退出期**（exit phase）：竞争对手因无法承受亏损而退出

3. **回收期**（recoupment phase）：在位者将价格提高到垄断水平，赚取足够多的利润以补偿牺牲期的损失——且净现值为正

如果没有回收期——即牺牲无法被回收——则掠夺定价在任何理性意义上都是不理性的。

8.4.2 芝加哥批判：掠夺”从不理性”

McGee (1958) 对 Standard Oil (1911) 案的经典重新评估。最高法院在 Standard Oil 案中指控 Standard Oil 通过掠夺性定价消灭竞争对手。McGee 审查了审判记录并得出结论：没有证据表明 Standard Oil 实际实施了掠夺性定价。**更根本的论证**：即使 Standard Oil 尝试过掠夺性定价，这也不会是利润最大化的策略——因为：

1. 掠夺者在牺牲期承担比被掠夺者**更大的损失**——掠夺者必须服务更大的市场份额
2. 即使被掠夺者退出，其资产不会消失——别人会以折扣价购买并重新进入（”有用资产不会闲置”）
3. **并购比掠夺更有效率**——与其承担掠夺的损失，不如直接收购竞争对手

Bork (1978) 在 *The Antitrust Paradox* 中进一步强化：错误成本框架——错误判定掠夺性定价的成本（阻止合法的激烈价格竞争）远大于错误不判定掠夺的成本（大多数”掠夺”实际上是竞争）。”掠夺性定价诉讼通常是被困于竞争市场的企业试图获得法院强制执行卡特尔的企图。”

Easterbrook (1984) 的错误成本框架：

- **假阳性**：将合法竞争误判为掠夺 → 保护无效率企业 → 消费者支付更高价格
- **假阴性**：未识别真实的掠夺 → 允许掠夺者在未来提高价格 → 消费者受损
- 核心论证：假阳性的预期成本 > 假阴性的预期成本，因为：(a) 真正的掠夺很少发生（如果理性），(b) 败诉企业经常指控掠夺以逃避竞争，(c) 市场通常会自我修正

芝加哥核心逻辑的”剪刀论证”

如果进入壁垒高 → 在位者已拥有垄断地位 → 不需要掠夺

如果进入壁垒低 → 回收不可能（新进入者会进来压低价格）

∴ 掠夺要么不必要，要么不可能回收

这是一个有力的逻辑——但它依赖于两个隐含假设：**完全信息**和**完美资本市场**。

8.4.3 Selten (1978)：连锁店悖论——博弈论”支持”了芝加哥

【核心】

现在用博弈论严格检验芝加哥的论证。Selten (1978) 的连锁店博弈用完全信息下的 SPE 逻辑**证明了**芝加哥的结论——掠夺在理性均衡中从不发生。

模型设定（七步框架）

- **玩家**：在位者（连锁店）+ N 个进入者（依次进入 N 个独立市场）
- **策略**：在市场 t，进入者 t 选择进入（IN）或不进入（OUT）；如果进入，在位者选择容纳（ACCOMMODATE）或掠夺（FIGHT）
- **时序**：N 期序贯博弈——进入者 1 先行动，然后进入者 2，……，进入者 N
- **信息**：完全信息——所有支付为共同知识
- **支付**（每个市场）：进入 + 容纳 $= (2, 2)$；进入 + 掠夺 $= (-1, -1)$；不进入 $= (5, 0)$
- **均衡概念**：SPE——逆向归纳

逆向归纳

- 市场 N：进入者 N 进入 $\rightarrow$ 在位者选择容纳（$2 > -1$） $\rightarrow$ 掠夺不发生。进入者预见到在位者将容纳 $\rightarrow$ 进入（$2 > 0$）。
- 市场 $N-1$：无论在位者在市场 $N-1$ 做什么，这不会改变市场 N 的结果——市场 N 中在位者总是容纳（最后一个市场没有未来需要保护）。$\rightarrow$ 在位者在市场 $N-1$ 也容纳 $\rightarrow$ 进入者 $N-1$ 进入。
- ……一直逆向归纳到市场 1：**在位者在每一个市场都容纳，每一个进入者都进入。掠夺从未发生。**

连锁店悖论的本质

直觉告诉我们：在位者应该在早期市场掠夺以建立" 强硬" 声誉，阻止后续进入者进入。但 SPE 的逻辑说：无论早期发生了什么，最后一个市场的进入者不会被阻挠 $\rightarrow$ 倒数第二个也不会 $\rightarrow$ …… $\rightarrow$ 第一个也不会。**声誉效应在完全信息博弈中不存在。**

Selten 本人对此感到" 不安"——他写道（大意）：" 我对这个博弈的逻辑分析得出的结论感到不安……我只能说，如果我是连锁店的决策者，我会在早期市场掠夺。" 这个" 悖论" 不是逻辑错误——它是**完全信息下 SPE 逻辑的正确推论**。它指向了完全信息假设的问题。

芝加哥 vs. Selten vs. 直觉：芝加哥说掠夺不理性（价格理论论证）；Selten 用博弈论**证明了**芝加哥在完全信息下是对的；但 Selten 自己也承认这个结论**感觉不对劲**。这个张力正是下一节的起点。

8.5 掠夺性定价 II：后芝加哥复兴

【核心】/【进阶】

Selten 悖论表明：在完全信息中，掠夺从不发生。但现实告诉我们——反垄断案件、商业史、以及我们的直觉——掠夺确实存在。这个矛盾如何解决？

答案：**引入不完全信息**。一旦进入者不确定在位者的类型、目标或约束，掠夺就从" 从不理性" 变成了" 在均衡中完全理性"。

8.5.1 声誉机制：" 四人帮"（Kreps-Milgrom-Roberts-Wilson 1982）

【核心】

核心创新：引入关于在位者**类型**的不完全信息。

模型设定（七步框架）

- **玩家**：在位者（连锁店）+ N 个进入者
- **类型空间**：在位者有两种可能类型——" 强硬型"（tough，从掠夺中获得正效用，以概率 ε）和" 软弱型"（weak/理性，支付与 Selten 博弈相同，以概率 $1-\varepsilon$）。在位者知道自己的类型；进入者只知道先验分布
- **策略**：与 Selten 博弈相同
- **时序**：N 期序贯——全部行动被所有后续进入者观察到
- **信息**：在位者类型是私人信息——进入者的信念通过 Bayes 规则更新
- **均衡概念**：完美 Bayesian 均衡（PBE）

关键结果——即使 ε 极小，声誉效应也" 爆炸"

考虑 $N = 2$ 的情况：

- 强硬型在位者：总是在被进入时掠夺（按定义）
- 软弱型在位者在市场 1 的行为：
  - 如果在市场 1 容纳 $\rightarrow$ 揭示自己是软弱的 $\rightarrow$ 市场 2 的进入者一定会进入
  - 如果在市场 1 掠夺 $\rightarrow$ 市场 2 的进入者观察到掠夺，但不确知在位者类型 $\rightarrow$ 如果 ε 足够大，进入者 2 可能不进入（风险太大）
  - 软弱型在位者在市场 1 掠夺的净收益：牺牲市场 1 利润 (-1 vs. 2) $\rightarrow$ 获取市场 2 中可能不被进入的收益 (5 vs. 2)

对于 $N > 2$，即使 $\varepsilon = 0.01$（1% 的概率在位者" 疯狂"），软弱型在位者在早期市场也有强烈的动机模仿强硬型——因为建立声誉的收益跨越多个未来市场。

为什么" 掠夺从不理性" 是错的

芝加哥的论证假设完全信息——进入者**确切知道**在位者是理性的利润最大化者。在现实中，进入者**不确定**在位者的目标函数（也许 CEO 是非理性的？也许在位者有不同目标？）。只要这种不确定性存在——哪怕是很小的——掠夺可能在均衡中出现。

芝加哥在完全信息的" 玩具世界" 中是对的——但在有信息摩擦的现实世界中是错的。

经济学含义：

1. 即使 ε 极小，在足够多市场中，软弱型在位者在早期有强烈动机模仿强硬型
2. **声誉是" 资产"**——它可以通过掠夺来积累，而容纳会消耗它
3. 在完全信息中（$\varepsilon = 0$） $\rightarrow$ 回到 Selten 悖论——掠夺从不发生
4. 一旦 $\varepsilon > 0$ $\rightarrow$ **从完全信息到不完全信息的定性跳跃**——声誉效应出现

8.5.2　深口袋机制：Bolton & Scharfstein (1990)

【进阶】

核心机制：即使关于" 类型" 没有不确定性（在位者是理性的），如果资本市场不完美，财务脆弱的进入者也可能被财务充裕的在位者掠夺出局。

模型直觉：

- 进入者需要外部融资来覆盖进入成本和营运资本
- 进入者与投资者之间存在**道德风险**（effort 不可观测 $\rightarrow$ 需要给进入者利润份额作为激励）或**信息不对称**（投资者不知道企业的真实前景）
- 这些代理问题 $\rightarrow$ 进入者的融资合同不完美：在某些世界状态下无法获得再融资
- 在位者知道这一点 $\rightarrow$ 掠夺（压低进入者的当前利润率） $\rightarrow$ 进入者的低利润使 (a) 其更难满足债务契约，(b) 向投资者发送负面信号，(c) 降低其内部资金 $\rightarrow$ 再融资成本上升或再融资不可能
- **即使在位者是完全理性的利润最大化者，掠夺仍然可能发生**——因为资本市场不完美允许在位者将进入者" 饿死"

深口袋的不对称性：

- 在位者有" 深口袋"——丰厚的内部资金，无需外部融资
- 进入者有" 浅口袋"——需要外部融资，受制于融资约束
- 掠夺放大了这种不对称性：在位者可以承受短期亏损；进入者的短期亏损触发融资约束 → 无法继续运营

与声誉机制的对比：

表 8.3: 掠夺性定价的机制对比

	声誉模型	**深口袋模型**
掠夺的目的	建立声誉以遏制**未来**进入者	消灭**当前**进入者
关键摩擦	信息不对称（在位者类型不确定）	资本市场不完美（融资约束）
关键条件	$\varepsilon > 0$ （进入者对类型有不确定）	进入者依赖外部融资
主要预测	掠夺在早期市场更可能	掠夺针对财务脆弱的进入者

两个机制并非互斥——在位者可以同时利用两者。

8.5.3 信号模型

【进阶】

两类掠夺性定价信号模型：

1. 信号干扰模型（Fudenberg & Tirole 1986 —"A 'Signal-Jamming' Theory of Predation"）：

- 进入者不确定市场需求（市场到底有多大？增长有多快?）
- 进入者只能观察市场价格和销量来推断需求
- 在位者可以在进入者进入初期压低价格 → 压低了进入者观察到的市场信号（" 信号干扰"） → 进入者推断市场前景黯淡 → 退出
- 这与声誉模型不同——在位者不是在建立声誉，而是在**干扰进入者获取信息的渠道**

2. 成本信号模型：

- 对称的 Milgrom-Roberts 逻辑——在位者通过低价来信号自己的低成本，或信号进入者的高成本

- 进入者根据观察到的价格更新信念 → 如果信念更新足够不利 → 退出

信号模型的统一特征：掠夺不是通过" 烧钱" 来证明自己的强硬——而是通过**扭曲进入者对关键市场变量的推断**来诱导退出。

8.5.4　法律标准：从 Areeda-Turner 到 AKZO

【核心】

Areeda-Turner (1975) 测试：

- 发表在 *Harvard Law Review*——可能是反垄断法中被引用最多的法学评论文章
- **AVC 规则**：价格 < 平均可变成本（AVC） → 推定掠夺性；价格 ≥ AVC → 推定合法
- 理由：价格 < AVC 意味着企业每销售一单位就在亏损——这只能被" 消灭对手后提价" 的预期来合理化；AVC 是客观的、可司法操作的测试
- 劣势（博弈论批判）：博弈论掠夺不一定需要价格 < AVC；未考虑回收；过于" 一刀切"

Brooke Group (1993) 标准（美国现行法律）：

- 最高法院在 *Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.* 中建立了两要素测试：**低于成本定价 + 回收的" 危险概率"**（dangerous probability of recoupment）
- 第二个要素的影响：原告在掠夺性定价案件中几乎不可能胜诉——自 Brooke Group (1993) 以来，美国最高法院**从未**在掠夺性定价案件中支持原告

EU 的 AKZO (1991) 测试：

- 一级规则：价格 < AVC → **自动推定**为滥用（不需要证明回收或意图）
- 二级规则：价格在 AVC 和 ATC 之间 → 如果**意图证据**表明掠夺性计划 → 滥用
- **不需要证明回收**——这是 EU 和美国标准的核心差异

8.6　合谋 I：囚徒困境、重复博弈与无名氏定理

【核心】

表 8.4: 三种法律标准对比

	Areeda-Turner (1975)	**Brooke Group (1993, US)**	**AKZO (1991, EU)**
成本基准	AVC	AVC 或 ATC	AVC（一级）/ ATC（二级）
回收要求	无明确要求	**必须证明**危险概率	**不需要证明**
意图证据	不需要	辅助性	在 AVC-ATC 区间 **决定性**
原告胜诉难度	低	极高	中等

表 8.5: 合谋的囚徒困境——Bertrand 双寡头（同质产品，边际成本 c）

	企业 2 合作（p^M）	**企业 2 背叛（$p^M - \varepsilon$）**
企业 1 合作（p^M）	$(\pi^M/2,\ \pi^M/2)$	$(0,\ \pi^M)$
企业 1 背叛（$p^M - \varepsilon$）	$(\pi^M,\ 0)$	$(0,\ 0)$

8.6.1 囚徒困境作为合谋的原始结构

其中 π^M 是垄断利润，p^M 是垄断价格。**一次性博弈：**(背叛, 背叛) 是唯一的 Nash 均衡——尽管 (合作, 合作) 使双方更好。**合谋的核心问题在此：**在没有外部强制执行的情况下，自私的企业如何维持合作？

8.6.2 无限重复博弈中的触发策略

【核心】

博弈结构：

- 同一阶段博弈（Bertrand 或 Cournot）无限次重复，$t = 1, 2, 3, \ldots \to \infty$
- 贴现因子 $\delta \in (0, 1)$ （共有）
- 企业 i 的目标：$\sum_{t=1}^{\infty} \delta^{t-1} \pi_i^t$

定义 8.4 (触发策略（Grim Trigger Strategy）).

- 开始：生产合谋产量 $q_i = q^M/2$ 或设定合谋价格 $p = p^M$
- 在每一期：如果所有企业在所有过去时期都遵守合谋 $\to$ 继续合谋
- 否则（任何企业在任何时期背叛了）$\to$ **永久**回到 *Nash* 均衡 $q_i = q^{Nash}$ 或 $p = c$

合谋可持续性条件

合谋的现值（贴现到 $t=1$）：

$$V^{\text{coll}} = \frac{\pi^{\text{coll}}}{1-\delta}$$

背叛的最优策略——在当期生产最优反应产量（给定对手继续合作），然后从下一期开始承受永久惩罚：

$$V^{\text{dev}} = \pi^{\text{dev}} + \frac{\delta \cdot \pi^{\text{Nash}}}{1-\delta}$$

合谋可持续当且仅当 $V^{\text{coll}} \geq V^{\text{dev}}$：

$$\frac{\pi^{\text{coll}}}{1-\delta} \geq \pi^{\text{dev}} + \frac{\delta \cdot \pi^{\text{Nash}}}{1-\delta}$$

重排后——**关键的临界贴现因子公式：**

$$\delta \geq \delta^* = \frac{\pi^{\text{dev}} - \pi^{\text{coll}}}{\pi^{\text{dev}} - \pi^{\text{Nash}}} \tag{8.1}$$

命题 8.1 (Cournot 双寡头的临界贴现因子). *对于线性需求 $P = a - Q$，边际成本 $c = 0$：*

- $\pi^M = a^2/4$，$\pi^{coll} = \pi^M/2 = a^2/8$
- *单期 Nash*：$q^N = a/3$，$\pi^{Nash} = a^2/9$
- *背叛*：$q^{dev} = 3a/8$（*给定对手生产 $a/4$*），$\pi^{dev} = 9a^2/64$
- $\delta^* = (9a^2/64 - a^2/8)/(9a^2/64 - a^2/9) = 9/17 \approx 0.529$

解读：*如果 $\delta \geq 0.529$（等价于年利率 $r \leq 89\%$），合谋可持续。这个临界值相当低——意味着在"正常"利率下，合谋是可持续的。*

8.6.3　无名氏定理（The Folk Theorem）

【核心】

定义 8.5 (无名氏定理). *在无限重复博弈中，对于 δ 足够接近 1，***任何***满足个人理性（individual rationality）的支付向量都可以在某个子博弈完美均衡中作为每一期的平均支付被支持。*

关键含义：

1. **合谋可以在任何水平上维持**——从刚刚超过 Nash 到完全垄断利润
2. 合谋水平不是由博弈的唯一均衡"给定"的——而是从**无限多个**可能均衡中选择的

3. 博弈论告诉我们合谋**可能**——但它没告诉我们合谋**会**以什么形式发生

无名氏定理的" 力量" 和" 弱点":

- **力量**：证明了即使没有显式合同或沟通，企业也可以通过" 默契" 维持任意高水平的合谋——只要它们足够耐心。这为反垄断担忧合谋提供了理论基础。
- **弱点**（Fisher 1989 批评，第 ~7~ 章 §7.7）：多重均衡意味着" 任何事都可以发生" ——博弈本身不能确定哪个均衡将被选择。" 一个能解释一切的理论什么都解释不了。"

8.6.4 最优惩罚：Abreu (1986)

【进阶】

问题：触发策略（永久 Nash 惩罚）真的是最优的惩罚吗？如果惩罚太温和 $\rightarrow$ 需要更高的 δ 来维持合谋。如果惩罚太严厉 $\rightarrow$ 惩罚本身不可信（企业进入惩罚期后，它们有动机重新合作而不是执行惩罚）。

Abreu 的贡献——" 胡萝卜加大棒"（stick-and-carrot）：

- 惩罚阶段 1（大棒）：执行一个**非常严厉**的行动（产量极高，价格极低——比 Nash 更严厉）
- 如果惩罚被正确执行 $\rightarrow$ 回到合作阶段（胡萝卜）
- 如果在惩罚阶段有企业偏离 $\rightarrow$ 重新开始惩罚阶段
- 最优惩罚 = 在可信性约束下最严厉的惩罚——惩罚阶段的均衡支付必须" 刚好" 满足参与约束
- **最优惩罚可以支持比触发策略更低 δ^* 的合谋**

8.6.5 Green & Porter (1984)：不完全监测下的价格战

【核心】/【进阶】

核心问题：前面的触发策略分析假设企业可以**完美观察**彼此的行动。但现实中，企业只能观察到自己的销量和市场价格——不能直接观察其他企业的定价。如果你的销量下降了，是因为对手在暗中降价（背叛），还是因为市场需求受到了负面冲击？

模型设定：

- n 家企业，同质产品，Cournot 竞争
- 需求受随机冲击影响：$P_t = \theta_t \cdot P(Q_t)$，其中 θ_t 是 i.i.d. 需求冲击

- **关键信息假设**：每家企业观察自己的产量 q_i 和市场价格 P_t——但**不观察**总产量 $Q_{-i,t}$ 或需求冲击 θ_t

推论：当企业观察到低价格时，它不能区分：(a) 某企业在背叛（生产了更多），还是 (b) 市场需求受到负面冲击（θ_t 低）。→ 合谋必然涉及**均衡路径上的价格战**。

Green-Porter 均衡：

- **合作期**：所有企业生产合谋产量 q^{coll}
- **触发价格** $\hat{p}$：如果市场价格 $P_t < \hat{p}$ → 触发惩罚期。$\hat{p}$ 被设定为使" 误报"（需求冲击导致 $P_t < \hat{p}$ 即使没人背叛）和" 漏报"（背叛发生但 $P_t \geq \hat{p}$）达到最优折衷
- **惩罚期**：T 期——所有企业生产 Nash 产量 q^{Nash}
- **恢复**：T 期后，回到合作期

价格战是均衡现象

Green-Porter 的核心洞见是：**价格战不是因为某家企业" 非理性" 或" 犯错"——它们是在不完美监测下维持合谋所必需的。**企业选择触发一个" 有时会错误惩罚所有人" 的机制——因为它是唯一可用的威慑工具。这解释了现实中的现象：卡特尔行业经常出现周期性价格战——看似" 愚蠢"（破坏了卡特尔利润），实际上可能正是维持卡特尔所必需的。

8.6.6　Rotemberg & Saloner (1986)：繁荣期的价格战

【进阶】

核心问题：如果需求是**可观察的**（无监测问题）但随时间变化呢？合谋应该在繁荣期还是衰退期更难维持？

Rotemberg-Saloner 的反直觉答案：繁荣期！

模型直觉：

- 假设需求在每期 t 可以是高（H）或低（L），独立同分布
- 在繁荣期（H 需求），背叛的即时利润更大——当前的" 蛋糕" 更大
- 但未来惩罚的现值不变——因为未来需求独立于当前需求，预期惩罚价值不变
- → δ^* 在繁荣期**更高**（需要更大的耐心来维持合谋）
- → 如果 δ 是固定的但不够高以在繁荣期维持完全合谋 → 企业必须在繁荣期**降低合谋价格**（增加产量）来减小背叛的诱惑
- → **价格在需求高时反而下降——反周期的加价**（counter-cyclical markups）

表 8.6: 价格战的两种模型

	Green-Porter (1984)	**Rotemberg-Saloner (1986)**
需求	i.i.d. 冲击，**不可观测**	i.i.d. 冲击，**可观测**
价格战的驱动因素	不完美监测——无法区分背叛和需求冲击	繁荣期背叛诱惑更大 → 需降低合谋价格
价格战是”错误”吗？	可能是——需求冲击导致误触发	不是——是有意的利润最大化行为

与 Green-Porter 的对比：

Haltiwanger & Harrington (1991) 的修正：如果需求冲击存在**时间序列依赖性** → 繁荣期不仅当前利润更高，而且未来预期利润也更高（需求持续繁荣） → 未来惩罚现值也上升了。如果需求高度持续性 → 未来惩罚增加 > 当期诱惑增加 → 繁荣期可能**更有利于**合谋。合谋的周期性取决于需求的时间序列性质。

8.7 合谋 II：实证识别、便利因素与瓦解

【核心】/【进阶】

8.7.1 Porter (1983)：联合执行委员会（JEC）——结构估计的先驱

【核心】

研究问题：19 世纪美国铁路卡特尔——联合执行委员会（Joint Executive Committee, JEC）——是否如 Green-Porter 模型预测的那样行动？

制度背景：

- 1880 年代，芝加哥和东海岸之间的铁路运输由 JEC 协调——一个**显式卡特尔**
- JEC 分配市场份额并设定运价
- 在 1887 年州际商业法案之前，没有联邦规制铁路
- JEC 每周公布每一条铁路的运输量 → 关于”谁背叛了”的信息相对透明

Porter 的方法论贡献——结构估计：

- Porter 没有做”跨行业 SCP 回归”——而是对一个特定行业（铁路）建立了**结构博弈模型**
- 模型核心：观测到的运价和运量序列由企业在两种”制度”（regimes）之间切换生成：

  - **合谋制度**：企业遵守配额，价格高
  - **价格战制度**：企业不遵守配额，价格低

- 制度之间转换由需求冲击和企业是否背叛驱动（如 Green-Porter 逻辑）

主要发现：

1. **存在制度切换**——数据确实显示了高价格期和低价格期（价格战）之间的切换，与 Green-Porter 理论一致
2. **价格战是短暂的**——平均持续时间远短于合作期
3. **价格战在需求意外低时更可能发生**——与 Green-Porter 逻辑一致

方法论意义——NEIO 预览

Porter (1983) 是 NEIO 的奠基性研究之一：不是问" 集中度与利润相关吗？"（SCP），而是问" 数据与一个特定的博弈论合谋模型一致吗？"——通过结构估计。它标志着从" 简约式回归" 到" 结构估计" 的范式转变。这将在第 ~10~ 章–第 ~12~ 章中成为主要方法论。

8.7.2　便利合谋的因素

【核心】/【进阶】

1. 多市场接触（Multimarket Contact）——Bernheim & Whinston (1990)：

- 机制：相同企业在多个市场中竞争 → 在一个市场中背叛的收益必须与**在所有市场**中被惩罚的未来损失相权衡
- → 合谋在多市场接触下更容易维持——惩罚更严厉（跨市场惩罚）
- 经验证据：航空公司研究——两方在多个航线上竞争时，票价更高

2. 信息交换：价格透明（公开宣布价格 → 更快检测背叛）；提前通知价格变动（允许企业在实施前调整）；行业协会发布详细数据（可能成为合谋的便利工具）。

3. 最惠国待遇条款（Most-Favored-Nation Clauses, MFN）：

- 卖方承诺向当前买方提供的价格不会高于向任何其他买方提供的价格
- 看似" 保护消费者"：确保你得到最低价
- **竞争效应**：MFN 条款实际上**降低了企业降价的激励**——因为降价不仅适用于特定买方，还适用于所有过去签订 MFN 的买方 → 降价成本更高 → 价格竞争软化 → 合谋更容易

4. 竞争性匹配条款（Meeting-Competition Clauses, MCC）：卖方承诺" 匹配任何竞争对手的价格"。看似促进竞争——实际上使降价更不吸引（降价企业知道对手会立即匹配） → 降价的动机消失 → 合谋更容易。

5. 交叉持股（Cross-Ownership）：企业 A 持有企业 B 的股份 → A 的部分内化 B 的利润 → 降价时会考虑对 B 的影响 → 共同利益 → 更软的价格竞争。

6. 对称性：对称企业（相同成本、产能、市场份额）更容易合谋。不对称 → 低成本企业更可能背叛（成本优势使其较少害怕惩罚）。

8.7.3 合谋的实证检测

【进阶】

1. 结构方法——Bresnahan (1982) 行为参数：估计企业的" 市场势力行为参数" θ。$\theta = 0$ → Bertrand；$\theta = 1/N$ → Cournot；$\theta = 1$ → 完全合谋。如果 θ 显著 $> 1/N$ → 证据表明合谋性定价（详见第 ~10~ 章）。

2. 结构突变检测：宽大计划引入或反垄断调查前后——价格是否出现**结构性下降**？卡特尔被打破时的价格变化幅度 = 合谋加价的估计。

3. 方差筛选（Variance Screens）：合谋期间价格应该更**稳定**——企业不会像竞争那样频繁调整价格。如果观察到某个行业的**价格方差异常低**（相对于合理成本波动预期） → 合谋的间接证据。

4. 机器学习与检测：当代前沿——使用异常检测、模式识别来筛查价格数据中的合谋模式。

8.7.4 宽大计划（Leniency Programs）——用博弈论瓦解合谋

【进阶】

核心机制：反垄断机构如何利用博弈论来瓦解合谋？

宽大计划的基本设计：

- 反垄断机构向**第一个**自愿报告卡特尔并向调查全面合作的企业提供完全的罚款豁免（及刑事责任豁免）
- 后来的合作者获得罚款减少（但不豁免）
- **这是囚徒困境的制度化**：给予每家企业一个" 先背叛，你先受益" 的新选项

博弈论分析——Motta & Polo (2003)：

- 没有宽大计划时：合谋企业面临被查处的概率 p 和罚款 F。只要 δ 足够大，合谋依然可持续
- 有宽大计划时：每家企业有一个**新选项**——背叛卡特尔并向机构举报。第一个举报企业获得豁免 + 仍可从未来合谋中获利（如果其他企业也被查处）

- 在宽大计划下，合谋需要满足**更严格**的条件——” 争先恐后背叛” 的激励

三个关键设计特征：

1. **” 先到先得”**：只有第一个报告企业获得完全豁免——最大化” 争先恐后背叛” 的激励
2. **” 主动揭露义务”**：企业必须提供机构尚未掌握的信息
3. **” 全面合作”**：获得豁免要求持续合作——提供文件、证人等

经验证据：美国 1993 年企业宽大计划修订 → 卡特尔举报剧增；EU 1996 年引入 → 大多数 EU 卡特尔案件现在由宽大举报引发。**这是博弈论指导” 制度设计” 的最成功案例之一。**

8.8　策略行为的中国脉络

【进阶】

8.8.1　中国《反垄断法》中的策略行为

2022 年修订的《反垄断法》第 22 条（原第 17 条）禁止具有市场支配地位的经营者从事下列滥用行为：

- (二) 没有正当理由，以低于成本的价格销售商品（**掠夺性定价**）
- (四) 没有正当理由，限定交易相对人只能与其进行交易（**排他性合同**——” 二选一”）

中国法律的独特性：比美国第 2 条 Sherman Act 更**具体**——直接列出了禁止行为；” 正当理由”（legitimate business justification）的举证责任分配在实践中非常关键。

8.8.2　标志性案例

1. Alibaba ” 二选一” 案 (2021)——排他性合同的现代版：

- 国家市场监管总局（SAMR）对阿里巴巴处以 **~182.28 亿**（约 $28 亿）罚款——中国反垄断史上最大罚单
- 核心指控：强迫商家在阿里巴巴平台和竞争对手平台之间” 二选一”——这正是 Aghion-Bolton (1987) 排他性合同逻辑：通过锁定足够多卖家，使竞争对手平台无法达到临界规模

- 网络效应放大了排他性合同的封锁效应：平台 A 排除了卖家 → 买家也去平台 A → 竞争对手更无法吸引卖家 → 排他性合同的影响在双边市场中是**超线性的**

2. 360 v. Tencent (2013)——滥用市场支配地位：

- 中国最高法院建立了中国互联网反垄断的多个先例
- 争议焦点：腾讯是否在即时通讯市场中具有支配地位？最高法院判决对 360 不利——裁定相关产品市场比即时通讯更广
- 该判决引发了关于中国数字市场反垄断标准的持续辩论

3. Didi-Kuaidi 补贴战 (2014–2015)——掠夺性定价的变体：

- 腾讯（支持滴滴）和阿里巴巴（支持快的）在打车应用市场进行了中国互联网史上规模最大的补贴战
- 从策略行为角度分析：这不是传统的掠夺性定价（一方消灭另一方后提价）——这是**同时在争夺市场控制权**
- 最终结果——滴滴和快的在 2015 年合并——" 打不过就买" 的策略
- **深口袋维度**：两个超级资本的" 深口袋" 竞争 → 任何一方都无法通过掠夺消灭另一方 → 合并成为均衡结果

8.8.3 中国合谋执法的特征

中国的卡特尔执法在传统制造业领域活跃——水泥、汽车零部件、医药、液晶面板。与 EU 类似，中国采用宽大计划（《反垄断法》第 46 条，2022 修订后进一步强化）。**中国特色的横向垄断协议执法**：行业协会经常充当卡特尔的组织者——这在西方相对少见。纵向垄断协议（RPM）的执法在 2013 年茅台/五粮液案后变得活跃——但纵向协议的" 当然违法 vs. 合理原则" 边界在中国法律中仍不明确。

8.9 本章小结

核心概念（10 个）：

1. **承诺与可信性**（commitment and credibility）——进入遏制、掠夺和合谋的共同基础：策略行动要有效率，必须在子博弈中是可信的（SPE/PBE）
2. **Dixit 产能承诺**——沉没投资 → 改变事后边际成本 → 事后最优产量改变 → 威胁可信

3. **Fudenberg-Tirole 策略动物分类学**——战略替代品/互补品 × 投资使在位者更强硬/更温和 → Top Dog, Puppy Dog, Lean and Hungry, Fat Cat

4. **Milgrom-Roberts 限价信号**——不完全信息 + 序贯行动 → 价格作为成本类型的信号 → 分离/混同 PBE → 限价成为理性的均衡现象

5. **连锁店悖论与声誉机制**——完全信息：掠夺从不发生（SPE）；不完全信息（$\varepsilon > 0$）：掠夺可能是均衡行为（PBE） → " 四人帮" 解决了 Selten 悖论

6. **深口袋掠夺**——资本市场不完美 → 在位者可以通过掠夺耗尽进入者的融资能力 → 即使芝加哥理性企业也会掠夺

7. **重复博弈与合谋可持续性**——$\delta \geq \delta^* = (\pi^{\text{dev}} - \pi^{\text{coll}})/(\pi^{\text{dev}} - \pi^{\text{Nash}})$ → 临界贴现因子

8. **无名氏定理与最优惩罚**——对于 $\delta \to 1$，任何高于 Nash 的支付向量都是 SPE → Abreu (1986) 最优惩罚可以支持比触发策略更低 δ^* 的合谋

9. **不完全监测下的价格战**——Green-Porter (1984) 价格战作为均衡现象；Rotemberg-Saloner (1986) 繁荣期反周期加价

10. **宽大计划**——囚徒困境的制度化：给予卡特尔成员" 先背叛先受益" 的新选项 → 瓦解合谋

核心公式（4 个）：

1. 合谋临界贴现因子：$\delta^* = \dfrac{\pi^{\text{dev}} - \pi^{\text{coll}}}{\pi^{\text{dev}} - \pi^{\text{Nash}}}$

2. 合谋可持续性条件：$V^{\text{coll}} = \dfrac{\pi^{\text{coll}}}{1-\delta} \geq \pi^{\text{dev}} + \dfrac{\delta \cdot \pi^{\text{Nash}}}{1-\delta} = V^{\text{dev}}$

3. Dixit 在位者事后边际成本函数：$MC_I(q_I) = w$ for $q_I \leq K$，$w + r$ for $q_I > K$

4. 单交叉条件（Milgrom-Roberts）：$\partial^2 \pi_I / \partial p_1 \partial c > 0$ → 低成本类型从降价中损失更小

本章位置：

- ← 第 ~7~ 章提供了所有工具（Nash, SPE, BNE, PBE, 重复博弈, Folk Theorem, Fudenberg-Tirole 分类学概要）——本章将每一个工具绑定到至少一个具体 IO 应用

- ← 第 ~5~ 章–第 ~6~ 章提供了智识张力（芝加哥掠夺批判、可竞争性进入假定）——本章用博弈论严格解决这些张力

- → 第 ~9~ 章将信息不对称从" 模型的一个特征" 提升为" 核心分析对象"（Laffont-Tirole 激励理论）
- → 第 ~10~ 章的 NEIO 方法（Conduct Parameter, Porter 结构估计）是 Ch.8 合谋检测的直接延伸
- → 第 ~13~ 章–第 ~15~ 章将策略行为分析扩展到平台市场、算法定价和当代反垄断辩论

8.10 习题

基础（概念理解）

习题 8.1 [承诺的可信性]

(a) 用博弈树表示 Dixit (1980) 的两阶段进入博弈。标注在位者的事前产能选择、进入者的进入决策、以及事后 Cournot 竞争。

(b) 为什么在 Spence (1977) 中" 过剩产能阻挠进入" 的威胁可能不可信，而在 Dixit (1980) 的 SPE 中变得可信？

(c) " 承诺必须同时是不可逆的**且**改变事后激励。" 给出一个反例：一个行动是不可逆的，但**不**改变事后激励——因此不构成有效的承诺。

习题 8.2 [**Fudenberg-Tirole 分类学**] 对于以下每个情景，判断在位者应该过度投资还是投资不足，并将每个案例与 Fudenberg-Tirole 分类学（Top Dog, Puppy Dog, Lean and Hungry, Fat Cat）匹配：

(a) 在位者投资于降低成本的技术，产品市场竞争是 Cournot（数量）。

(b) 在位者投资于降低成本的技术，产品市场竞争是 Bertrand（价格，差异化产品）。

(c) 在位者投资于广告以提升消费者支付意愿，产品市场竞争是 Bertrand。

(d) 在位者投资于广告以提升消费者支付意愿，产品市场竞争是 Cournot。

习题 8.3 [连锁店悖论]

(a) 在 Selten (1978) $N = 3$ 的连锁店博弈中（完全信息），用逆向归纳写出 SPE。掠夺在任何一个市场中发生吗？

(b) 现在假设进入者认为在位者有 10% 的先验概率是" 强硬型"（总是掠夺）。对于 $N = 2$，求 PBE。掠夺是否可能在均衡中发生？为什么？

(c) ” 芝加哥学派说掠夺从不理性。Selten (1978) 用博弈论**证明**了芝加哥是对的。但 Kreps-Wilson (1982) 用博弈论**证明**了芝加哥是错的。”——解释这个表面矛盾。

习题 8.4 [合谋的囚徒困境结构] 两家航空公司服务于同一条航线。需求为 $Q = 100 - P$，边际成本为 20。如果合谋，它们平分垄断利润。如果一家背叛（设定比合谋价格低 ε 的价格），它捕获整个市场。如果两家都背叛，它们进行 Bertrand 竞争。

(a) 将此博弈表示为 2×2 矩阵（策略：合作/背叛），填入支付。

(b) 一次性博弈的 Nash 均衡是什么？它 Pareto 有效吗？

(c) 如果博弈无限重复且 $\delta = 0.6$，触发策略能否维持合谋？（提示：$\pi^{\text{coll}} = \pi^M/2$，$\pi^{\text{dev}} = \pi^M$，$\pi^{\text{Nash}} = 0$）

中级（模型推导）

习题 8.5 [Dixit 产能承诺的数值求解] 考虑 Dixit (1980) 模型。需求 $P = 100 - q_I - q_E$。在位者的产能 K 成本为每单位 $c_0 = 5$（沉没）。对于 $q_I \le K$，可变成本 $w = 0$。对于 $q_I > K$，超额产量边际成本 $w + r = 10$。进入者（如果进入）的可变边际成本 $= 10$。进入的固定成本 $F = 200$。

(a) 给定 K，如果进入发生，在位者的事后边际成本函数 $MC_I(q_I, K)$ 是什么？在位者的反应函数在 $q_I = K$ 处有何特征？

(b) 固定 $K = 20$。如果进入者进入，求事后 Cournot 均衡的 (q_I, q_E, P)。进入者的利润是多少？它会进入吗？

(c) 求阻挠进入所需的最小 K^*。计算在位者在 K^* 处的利润，并与容纳进入时的利润比较。

(d) 如果 K 的沉没成本消失（c_0 可在退出时回收），进入可以被阻挠吗？与可竞争性理论的关系是什么？

习题 8.6 [合谋的临界贴现因子与比较静态] n 家相同企业在无限重复的 Cournot 博弈中互动。需求 $P = a - bQ$，边际成本 $c \ge 0$。

(a) 推导单期 Nash 均衡、合谋（联合垄断，n 家企业均分）和最优背叛（在对手合作时）的表达式。

(b) 推导临界贴现因子 $\delta^*(n)$ 作为 n 的函数。计算 $\partial \delta^*/\partial n$ 的符号并给出经济学解释。

(c) 计算 $\lim_{n\to\infty} \delta^*(n)$。当 n 非常大时，这对合谋可能性意味着什么？

(d)（加分项）现假设企业使用 Abreu (1986) 最优惩罚而非触发策略。相对于触发策略，最优惩罚下的 $\delta^*(n)$ 是更高还是更低？为什么？

习题 8.7 [Milgrom-Roberts 限价信号的简化版] 在位者的成本为 $c \in \{c_L = 0, c_H = 10\}$，先验概率 $\Pr(c = c_L) = 1/3$。阶段 1：在位者选择 p_1。阶段 1 的需求为 $q = 100 - p_1$。阶段 2：进入者观察 p_1，决定进入/不进入。如果进入，阶段 2 利润是：如果 $c = c_L \to \pi_E = -50$；如果 $c = c_H \to \pi_E = 50$。进入者仅当预期利润 ≥ 0 时才进入。贴现因子 $\delta = 1$。

(a) 在位者的完全信息阶段 1 最优垄断价格是什么（分别对于两种类型）？阶段 1 的相应利润是多少？

(b) 在什么条件下存在分离均衡？求支持分离的最低和最高 p_1^L（低成本在位者的信号价格）。

(c) 存在混同均衡吗？在什么条件下？

进阶（概念延伸）

习题 8.8 [统一框架——进入遏制、掠夺与合谋的共同结构] 本章的三部分（进入遏制、掠夺、合谋）可能看起来像三个不同的主题。论证它们共享一个深层共同结构：

(a) 每个部分中，" 囚徒困境" 结构在哪里？（提示：对于进入遏制，在位者和进入者的" 合作" 和" 背叛" 分别是什么？）

(b) 每个部分中，" 承诺" 或" 可信性" 问题是如何被解决的？

(c) 这三个部分如何映射到第 ~7~ 章的七步"Tirole 框架"？制作一个对照表。

(d) 基于以上分析，你认为进入遏制、掠夺和合谋是三个不同现象，还是同一策略行为逻辑在三个不同情境中的应用？论证你的立场。

习题 8.9 [平台市场中的策略行为] 考虑以下情景：Uber 进入了一个 DiDi 已经运营的城市的打车市场。两家平台都需要吸引乘客（价格敏感，可多归属）和司机（通常单归属，对总行程量和价格敏感）。

(a) DiDi 可以使用什么进入遏制策略？区分" 降价" 和" 排他性合同"（与司机签订排他性合同——类似于 Aghion-Bolton）——各自的 Fudenberg-Tirole 分类学位置是什么？

(b) 双边市场的网络效应如何改变标准 Dixit 模型的逻辑？（提示：当在位者" 过度投资" 于司机端时，这会通过跨边网络效应影响乘客端吗？）

(c) 中国《反垄断法》2022 年修订明确禁止平台企业使用" 数据与算法、技术、资本优势以及平台规则" 进行排他性行为。这个条款在多大程度上解决了 Aghion-Bolton (1987) 排他性合同模型中识别的市场失灵？

习题 8.10 [算法合谋与当代反垄断前沿] 现代定价算法正在改变合谋的动力学：

(a) 如果两家企业使用相同的第三方定价算法（已被 DOJ 起诉），在什么意义上这构成了" 合谋"——即使企业之间没有显式沟通？这与重复博弈模型有什么不同？

(b) 如果定价算法是强化学习（reinforcement learning）而非" 显式" 合谋——算法通过试错" 学到" 了合谋性定价——这属于哪种法律范畴？" 意图" 在这个情境中意味着什么？

(c) 假设你要为反垄断机构设计一个检测算法合谋的" 筛选测试"。基于 Green-Porter (1984) 和 Rotemberg-Saloner (1986) 的模型，你会寻找什么样的数据特征？什么将算法合谋与" 正常的" 寡头相互依赖区分开来？

8.11　延伸阅读

★ 必读（核心）

Tirole (1988) *The Theory of Industrial Organization*, Ch.8 (Entry, Accommodation and Exit) & Ch.6 (Short-Run Price Competition, 重复博弈合谋部分)。进入遏制/容纳/合谋的标准研究生级处理——Dixit, Fudenberg-Tirole, 无名氏定理。

Dixit (1980) "The Role of Investment in Entry-Deterrence," *Economic Journal*, 90, 95–106. 产能承诺模型的原始论文——SPE 在 IO 中的经典应用。

Fudenberg & Tirole (1984) "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look," *AER P&P*, 74, 361–366. 策略动物分类学的原始陈述——四象限框架。

Kreps, Milgrom, Roberts & Wilson (1982) "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *JET*, 27(2), 245–252. "四人帮"——声誉如何解决连锁店悖论。

Porter (1983) "A Study of Cartel Stability: The Joint Executive Committee, 1880–1886," *Bell Journal*, 14(2), 301–314. JEC 结构估计——合谋理论与实证之间的桥梁（NEIO 先驱）。

★★ 进阶（深化理解）

Milgrom & Roberts (1982) "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis," *Econometrica*, 50(2), 443–459. 限价信号模型的完整处理——PBE 推导。

Bolton & Scharfstein (1990) "A Theory of Predation Based on Agency Problems in Financial Contracting," *AER*, 80(1), 93–106. 深口袋掠夺——资本市场不完美的角色。

Green & Porter (1984) "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 52(1), 87–100. 不完全监测下的价格战。

Rotemberg & Saloner (1986) "A Supergame-Theoretic Model of Price Wars During Booms," *AER*, 76(3), 390–407. 繁荣期价格战——反周期加价分析。

Abreu (1986) "Extremal Equilibria of Oligopolistic Supergames," *JET*, 39, 191–225. 最优惩罚——" 胡萝卜加大棒" 策略。

Aghion & Bolton (1987) "Contracts as a Barrier to Entry," *AER*, 77(3), 388–401. 排他性合同阻挠进入的形式化。

Salop & Scheffman (1983) "Raising Rivals' Costs," *AER P&P*, 73(2), 267–271. 提高对手成本的开创性论文。

Bernheim & Whinston (1990) "Multimarket Contact and Collusive Behavior," *RAND Journal*, 21(1), 1–26. 多市场接触便利合谋的正规理论。

★★★ 前沿（研究前沿与中国语境）

Harrington (2006) "How Do Cartels Operate?" *Foundations and Trends in Microeconomics*, 2(1), 1–105. 卡特尔如何实际运作的全面调查。

Motta & Polo (2003) "Leniency Programs and Cartel Prosecution," *Int'l J. of Industrial Organization*, 21(3), 347–379. 宽大计划的博弈论分析。

Salop (2024) "The 2023 Merger Guidelines: A Post-Chicago and Neo-Brandeisian Integration," *Review of Industrial Organization*, 65(1), 79–128. 当代并购指南整合策略行为逻辑的方式。

张维迎 《博弈论与信息经济学》，格致出版社。中文博弈论标准教材——与本章互读。

Ezrachi & Stucke (2016) *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*, Harvard University Press. 算法合谋的全面分析。

国家市场监管总局 (2021) 阿里巴巴" 二选一" 行政处罚决定书（国市监处〔2021〕28号）。中国反垄断标志性案例。

→ 第 ~9~ 章：不对称信息与规制——Laffont-Tirole 激励理论

本章展示了当在位者和进入者之间存在信息不对称时，价格可以作为信号（Milgrom-Roberts 限价）而声誉可以遏制进入（Kreps-Wilson 掠夺）。但所有这些模型中，信息不对称是*模型的一个特征*——分析和结论集中在企业之间（B2B）的策略互动上。

第 ~9~ 章将信息不对称从" 模型的一个特征" 提升为*核心分析对象*——并引入规制者作为第三方参与者。关键过渡：Ch.8 处理的是**企业与企业之间**的信息不对称（在位者 vs. 进入者）；Ch.9 处理的是**企业与规制者之间**的信息不对称（被规制企业 vs. 规制者）。Laffont-Tirole 激励理论的核心问题——当规制者不知道企业的真实成本时，如何设计规制方案？——是本章信号传递逻辑的直接延伸。

第九章　不对称信息与规制：Laffont-Tirole 激励理论

章首语

1993 年，Jean-Jacques Laffont 和 Jean Tirole 出版了 *A Theory of Incentives in Procurement and Regulation*。这本书的开篇不是关于规制或政府采购——而是关于一个更根本的问题：**当你不知道对方知道什么的时候，你如何设计合同？**

二十年后，Tirole 在诺贝尔奖演讲中回顾了这一研究纲领。他说，机制设计——特别是在规制中的应用——可能是经济学对公共政策最直接、最重要的贡献之一。当英国在 1984 年将 British Telecom 私有化并引入 RPI-X 价格上限规制时，当中国在 1990 年代推行国企承包制改革时，当欧盟在 2000 年代设计电信接入定价框架时——这些制度设计的背后都有一个共同的理论逻辑：**如何在信息不对称的条件下激励效率，同时控制信息租金的分配效应。**

本章讲述这个逻辑。它从最简单的信息不对称出发，逐步构建 Laffont-Tirole 激励规制框架。第 ~8~ 章处理的是企业与企业之间的信息不对称——在位者知道自己的成本而进入者不知道。本章将镜头转向企业与规制者之间的信息不对称——并引入机制设计作为全新的分析工具。

9.1　核心问题

读完本章，你应该能够回答以下问题：

1. **规制为什么难？**规制者面临的根本困境是什么？逆向选择（企业知道自己的成本，规制者不知道）和道德风险（企业降低成本的努力不可观测）如何相互作用？为什么”让企业报告成本，然后据此定价”行不通？

2. **Laffont-Tirole 框架如何形式化这个困境？** $C = \beta - e$ 的成本结构捕捉了什么经济学直觉？规制者目标函数中的 λ（公共资金的影子成本）为什么是决定最优规制的关键参数？核心权衡——激励 vs. 抽取租金——如何在数学上精确表达？

3. **什么是”菜单规制”？**为什么给企业提供一组合同菜单（固定价格 + 成本加成 +

中间选项）比单一合同更好？显示原理（Revelation Principle）如何保证这种” 自我选择” 机制实现了最优？

4. **信息租金的本质是什么？**高效率企业为什么必然获得正的信息租金？” 无扭曲在顶部，扭曲在底部” 为什么是几乎所有甄别模型的通则？单交叉条件（single-crossing property）如何保证分离是可行的？

5. **机制设计的边界在哪？**防合谋原理如何将两层委托代理扩展为三层？不完备合同在什么条件下不损害效率（Maskin-Tirole 定理）？标尺竞争如何用” 相对绩效评估” 替代” 绝对成本信息”？

6. **这些理论如何落地？** RPI-X 价格上限、回报率规制、接入定价的制度设计逻辑是什么？中国的国企改革和自然垄断规制在什么意义上体现了激励规制的核心逻辑？

9.2　规制面临的根本问题：信息不对称

9.2.1　自然垄断与规制的必要性

为什么需要规制？回顾第 ~2~ 章–第 ~6~ 章中分散出现的论证：

定义 9.1 (自然垄断). *一个行业是自然垄断（natural monopoly）当且仅当成本函数在整个相关产量范围内是次可加的（subadditive）：*

$$C(Q) < C(q_1) + C(q_2) + \cdots + C(q_n), \quad \sum_{i=1}^{n} q_i = Q$$

即由一家企业生产整个市场产量比由多家企业共同生产更便宜。

自然垄断的充分条件（也是最常见的条件）是规模经济：平均成本 $AC(Q)$ 在整个相关产量范围内递减。但在某些情况下，即使平均成本上升，成本函数仍可能是次可加的——次可加性是比规模经济更一般的条件。

在自然垄断下，自由进入导致” 破坏性竞争”（多方重复投入固定成本）或自然垄断者行使市场势力（价格远高于边际成本）。这两种结果都是社会低效的。政府面临两种基本政策选择：

(a) **公有制**（government ownership）：政府直接拥有并经营企业

(b) **私有制 + 规制**（private ownership with regulation）：企业私有，但政府通过规制合同约束其行为

规制者的工具包：价格规制（设定最高价格或收益率）、进入规制（控制市场中企业的数量）、质量规制（设定最低服务标准）、转移支付（税收补贴或征税）。

本章焦点：激励规制（incentive regulation）——不是”规制者命令企业做什么”，而是”规制者设计合同使企业在追求自身利润最大化的过程中做规制者想做的事”。这是 Laffont-Tirole 框架的核心创新：将规制问题重新表述为不对称信息下的合同设计问题。

9.2.2 逆向选择与道德风险：两种信息不对称的交织

规制者不知道关于被规制企业的两件事：

(1) 逆向选择（adverse selection）——隐藏信息（hidden information）：

- 企业知道自己的技术效率参数 β——边际成本的”固有”成分（技术水平、管理能力、地理位置等）
- 规制者不知道 β——只能知道 β 的可能范围（如 $\beta \in [\beta_{\min}, \beta_{\max}]$）及其先验分布 $F(\beta)$
- 如果规制者直接问”你的 β 是多少？”——高效率企业（低 β）会报告自己是低效率的（高 β）以获得更优惠的规制条件
- 这是签约前的信息不对称——企业在合同签订之前已经知道自己的类型

(2) 道德风险（moral hazard）——隐藏行动（hidden action）：

- 企业可以通过努力 e 来降低成本——但努力对企业管理者有负效用 $\psi(e)$
- 规制者观察不到 e——只能观察到实现的总成本 $C = \beta - e$
- 因为规制者不能区分”低成本是因为高努力”还是”低成本是因为高固有效率”——企业可以通过减少努力来偷懒
- 这是签约后的信息不对称——企业在合同签订之后选择努力水平

核心创新：两种信息不对称的相互作用

在经典的委托–代理文献中，逆向选择（Mirrlees 1971, Baron-Myerson 1982）和道德风险（Holmström 1979）通常被分开处理。Laffont-Tirole 的关键洞见是：**在规制中，两者同时存在且相互作用**——因为同一个可观测变量 C 同时受到隐藏类型 β 和隐藏行动 e 的影响。这使规制设计比”纯”逆向选择或”纯”道德风险更难——也更有趣。

为什么不能简单地”问企业”？三种天真的方案及其失败：

- *方案 A*：规制者问：”你的 β 是多少？” $\to$ 高效率企业（低 β）会报告自己是高 β $\to$ 获得更慷慨的规制条件

- 方案 B：规制者命令：”你必须努力 $e = e^*$” → 企业可能同意，但合同签订后选择 $e < e^*$（规制者看不到 e）
- 方案 C：规制者报销全部成本 + 给固定费 → 企业没有任何激励降低成本（$e = 0$）→ 成本膨胀

核心困境的精确表述：

规制的基本权衡

- **高能激励**（让企业承担成本）→ 企业积极降低成本 → 但高效率企业获得大量信息租金（因为它以低成本完成任务所获报酬超过保留效用）
- **低能激励**（报销成本）→ 信息租金消失 → 但企业没有激励降低成本 → 生产成本过高

最优规制在高能激励的效率收益和低能激励的租金节省之间取得平衡。

9.2.3　公共资金的影子成本 λ：连接规制与宏观财政

什么是 λ？政府通过税收筹集补贴给企业的每一元钱，其社会成本不是 1 元——因为税收造成无谓损失（deadweight loss）。设筹集 1 元税收的社会成本为 $1 + \lambda$ 元。λ 取决于：

- 税收体系的效率（宽税基、低税率的体系具有较低的 λ）
- 经济的总体税负水平（在已经高税负的经济中，λ 更高——拉弗曲线逻辑）
- 政府支出的其他竞争性需求（在财政紧张时期，λ 更高）

λ 是决定最优规制”斜率”（incentive power）的核心参数：

- $\lambda \to 0$（公共资金无影子成本）→ 规制者可通过一次性转移支付完全消除信息租金 → 最优规制趋近固定价格合同
- $\lambda \to \infty$（极高公共资金成本）→ 规制者极度关注租金抽取 → 最优规制趋近成本加成合同

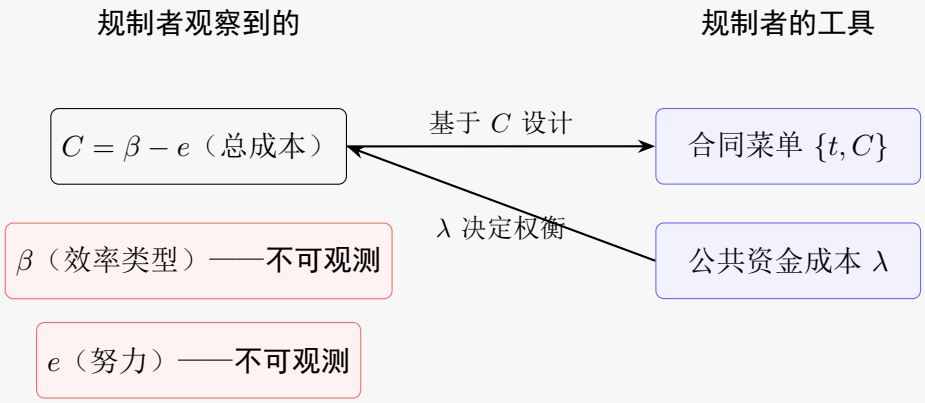
现实中的 λ 估计：发达国家通常 $\lambda \approx 0.3$（Ballard-Shoven-Whalley 1985）；发展中国家 λ 可能 > 1.0（Ahmad-Stern 1991）。

λ 与发展阶段：这是将中国经验纳入理论框架的关键接口。在中国改革开放的不同阶段，λ 的有效值不同：

- 1980 年代（财政包干制 + 大规模基础设施需求）→ λ 非常高 → 倾向成本节约（承包制——本质是高能激励）

- 2000 年代至今（财力充裕 + 关注分配）→ λ 相对降低 → 高能激励仍重要但更关注租金控制和服务质量

图 9.1 规制信息不对称的三维分解



核心困境：规制者只能看到总成本 C——但无法分解为 β（需要区分以控制租金）和 e（需要激励以降低成本）。λ 决定了规制者在激励和租金抽取之间的平衡点。

9.3 Laffont-Tirole (1986) 基准模型

【核心】

本节完整展开 Laffont-Tirole (1986) 的基准模型——全章的核心。这是读者第一次看到完整的机制设计推导，每一步都配以经济学直觉。

9.3.1 机制设计五步框架

在进入具体模型之前，先建立一个与第 ~8~ 章的"Tirole 七步建模法" 并列的机制设计五步框架：

Step 1 界定经济环境：类型空间、行动空间、支付函数、保留效用

Step 2 确定信息结构：谁在什么时候知道什么——可观测 vs. 可验证

Step 3 写出 IC + PC：激励相容约束和参与约束——代理人的行为空间

Step 4 应用显示原理：将注意力限制在直接显示机制上

Step 5 求解最优机制：目标函数最大化，受约束于 IC + PC → 比较静态

与 Ch.8 的" 七步框架" 的对照：步骤 1–2 对应 Ch.8 的" 玩家/策略/时序/信息/支付"；步骤 3–4 是机制设计特有（Ch.8 没有对应项——因为 Ch.8 处理的是博弈均衡，而非规则设计）；步骤 5 对应 Ch.8 的" 均衡概念/比较静态"。

9.3.2　模型设定——两种类型情形

Step 1 —经济环境：

企业生产一项公共事业产品（电力、电信、铁路等）。企业的总成本：

$$C = \beta - e$$

其中：

- $\beta =$ 企业的" 效率类型"——$\beta \in \{\beta_L, \beta_H\}$，$\beta_L < \beta_H$（低 $\beta =$ 高效率）
- 先验概率：$\Pr(\beta = \beta_L) = \nu$，$\Pr(\beta = \beta_H) = 1 - \nu$
- $e =$ 企业管理者降低成本的" 努力"——不可被规制者观测
- $\psi(e) =$ 努力的负效用——$\psi(0) = 0$，$\psi' > 0$，$\psi'' > 0$（递增的边际负效用）

企业获得总支付：t（规制者给企业的净转移支付——包括成本报销和固定费）。企业的效用/租金：

$$U = t - \psi(e)$$

企业的保留效用（参与约束）：$U \geq 0$（企业总可以拒绝合同）。

Step 2 —规制者的目标函数：

规制者最大化社会福利 W：

$$W = S - (1+\lambda)(t+C) + U = S - (1+\lambda)[C + \psi(e)] - \lambda U \tag{9.1}$$

对该目标函数的逐项解释：

- S：消费者从公共服务中获得的毛剩余（本章假设产出固定，焦点在成本效率）
- $(1+\lambda)(t+C)$：消费者通过税收/价格支付给企业的总成本——公共资金的影子成本 λ 放大了每一元支付的社会成本
- $+U$：企业的租金也算入社会福利（企业由公民所有/雇员由公民组成）——但规制者给企业租金的权重为 1，而给消费者的权重为 $(1+\lambda) \rightarrow$ 规制者更偏爱消费者剩余（因为公共资金昂贵）

关键的福利分解：为什么规制者"不喜欢"企业租金

$$W = \underbrace{S}_{\text{固定毛剩余}} - \underbrace{(1+\lambda)[C+\psi(e)]}_{\text{加权的总生产成本}} - \underbrace{\lambda U}_{\text{信息租金的额外社会成本}}$$

给企业 1 元租金 → 社会成本是 λ 元（因为必须通过扭曲性税收筹集 1 元 → 社会成本 $1+\lambda$；企业获得 1 元 → 净社会损失 $=\lambda$）。规制者确实关心控制企业的租金——不是因为它"不喜欢"企业盈利——而是因为向企业转移 *1* 元租金的边际社会成本 >0。

Step 3 —信息结构：

- 规制者观察：总成本 C（事后——合同执行后）
- 规制者不观察：β 和 e 分别
- 规制者知道：函数形式 $C=\beta-e$，努力的负效用 $\psi(e)$，β 的先验分布

Step 4 —规制者的工具：规制者根据观测到的 C 向企业提供一个转移支付方案 $t(C)$。因为 C 是可观测和可验证的，合同可以写成 C 的函数。

Step 5 —时序：

自然选择β → 规制者提供合同菜单$\{t(\hat{\beta}), C(\hat{\beta})\}$ → 企业选择合同（宣布$\hat{\beta}$） → 企业选择e（不可观测） → $C=\beta-e$实现 $t(C)$执行

9.3.3 完全信息基准（First-Best）

在引入信息不对称之前，先求解完全信息下的最优——作为基准。

设定：规制者可以观测 β 和 e（无信息不对称）。对于类型 β 的企业，规制者的规划为：

$$\max_{e,U} \; S-(1+\lambda)[\beta-e+\psi(e)]-\lambda U \quad \text{s.t.} \quad U\geq 0$$

因为 $-\lambda U$ 在目标函数中系数为负，规制者设 $U=0$（企业恰好获得保留效用）。一阶条件：

$$\frac{\partial W}{\partial e} = -(1+\lambda)[-1+\psi'(e)] = 0 \quad \Rightarrow \quad \boxed{\psi'(e^{FB})=1}$$

完全信息下的三个特征

(1) **最优努力：**$\psi'(e^{FB})=1$ ——努力的边际负效用恰好等于边际成本节约（1 元努力降低 1 元成本）

(2) **零租金：**$U^{FB}(\beta)=0$ ——规制者可以精确抽取企业的所有剩余

(3) **类型独立的最优努力：**$e^{FB}(\beta_L)=e^{FB}(\beta_H)=e^{FB}$ ——努力的边际收益（1 元成本节约）与 β 无关

9.3.4　不对称信息下的最优规制

设定：规制者无法观测 β 和 e，只能观测 C。规制者提供一个合同菜单：$\{t_L, C_L\}$（面向类型 β_L 的合同）和 $\{t_H, C_H\}$（面向类型 β_H 的合同）——企业选择其中一个合同。

如果类型 β 的企业选择合同 $\{t, C\}$，它的效用为 $U = t - \psi(\beta - C)$（因为 $C = \beta - e \Rightarrow e = \beta - C$）。

激励相容约束（Incentive Compatibility, IC）：

类型 β_L 的企业必须偏好自己的合同而非 β_H 的合同：

$$IC_L: \quad t_L - \psi(\beta_L - C_L) \geq t_H - \psi(\beta_L - C_H) \tag{9.2}$$

同理：

$$IC_H: \quad t_H - \psi(\beta_H - C_H) \geq t_L - \psi(\beta_H - C_L) \tag{9.3}$$

参与约束（Participation Constraints, PC）：

$$PC_L: \quad U_L = t_L - \psi(\beta_L - C_L) \geq 0 \tag{9.4}$$

$$PC_H: \quad U_H = t_H - \psi(\beta_H - C_H) \geq 0 \tag{9.5}$$

规制者的规划：

$$\max_{\{C_L, U_L, C_H, U_H\}} \nu\big\{S - (1+\lambda)[C_L + \psi(\beta_L - C_L)] - \lambda U_L\big\} + (1-\nu)\big\{S - (1+\lambda)[C_H + \psi(\beta_H - C_H)] - \lambda U_H\big\}$$

受约束于：IC_L, IC_H, PC_L, PC_H。

9.3.5　哪些约束是紧的？——机制设计的标准逻辑

推论 1 PC_H **紧，**PC_L **松。**低效率类型（β_H）虽然想获得高效率类型的合同待遇，但高效率类型的合同要求更低的成本（更难达到），因此它恰好无差异于参与与否——参与约束为紧。高效率类型则严格偏好参与——其参与约束是松的。

推论 2 IC_L **紧，**IC_H **松。**高效率类型（β_L）有激励模仿低效率类型（β_H）：如果类型 β_L 选合同 H，它可以花费努力 $e = \beta_L - C_H < \beta_H - C_H$（比低效率类型需要更少的努力来达到相同的 C_H）。其租金为：

$$U_L(C_H) = t_H - \psi(\beta_L - C_H) > t_H - \psi(\beta_H - C_H) = U_H$$

高效率类型模仿低效率类型的额外租金：

$$\Phi(C_H) \equiv \psi(\beta_H - C_H) - \psi(\beta_L - C_H)$$

这是信息租金的根本来源：**高效率类型的"努力节省"优势**——做同一件事（达到相同的成本 C_H），高效率类型付出的努力负效用更少。

推论 3 在最优规制下：

- PC_H 紧 $\Rightarrow U_H^{SB} = 0$
- IC_L 紧 $\Rightarrow U_L^{SB} = \Phi(C_H^{SB}) = \psi(\beta_H - C_H^{SB}) - \psi(\beta_L - C_H^{SB})$

9.3.6　最优解与核心结果

将紧约束代入目标函数并对 C_L 和 C_H 求一阶条件。设 $e_L = \beta_L - C_L$，$e_H = \beta_H - C_H$：

高效率类型（β_L）：

$$\frac{\partial W}{\partial e_L} = 0 \quad \Rightarrow \quad \boxed{\psi'(e_L^{SB}) = 1}$$

高效率类型的努力 = 完全信息最优水平。**”无扭曲在顶部”（no distortion at the top）**。

低效率类型（β_H）：

$$\boxed{\psi'(e_H^{SB}) = 1 - \frac{\lambda}{1+\lambda} \cdot \frac{\nu}{1-\nu} \cdot \left[\psi'(e_H^{SB}) - \psi'(e_H^{SB} - \Delta\beta)\right] < 1} \tag{9.6}$$

其中 $\Delta\beta = \beta_H - \beta_L > 0$ 是类型的”散度”。因为 $\psi'' > 0$，有 $\psi'(e_H) > \psi'(e_H - \Delta\beta)$，故方括号项为正 $\Rightarrow \psi'(e_H^{SB}) < 1$。低效率类型的努力低于完全信息最优水平。**”扭曲在底部”（distortion at the bottom）**。

最优租金：

$$U_H^{SB} = 0 \qquad \text{（低效率类型——零租金）} \tag{9.7}$$

$$U_L^{SB} = \psi(e_H^{SB}) - \psi(e_H^{SB} - \Delta\beta) \equiv \Phi(e_H^{SB}) > 0 \qquad \text{（高效率类型——正信息租金）} \tag{9.8}$$

核心权衡的精确数学表达

提高低效率类型的努力 e_H 的边际效应：

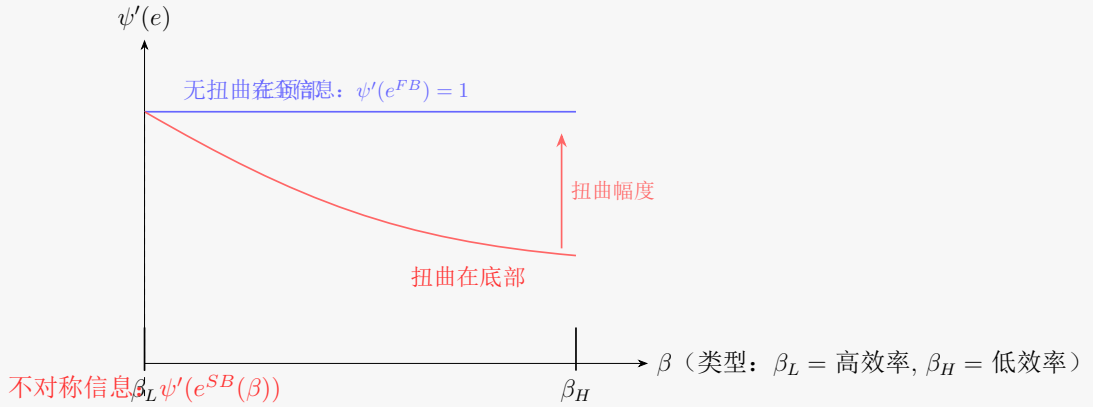
- **收益**：社会生产成本降低——$(1+\lambda) \cdot [1 - \psi'(e_H)]$
- **成本**：增加高效率类型的信息租金——$\lambda \cdot \frac{\nu}{1-\nu} \cdot [\psi'(e_H) - \psi'(e_H - \Delta\beta)]$

在最优解处，边际收益 = 边际成本。

经济学直觉

让低效率类型更加努力（$e_H \uparrow$）$\rightarrow C_H \downarrow \rightarrow$ 低效率类型的合同变得”更苛刻” $\rightarrow$ 高效率类型模仿低效率类型的吸引力下降（因为低效率类型的合同现在要求更低的成本、更多的努力）$\rightarrow$ 高效率类型的信息租金减少。但让低效率类型更努力本身是好事（降低生产成本）。在最优处，这两者的边际效应恰好平衡。

图 9.2　”无扭曲在顶部，扭曲在底部”的图形展示



在顶部（$\beta = \beta_{\min}$），不对称信息下的最优努力等于完全信息水平。随着 β 增加（更低效率），努力被向下扭曲——扭曲幅度随 β 增加而增大。

9.3.7　最优线性成本报销规则

以上推导的是一般最优机制。当规制者受制度约束只能提供线性合同时，考虑线性成本报销规则（linear cost-reimbursement rule）：

$$t(C) = a - b \cdot C$$

其中 $b \in [0, 1]$ 是合同的*激励强度*（incentive power）：

- $b = 1$：固定价格合同（fixed-price contract）——企业承担全部成本节约
- $b = 0$：成本加成合同（cost-plus contract）——规制者报销全部成本
- $b \in (0, 1)$：激励性规制合同——部分成本分担

在连续类型情形下，最优 b 的公式为：

$$b^* = \frac{1}{1 + \lambda \cdot \frac{F(\beta)}{f(\beta)} \cdot \psi''(e(\beta))}$$

直观地：

- $\lambda \to 0 \Rightarrow b^* \to 1$（固定价格——不关心租金）

- $\lambda \to \infty \Rightarrow b^* \to 0$（成本加成——极度关心租金抽取）
- $F(\beta)/f(\beta)$ 越大（" 风险率" 越高）$\Rightarrow b^*$ 越小（对" 尾部" 类型降低激励以节省租金）

9.3.8 比较静态与实证含义

表 9.1: 最优规制与比较静态

参数变化	对 e_H^{SB}（低效率努力）	对 U_L^{SB}（信息租金）	对 b^*（激励强度）
$\lambda\uparrow$（公共资金更昂贵）	$\downarrow$（更多扭曲）	$\downarrow$	$\downarrow$（向成本加成移动）
$\nu\uparrow$（高效率类型更常见）	$\downarrow$（降低租金收益更大）	$\uparrow$（每企业租金可能更高）	$\uparrow$（对常见类型更强激励）
$\Delta\beta\uparrow$（类型散度更大）	$\downarrow$（租金" 成本" 更高）	$\uparrow$（模仿收益更大）	$\downarrow$

训练读者的比较静态直觉：每一个比较静态都应该有经济学解释，而不仅仅是数学。$\lambda\uparrow \to$ " 政府缺钱" $\to$ 不能给企业太多租金 $\to$ 牺牲效率（容忍更高成本）以节省租金 $\to$ 合同更像成本加成。

9.4 菜单规制与显示原理

【核心】

9.4.1 从两种类型到连续类型：直觉过渡

两种类型模型给出了" 无扭曲在顶部，扭曲在底部" 的核心洞见，但现实中 β 通常是连续的。设 $\beta \in [\beta_{\min}, \beta_{\max}]$，分布函数 $F(\beta)$，密度 $f(\beta)$。

通过显示原理将任意机制转化为直接显示机制，规制者的规划变为：

$$\max_{e(\beta),U(\beta)} \int_{\beta_{\min}}^{\beta_{\max}} \left[S - (1+\lambda)(\beta - e(\beta) + \psi(e(\beta))) - \lambda U(\beta)\right] dF(\beta) \tag{9.9}$$

受约束于：

$$U'(\beta) = -\psi'(e(\beta)) \qquad \text{（IC 一阶条件——包络定理）} \tag{9.10}$$

$$e'(\beta) \leq 0 \qquad \text{（IC 二阶条件——单调性）} \tag{9.11}$$

$$U(\beta_{\max}) \geq 0 \qquad \text{（参与约束——最低效率类型）} \tag{9.12}$$

包络定理的推导（直观）：类型 β 的企业如果报告 $\hat{\beta}$，获得的租金为 $U(\hat{\beta},\beta) = t(\hat{\beta}) - \psi(\beta - C(\hat{\beta}))$。激励相容要求真实报告是最优的。包络定理给出：

$$\frac{dU(\beta)}{d\beta} = \left.\frac{\partial U}{\partial \beta}\right|_{\hat{\beta}=\beta} = -\psi'(e(\beta))$$

积分得：

$$U(\beta) = U(\beta_{\max}) + \int_{\beta}^{\beta_{\max}} \psi'(e(\tau))\, d\tau$$

关键洞见：信息租金完全由努力分配 $\{e(\beta)\}$ 决定——一旦规制者选择了努力的路径，租金就由激励相容约束自动决定。规制者不能独立选择" 努力" 和" 租金"——它们被 IC 约束捆绑在一起。

连续类型下的最优条件：

$$\psi'(e(\beta)) = 1 - \frac{\lambda}{1+\lambda} \cdot \frac{F(\beta)}{f(\beta)} \cdot \psi''(e(\beta)) \tag{9.13}$$

与两种类型情形一致的洞见：

- 在顶部（$\beta = \beta_{\min}$, $F(\beta_{\min}) = 0$）：$\psi'(e(\beta_{\min})) = 1$ ——无扭曲在顶部
- 对 $\beta > \beta_{\min}$：$F(\beta)/f(\beta) > 0 \Rightarrow \psi'(e(\beta)) < 1$ ——扭曲随 β 增加

9.4.2　显示原理（Revelation Principle, Myerson 1979）

【核心】

定义 9.2 (显示原理). 对于任何一个机制和该机制的任何 *Bayesian* 均衡，存在一个等价的直接显示机制（*direct revelation mechanism*），其中：

- 代理人同时且诚实地报告自己的类型（$\hat{\beta}_i = \beta_i$）
- 结果与原始均衡相同

含义：规制者在设计最优规制时，可以不失一般性地将注意力限制在直接显示机制上——即要求企业报告其类型 $\hat{\beta}$，并承诺根据报告实施结果 $\{t(\hat{\beta}), C(\hat{\beta})\}$。

在 Laffont-Tirole 框架中的操作：

(1) 规制者设计一个合同菜单 $\{t(\hat{\beta}), C(\hat{\beta})\}$，企业通过选择合同来隐式地报告其类型

(2) 显示原理保证：任何最优规制等价于一个激励相容的直接显示机制

(3) → 规制者的设计问题简化为：在 IC 和 PC 约束下选择 $\{t(\beta), C(\beta)\}$

”菜单”的经济学解释

合同菜单 $=\{$固定价格合同, 激励性合同, 成本加成合同$\}$ 的连续族：

- **高效率企业**（低 β）选择高能激励合同（b 接近 1）——因为它知道可以通过努力大幅降低成本
- **低效率企业**（高 β）选择低能激励合同（b 接近 0）——因为降低成本很难（边际负效用高），选择高能激励导致的租金太少

这是”自我选择”（*self-selection*）的精髓：企业通过选择不同合同的行动，诚实地揭示了规制者无法直接获取的关于其类型的信息。

9.4.3 单交叉条件与分离的可行性

企业的效用：$U=t-\psi(\beta-C)$。选择变量为 $\{C,t\}$：

$$\frac{\partial U}{\partial C}=\psi'(\beta-C),\quad \frac{\partial U}{\partial t}=1$$

边际替代率：$MRS_{C,t}(\beta)=\psi'(\beta-C)$。

$$\frac{\partial MRS}{\partial \beta}=\psi''(\beta-C)>0$$

因为 $\psi''>0$（递增边际负效用），MRS 随 β 递增 $\rightarrow$ **单交叉条件（single-crossing property）成立** $\rightarrow$ 分离是可行的。

经济学直觉：低 β 企业（高效率）和高 β 企业（低效率）在 C 和 t 之间的”边际替代率”不同——高效率企业更愿意接受更低的 C（更多的努力）以换取更高的 t——因为努力对它更”便宜”。注意：在此设定中，C 是企业想要的”坏物品”（higher cost = less effort needed），t 是”好物品”，因此无差异曲线在 (C,t) 空间中向上倾斜——这是单交叉条件几何意义的关键。

9.4.4 与 Ch.8 信号模型的对比

统一洞见：Ch.8 和 Ch.9 分别是信息不对称在企业–企业界面和企业–政府界面上的两种不同制度形式。两者共享相同的数学基础（单交叉条件、包络定理），但制度背景不同。

9.5 机制设计基础：Baron-Myerson 与最优定价

【核心/进阶】

本节将激励规制的逻辑从”成本报销”扩展到”最优产出/定价”——展示相同的机制设计逻辑如何在不同的政策工具上运行。

表 9.2: Ch.8 信号传递 vs. Ch.9 机制甄别

维度	Ch.8 Milgrom-Roberts 限价信号	Ch.9 Laffont-Tirole 菜单规制
互动方向	知情者（在位者）先行动 → 信号传递（signaling）	不知情者（规制者）先行动 → 甄别（screening）
工具变量	价格 P	合同菜单 $\{t, C\}$
私信息内容	在位者的成本类型	企业的效率类型
均衡概念	PBE（完美贝叶斯均衡）	BNE（贝叶斯纳什均衡）——显示原理
关键条件	单交叉条件（低成本在位者承担低价更” 便宜”）	单交叉条件（高效率企业承担高激励更” 便宜”）
效率结果	限价可能低效（低于短期最优价格）	扭曲在底部（低效率类型的努力低于最优）
信息租金	低成本在位者通过” 分离” 获得租金	高效率企业通过 IC 约束获得信息租金

9.5.1 Baron-Myerson (1982) 模型

与 Laffont-Tirole 的关键区别：

- Laffont-Tirole：成本报销问题——焦点在最小化生产成本
- Baron-Myerson：定价问题——焦点在最优产出/价格，给定未知的成本

模型设定：

- 垄断企业生产 q 单位产品，成本 $= \theta \cdot q + K$（θ = 边际成本，为企业的私人信息；K = 固定成本）
- 需求：$P(q)$（向下倾斜）
- 规制者选择价格 P（或等价地，产出 q）和转移支付 t
- 消费者剩余：$S(q) = \int_0^q P(x)\,dx - P(q) \cdot q$

社会福利：

$$W = S(q) - (1+\lambda)(t + \theta q + K) + U$$

其中 $U = t + P(q)q - \theta q - K$ 是企业的利润（租金）。

消去 t，得到：

$$W = S(q) + P(q)q - (1+\lambda)(\theta q + K) - \lambda U$$

核心权衡：

- 对低成本企业（低 θ），最优价格应接近边际成本（$P \approx \theta$）以实现配置效率
- 但低价格意味着低利润 $\rightarrow$ 低成本企业有动机谎报自己是高成本（以获取更高价格）
- $\rightarrow$ 规制者必须在配置效率（价格接近边际成本）和信息租金（低成本企业获得租金以说真话）之间权衡

9.5.2 最优定价公式

在需求独立于成本类型且消费者同质的简化设定下，应用显示原理和机制设计五步法，得到 Baron-Myerson 最优定价公式：

$$\boxed{P(q^*(\theta)) = \theta + \frac{\lambda}{1+\lambda} \cdot \frac{F(\theta)}{f(\theta)}} \tag{9.14}$$

- $F(\theta)/f(\theta)$ 是" 风险率"（hazard rate）——衡量在类型 θ 处存在企业的条件概率密度
- 对于最高效类型（$\theta = \theta_{\min}$, $F(\theta_{\min}) = 0$）：$P = \theta$ ——边际成本定价——无扭曲在顶部
- 对于 $\theta > \theta_{\min}$：$P > \theta$ ——价格高于边际成本——产出向下扭曲

与 Laffont-Tirole 的平行结构：

表 9.3: Laffont-Tirole vs. Baron-Myerson

	Laffont-Tirole (1986)	Baron-Myerson (1982)
决策变量	e（努力——降低成本）	q 或 P（产出/定价）
私信息	β（效率类型）	θ（边际成本类型）
无扭曲在顶部	$\psi'(e_L^*) = 1$	$P(\theta_{\min}) = \theta_{\min}$
扭曲在底部	$\psi'(e_H^{SB}) < 1$	$P(\theta) > \theta$ for $\theta > \theta_{\min}$
信息租金	$U_L = \Phi(e_H)$	$U(\theta) = \int_\theta^{\theta_{\max}} q(\tau)\, d\tau$

机制设计作为”统一方法论”

Laffont-Tirole（成本报销）和 Baron-Myerson（最优定价）是同一机制设计逻辑在两个不同政策工具上的应用。两者的共同结构：

(1) 代理人拥有关于自己类型的私人信息（β 或 θ）

(2) 委托人设计一个合同菜单，让代理人通过自我选择来诚实地揭示类型

(3) 最优机制在”顶部”不产生扭曲（最高效类型获得最优待遇）

(4) 对”底部”类型引入扭曲以减少高效类型的租金

这是一个通用模板——可以应用于碳减排、政府采购、医疗支付等任何存在不对称信息的规制场景。

9.5.3　单交叉条件与单调性

【进阶】

激励相容的二阶条件：$e'(\beta) \leq 0$——即努力必须随 β（低效率）非增。等价地：更高 β（更低效率）的企业被分配更低的成本降低目标。

如果单调性违反 $\rightarrow$ **混同**（bunching）——多个类型被分配相同的合同——激励相容的二阶条件在该处为紧。

在 Laffont-Tirole 中：

$$\frac{\partial}{\partial \beta}\left[\frac{\partial U/\partial C}{\partial U/\partial t}\right] = \frac{\partial}{\partial \beta}[\psi'(\beta - C)] = \psi''(\beta - C) > 0$$

单交叉条件成立 $\rightarrow$ 分离是可行的 $\rightarrow$ 不需要混同。

9.6　防合谋原理与规制俘获

【进阶】

9.6.1　从两层到三层：谁来监督监督者？

标准的两层框架：Principal（规制者）$\rightarrow$ Agent（被规制企业）

现实中的三层结构：Principal（国会/政府）$\rightarrow$ Supervisor（规制机构——如证监会、电监会）$\rightarrow$ Agent（被规制企业）

核心问题：如果监督者（规制机构）可以被企业”收买”（bribe/capture），两层框架的最优机制是否仍然成立？

Tirole (1986, 1992) 的三层模型：

- 监督者观测到关于企业类型的含噪信号 σ（可能等于也可能不等于真实的 β）
- 监督者可以：(a) 诚实报告 σ 给委托人 $\rightarrow$ 获得信息租金 $R(\sigma)$；(b) 与企业合谋——隐匿信息或报告伪造的 σ ——以换取企业的" 贿赂"（side transfer）s
- 企业和监督者之间存在合谋合同——不可被委托人直接观测

9.6.2 防合谋原理（Collusion-Proofness Principle）

定义 9.3 (防合谋原理). 对于任何一个可能导致合谋的" 大机制"（*grand mechanism*），存在一个等价的防合谋机制（*collusion-proof mechanism*），该机制在事前给定时，企业–监督者联盟没有激励偏离真实报告——且该机制产生的配置结果等价于原始机制。

含义：

(1) 规制者在设计规制时可以不失一般性地将注意力限制在防合谋机制上

(2) 防合谋的" 成本" 是必须支付给监督者一个额外的激励支付（等于他从合谋中能获得的最大化收益）

(3) 这增加了规制成本 $\rightarrow$ 最优产出进一步向下扭曲

(4) 但结构不变——防合谋机制的最优形式仍然是" 让低效率类型承担更多扭曲"

9.6.3 规制俘获：从 Stigler 描述到 Laffont-Tirole 微观基础

Stigler (1971) 的俘获理论（已在第 ~5~ 章中介绍）：规制者被规制产业" 俘获"——规制服务于产业而非公共利益。机制：政治献金、旋转门、信息优势。这是实证观察——缺乏微观基础。

Laffont-Tirole 的贡献：为俘获提供微观基础：

- 俘获不是" 规制者被产业说服"——而是在给定信息结构下，企业和监督者之间存在可执行的双边合同（side contract）空间
- 当监督者的信号 σ 是硬信息（hard information——可验证）且关于企业类型的信息——企业愿意为监督者隐匿信息而支付最多 $U_L - U_H$（高效率企业的信息租金）
- 当监督者的信号是软信息（soft information——不可验证）——合谋需要" 贿赂"（monetary transfer）或" 未来利益的承诺"（旋转门）

从描述到分析：俘获的微观基础

芝加哥学派（Stigler, Peltzman, Becker）告诉我们俘获发生了。Laffont-Tirole 告诉我们俘获在什么信息结构下是可行的、俘获的成本是多少、以及如何设计制度来最小化俘获的损害。这是" 描述性理论" 和" 分析性理论" 之间的根本差异——正如 Ch.7 展示了博弈论如何将 IO 从" 描述和分类" 转变为" 模型化并求解"。

政策含义：

- 降低俘获风险 → 减少监督者与企业之间的信息不对称——通过公开透明的信息披露制度
- 增加合谋难度 → 降低监督者的" 可收买性"——竞标流程、审计、举报人保护
- 减少俘获的潜在收益 → 宽大计划（leniency programs）——使企业与监督者之间的" 合谋博弈" 变成" 囚徒困境"

9.7　不完备合同与 Maskin-Tirole 定理

【进阶】

9.7.1　不完备合同问题

什么是合同的" 不完备性"？一个合同是完备的（complete）如果它对所有可能的未来状态都规定了各方的行动和支付。现实中，合同不完备的三个来源：

(1) **不可预见的偶然性**（unforeseen contingencies）：合同签订时无法预见所有未来可能的事件

(2) **描述成本**（cost of description）：即使可以预见，描述所有可能状态的成本非常高

(3) **验证成本**（cost of verification）：即使可以描述，向第三方（如法院）证明某个状态已发生的成本非常高

在规制中的体现：一个 25 年的电力特许经营合同——如何预见 25 年间的技术变化（光伏、储能、智能电网）、需求变化（电动汽车）、环境规制变化（碳税）？规制合同必然是高度不完备的——需要" 重新谈判"（renegotiation）机制。

产权理论的论证（Grossman-Hart 1986, Hart-Moore 1990）：当合同不完备时，剩余控制权（residual control rights）的分配——即谁有权决定合同未规定事项——是关键的。这为在私有化和公有化之间的选择提供了理论基础——不同的所有权结构分配了剩余控制权给不同的主体。

9.7.2 Maskin-Tirole (1999) 定理

核心问题：不完备合同是否必然导致 Pareto 效率损失？

Hart-Moore 的观点：是的——当存在不可预见的偶然性时，合同必然是不完备的，这必然导致事后低效。

Maskin-Tirole 的反论证：

命题 9.1 (Maskin-Tirole 定理，简化版). 如果代理人可以做出描述性决策（*describe their decisions*）——即代理人不仅仅选择"接受/拒绝"合同（二元选择），而是可以选择任意的行动并描述该行动——那么即使合同是不完备的（不能预见所有状态），*Pareto* 效率仍然可以在均衡中达到。

直觉：

- 不完备合同的"问题"在于双方不能事先将支付与所有状态挂钩
- 但如果代理人可以事后（在不确定性解决后）"描述"自己所处的状态——并承担特定的行动——那么代理人可以通过"选择性执行"来复制完备合同的结果
- 关键是：代理人不仅在"接受"和"拒绝"之间选择——代理人可以行动且这个行动是可描述的

定理前提：

(1) 双方可以事后描述已经发生的事件（ex post describability）

(2) 双方可以事后承诺（可以签订事后合同）

(3) 效用可以在双方之间转移（转移支付是可行的）

与 Hart-Moore 的对比：

- Hart-Moore：假设不可预见 $\rightarrow$ 必然不完备 $\rightarrow$ 必然低效
- Maskin-Tirole：不可预见 $\neq$ 不可描述 $\rightarrow$ 事后可描述性打破了"不完备 = 低效"的必然联系
- **产权的重要性取决于"事后可描述性"的假设是否成立**——如果事后描述是容易的，产权不那么重要；如果事后描述是困难的（如创新活动），产权至关重要

对规制实践的含义

- **短期合同 vs. 长期合同**：如果事后可描述性成立 → 短期合同 + 事后重新谈判 ≈ 长期完备合同的结果 → 不需要锁定 25 年合同
- **重新谈判的设计**：如果重新谈判被预期到 → 初始合同应故意留空某些条款 → 为事后信息揭示和适应性调整留出空间
- **中国 SOE 的" 双轨制" 合同**：国企经营责任合同（承包制）→ 3–5 年期限 → 到期前重新谈判——这实际上是" 不完备合同 + 定期重新谈判" 的制度化——但棘轮效应（ratchet effect）使重新谈判变成了" 鞭打快牛"

9.8　应用与实践：全球激励规制改革

【核心】

9.8.1　价格上限规制（RPI-X）vs. 回报率规制

起源：Littlechild (1983) 为英国电信私有化提出的规制方案。

机制：

$$P_t = P_{t-1} \cdot (1 + \text{RPI} - X)$$

其中：RPI = 零售价格指数（总体通胀率），X = 预期的" 生产力增长"（expected productivity growth）。

经济学直觉：

- 企业可以保留 RPI-X 上限以下的任何成本节约（高能激励，$b = 1$——在企业不" 触碰" 价格上限的范围内）
- 价格上限在两次审查（review）之间固定（通常 4–5 年）→ 企业在此期间是" 剩余索取者"（residual claimant）
- 审查时，规制者根据企业的实际成本和生产力表现重新设定 X 和初始价格水平

RPI-X 的悖论——棘轮效应：

- 价格上限的" 高能激励" 只在审查间隔期内有效
- 如果规制者在审查时会收回企业过去的超额利润（通过设定更激进的 X）→ 价格上限退化回回报率规制
- **价格上限的信誉（规制者能否承诺不" 征用" 过去的高利润）是决定价格上限是否真正有效的关键条件**

表 9.4: 价格上限 vs. 回报率规制

	价格上限（RPI-X）	回报率规制（Rate-of-Return）
激励强度	高能激励（在审查间隔期内，$b \approx 1$）	低能激励（$b \approx 0$——成本全部报销）
成本降低激励	强——企业保留全部节省	弱——节省立即传递给消费者
信息租金	高——高效率企业获得大量租金	低——成本报销消除了租金
信息需求	低（仅需 RPI 和 X）——但审查时需要大量信息	高（需要详细的成本核算）
投资激励	强——降低成本创新反映在利润中	弱——Averch-Johnson (1962)" 金边" 效应：过度资本化

9.8.2 标尺竞争（Yardstick Competition, Shleifer 1985）

核心问题：规制者不知道企业的真实成本 β——但如果存在多家处于类似环境的企业呢？规制者可以通过比较它们之间的表现来推断成本。

Shleifer (1985) 的标尺竞争方案：

- 假设有 n 家被规制企业在独立但同质的环境中运营
- 规制者向企业 i 提供的转移支付：

$$t_i = a_i - b \cdot (C_i - \bar{C}_{-i})$$

其中 $\bar{C}_{-i} = \frac{1}{n-1} \sum_{j \neq i} C_j$（其他企业的平均成本）

标尺竞争的核心洞见

企业 i 的转移支付取决于其他企业的成本而非自身的成本 $\rightarrow$ 企业 i 减少成本 1 元 $\rightarrow$ t_i 增加 b 元（如果 $b = 1$，企业获得全额成本节省）$\rightarrow$ **高能激励**。但信息租金被消除！因为 $\bar{C}_{-i}$ 是外生于企业 i 的行动的（无法被企业 i 操控）。

标尺竞争实现了" 高能激励 + 零信息租金"——这是独立合同无法实现的。

适用条件：

(1) 被比较的企业必须在类似环境下运营（否则" 标尺" 是有偏的）

(2) 企业之间不能合谋（否则它们可以协调高成本）

(3) 企业数量 n 足够大（$n = 2$ 时，每家企业都有 50% 的权重在" 标尺" 中）

中国应用：省级电网公司的对标管理——比较不同省份的单位供电成本；城市供水/燃气公司的区域比较；移动通信运营商的网络质量对标。

9.8.3 接入定价（Access Pricing）

核心问题：在自然垄断行业，上游基础设施是垄断的（拥有" 必要设施"），但下游服务可以竞争。如何为下游竞争者接入上游设施定价？

三种方法：

(1) **有效成分定价规则（ECPR, Baumol-Willig 1983）**：

接入价格 = 提供接入的增量成本+在位者因接入而损失的下游利润（机会成本）

注：ECPR 的" 增量成本" 是指为接入者提供服务的直接增量成本，不等同于在位者自身下游服务的边际成本。性质：保证生产效率——只有在进入者比在位者更高效时进入才会发生。

(2) **Ramsey 接入定价**：

$$\text{接入价格} = \text{上游边际成本} + \text{Ramsey 加成}$$

加成与下游服务的需求弹性成反比。性质：最大化社会福利受制于企业收支平衡。

(3) **长期增量成本（LRIC/TELRIC）**：

$$\text{接入价格} = \text{提供接入服务的长期增量成本}$$

性质：最接近" 竞争市场" 价格——但可能无法补偿在位者的固定和沉没成本。

9.8.4 全球规制改革浪潮

关键教训：

- **制度禀赋**（institutional endowment）至关重要——同样的 RPI-X 在英国成功，在缺乏制度能力的国家可能失败
- **承诺问题**——如果被规制企业是固有的，政府更难以可置信地承诺不征用企业的事后利润
- **改革的顺序**——先建立独立规制机构 + 信息基础设施 → 再引入高能激励合同

表 9.5: 全球关键激励规制改革事件

国家/地区	行业	改革类型	时间	关键特征
英国	电信	私有化 + RPI-X	1984	第一个大规模 RPI-X 实验——Oftel 作为独立规制机构
美国	电信	接入定价 + 分拆	1984, 1996	AT&T 分拆 → 1996 电信法引入竞争
欧盟	多个网络行业	自由化 + 独立规制	1990s–2000s	每个成员国建立独立规制机构 + 接入指令
智利	电力/电信	私有化 + 激励规制	1982	最早的发展中国家实验
中国	电信/电力/铁路	政企分开 + 竞争引入	1990s–至今	见 §~9.9

9.9 中国语境：从国企承包制到现代规制体系

【核心】

本节将 Ch.9 的理论框架系统地映射到中国从计划经济到” 社会主义市场经济” 的规制演进。

9.9.1 SOE 改革的三个阶段：一个激励规制的视角

第一阶段：放权让利（1978–1986）

- 核心逻辑：将经营自主权从政府下放给企业 → 扩大企业的决策空间
- 利润留成制度（profit retention）：企业可以保留一部分利润 → 增加了 e 的边际收益
- **映射到 Laffont-Tirole**：从 $b = 0$（全额上缴利润）向 $b > 0$（企业留存部分利润）移动 → 增加了激励强度

第二阶段：承包经营责任制（1987–1993）

- 核心逻辑：企业与政府签订 3–5 年的” 承包合同”——约定利润上缴基数和递增比例
- ” 包死基数、确保上缴、超收多留、欠收自补”

- **映射到 Laffont-Tirole**：承包制 = 部分固定价格合同（$b > 0$ 但 $b < 1$）
- **承包制的机制设计问题**：

  (1) **基数谈判**——政府不知道企业的真实 $\beta \to$ 承包基数通过" 一对一谈判" 确定 $\to$ 高效率企业获得大量信息租金

  (2) **棘轮效应**——政府观察企业表现后提高下一期承包基数 $\to$ 企业预期到这一点 $\to$ 隐藏成本降低潜力

  (3) **软预算约束**（Kornai 1980）——企业亏损时政府变更原合同提供补贴 $\to b$ 被事后" 重新谈判" 降低

第三阶段：现代企业制度与国资委（1993–至今）

- 1993 年公司法 $\to$ 国有企业公司化改造
- 2003 年国务院国资委（SASAC）成立——将" 出资人" 职能从多个部委集中到一个机构
- SASAC 的**绩效合同**——与央企负责人签订年度和任期经营业绩考核责任书
- **映射到 Laffont-Tirole**：
  - SASAC 绩效合同 = 显性的激励合同——基本指标（利润总额、净资产收益率）+ 分类指标
  - 但 SASAC 同时是出资人（股东）和事实上的规制者 $\to$ 角色的天然冲突
  - 绩效合同的" 基数" 仍通过谈判而非菜单自选择 $\to$ 信息租金问题仍未解决

9.9.2　自然垄断行业的规制改革

电信改革（1994–2008）：

- 1994：中国联通成立（打破邮电部垄断）
- 1999：原中国电信分拆为四——中国电信（固网）、中国移动（移动）、中国卫通、国信
- 2008：三足鼎立重组——中国电信、中国移动、中国联通
- **规制维度**：接入定价（网间结算）、号码携带、铁塔共享（2014 中国铁塔成立）
- **理论映射**：频率牌照拍卖 = 机制设计；网间结算 = 接入定价的 ECPR/LRIC 问题

电力改革（2002–至今）：

- 2002：”厂网分开”——发电（五大发电集团）与电网（国家电网、南方电网）分离
- 2015：”管住中间、放开两头”——电网从”购售电”转变为”输配电”
- **规制维度：**输配电价核定 = 回报率规制（”准许成本 + 合理收益”）↔ 与 RPI-X 的张力
- **理论映射：**电网自然垄断的”必要设施”接入定价 → LRIC vs. 实际成本

铁路改革（2013–至今）：

- 2013：铁道部撤销 → 国家铁路局（规制）+ 中国铁路总公司（运营）
- **规制维度：**路网接入定价、客运票价形成机制
- **理论映射：**不完备合同——铁路特许经营合同的期限、审查和重新谈判

中国自然垄断规制的三个独特问题

(1) **规制者-出资人合一：**SASAC 既是规制者又是股东 → 三层委托代理模型中”委托人”和”监督者”角色重叠 → 俘获不是”外部收买”而是”内在利益一致”

(2) **国企的”双重目标”：**中国 SOE 同时追求经济利润和公共政策目标（就业、战略安全、产业政策）→ 目标函数不仅包括 S 和 λ，还包括”国家战略价值”

(3) **软预算约束的持续性：**尽管 1994 年银行商业化改革、2015 年”僵尸企业”清理——但地方政府和银行对国企的隐性担保仍在 → λ 的有效值在中央和地方政府层面不同

9.9.3 张维迎-林毅夫辩论：一个激励规制理论的解读

辩论背景（1990s 至今）：

- **张维迎**（牛津-trained，Laffont-Tirole 框架的早期中国引入者）：主张国企的根本问题是*产权不清晰* → 需要私有化（或至少混合所有制）
- **林毅夫**（芝加哥-trained，比较优势框架）：主张国企的根本问题是*政策性负担*（policy burden）→ 剥离负担后，在竞争性市场中，所有制不是首要问题

用 Laffont-Tirole 框架解读：

- **张维迎的论证（产权视角）**：剩余控制权的分配决定了事后重新谈判中的讨价还价权 → 国有企业中，政府拥有剩余控制权 → 经理没有高能激励 → 产权改革是" 基础制度"

- **林毅夫的论证（政策性负担视角）**：即使有良好的产权结构，如果企业被迫承担非商业目标，这等价于在目标函数中增加了一个外生项 → ψ 和 C 不再是企业管理者可以完全控制的 → 剥离负担是" 前提条件"

- **统一框架**：两种论证在同一框架中分析——(a) 产权改革改变了 λ 的有效值（剩余控制权影响重新谈判预期），(b) 政策性负担是额外的参与约束（企业必须获得足够多的转移支付以覆盖负担成本）→ 两者都影响最优合同的激励强度——但作用于不同的" 参数" 上

9.10　现代前沿：数字平台、数据与气候规制

【前沿】

9.10.1　数字平台的规制

新型信息不对称：

- 平台拥有数据的数量和质量远超规制者——规制者不知道平台的真实成本和用户价值

- 平台的算法是" 黑箱"——规制者无法观测算法的运作（类似于规制者无法观测 e）

- **数据作为新的"β"**：平台的数据资产决定了其效率——但数据是不可观测和不可验证的

菜单规制的数字版本：向平台提供" 数据访问菜单"——平台可以选择不同级别的数据开放义务，换取不同的规制待遇。

DMA（欧盟数字市场法, 2022）的激励规制逻辑：

- 互操作性（interoperability）义务 = 接入定价问题——主导平台必须以 FRAND 条件允许第三方接入其生态系统的核心功能

- 自我优待（self-preferencing）禁止 = 防止" 规制俘获" 的算法版——平台不能操纵排名和展示以偏袒自己的服务

9.10.2 AI 与算法规制

- 算法的" 可解释性"（explainability）= 信息不对称的降低——规制者能" 看懂" 算法在做什么
- 算法审计（algorithmic auditing）= 类似于标尺竞争——用" 其他算法" 作为标尺来评估目标算法
- AI 定价的合谋风险——多智能体强化学习可能演化出默契合谋策略 → 超出传统合谋检测的范畴

9.10.3 气候规制中的激励设计

- 碳交易（ETS）= 类似于" 标尺竞争"——通过市场化交易发现碳的" 价格"
- 碳边境调节机制（CBAM）= 跨国规制中的" 参与约束"——不参与国际碳定价的国家面临" 边境税"
- 绿色补贴的机制设计——政府推行碳减排，但不知道企业的真实减排成本 → 这是 Laffont-Tirole 模型的直接应用——β = 减排成本类型，e = 减排努力

9.11 章末总结与通往 NEIO 的桥梁

9.11.1 核心概念回顾

(1) **逆向选择 + 道德风险**：规制中的双重信息不对称——同一个可观测变量 C 同时受隐藏类型 β 和隐藏行动 e 影响

(2) **公共资金的影子成本 λ**：连接规制与宏观公共财政——每 1 元转移支付的社会成本为 $1+\lambda$

(3) **显示原理（Myerson 1979）**：" 只需考虑直接显示机制"——将注意力限制在激励相容的合同菜单上

(4) $C = \beta - e$：Laffont-Tirole 成本结构——用同一个公式捕捉隐藏类型和隐藏行动

(5) **核心权衡：激励 vs. 抽取租金**：高能激励（b 接近 1）→ 效率 ↑ + 信息租金 ↑；低能激励（b 接近 0）→ 租金 ↓ + 效率 ↓

(6) **菜单规制与自我选择**：企业通过选择合同来隐式地揭示自己的类型——合同菜单替代了信息获取

(7) **" 无扭曲在顶部，扭曲在底部"**：$\psi'(e_L^*) = 1$，$\psi'(e_H^{SB}) < 1$——最优规制的核心结构特征

(8) **防合谋原理**：三层委托代理的统一框架——规制者可防患合谋而不失一般性

(9) **Maskin-Tirole 不完备合同定理**：不可预见 $\neq$ 不可描述 $\rightarrow$ 效率可能在事后可描述性下达到

(10) **标尺竞争（Shleifer 1985）**：用相对绩效替代绝对成本信息——实现了" 高能激励 + 零信息租金"

9.11.2　核心公式

(1) $C = \beta - e$ ——成本结构（连接隐藏类型和隐藏行动）

(2) $W = S - (1+\lambda)[C + \psi(e)] - \lambda U$ ——规制者的目标函数

(3) $\psi'(e_L^*) = 1,\ \psi'(e_H^{SB}) < 1$ ——最优努力（无扭曲在顶部，扭曲在底部）

(4) $U_L^{SB} = \Phi(e_H^{SB}) = \psi(e_H^{SB}) - \psi(e_H^{SB} - \Delta\beta)$ ——高效率类型的信息租金

(5) $P(q^*(\theta)) = \theta + \frac{\lambda}{1+\lambda} \cdot \frac{F(\theta)}{f(\theta)}$ ——Baron-Myerson 最优定价

9.11.3　方法论总结：机制设计五步法 vs. Tirole 七步建模法

表 9.6: 两种方法论框架的对比

	机制设计五步法（Ch.9）	Tirole 七步建模法（Ch.8）
适用场景	不知情者设计规则	给定规则下求解均衡
焦点	**最优制度设计**	**均衡行为预测**
步骤 1	界定经济环境	玩家
步骤 2	确定信息结构	策略
步骤 3	写出 IC + PC	时序
步骤 4	应用显示原理	信息
步骤 5	求解最优机制	支付 + 均衡概念 + 比较静态

本章的方法论课程：博弈论（第 ~7~ 章）处理的是" 给定规则下的均衡行为"；机制设计（本章）处理的是" 最优规则的设计本身"。这是两个互补的视角——分别对应" 实证 IO"（给定市场结构，预测企业行为）和" 规制 IO"（给定信息约束，设计最优规制）。

9.11.4 通往 Ch.10 的桥梁

第 ~7~ 章给了我们博弈论语言。第 ~8~ 章用这种语言讲述了企业之间的策略行为故事。本章展示了当信息不对称发生在企业和政府之间时，如何用机制设计——而非博弈论均衡——来设计最优规制。这三章共同构成了模块 4 的" 理论 IO" 三部曲。

但现在有一个问题：**这些理论模型的参数如何从数据中估计？**在本章中，我们假设规制者知道 β 的分布 $F(\beta)$——但在现实中，$F(\beta)$ 本身需要从数据中估计。Laffont-Tirole 框架的最优合同依赖于 $\psi(e)$、λ、$F(\beta)$ 等参数——这些都是未知的，需要在实证中被赋予数值。

这正是新实证产业组织（NEIO）要解决的问题。第 ~10~ 章–第 ~12~ 章将展示如何将博弈论模型和机制设计模型" 带到数据中"——用结构估计方法来识别和估计 IO 模型的核心参数。

模块 4 三部曲完成：从博弈论到机制设计

Ch.7 —博弈论重塑 IO：提供了分析策略互动的统一语言（Nash → SPE → BNE → PBE → 重复博弈）。

Ch.8 —策略行为的微观基础：企业之间的信息不对称 → 进入遏制、掠夺、合谋——信号传递（signaling）逻辑。

Ch.9 —不对称信息与规制：企业与政府之间的信息不对称 → 最优规制设计——机制甄别（screening）逻辑。

三个章节共享一个深层结构：**信息不对称是理解现代 IO 的核心**——无论它发生在企业之间还是企业与政府之间。博弈论提供了分析企业间信息不对称的工具（预测均衡行为）；机制设计提供了分析企业-政府间信息不对称的工具（设计最优制度）。

NEIO（第 ~10~ 章–第 ~12~ 章）将这些理论模型带到数据中——用结构估计方法来检验、量化和应用这些模型于真实世界的市场和规制问题。

习题

基础题

Ex 9.1 完全信息基准的求解。考虑 Laffont-Tirole 基准模型（§~9.3.3）的完全信息设定：$C = \beta - e$，$\psi(e) = e^2/2$，规制者目标函数 $W = S - (1+\lambda)(C + \psi(e)) - \lambda U$，企业保留效用为 0。

(a) 求解完全信息下的最优努力 e^{FB} 和最优租金 U^{FB}。

(b) 当 $\lambda = 0.3$ 时，e^{FB} =?（数值答案）

(c) 当 $\lambda = 0.3$ 时，完全信息下规制者的最大社会福利 W^{FB} =?（用 S 和 β 表示）

Ex 9.2 两种类型模型的具体求解。考虑 §~9.3.4的两种类型模型。设 $\beta_L = 2$, $\beta_H = 5$, $\Delta\beta = 3$。$\psi(e) = e^2/2$。$\nu = \Pr(\beta = \beta_L) = 0.4$。$\lambda = 0.3$。

(a) 验证在完全信息下，两种类型的最优努力均为 $e^* = 1$。

(b) 写出 IC 和 PC 约束。

(c) 假设只有 PC_H 和 IC_L 是紧的。求解不对称信息下的最优努力 e_L^{SB} 和 e_H^{SB}。

(d) 高效率类型的信息租金 U_L^{SB} =?（数值答案应包括推导过程）

(e) 比较 (c) 与 (a)：低效率类型的努力被向下扭曲了多少？为什么？

Ex 9.3 线性合同的比较静态。考虑线性成本报销规则 $t(C) = a - b \cdot C$（§~9.3.7）。设 $\lambda = 0.3$, $\nu = 0.5$, $\Delta\beta = 2$, $\psi(e) = e^2/2$。

(a) 推导最优激励强度 b^* 的表达式（在连续类型下）。

(b) 如果 λ 从 0.3 增加到 1.0，b^* 如何变化？提供经济学直觉。

(c) 如果 $\Delta\beta$ 从 2 增加到 4，b^* 如何变化？提供经济学直觉。

Ex 9.4 显示原理的应用。规制者面对一家企业，企业的边际成本 θ 可能是 1 或 3，各以概率 0.5。需求为 $P(q) = 10 - q$。规制者想最大化期望社会福利：消费者剩余 + 企业利润 − 转移支付成本 $(1 + \lambda)t$，$\lambda = 0.3$。

(a) 写出规制者的直接显示机制设计问题——包括目标函数、IC 和 PC。

(b) 运用显示原理，求解最优定价 $P^*(\theta)$ 和最优转移支付 $t^*(\theta)$。

(c) 验证在最优机制中，低边际成本类型（$\theta = 1$）的信息租金 > 0。

进阶题

Ex 9.5 标尺竞争与独立合同。两家被规制企业在独立市场中运营。各自的成本：$C_i = \beta_i - e_i$，其中 $\beta_i \in \{2, 5\}$，独立同分布，各概率 0.5。$\psi(e) = e^2/2$。$\lambda = 0.3$。

(a)（独立合同）求解独立合同下的最优规制——$e_L^{SB}, e_H^{SB}, U_L^{SB}$ 和期望社会福利。

(b)（标尺竞争）规制者使用标尺竞争方案：$t_i = a - (C_i - C_j)$。求解均衡努力和期望社会福利。与 (a) 比较。

(c) 现在假设企业可以合谋——它们可以协调 $e_i = e_j$。标尺竞争方案下，合谋的可持续性条件是什么？

Ex 9.6 防合谋原理的三层模型。考虑三层结构：委托人 → 监督者 → 企业。监督者观测到关于企业类型的完美信号 $\sigma = \beta$（概率 q）或无信号（概率 $1-q$）。企业和监督者之间可以进行转移支付 s。监督者的保留效用为 0。

(a) 写出企业–监督者联盟的 IC 和 PC。

(b) 应用防合谋原理，推导最优防合谋规制机制。与两层模型比较——防合谋的额外社会成本是多少？

(c) 在什么条件下，完全禁止" 旋转门" 是防合谋的最优策略？

Ex 9.7 Maskin-Tirole 不完备合同的应用。规制者和企业签订一个两期合同。第 1 期技术类型 $\theta_1 = 5$ 已知。第 2 期的技术类型 θ_2 在合同签订时不可预见——但将在第 2 期之前实现（$\theta_2 \in \{3, 7\}$，各概率 0.5）。企业成本 $= \theta_t - e_t$。$\psi(e) = e^2/2$。双方在每期可以重新谈判合同。$\lambda = 0.3$。

(a)（不完备合同）假设第 1 期签订的合同不能将支付与 θ_2 挂钩。事后双方可重新谈判第 2 期合同。求解第 2 期最优努力 $e_2^*(\theta_2)$。福利是否达到 Pareto 最优？

(b)（与产权理论比较）现在假设企业拥有第 2 期决策的" 剩余控制权"——企业在重新谈判中可以提出 take-it-or-leave-it 合同。这对第 2 期的最优努力和福利有什么影响？

(c) Maskin-Tirole 定理在此是否成立？为什么（或为什么不）？

应用/实证题

Ex 9.8 中国电信规制改革的分析。中国电信行业在 1994 年引入竞争（联通），1999 年第一次分拆，2008 年重组为三家全业务运营商。

(a) 1994 年之前，中国电信（原邮电部电信总局）是唯一的电信运营商。用 Laffont-Tirole 框架分析：当规制者同时也是运营者时，" 委托人" 的目标函数是什么？与 §~9.3.3中的社会福利函数有何不同？

(b) 1999 年分拆后，规制者（信息产业部）和运营者（中国电信、中国移动等）分离。用防合谋原理（§~9.6.2）分析：事前和事后的" 规制俘获" 风险有何不同？

(c) 讨论：中国为什么不采用" 纯"RPI-X 价格上限来规制电信价格？

Ex 9.9 全球激励规制改革的数据分析。收集英国、美国、法国、德国、巴西、印度、中国在 1990–2020 年激励规制采用的数据。

(a) 创建事件时间线——每个国家在电信、电力、铁路三个行业首次采用激励规制的年份。

(b) 基于 λ 的文献估计，讨论：λ 较高的国家（发展中国家）是否倾向于更早或更晚采用高能激励规制？与 Laffont-Tirole 模型的预测一致吗？

(c) 讨论：中国的激励规制采用时间线是否与中国的财政改革阶段（1994 年分税制改革）一致？

Ex 9.10 中国 SOE 绩效合同的数据模拟。SASAC 的《中央企业负责人经营业绩考核办法》规定了年度和任期考核指标。

(a) 假设 SASAC 知道 $F(\beta)$ 但不知道每家企业的 β_i。根据连续类型模型（§~9.4.1），计算最优成本目标 $C^*(\beta)$ 和最优转移支付 $t^*(\beta)$。（参数化：$\psi(e) = e^2/2$，$\lambda = 0.5$ 作为基准。）

(b) SASAC 的实际考核制度使用”行业对标”——每家企业与同行业的历史最好水平和行业平均水平比较。这种制度在什么意义上近似于标尺竞争（§~9.8.2）？在什么意义上偏离了标尺竞争？

(c) 模拟当 SASAC 不能承诺长期合同时（棘轮效应），最优合同的激励强度如何随时间衰减。模拟 $n = 3$ 期的动态。

推荐阅读

[1] **Laffont, J.-J. & Tirole, J. (1993).** *A Theory of Incentives in Procurement and Regulation.* MIT Press. —Ch.1–3。这是本章的基础文本。

[2] **Laffont, J.-J. & Tirole, J. (1986).** "Using Cost Observation to Regulate Firms." *Journal of Political Economy*, 94(3), 614–641.

[3] **Baron, D. P. & Myerson, R. B. (1982).** "Regulating a Monopolist with Unknown Costs." *Econometrica*, 50(4), 911–930.

[4] **Myerson, R. B. (1979).** "Incentive Compatibility and the Bargaining Problem." *Econometrica*, 47(1), 61–73.

[5] **Shleifer, A. (1985).** "A Theory of Yardstick Competition." *RAND Journal of Economics*, 16(3), 319–327.

[6] **Tirole, J. (2014).** "Market Power and Regulation." Nobel Prize Lecture.

[7] **Tirole, J. (1992).** "Collusion and the Theory of Organizations." In J.-J. Laffont (ed.), *Advances in Economic Theory*, Vol. 2, Cambridge University Press, 71–155.

[8] **Maskin, E. & Tirole, J. (1999).** "Unforeseen Contingencies and Incomplete Contracts." *Review of Economic Studies*, 66(1), 83–114.

[9] **Laffont, J.-J. & Tirole, J. (2000).** *Competition in Telecommunications.* MIT Press.

[10] **Armstrong, M. & Sappington, D. E. M. (2007).** "Recent Developments in the Theory of Regulation." In *Handbook of IO*, Vol. 3, 1557–1700.

[11] **Laffont, J.-J. (2005).** *Regulation and Development.* Cambridge University Press.

[12] **Joskow, P. L. (2008).** "Incentive Regulation and Its Application to Electricity Networks." *Review of Network Economics*, 7(4), 547–560.

[13] **Laffont, J.-J. & Martimort, D. (2002).** *The Theory of Incentives: The Principal-Agent Model.* Princeton University Press.

[14] **张维迎 (1999).** 《企业理论与中国企业改革》. 北京大学出版社。

[15] **林毅夫、蔡昉、李周 (1997).** 《充分信息与国有企业改革》. 上海三联书店。

[16] **Bai, C.-E., Lu, J., & Tao, Z. (2006).** "The Multitask Theory of State Enterprise Reform." *AER*, 96(2), 353–357.

[17] **Naughton, B. (2007).** *The Chinese Economy: Transitions and Growth.* MIT Press.

[18] **OECD (2015).** *OECD Reviews of Regulatory Reform: China.* OECD Publishing.

[19] **Mirrlees, J. A. (1971).** "An Exploration in the Theory of Optimum Income

Taxation." *Review of Economic Studies*, 38(2), 175–208.

[20] **Holmström, B. (1979).** "Moral Hazard and Observability." *Bell Journal of Economics*, 10(1), 74–91.

[21] **Ballard, C. L., Shoven, J. B., & Whalley, J. (1985).** "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States." *American Economic Review*, 75(1), 128–138.

[22] **Ahmad, E. & Stern, N. (1991).** *The Theory and Practice of Tax Reform in Developing Countries.* Cambridge University Press.

[23] **Tirole, J. (1986).** "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." *Journal of Law, Economics, & Organization*, 2(2), 181–214.

[24] **Stigler, G. J. (1971).** "The Theory of Economic Regulation." *Bell Journal of Economics and Management Science*, 2(1), 3–21.

[25] **Peltzman, S. (1976).** "Toward a More General Theory of Regulation." *Journal of Law and Economics*, 19(2), 211–240.

[26] **Becker, G. S. (1983).** "A Theory of Competition Among Pressure Groups for Political Influence." *Quarterly Journal of Economics*, 98(3), 371–400.

[27] **Grossman, S. J. & Hart, O. D. (1986).** "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94(4), 691–719.

[28] **Hart, O. & Moore, J. (1990).** "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98(6), 1119–1158.

[29] **Littlechild, S. C. (1983).** *Regulation of British Telecommunications' Profitability.* Department of Industry, London.

[30] **Averch, H. & Johnson, L. L. (1962).** "Behavior of the Firm Under Regulatory Constraint." *American Economic Review*, 52(5), 1052–1069.

[31] **Baumol, W. J. & Willig, R. D. (1983).** "Fixed Costs, Sunk Costs, Entry Barriers, and Sustainability of Monopoly." *Quarterly Journal of Economics*, 98(3), 405–431.

[32] **Kornai, J. (1980).** *Economics of Shortage.* North-Holland.

第十章　从 SCP 到结构估计：Bresnahan 纲领与新实证产业组织

章首语

1982 年，Timothy Bresnahan 在 *Economics Letters* 上发表了一篇仅 4 页的短文，标题是 “The Oligopoly Solution Concept is Identified”。这篇文章提出了一个看起来不可能完成的任务：从市场价格和数量的数据中，**同时**识别企业的边际成本和企业的竞争行为——而这两个东西都是不可观测的。

这个问题的困难可以用一个简单的类比来理解。假设你观察到一个房间里的温度和湿度。温度和湿度是由暖气和空调同时决定的——但你看不到暖气和空调的 “设定”。你只能观察到结果。你怎么知道房间是 “太热”（垄断高价）还是 “太冷”（竞争性低价）？

Bresnahan 的答案是：你需要找到一个会**改变房间保温性能**的变量——而不只是改变外部温度。在经济学中，这意味着需要找到一个不只是移动需求截距、而是移动需求**斜率**的外生变量。为什么是斜率？因为在不同的竞争行为下，企业对需求斜率变化的反应不同——这提供了识别市场势力的必要 “杠杆”。

这一洞见——被称为 **Bresnahan 纲领**——开启了新实证产业组织（New Empirical Industrial Organization, NEIO）的革命。从 Porter (1983) 的铁路卡特尔 regime-switching 检验，到 Genesove-Mullin (1998) 的糖业卡特尔 “黄金标准” 验证，到 Olley-Pakes (1996) 的生产函数生产率估计，到 De Loecker-Warzynski (2012) 的全球加成率趋势——NEIO 在过去四十年中彻底改变了我们如何 “看见” 市场势力。

本章是模块 5（NEIO）的开篇。第 ~7~ 章–第 ~9~ 章建立了博弈论和机制设计的理论模型；本章回答那个被第 ~9~ 章明确提出但尚未回答的问题：**这些理论模型的参数如何从数据中估计？**

10.1　核心问题

读完本章，你应该能够回答以下问题：

1. **NEIO 为什么需要“革命”？** SCP 跨行业回归的根本缺陷是什么？为什么“利润 = 市场势力”这个等式在理论和实证上都是错误的？Demsetz (1973) 的“效率替代假说”如何颠覆了 SCP 的推断逻辑？

2. **什么是“结构估计”？**“结构”（structural）在 NEIO 中意味着什么？它如何区别于“简约式”（reduced-form）方法？为什么 NEIO 坚持从经济理论的一阶条件出发来推导估计方程？

3. **Bresnahan 的 θ 参数如何统一不同的竞争模型？**为什么 $P+\theta\cdot P'(Q)\cdot Q=MC$ 可以在同一个框架内捕捉完全竞争、Cournot、Bertrand 和合谋？“嵌套”（nesting）为什么是计量检验的关键？

4. **需求旋转为什么是识别的关键？**为什么仅移动需求截距的外生变量不能识别 θ？为什么需要移动需求**斜率**？如何直观地（及形式化地）理解这一识别条件？

5. **如何在实践中实施 NEIO？**从 Porter (1983) 的 regime-switching 到 Genesove-Mullin (1998) 的直接成本数据——NEIO 的经典应用展示了哪些不同的识别策略？各自在什么条件下有效？

6. **生产函数方法如何“绕过”需求估计？**为什么 Olley-Pakes (1996) 和 De Loecker-Warzynski (2012) 的方法可以在不估计需求系统的情况下估计加成率？这种方法的“代价”是什么？

10.2　SCP 的黄昏：为什么需要 NEIO

NEIO 的革命不是凭空产生的——它是对旧实证产业组织（SCP）范式系统性失败的反应。本节回顾 SCP 的实证方法论，呈现其根本缺陷，并建立 NEIO 的智识动机。

10.2.1　SCP 跨行业回归的逻辑与根本局限

SCP 实证研究的标准框架。在第 ~3~ 章–第 ~4~ 章中，我们详细讨论了哈佛学派的 SCP 范式及其衰落。这里从计量经济学的角度回顾其核心方程。典型的 SCP 跨行业回归形式为：

$$\pi_i=\alpha+\beta\cdot \text{CR4}_i+\gamma\cdot \text{Barrier}_i+\delta\cdot \text{AdIntensity}_i+\varepsilon_i$$

其中 i 索引行业，π_i 是行业 i 的平均利润率（或价格-成本差，PCM），CR4_i 是四企业集中度，Barrier_i 是进入壁垒的度量（如最小有效规模、广告支出），AdIntensity_i 是广告强度。

核心识别假设：CR4 和其他“结构”变量外生于利润——即不存在同时影响集中度和利润的遗漏变量。如果这个假设不成立，β 不能被解释为“集中度对利润的因果效应”。

三个根本性问题。即使忽略识别假设，SCP 回归面临三个更深层的缺陷：

(1) 会计利润 $\neq$ 经济利润（Fisher & McGowan 1987）。会计准则下的利润不反映经济利润（经济利润 = 收入 − 经济成本，包括资本的机会成本）。无形资产（品牌、研发、专利）在会计上通常被费用化而非资本化——拥有大量无形资产的企业，其会计利润被系统性地**低估**。这导致 SCP 回归的因变量本身就充斥着测量误差，且该误差与行业特征（无形资产密集度）系统相关。

(2) 效率替代假说（Demsetz 1973）。假设行业中有几家高效率企业（低边际成本）和多家低效率企业（高边际成本）。高效率企业获得高市场份额（$\to$ 高 CR4）**且**获得高利润（因为成本低）。低效率企业低市场份额 + 低利润。于是我们在数据中观察到：CR4 与行业平均利润正相关。但这反映的是**效率差异**（高效率企业既获得了高份额也获得了高利润），而非**市场势力**（合谋导致高价格）。SCP 的推断——“高集中度 $\to$ 合谋 $\to$ 高利润”——可能完全颠倒了因果方向。

(3) 芝加哥的反驳（回顾第 ~5~ 章）。Peltzman (1977) 和 Brozen (1982) 发现，早期 SCP 研究中的利润率-集中度正相关在控制了效率差异后**消失**，且在更长的面板中**不持续**。如果高集中度反映了高利润行业吸引了新进入者，那么在长期中，进入会侵蚀利润——“高利润”是暂时性的，而非结构性的。

SCP 范式的根本困境

SCP 的核心推断链——“结构 $\to$ 行为 $\to$ 绩效”——从未被直接检验。SCP 实证文献检验的是**结构与绩效的相关性**（CR4 $\to \pi$），但将中间的“行为”环节**假设**掉了（高 CR4 $\to$ 合谋 $\to$ 高利润）。NEIO 要做的，恰恰是打开这个黑箱——**直接估计行为**。

10.2.2 从“行业间比较”到“行业内估计”：NEIO 的方法论转向

NEIO 的三个方法论承诺。 Bresnahan (1989) 在 *Handbook of IO* 第 17 章中系统阐述了 NEIO 的方法论纲领：

(1) 经济原语（economic primitives）：估计的是**结构参数**（需求弹性、边际成本、行为参数）——而非仅仅是“集中度与利润的相关性”。结构参数在政策变化时是**不变的**（invariant to policy）——这赋予了结构估计进行反事实模拟的能力（Reiss & Wolak 2007）。

(2) 单个行业深度研究：NEIO 通常聚焦于单个行业的时间序列或面板数据——而非跨行业的横截面。这使研究者可以：(a) 利用特定行业的制度知识来建立识别策略（如 Porter 对五大湖通航状态的利用），(b) 允许行为参数随行业环境变化（而非假设所有行业共享同一个 “合谋程度”），(c) 避免跨行业比较中 “苹果和橙子” 的可比性问题。

(3) 行为假设的直接检验：不去推断 “是否存在市场势力”，而是直接检验 “企业遵循什么行为规则”。在同质产品行业，$\theta = 0$（完全竞争）、$\theta = 1/n$（Cournot）、$\theta = 1$（完全合谋）是统计上可检验的嵌套假设。这使 “市场势力” 从模糊的概念变成了**可检验的参数**。

表 10.1: SCP（旧实证 IO）与 NEIO（新实证 IO）的核心对比

	SCP（旧实证 IO）	**NEIO（新实证 IO）**
分析单位	跨行业横截面（几十到几百个行业）	单个行业时间序列/面板（一个行业，数百期/家企业）
核心参数	集中度-利润率的相关系数 β	行为参数 θ，需求弹性 η，边际成本 MC
理论基础	松散的 “结构 $\to$ 行为 $\to$ 绩效” 链	严格的企业 FOC（博弈论模型——第 ~7~ 章）
识别来源	行业间的集中度变异	时间序列中的需求/成本冲击 + 需求旋转变量
反事实能力	无（β 不是结构参数——政策变化时 β 可能改变）	有（结构参数在政策变化时不变——可进行合并不批模拟）
行为检验	不可行（行为被 “假设” 而非 “检验”）	Wald/LR 检验：$H_0 : \theta = 0$, $H_0 : \theta = 1/n$, $H_0 : \theta = 1$
主要局限	利润的测量误差 + 遗漏变量（效率）	函数形式假设 + 计算复杂度 + 排除限制

SCP vs. Chicago vs. NEIO：三位辩证法。 SCP 说 “高集中度导致高利润”（结构论），芝加哥说 “高利润反映高效率而非市场势力”（效率论）。NEIO 绕过了这场辩论——不去争论 “利润从何而来”（因为这个问题的答案依赖于不可观测的成本和利润分解），而是直接估计 “企业的行为参数 θ 是什么”。θ 不依赖于会计利润的测量——它来自于企业的一阶条件和需求/成本估计。如果说 SCP 和芝加哥在争论一枚硬币是正面还是反面，NEIO 则发明了天平来直接称量硬币的重量。

10.2.3　芝加哥学派对 SCP 的挑战与 NEIO 的回应

芝加哥学派的实证批判。回顾第 ~5~ 章的讨论——芝加哥学派对 SCP 发起了三个层面的攻击：

- **Demsetz (1973)：效率替代假说。**高集中度 ← 高效率企业扩大市场份额 → 高利润（效率导致的结果）。集中度和利润的关联是**伪相关**——两者都由效率驱动。

- **Peltzman (1977)：控制了效率后，利润-集中度正相关消失。**在回归中加入反映效率差异的变量（如不同规模企业的利润率之差）后，SCP 的核心发现（$\beta > 0$）不再统计显著。

- **Brozen (1982)：SCP 的发现不持久。**在更长的面板数据中，SCP 发现的利润-集中度正相关**随时间衰减**——这与“高利润吸引进入”的芝加哥世界观一致，但不符合“高利润反映持久市场势力”的 SCP 世界观。

NEIO 如何超越这场辩论。 NEIO 不是选择 SCP 或芝加哥的一方——它认可芝加哥的大部分批评（SCP 的“会计利润”方法有问题），但拒绝芝加哥的默认假设（“市场通常是竞争的”）。NEIO 的立场是：市场势力是否存在、程度多大，是一个**需要被检验的实证问题**——而不是一个可以被“假设”的前提。SCP 假设了高集中度意味着市场势力（错了），芝加哥假设了市场通常是竞争的（也可能错）。NEIO 说：让我们从数据中**直接检验**企业的行为规则——用 θ 来判定。

历史的讽刺

NEIO 的方法论革命在很大程度上**印证**了芝加哥的一个核心洞见：经济学应该从个体的最优化行为出发来理解市场结果（方法论个体主义）。Bresnahan 的供给关系正是从企业的 FOC 推导出来的——而不是从行业层面的“结构-绩效”相关性推导出来的。NEIO 在**方法论**上是芝加哥式的（微观基础），但在**结论**上不预设芝加哥的世界观（市场是否竞争是一个实证问题）。

10.3 Bresnahan (1982, 1989)：行为参数与需求旋转

【核心】

这是全章的心脏。本节完整展开 Bresnahan 的识别框架：从企业一阶条件推导行为参数 θ，陈述识别问题，展示需求旋转的解决方案，并提供形式化的识别条件和估计方法。每一个步骤都配以经济学直觉。

10.3.1 从企业 FOC 到行为参数 θ：一个统一的定价方程

设定。考虑一个有 n 家企业的同质产品行业（n 已知或可推断）。

- 企业 i 生产 q_i 单位，行业总产量 $Q = \sum_{i=1}^{n} q_i$

- 反需求函数：$P = P(Q, Z)$，其中 Z 是外生需求移动变量（如收入、天气、节假）

- 企业 i 的总成本：$C_i = C(q_i, W_i)$，其中 W_i 是外生成本移动变量（如要素价格）

- 边际成本：$\text{MC}_i = \partial C_i / \partial q_i$

企业 i 的利润最大化问题：

$$\max_{q_i} \ \pi_i = P(Q, Z) \cdot q_i - C_i(q_i, W_i)$$

一阶条件（FOC）：

$$\frac{\partial \pi_i}{\partial q_i} = P(Q, Z) + \frac{\partial P(Q, Z)}{\partial Q} \cdot \frac{\partial Q}{\partial q_i} \cdot q_i - \text{MC}_i(q_i, W_i) = 0$$

其中 $\partial Q / \partial q_i$ 是**推测变差**（conjectural variation）——当企业 i 增加 1 单位产量时，它“推测”行业总产量会增加多少。这个术语起源于 Bowley (1924)，在现代 IO 中被重新解释为*行为参数*（conduct parameter）——它捕捉的是企业竞争的“强度”，而非字面意义上的“推测”。

关键一步——定义行业层面的行为参数 θ。Bresnahan 的关键简化是：在对称均衡中（$q_i = Q/n$ 对所有 i），定义行业层面的行为参数 $\theta = (\partial Q / \partial q_i) \cdot (q_i / Q) = (\partial Q / \partial q_i)/n$（即份额加权的推测变差，share-weighted conjectural variation）。这一定义将企业层面的推测变差转换为行业层面的竞争强度度量：

$$\boxed{P(Q, Z) + \theta \cdot P'(Q, Z) \cdot Q = \text{MC}(Q/n, W)}$$

其中 $P'(Q, Z) = \partial P / \partial Q < 0$（需求曲线向下倾斜），MC 被假设为对所有企业对称。

表 10.2: 行为参数 θ 在不同竞争模型中的取值

竞争模式	企业行为	θ 的取值
完全竞争（Perfect Competition）	每家企业视市场价格为给定，不认为自己的产量影响 P	$\theta = 0$
Bertrand 同质产品（Bertrand, homogeneous）	价格竞争 + 同质产品 $\to$ 价格战 $\to$ $P = \text{MC}$	$\theta = 0$
对称 Cournot（Symmetric Cournot）	每家企业选择产量，视对手产量为给定	$\theta = 1/n$
Stackelberg 领导者	领导者内化其产量对追随者产量的影响	$\theta > 1/n$（具体取决于 n 和需求曲率）
完全合谋/联合垄断（Perfect Collusion）	企业联合选择产量使联合利润最大化	$\theta = 1$
部分合谋（Partial Collusion）	企业在 Cournot 和完全合谋之间	$1/n < \theta < 1$

经济学直觉：θ 衡量企业“内化”其产量决策对市场价格影响的程度。

- 完全竞争企业“忽视”其产量对价格的影响（$\theta = 0$）——它视 P 为给定。
- Cournot 企业部分内化（$\theta = 1/n$）——因为 n 家企业分享市场，每家企业的产量占市场的 $1/n$。
- 合谋企业完全内化（$\theta = 1$）——因为合谋联盟像一个垄断者一样考虑其产量对价格的完整影响。

θ 的嵌套性质——这是 NEIO 最关键的技术优势

因为 $\theta = 0$、$\theta = 1/n$、$\theta = 1$ 都是**同一个模型的特例**（参数约束），“区分竞争和合谋”从模糊的概念变成了一个**可检验的嵌套假设**：我们估计 $\hat{\theta}$，然后用 Wald 检验或似然比检验来检验 $H_0 : \theta = 0$、$H_0 : \theta = 1/n$ 或 $H_0 : \theta = 1$。这是 SCP 范式永远无法做到的。

10.3.2　识别问题的精确陈述：为什么 θ 和 MC 不可分离？

根本问题。考虑简化的情况以暴露识别问题的本质。设反需求函数为线性：

$$P = a - bQ + \varepsilon \qquad (\text{需求}), \qquad \varepsilon \sim \text{需求冲击}$$

边际成本为常数：$\text{MC} = c + \nu$，其中 ν 是成本冲击。

供给关系（企业 FOC）：

$$P = c + \nu + \theta \cdot b \cdot Q$$

这个式子看起来像一条 "供给曲线"——P 是 Q 的线性函数，截距为 $c + \nu$，斜率为 θb。

均衡。令需求等于供给关系：

$$a - bQ + \varepsilon = c + \nu + \theta bQ \quad \Rightarrow \quad Q^* = \frac{a - c + \varepsilon - \nu}{b(1+\theta)}$$

$$P^* = c + \nu + \theta b \cdot \frac{a - c + \varepsilon - \nu}{b(1+\theta)} = \frac{c + \nu + \theta(a + \varepsilon)}{1 + \theta}$$

为什么 θ 和 b 无法分离？供给关系的斜率是 θb——但 b 来自需求。如果只有截距在变化（ε 或 ν 的变化），均衡点 (Q^*, P^*) 的散点图是一条曲线，我们只能识别**供给关系的斜率**（即 θb），无法分离 θ 和 b。

Lau (1982) "不可能定理" 的直觉：如果需求函数在所有参数上是线性的，而边际成本函数也是线性的，且不存在一个变量可以**旋转**需求曲线（改变 b），那么 θ 和 b 在结构上是**不可识别**的——无论数据有多好，样本量有多大，有多少个外生冲击。

用图形理解：想象我们观察到市场在不同时期的 (Q_t, P_t)。需求冲击 ε_t 移动需求曲线——均衡点沿供给关系移动，描绘出供给关系的形状。成本冲击 ν_t 移动供给关系——均衡点沿需求曲线移动，描绘出需求的形状。但供给关系的斜率中包含的 θ 和 b 始终 "乘在一起"：供给关系的斜率 $= \theta b$。除非有什么东西可以独立地揭示 b 或 θ，否则两者无法被分开。

10.3.3　解决方案：需求旋转（Demand Rotation）

Bresnahan 的核心洞见。需要一个外生变量，它不仅移动需求截距，还移动需求**斜率**。具体地，令需求函数为：

$$\boxed{P = a_0 + a_1 R + (b_0 + b_1 R) \cdot Q + Z'\gamma + \varepsilon}$$

其中 R 是外生变量（"旋转变量"），它：

- 移动截距：$a_1 \neq 0$（R 的变化移动需求曲线）
- **旋转斜率**：$b_1 \neq 0$（R 的变化改变需求曲线的斜率——这是关键）

- 不进入 MC 函数：R 被排除于成本方程

供给关系仍然来自企业 FOC：

$$P = \mathrm{MC}(Q/n, W) - \theta \cdot \underbrace{(b_0 + b_1 R)}_{=P'(Q)<0} \cdot Q + \nu$$

为什么现在 θ 可以被识别？当 R 变化时：

(1) R 的变化改变了需求斜率（$b_0 + b_1 R$ 变化 $\to$ 需求曲线“旋转”）

(2) 需求斜率的变化改变了供给关系中的项 $-\theta \cdot P'(Q) \cdot Q$——因为 $P'(Q) = b_0 + b_1 R < 0$

(3) **在不同的 R 下，供给关系的有效斜率不同**（θb_0 vs. $\theta(b_0 + b_1 R)$——因为 $b_1 R$ 不同）

(4) 通过观察 P^* 和 Q^* 如何随 R 的变化而变化（同时控制了 R 的截距效应），我们可以分离 θ 和 b

识别 θ 的直观公式。在均衡中，供给关系的有效斜率为 $-\theta \cdot (b_0 + b_1 R)$。当 R 变化一个单位，有效斜率变化 $-\theta b_1$。由于 b_1 可以从需求估计中识别（通过观察 R 如何影响需求斜率），而有效斜率的变化可以从供给关系中识别，θ 就可以被“除以”出来：

$$\theta = -\frac{\partial(\text{供给关系有效斜率})/\partial R}{b_1}$$

需求旋转的经济学直觉——全章最重要的直觉

- **截距变化**（a_1）：移动需求曲线但不旋转——均衡沿供给关系滑动——只能描绘出供给关系的“位置”（即 MC 的水平），不能描绘出供给关系的“斜率”（即 θ）。
- **斜率变化**（b_1）：**旋转**需求曲线——如果 $\theta > 0$，企业 FOC 中包含 $P'(Q)$ 项——需求斜率的改变改变了供给关系的有效斜率——θ 越大的行业，供给关系对需求旋转越“敏感”。
- **如果 $\theta = 0$**（完全竞争）：供给曲线就是 MC 曲线——不受需求斜率影响——需求旋转不改变供给曲线的形状。
- **如果 $\theta > 0$**：供给关系中的 $\theta \cdot P'(Q) \cdot Q/n$ 项随需求斜率变化——θ 越大，供给关系对需求旋转的“反应”越大。

结论：需求旋转提供了一个“分离 和 MC”的天然实验——这是在不需要直接观测 MC 的情况下识别市场势力的唯一方法。

10.3.4　识别条件的正式陈述

识别条件（Bresnahan 1989, 基于 Lau 1982）。在以下条件下，θ 可从市场数据中被识别：

(C1) 需求旋转（Demand Rotation）：存在至少一个外生变量 R，它进入需求函数且与内生变量（Q 或 P）的交互项系数非零——即 $b_1 \neq 0$。R 必须不仅移动需求截距，还要旋转需求斜率。

(C2) 排除限制（Exclusion Restriction）：R 不进入边际成本函数。如果 R 同时影响需求和 MC，则 R 的变化既移动了需求也移动了供给——混淆了识别。

(C3) 非极端弹性（Non-extreme Elasticities）：需求函数和 MC 函数不能是 "完全弹性的" 或 "完全无弹性的"——否则 θ 不进入 FOC（FOC 退化为 $P = \text{MC}$ 或 Q 固定）。

条件 (C1) 是 Bresnahan 的核心贡献——也是 NEIO 与 SCP 的最根本区别。SCP 寻找的是 "集中度与利润的截距关系"（即截距移动）；NEIO 寻找的是 "需求旋转"（即斜率变化）——后者在计量经济学上提供了识别 θ 所需的 "额外维度"。

Lau (1982) 的 "不可能定理"——为什么需求旋转是必要的。 Lau (1982, *Economics Letters*) 证明：

> 如果需求函数在参数上是线性的（$P = a - bQ + Z\gamma$），而边际成本也是线性的，且不存在 "旋转变量"（即没有变量可以与 Q 交互来改变 b），那么 θ 和 b 在结构上不可区分——无论样本量多大。

这意味着：识别 θ 需要**非线性性**——要么需求在变量上是非线性的，要么存在一个 "交互项"（$R \times Q$）来旋转需求。仅靠截距的变化永远不够。

10.3.5　估计方法：GMM 与极大似然

估计策略——两步法或系统估计。

步骤 1：需求估计。需求函数（反需求形式）：

$$P_t = a_0 + a_1 Q_t + a_2 Z_t + a_3 R_t + a_4 (R_t \times Q_t) + \varepsilon_t^D$$

其中 a_4 捕捉需求旋转（$a_4 \neq 0$ 是识别 θ 的必要条件）。**内生性问题：**Q_t 与 ε_t^D 相关（需求冲击同时影响 P 和 Q）——OLS 不一致。需要工具变量：Z_t（需求移动变量，不进入 MC）、R_t（旋转变数，不进入 MC）、W_t（成本移动变量，不进入需求）。

步骤 2：供给关系估计。从企业 FOC：

$$P_t = \text{MC}(Q_t/n, W_t) - \theta \cdot P'(Q_t, Z_t, R_t) \cdot Q_t/n + \nu_t$$

代入步骤 1 估计的需求斜率 $\widehat{P}'(Q_t) = a_1 + a_4 R_t$（在反需求规格下）：

$$P_t = c_0 + c_1 Q_t + c_2 W_t - \theta \cdot (a_1 + a_4 R_t) \cdot Q_t/n + \nu_t$$

在这个方程中：

- c_0, c_1, c_2 是 MC 参数（待估计）
- θ 是行为参数（待估计——本章的核心参数）
- $(a_1 + a_4 R_t)$ 来自需求估计——R_t 的变异使 θ 的效应可以被区分于 MC 参数

为什么需要 GMM 而非 OLS？ 在供给关系中，Q_t 与 ν_t 相关（成本冲击 $\uparrow \to Q \downarrow$）——OLS 不一致。供给关系中的工具变量包括 W_t（成本移动变量，不进入需求）和 R_t（旋转变量——进入需求但不进入 MC）。系统 GMM（Hansen 1982）同时使用需求方程和供给方程的矩条件：

$$\mathbb{E}[\varepsilon_t^D \mid Z_t, R_t, W_t] = 0, \quad \mathbb{E}[\nu_t \mid Z_t, R_t, W_t] = 0$$

这比两步法更有效（因为需求和供给的残差通常相关——需求和成本冲击可能同时发生）。

极大似然估计（FIML）。 如果假设 (ε_t^D, ν_t) 服从联合正态分布，可以使用全信息极大似然（Full Information Maximum Likelihood, FIML）同时估计需求参数、MC 参数和 θ。FIML 比 GMM 更有效（当分布假设正确时），但对分布假设更敏感。

10.3.6 嵌套检验与非嵌套检验：从 $\hat{\theta}$ 到行为推断

嵌套检验（Nested Tests）——区分行为假设。 估计 $\hat{\theta}$ 及其标准误后，用标准检验区分竞争模式：

- $H_0 : \theta = 0$（完全竞争/Bertrand 同质）——检验市场势力是否存在
- $H_0 : \theta = 1/n$（对称 Cournot）——检验 Cournot 行为是否成立
- $H_0 : \theta = 1$（完全合谋/联合垄断）——检验完全合谋是否成立

这些是嵌套检验——因为每个 H_0 都是 "θ 自由估计" 模型的特例。可以用 Wald 检验（估计 $\hat{\theta}$，检验约束）或似然比检验（比较限制和非限制模型的似然值）。

非嵌套检验（Non-Nested Tests）——比较不同的需求/MC 规格。 如果两个模型是**非嵌套**的（如线性需求 vs. 对数线性需求——一个不能通过参数约束从另一个导出），则需要：

- **Vuong (1989) 检验：** 比较两个非嵌套模型哪个更接近 "真实" 数据生成过程。检验统计量 $V \sim N(0,1)$——正的 V 支持模型 1，负的支持模型 2。

- **信息准则：** AIC、BIC——BIC 对参数数量的惩罚更重（通常偏好更简约的模型），在结构 IO 中更为常用（因为结构模型通常有较多参数）。

实践中的模型选择策略：

(1) 首先估计一般的 θ 模型（θ 自由估计）

(2) 用 Wald/LR 检验检验 $H_0 : \theta = 0$、$H_0 : \theta = 1/n$、$H_0 : \theta = 1$

(3) 如果所有嵌套检验都被拒绝 $\rightarrow \theta$ 是 “中间值”——部分合谋、不对称行为、或模型误设

(4) 进行稳健性分析：在不同需求函数形式、不同 MC 规格下重新估计 θ——如果 $\hat{\theta}$ 在不同规格下变化很大，结论不可靠

10.4　经典应用 I：Porter (1983) 与铁路卡特尔

【核心】

Porter (1983) 是 Bresnahan 框架的第一个、也是最具影响力的实证应用。它不仅展示了 NEIO 方法的可行性和威力，还开创了结构 IO 实证研究的写作范式。

10.4.1　制度背景：19 世纪末 JEC 铁路卡特尔

JEC（Joint Executive Committee）的运作。美国东部的铁路运输卡特尔——Joint Executive Committee——成立于 1879 年，持续到 1886 年。成员包括从芝加哥到东海岸的三条主要铁路干线和多家支线铁路。JEC 的核心职能是：

- **分配市场份额：** JEC 的中央办公室根据历史份额、运力和协商结果分配各成员在不同路线上的市场份额
- **设定价格：**统一的运费率（每吨英里的美元价格）
- **监督与执行：**成员报告其实际货运量和收入——如果某成员超过了分配的份额，需向其他成员支付 “补偿”

JEC 合法性的关键： 1887 年 Interstate Commerce Act（第一部美国联邦反垄断法）之前，铁路卡特尔在美国是**合法**的——卡特尔协议、价格设定会议和市场份额分配的详细记录被保存下来。这意味着 Porter 可以接触到一个极罕见的**公开的卡特尔运营数据**——在大多数情况下，卡特尔是秘密运作的，研究者无法获得其内部记录。

数据：周度数据，1880–1886 年（约 350 周）：

- 价格（P_t）：每吨英里的平均运费（美分）

- 数量（Q_t）：总吨英里货运量
- 五大湖通航状态（LAKE_t）：$= 1$ 如果五大湖可通航（夏季——约 4 月至 11 月），$= 0$ 如果结冰（冬季——约 12 月至 3 月）
- 进入/退出事件：铁路支线的开设和关闭

10.4.2 识别策略：五大湖通航状态作为需求旋转变量

为什么五大湖通航状态是理想的需求旋转变量？

- **需求侧：**夏季五大湖可通航 → 湖泊运输（steamships）是铁路运输的替代 → 铁路需求面临弹性替代——铁路的剩余需求变得**更有弹性**（$|b|$ 更小）。冬季湖面结冰 → 湖泊无法通航 → 铁路是唯一运输方式 → 铁路需求**更缺乏弹性**（$|b|$ 更大）。这不仅是截距的变化——通航状态改变了需求的**斜率**。
- **供给侧（排除限制）：**五大湖的结冰/解冻是**气象学**现象——与铁路的运营成本（劳动力、煤炭、铁轨维护）无关。通航状态不进入 MC 函数。这是关键的排除限制——通航状态只通过需求侧影响市场。
- **外生性：**五大湖的结冰/解冻完全由季节性气温决定——不受铁路企业或 JEC 的任何决策影响。

需求函数（反需求形式）：

$$\log P_t = \alpha_0 + \alpha_1 \log Q_t + \alpha_2 \text{LAKE}_t + \alpha_3 (\text{LAKE}_t \times \log Q_t) + \alpha_4 S_t + \varepsilon_t^D$$

其中 α_3 是需求旋转参数——$H_0 : \alpha_3 = 0$ 意味着无需求旋转（此时 θ 不可识别）。S_t 是季节和需求移动变量（如收获季节的粮食运输需求）。

10.4.3 Regime-Switching 模型：合谋与价格战的切换

Porter 的关键创新——允许 θ 在两个 regime 之间切换。与“估计一个平均 θ”不同，Porter 认识到 JEC 并非始终处于合谋或始终处于价格战——它在这两种状态之间**切换**。这需要一种计量模型来识别**不可观测的 regime：**

$$\text{供给关系：} \quad P_t = \text{MC}_t - \theta^{S_t} \cdot P'(Q_t) \cdot Q_t / n + \nu_t$$

其中 $S_t \in \{C, W\}$（C = 合谋 regime，W = 价格战 regime），且这两个 regime 中 $\theta^C > \theta^W$（合谋期的行为参数大于价格战期）。Regime S_t 本身是**不可观测的**——计量经济学家只知道 P_t 和 Q_t，不知道当前是合谋还是价格战期。

切换模型的结构： Regime 切换由一个一阶马尔可夫链驱动：

$$\Pr(S_t = C \mid S_{t-1} = C) = p \quad \text{(合谋持续概率)}$$

$$\Pr(S_t = W \mid S_{t-1} = W) = q \quad \text{（价格战持续概率）}$$

p 和 q 也是待估计的参数。如果 $p > 1 - q$（即 $\Pr(C \mid C) > \Pr(C \mid W)$），则合谋期倾向于比价格战期更持久——这与“合谋是常态，价格战是短暂偏离”的直觉一致。

似然函数。对每个 t，我们有来自两个 regime 成分的混合密度：

$$\mathcal{L} = \prod_{t=1}^{T} \Big[\lambda_t \cdot f(P_t, Q_t \mid S_t = C) + (1 - \lambda_t) \cdot f(P_t, Q_t \mid S_t = W) \Big]$$

其中 $\lambda_t = \Pr(S_t = C \mid \text{历史数据})$ 是基于过去数据的滤波概率（filtered probability）——通过 Hamilton (1989) 滤波器（或其前身）迭代计算。

10.4.4　核心结果：合谋 $\theta \approx 0.9$，价格战 $\theta \approx 0.1$

Porter 的主要发现：

(1) 行为参数在两个 regime 间差异显著：

- 合谋 regime：$\hat{\theta}^C \approx 0.9$（标准误约 0.1）——非常接近完全合谋（$\theta = 1$）。Wald 检验不能拒绝 $H_0 : \theta^C = 1$。
- 价格战 regime：$\hat{\theta}^W \approx 0.1$（标准误约 0.05）——非常接近完全竞争（$\theta = 0$）。Wald 检验不能拒绝 $H_0 : \theta^W = 0$。
- **两个 $\hat{\theta}$ 统计上显著不同**（Wald 检验 $p < 0.01$）——行为确实在两个 regime 之间切换。

(2) 价格战是合谋的均衡特征——而非合谋的“失败”：

- 价格战倾向于在需求低时发生（LAKE = 1——五大湖通航，铁路面临湖泊运输竞争替代）——这与 Green & Porter (1984) 的“不完全监测下的合谋”理论一致（第 ~8~ 章 §~9.6）：当需求下降时，企业更难区分“我的市场份额下降是因为需求下降”和“我的市场份额下降是因为对手背叛”——最优合谋方案包含**周期性价格战**作为“惩罚机制”的一部分。
- 价格战的“成本”（价格战期间的利润损失）小于合谋的“收益”（合谋期的高利润）——因此合谋的**净现值**仍然为正。

(3) 价格战通常持续 2–8 周，然后回归合谋。合谋期的平均长度远大于价格战期——这与 p（合谋持续概率）和 q（价格战持续概率）的估计一致：$\hat{p} \approx 0.85$（每周期有 85% 概率继续合谋），$\hat{q} \approx 0.70$（每周期有 70% 概率继续价格战）。这意味着：合谋期的预期持续时间为 $1/(1 - 0.85) \approx 6.7$ 周，价格战期的预期持续时间为 $1/(1 - 0.70) \approx 3.3$ 周。

(4) 价格战期间价格降幅：约 30%–50%。这一幅度与“价格降至 MC 附近”一致——因为如果合谋价格包含约 50% 的 markup（$\theta = 1$ 时，$P = \text{MC}/(1 - 1/\eta)$，若 $\eta \approx 2$，markup $\approx 100\%$），价格战的回归将价格拉回了 MC 附近。

与 Ch.8 的连接： Porter (1983) 的实证结果为 第 ~8~ 章的合谋理论提供了**直接的实证验证**。特别是，第 ~8~ 章 §~9.6 中介绍的 Green & Porter (1984) 模型——不完全监测下，合谋的最优执行方案包含周期性价格战——在 Porter 的 JEC 数据中得到了显式的支持。这是一个理论-实证“闭环”的教科书案例：理论预测了周期性价格战，实证研究证实了价格战的发生以及它们的触发条件（需求下降），并与理论的数值预测（θ 从 ≈ 1 下降到 ≈ 0）一致。

10.4.5 Porter 的方法论遗产

方法论贡献：

(1) 第一个将 Bresnahan 框架应用于真实数据的实证研究——证明了 NEIO 不仅是理论上的可能，也是实证上可操作的。

(2) 将 regime-switching 引入结构 IO——这至今仍是检测“行为变化”的核心计量工具（在当代应用于检测合谋瓦解、竞争政策冲击、技术变革等的产业效应）。

(3) 展示了如何利用特定行业的**制度特征**（五大湖通航状态作为气象学外生冲击）来建立识别——NEIO 的“每个行业有自己的识别策略”传统由此开始。

(4) 开创了“结构 IO 实证写作范式”——制度背景 $\to$ 理论模型 $\to$ 识别策略 $\to$ 估计方法 $\to$ 结果 $\to$ 经济学解释 $\to$ 与理论对话。

局限：

- Regime 的数量（2）和性质（合谋 vs. 价格战）是预先指定的——如果存在第三种 regime（如“部分合谋”），模型会错误分类。
- Regime 切换被建模为外生的（常数转移概率 p 和 q）——在现实中，regime 切换可能受**内生**因素影响（企业策略性地选择何时发动价格战，何时回归合谋）。
- 同质产品假设——在铁路大宗商品运输中可能基本成立，但在大多数其他行业中不成立（为 Ch.11 的 BLP 方法埋下伏笔）。

10.5 经典应用 II：更多识别策略

【核心】

Porter (1983) 展示了需求旋转 + regime-switching 的组合策略。但 NEIO 的核心信息是：**没有一种适用于所有行业的识别策略——每个行业需要自己的制度知识来建立识别**。本节介绍另外两个经典应用，展示了不同的识别路径。

10.5.1 Bresnahan (1987)：美国汽车业——区分 Bertrand、Cournot 与合谋

背景与核心问题。1950s–1970s 的美国汽车市场——三巨头（GM、Ford、Chrysler）加上进口品牌。核心问题：美国汽车行业是 Bertrand 竞争（价格竞争——差异化产品下的行为是什么？）、Cournot 竞争（数量竞争——经销商分配量由企业决定？），还是部分合谋？

产品差异化带来的挑战。汽车是差异化产品（不同的马力、重量、燃油效率、尺寸、品牌、目标消费者群体）。Bresnahan 之前的 NEIO 框架假设同质产品——这意味着单一市场和单一价格。在差异化产品下，每家企业的需求系统包含多个产品的多个价格——“维度爆炸”（curse of dimensionality）。

识别策略——石油危机作为需求旋转。 Bresnahan 使用了一个巧妙的需求旋转来源：**石油价格的急剧变化（1973, 1979）**。石油价格的变化不仅改变了消费者对汽车的总体需求（截距变化）——更重要的是，它改变了消费者在**大型/耗油车型**和**小型/省油车型**之间的**替代弹性**（斜率变化）：

- 石油危机前（低油价）：消费者的替代弹性主要受品牌、马力、舒适性等特征驱动——大型车和小型车之间的替代弹性较低
- 石油危机后（高油价）：燃料经济性成为核心关切——消费者在大型车和小型车之间的替代弹性显著上升——需求系统被 “旋转”

这提供了一个识别行为参数 θ 的 “天然实验”——在不同的替代弹性下，企业对竞争对手产品价格变化的反应不同。

模型结构（简化版）： Bresnahan 使用了一个**多产品企业 FOC**——每家三巨头生产多个车型，企业选择价格（而非产量）最大化其**所有产品**的联合利润：

$$\mathbf{p}_j = \mathbf{mc}_j - \theta \cdot \left(\frac{\partial \mathbf{q}_j}{\partial \mathbf{p}_j} \right)^{-1} \mathbf{q}_j$$

其中 $\mathbf{p}_j$ 是企业 j 的产品价格向量，$\mathbf{q}_j$ 是相应的需求向量，$\partial \mathbf{q}_j / \partial \mathbf{p}_j$ 是需求系统的 Jacobian（自身和交叉价格效应）。θ 捕捉企业 j 在设定价格时 “内化” 竞争对手利润的程度——$\theta = 0$ 为 Bertrand（视对手价格为给定），$\theta > 0$ 为部分合谋，$\theta = 1$ 为完全合谋（联合价格设定）。

核心发现：不能拒绝 $H_0 : \theta = 0$（Bertrand 价格竞争）。即美国汽车行业的行为接近 Bertrand 差异化产品竞争——而非 Cournot 或合谋。但结果对需求系统的函数形式**高度敏感**——在不同需求规格下（对数线性 vs. AIDS 需求系统），$\hat{\theta}$ 的范围较宽。

方法论教训——为什么我们需要 BLP：Bresnahan (1987) 是 NEIO 在差异化产品行业中的**先驱**——它展示了 NEIO 可以在多产品差异化行业中运作。但它也暴露了 Bresnahan 框架在多产品环境下的根本局限：

- **维度爆炸**：对 J 个产品的需求系统，需要估计 $J \times J$ 个弹性——参数数量以 J^2 增长。即使聚合车型，也需要估计大量参数。

- **对需求函数形式高度敏感**：$\hat{\theta}$ 取决于需求的 "正确" 规格——但关于同一数据的最优需求规格可能没有共识。

- **价格内生性**：产品价格与未观测的产品质量相关——需要工具变量，但哪些变量可以作为 "多产品许多价格" 的工具变量？

这些局限正是 Berry, Levinsohn & Pakes (1995)——Ch.11 的核心——试图通过随机系数离散选择模型来解决的问题。Bresnahan (1987) 为 BLP 的 "革命" 铺平了道路。

10.5.2 Genesove-Mullin (1998)：糖业卡特尔的 "黄金标准" 验证

核心创新——直接观测边际成本。 Genesove 和 Mullin (1998, *RAND Journal of Economics*) 研究了 19 世纪末/20 世纪初的美国糖业。美国糖精炼公司（American Sugar Refining Company, ASRC）控制了美国精炼糖市场的大部分份额（在 1891 年控制了约 98% 的产能），在 1890–1914 年间是一个事实上的卡特尔。ASRC 的详细成本会计记录被保存下来——包括每磅糖精炼的原材料（甘蔗/甜菜）、劳动力、能源和运输成本。这意味着研究者可以**直接计算**边际成本 MC——MC 是**数据**而非**待估计参数**。

这为什么是 "革命性" 的？ Bresnahan 框架的 "阿喀琉斯之踵"（Achilles' heel）是需要同时估计 θ 和 MC——两个都是未知的。如果 MC 是数据，θ 可以从企业 FOC 中直接 "反向推导"：

$$\theta = \frac{P - \text{MC}}{-P'(Q) \cdot Q/n}$$

其中 $P - \text{MC}$ 是数据直接可计算的（价格减去已知的边际成本），$P'(Q)$ 来自需求估计（需求弹性 η）。因为只需要估计需求（而非同时估计需求和 MC），识别问题的难度大幅降低——不再需要需求旋转（因为 θ 不再与 MC 参数纠缠在一起）。

实证策略（三步骤）：

(1) 用成本会计数据计算每一期的 MC_t（直接成本 + 间接成本的分摊——Genesove 和 Mullin 对成本分解进行了大量的历史研究）

(2) 估计需求函数 $Q = D(P, Z)$ 得到需求弹性 $\hat{\eta}$

(3) 计算 "直接" 行为参数：$\hat{\theta}_t = (P_t - \text{MC}_t)/(P_t/\hat{\eta}_t \cdot Q_t/n)$ 或更准确地，从企业 FOC 的精确形式解出 θ

核心发现：

- **合谋期验证“完美”：**在公认的卡特尔存续期间（1890–1914），$\hat{\theta} \approx 1.0$——几乎精确等于完全合谋的预测。历史学家基于档案证据确认 ASRC 在此期间行使了市场势力——这与 $\hat{\theta}$ 的估计完全一致。
- **竞争期验证“完美”：**在卡特尔瓦解后的时期，$\hat{\theta} \approx 0$（完全竞争）或 $\approx 1/n$（Cournot）。在糖精炼行业中，不同时期的竞争模式有显著差异——而这些差异与行业结构的变化（进入、退出、反垄断行动）一致。
- **$\hat{\theta}$ 在 regime 之间的切换是急剧的：**而非渐进过渡——这表明行业在“合谋”和“竞争”之间跳跃，而非在连续体上平滑移动。这与 regime-switching 行为一致（呼应 Porter 1983 的发现）。
- **Lerner Index 的验证：**在合谋期，Lerner Index $\approx 1/|\eta|$（完全合谋的预测——$\mu = P/\text{MC} = \eta/(\eta-1)$）。在竞争期，Lerner Index $\approx 1/(n|\eta|)$（Cournot）或 0（完全竞争）。

方法论遗产——“黄金标准”验证：Genesove-Mullin (1998) 被视为 NEIO 的“黄金标准”（gold standard）验证。它证明了**当你有直接成本数据时，Bresnahan 框架运作得非常完美**——估计的 $\hat{\theta}$ 与历史学家基于档案证据判断的行为 regime 完全一致。它同时也暗示了 **NEIO 的弱点不在于理论——理论框架在理想条件下运作良好。**NEIO 的弱点是**数据**——没有直接成本数据时，结果依赖于 MC 函数形式的假设。这为后来的“生产函数方法”（不需要成本数据来估计加成率——见 §~10.7 和 §~10.8）提供了强有力的动力。

表 10.3: 三个经典 NEIO 应用的对比总结：Porter (1983)、Bresnahan (1987) 与 Genesove-Mullin (1998)

	Porter (1983)	**Bresnahan (1987)**	**Genesove-Mullin (1998)**
行业	铁路运输	汽车	糖精炼
产品类型	同质（大宗商品运输）	差异化（不同车型）	同质（糖）
核心问题	合谋 vs. 价格战	Bertrand vs. Cournot vs. 合谋	验证 Bresnahan 框架
识别策略	需求旋转（五大湖通航）+ regime-switching	石油危机改变替代弹性	**直接 MC 数据**——避免识别问题
$\hat{\theta}$	合谋 0.9，价格战 0.1	≈ 0（Bertrand）	合谋 ≈ 1.0，竞争 ≈ 0
关键贡献	第一个 NEIO 实证应用	将 NEIO 扩展到差异化产品	“黄金标准”验证
关键局限	同质产品假设	对需求函数形式高度敏感	要求直接成本数据（极少可得）

10.6 识别问题的深化：从需求旋转到 Corts 批判

【进阶】

10.6.1 识别条件的正式证明（简化版）

（本节为进阶读者提供识别 θ 的形式化证明。核心逻辑已在 §~10.3.3 中给出了直观说明——这里提供更严谨的推导。）

设定。延续 §~10.3.2 的线性需求 + 常数 MC 框架，加入旋转变量 R：

$$\text{反需求：} \quad P = a_0 + a_1 R + (b_0 + b_1 R)Q + \varepsilon$$

$$\text{边际成本：} \quad \text{MC} = c + \nu$$

$$\text{供给关系：} \quad P = c + \nu + \theta \cdot (b_0 + b_1 R) \cdot Q/n$$

均衡求解。令需求 = 供给关系：

$$a_0 + a_1 R + (b_0 + b_1 R)Q + \varepsilon = c + \nu + \theta(b_0 + b_1 R)Q/n$$

$$\Rightarrow \quad Q^* = \frac{a_0 + a_1 R - c - \nu + \varepsilon}{(b_0 + b_1 R)(1 - \theta/n)}$$

$$P^* = c + \nu + \theta(b_0 + b_1 R)Q^*/n$$

约简式（Reduced Form）——P^* 和 Q^* 如何随 R 变化。R 的变化同时移动了 Q^* 和 P^*。在均衡中：

$$\frac{\partial P^*}{\partial R} = \frac{\partial}{\partial R}\left[c + \nu + \frac{\theta}{n}(b_0 + b_1 R)Q^*\right]$$

P^* 随 R 的变化有两个成分：(a) 旋转变量对 Q^* 的影响（在均衡中），(b) 旋转变量对供给关系斜率 $\theta(b_0 + b_1 R)/n$ 的直接影响。通过观察这两个成分（在工具变量识别中），可以将 θ 从 b_0 和 b_1 中分离出来。

核心——为什么 $b_1 \neq 0$ 是识别 θ 的必要条件。如果 $b_1 = 0$（无需求旋转），则 $b_0 + b_1 R = b_0$ 对所有 R 都是常数。供给关系的有效斜率 $\theta b_0/n$ 不随 R 变化——R 仅移动需求截距——均衡点沿同一条供给关系滑动——供给关系的形状不包含关于 θ 的独立信息。θ 和 b_0 始终“乘在一起”而无法分离。

10.6.2 需求旋转变量的经典选择

选择 R 的标准：(1) $b_1 \neq 0$——变量确实旋转了需求（与内生数量/价格的交互项统计显著）；(2) 排除性——变量不进入 MC；(3) 外生性——变量不内生于市场的未观测冲击。

表 10.4: 需求旋转变量的经典选择：行业、变量与排除限制论证

行业	旋转变量 R	为什么旋转需求？	为什么满足排除限制？
铁路运输（Porter 1983）	五大湖通航状态	替代运输方式的可用性改变需求弹性	通航状态不影响铁路运营成本（天气）
汽车（Bresnahan 1987）	石油危机/燃料价格	燃料价格改变消费者在大小车型间的替代弹性	燃料价格不进入汽车制造成本
咖啡烘焙（Bettendorf-Verboven 2000）	天气冲击	不同品牌的替代弹性随天气变化	天气不影响烘焙 MC
啤酒（Hausman-Leonard-Zona 1994）	城市间零售竞争强度	不同城市的零售结构改变品牌间替代弹性	零售结构不进入酿造 MC
水泥（Jans-Rosenbaum 1997）	运输距离	距离改变不同水泥厂间的替代弹性	运输距离不进入工厂 MC 函数
手机（Nevo 2000）	产品线扩展	新产品引入改变品牌间替代模式	新产品引入时间外生于已有产品 MC

每个行业的旋转变量都利用了该行业**独特的制度特征**——这正是 NEIO 的 “行业深度研究” 方法论的核心：不存在 “一刀切” 的需求旋转变量，需要研究者对具体行业的制度和技术有深入理解。

10.6.3　Corts (1999) 批判：当 MC 也是弹性的

Corts 的核心论文。Corts (1999, *Journal of Econometrics*, “Conduct Parameters and the Measurement of Market Power”) 对 Bresnahan 框架提出了一个根本性的方法论批判。

Corts 的论证： Bresnahan 框架的识别依赖于一个关键假设：MC 的 “有效” 水平独立于需求旋转变量 R——即 R 不进入 MC 函数（排除限制 C2）。但如果企业在不同的需求条件下使用**不同的生产技术**——即 MC 函数本身随 R 变化——那么 $\hat{\theta}$ 会 “吸收”MC 的变异，导致有偏的推断。

形式化：假设真实的 MC 是 $\text{MC} = c_0 + c_1 R + \nu$（即 R 影响 MC），但研究者错误地假设 MC 独立于 R（$\text{MC} = c_0 + \nu$）。那么真实的供给关系是：

$$P = c_0 + c_1 R + \nu + \theta_{\text{true}} \cdot P'(Q, R) \cdot Q/n$$

但研究者估计的模型是：

$$P = c_0 + \nu + \theta_{\text{est}} \cdot P'(Q, R) \cdot Q/n$$

$c_1 R$ 的效应会被 θ_{est} "吸收"——即 $\hat{\theta}_{\text{est}} \neq \theta_{\text{true}}$。**$\hat{\theta}_{\text{est}}$ 是"行为参数"还是"MC 误设参数"？**——无法区分。

对 NEIO 实践的深远影响。 Corts 批判之后，NEIO 论文通常包含以下至少一项：

(1) Corts 稳健性检验：在不同 MC 规格下（$c_1 = 0$、$c_1 \neq 0$、半参数 MC）重新估计 θ——如果 $\hat{\theta}$ 在这些不同规格下稳定（变化范围小），结论可信；如果 $\hat{\theta}$ 变化很大，对 θ 的解释应极为谨慎。

(2) 直接成本数据（Genesove-Mullin 路径）——完全避免 MC 误设问题。

(3) 生产函数方法（§~10.7, §~10.8）——从完全不同的数据（企业投入品而非市场均衡价格）推断市场势力，不依赖于 MC 函数形式假设。

Corts 批判与 SCP 困境的结构相似。有趣的是，Corts 批判和 Demsetz (1973) 对 SCP 的批判在结构上是相似的：

- SCP 问题：利润 = 经济利润 + 效率差异——将"效率差异"误归为"市场势力"
- Corts 问题：$P - \text{MC}$ = 市场势力 markup + MC 误设——将"MC 误设"误归为"市场势力"
- 两种情况下，不可观测的"成分"都混淆了核心参数的推断

这解释了为什么 NEIO 学者如此重视生产函数方法——它通过从完全不同的数据中推断 markup（不需要估计 MC 函数）来绕过了这个问题。

10.7 生产函数方法 I：生产率估计

【核心】

Bresnahan 框架需要三个条件：同质产品（或至少可定义"市场"）、需求旋转变量 R、合理的 MC 函数形式。当这些条件难以满足时，NEIO 提供了一条"第二条道路"：用生产函数估计方法从企业投入产出数据中推断企业的**生产率**和**加成率**——**不需要估计需求系统，也不需要需求旋转变量**。

10.7.1　生产函数估计的核心问题：未观测的生产率 ω

标准的 Cobb-Douglas 生产函数（对数形式）：

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \varepsilon_{it}$$

其中：

- $y_{it} = \ln$（增加值/产出）——企业 i 在 t 期的产出
- $l_{it} = \ln$（劳动力投入）——劳动力
- $k_{it} = \ln$（资本投入）——资本存量
- ω_{it} = **未观测的全要素生产率**（Total Factor Productivity, TFP）——企业管理者知道自己的生产率，但计量经济学家不知道
- ε_{it} = 真正的随机冲击——i.i.d. 测量误差或未预期的生产中断——**企业管理者也不知道**

同时性偏差（Simultaneity Bias）——为什么 OLS 不一致。企业的管理者在决定当期的劳动力投入 l_{it} 和中间投入 m_{it} 时，**观察到了** ω_{it}（管理者知道自己的生产率水平）。因此：

- l_{it} 与 ω_{it} 正相关——高生产率企业选择高劳动力投入
- m_{it} 与 ω_{it} 正相关——高生产率企业使用更多中间投入
- k_{it} 可能与 ω_{it} 相关——高生产率企业投资更多

$\rightarrow$ OLS 将 ω_{it} 对 y_{it} 的正效应归因于 l_{it} 和 k_{it} $\rightarrow$ OLS 高估 β_l 和 β_k（向上偏误）。这是生产函数估计的核心问题——我们需要估计 ω_{it} 以得到一致的 β_l 和 β_k，但我们不能直接观测 ω_{it}。这是*不可观测的异质性*问题——在所有面板数据模型中广泛存在，但在生产函数估计中尤为严重（因为 ω_{it} 与企业投入决策的相关性很强）。

固定效应（FE）估计为什么不够？企业固定效应 α_i 仅控制了*时不变的*未观测生产率——但现实中生产率随时间变化（ω_{it}，而非 ω_i）。如果企业投资于研发或学习新技术，其生产率在面板期间可能显著改变。FE 估计量在 ω_{it} 时变时仍然不一致。

10.7.2　Olley-Pakes (1996)：用投资作为生产率代理

OP 的核心洞见。企业的**投资** i_{it} 是管理者基于当前生产率 ω_{it} 和当前资本 k_{it} 的决策：

$$i_{it} = f_t(\omega_{it}, k_{it})$$

如果 f_t 在 ω_{it} 上是严格递增的（更高的生产率 $\to$ 更多的投资——这是合理的假设），则可以**反解**出生产率：

$$\omega_{it} = f_t^{-1}(i_{it}, k_{it}) \equiv h_t(i_{it}, k_{it})$$

将 ω_{it} 替换到生产函数中：

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + h_t(i_{it}, k_{it}) + \varepsilon_{it}$$

OP 两阶段估计：

第一阶段——估计 β_l：

$$y_{it} = \beta_l l_{it} + \phi_t(i_{it}, k_{it}) + \varepsilon_{it}$$

其中 $\phi_t(i_{it}, k_{it}) \equiv \beta_k k_{it} + h_t(i_{it}, k_{it})$ 是一个**非参数函数**。在实践中，$\phi_t(\cdot)$ 通常用多项式近似：$\phi_t(i,k) = \sum_p \sum_q \gamma_{pq} i^p k^q$。第一阶段 OLS（或半参数回归）给出 $\hat{\beta}_l$ 的一致估计（因为 l_{it} 和 ϕ_t 之间没有函数依赖——在 OP 假设下）。**注意：** β_k 在**第一阶段没有被估计**——它隐藏在 $\phi_t(\cdot)$ 中，因为 k_{it} 同时出现在 $\beta_k k_{it}$ 和 $h_t(i_{it}, k_{it})$ 中，两者不能在同一阶段中被分离。

第二阶段——估计 β_k： OP 利用了两个关键假设：

(A1) ω_{it} 遵循一阶马尔可夫过程：$\omega_{it} = g(\omega_{i,t-1}) + \xi_{it}$，其中 ξ_{it} 是生产率的 “创新”（innovation）

(A2) 资本在 $t-1$ 期决定（即 k_{it} 不响应 ξ_{it}——资本是 “前定的”，predetermined）

从假设 (A2)，k_{it} 与 ξ_{it} 正交（当期的生产率创新不影响当期的资本——因为资本在前期已确定）：

$$\mathbb{E}[\xi_{it} \mid k_{it}] = 0$$

这是识别 β_k 的矩条件。从第一阶段获得 $\hat{\phi}_t$ 后，用这个矩条件通过 GMM 估计 β_k。

OP 的关键假设：

- **投资的严格单调性：** $i_{it} = f_t(\omega_{it}, k_{it})$ 在 ω_{it} 上严格递增——确保可逆性
- **投资为正：**只有 $i_{it} > 0$ 的观测值进入第一阶段——零投资的企业被排除（可能导致选择偏差——OP 在第二阶段对此进行修正，用企业存续概率模型来调整选择偏差）
- **资本是前定的：** k_{it} 在 $t-1$ 期决定——不响应 ξ_{it}
- **劳动力是 “自由变量”：** l_{it} 可以在当期响应 ω_{it}（包括 ξ_{it}）——这就是为什么 β_l 必须在第一阶段估计（l_{it} 与 ξ_{it} 相关）

10.7.3　Levinsohn-Petrin (2003)：用中间投入替代投资

LP 的动机——OP 的“零投资”问题。 OP 要求 $i_{it} > 0$——投资为零的企业被排除。在许多发展中国家的数据中，相当大比例的企业在给定年份报告零投资（中小企业、信贷约束企业、国有企业——在中国工业企业数据库中尤其常见）。排除零投资企业可能导致严重的**样本选择偏差**（不仅丢失样本量，而且排除的企业是系统性地不同的）。

LP 的替代方案——用中间投入 m_{it} 作为生产率的代理。中间投入（材料成本、电力消耗、原材料采购）几乎**总为正**——极少有企业报告零中间投入。Levinsohn 和 Petrin (2003) 的关键观察是：中间投入需求 m_{it} 也是生产率 ω_{it} 的严格递增函数：

$$m_{it} = f_t(\omega_{it}, k_{it}) \quad \Rightarrow \quad \omega_{it} = f_t^{-1}(m_{it}, k_{it}) \equiv h_t(m_{it}, k_{it})$$

第一阶段：$y_{it} = \beta_l l_{it} + \phi_t(m_{it}, k_{it}) + \varepsilon_{it}$——给出 $\hat{\beta}_l$。第二阶段：用 $\mathbb{E}[\xi_{it} \mid k_{it}] = 0$ 的矩条件 GMM 估计 β_k。

表 10.5: Olley-Pakes (1996) 与 Levinsohn-Petrin (2003) 生产函数估计方法对比

	Olley-Pakes (1996)	**Levinsohn-Petrin (2003)**
代理变量	投资 i_{it}	中间投入 m_{it}
数据要求	投资数据——> 0 的观测	中间投入数据——几乎总 > 0
样本丢失	排除零投资企业（可能 $> 50\%$）	不排除——几乎所有企业使用中间投入
单调性条件	投资在生产率上递增	中间投入在生产率上递增
适用场景	大企业为主（持续有投资）	普适——大企业和小企业
选择偏差修正	第二阶段的 probit 生存模型	一般不包含（若需要，可加）

LP 在 IO 实证中比 OP 更广泛地被采用——因为中间投入数据更容易获得，样本丢失更少，且“投资为正”的限制对中国和印度等发展中国家的企业数据尤其不友好。

10.7.4　Ackerberg-Caves-Frazer (2015)：函数依赖性批判

ACF 的核心批判。 OP 和 LP 的第一阶段：

$$y_{it} = \beta_l l_{it} + \phi_t(\text{proxy}_{it}, k_{it}) + \varepsilon_{it}$$

其中 $\text{proxy}_{it} = i_{it}$（OP）或 m_{it}（LP）。问题在于：如果 l_{it} 也是 ω_{it} 和 k_{it} 的函数——即劳动力需求：$l_{it} = f_t^L(\omega_{it}, k_{it})$——那么 l_{it} 和 $\phi_t(\text{proxy}_{it}, k_{it})$ 之间可能存在**函数依赖性**（functional dependence）：

$$l_{it} = f_t^L(\omega_{it}, k_{it}) = f_t^L(h_t(\text{proxy}_{it}, k_{it}), k_{it}) \equiv g_t(\text{proxy}_{it}, k_{it})$$

如果 $g_t(\cdot)$ 和 $\phi_t(\cdot)$ 有相同的函数形式（如都是 $(\text{proxy}_{it}, k_{it})$ 的多项式），l_{it} 和 ϕ_t 在第一阶段是**完全共线的**——β_l 在第一阶段**不能被估计**。这是 OP/LP 的一个潜在的 "致命" 问题。

ACF 的修正方案：

(1) 第一阶段（"净化" 阶段）：不试图估计 β_l——只用非参数 $\phi_t(\text{proxy}_{it}, k_{it})$ 去除 ε_{it}：

$$\hat{y}_{it} = y_{it} - \hat{\varepsilon}_{it} = \beta_l l_{it} + \beta_k k_{it} + \omega_{it}$$

(2) 第二阶段（GMM）：用**两个**矩条件联合估计 (β_l, β_k)：

$$\mathbb{E}[\xi_{it} \mid k_{it}, l_{i,t-1}] = 0 \quad \text{(生产率创新与当期资本和滞后劳动力正交)}$$
$$\mathbb{E}[\xi_{it} \mid k_{it}] = 0 \quad \text{(生产率创新与当期资本正交)}$$

使用 $\omega_{it} = g(\omega_{i,t-1}) + \xi_{it}$ 和 $\omega_{i,t-1} = \hat{\phi}_{t-1} - \beta_l l_{i,t-1} - \beta_k k_{i,t-1}$（来自第一阶段的净化残差）。

ACF 修正的含义： β_l 不再在第一阶段被估计——第一阶段只 "净化"ε_{it}。β_l 和 β_k 在第二阶段被**联合**估计——需要额外的矩条件（引入 $l_{i,t-1}$ 作为工具变量）。这需要更长的面板数据（至少 $T \geq 3$ 期），但避免了函数依赖性问题。ACF (2015) 之后，大多数 IO 生产函数估计使用 ACF 修正或类似方法。

10.8 生产函数方法 II：加成率估计（De Loecker-Warzynski 2012）

【进阶】

10.8.1 核心公式：$\mu = \theta^V / \alpha^V$

De Loecker-Warzynski 的核心洞见。 De Loecker & Warzynski (2012, *AER*) 提出了一个极其简洁的加成率（markup）公式——这是 NEIO 生产函数分支的奠基性贡献：

$$\boxed{\mu_{it} = \frac{\theta_{it}^V}{\alpha_{it}^V}}$$

其中：

- $\mu_{it} = P_{it}/\text{MC}_{it}$ ——加成率（markup）
- $\theta_{it}^V = \frac{\partial F}{\partial V_{it}} \cdot \frac{V_{it}}{Q_{it}}$ ——可变投入 V 的**产出弹性**（output elasticity）——来自生产函数估计

- $\alpha_{it}^V = \frac{w_{it}^V \cdot V_{it}}{P_{it} \cdot Q_{it}}$ ——可变投入 V 的**收入份额**（revenue share）——直接来自数据（可观测）
- $w_{it}^V =$ 可变投入 V 的价格

经济学直觉：

- 在完全竞争下（$P = \text{MC}$），企业雇用投入品直到其边际产出价值等于其价格：$\partial F/\partial V \cdot P = w^V$。这意味着 $\theta^V = \alpha^V$，即 $\mu = 1$（零加成）。
- 在市场势力下（$P > \text{MC}$），$\partial F/\partial V \cdot P > w^V$——因为边际产出价值的收益超过了投入品的成本。这意味着 $\theta^V > \alpha^V$，即 $\mu > 1$（正加成）。
- μ **不依赖于需求函数形式或竞争模式假设——仅依赖于成本最小化。**这是 De Loecker-Warzynski 公式的 "魔力"——它从标准的成本最小化一阶条件中推导出来，不需要假设特定的需求系统或特定的行为模式。

推导（成本最小化）：企业最小化成本：$\min_V w^V V$ s.t. $F(V,K) \geq \bar{Q}$。拉格朗日函数：$\mathcal{L} = w^V V + \lambda(\bar{Q} - F(V,K))$。一阶条件：$w^V = \lambda \cdot \partial F/\partial V$。但 $\lambda = \partial \text{Cost}/\partial Q = \text{MC}$（拉格朗日乘子是边际成本——包络定理）。因此：$\partial F/\partial V = w^V/\text{MC}$。两边乘以 V/Q：$\underbrace{(\partial F/\partial V \cdot V/Q)}_{\theta^V} = w^V V/(\text{MC} \cdot Q)$。除以 $\alpha^V = w^V V/(P \cdot Q)$：$\theta^V/\alpha^V = P/\text{MC} = \mu$。

可变投入 V 的选择——为什么用中间投入 M 而非劳动力 L？

- **中间投入 M（推荐）：**材料/能源投入通常是 "无摩擦" 的——没有调整成本，静态 FOC 成立。企业可以在当期自由调整 M 以响应生产率冲击。
- **劳动力 L（不推荐）：**劳动力可能有调整成本（雇佣/解雇成本 + 劳动合同保护 + 培训成本）。如果劳动力有调整成本，静态 FOC 不成立——θ^L/α^L 是**有偏的**加成率度量（通常会高估 μ，因为调整成本使实际劳动力投入低于静态最优）。
- **资本 K（不适用）：**资本是动态变量（调整成本 + 沉没性）——资本不是 "可变" 投入——不能用这个公式。

10.8.2　实施细节与挑战

产出弹性 θ^V 的估计：

- 需要一致的生产函数估计（OP/LP/ACF——§~10.7）来获得 $\hat{\theta}^V$
- 产出弹性可能是**企业特定的**（随企业规模、技术、所有制变化）——在实践中通常假设为行业-年份常数（θ_j^V）或使用企业特定估计（但需要更多的参数假设）
- ACF (2015) 方法是当前的标准做法——但许多应用仍使用 LP (2003)，特别是在面板较短时

收入份额 α^V 的计算：

- α^V = 材料成本/总收入——在理想情况下，材料成本直接来自企业的成本会计记录
- **关键实践问题：**需要区分 "材料成本" 和 "其他成本"（劳动力、资本折旧、间接费用）——在企业层面的会计数据中，这种区分可能模糊或标准不统一
- 许多研究使用 "销售成本"（Cost of Goods Sold, COGS）作为材料成本的代理——但 COGS 可能包含劳动力成本（直接劳动力）和折旧（制造费用）

价格平减问题（De Loecker-Eeckhout-Unger 2020 的关键挑战）：

- 生产函数估计需要**数量**数据——但通常只能观测到**收入**（$P_{it} \cdot Q_{it}$）和**支出**（$w_{it}^V \cdot V_{it}$）
- 实践中：使用行业层面的价格平减指数（PPI）从收入中 "提取" 数量：$\hat{Q}_{it} = \text{Revenue}_{it}/\text{PPI}_{jt}$
- **但：**如果企业层面的价格与行业平均价格**不同**（因为企业有市场势力——$\mu_{it} > 1 \to P_{it} > \bar{P}_{jt}$），行业平减会**系统性地低估**企业层面的产出数量 $\to$ 高估生产率 $\omega_{it} \to$ 高估产出弹性 $\theta^V \to$ **高估加成率** μ
- 这是 De Loecker 方法面临的最大方法论挑战——如何在不需要企业级价格数据的情况下区分 "生产率" 和 "加成率"

10.8.3 De Loecker-Eeckhout-Unger (2020)：全球加成率上升趋势

核心发现（De Loecker, Eeckhout & Unger 2020, *QJE*）：

(1) 全球加成率自 1980 年以来显著上升。美国：平均加成率从 1980 年的 ≈ 1.21（P 比 MC 高 21%）上升到 2016 年的 ≈ 1.61（P 比 MC 高 61%）。全球趋势相似——但幅度和时机在不同地区有差异。

(2) 上升主要集中在 "超级明星企业"（superstar firms）。这不是 "所有企业都在提价"——而是**高加成率的少数企业获得了越来越大的市场份额**。加成率分布的上尾在 "拉长"——极高加成率的企业越来越重要。这与 Autor, Dorn, Katz, Patterson & Van Reenen (2020) 的 "超级明星企业" 命题一致：高效率/高加成企业获得更大市场份额。

(3) 高加成企业集中在：无形资产密集型行业（科技、制药、媒体）、数字密集型行业（软件、平台、电子商务）、"赢家通吃" 市场（社交网络、搜索引擎、操作系统）。

(4) **宏观含义**：平均加成率 ↑ ↔ 劳动收入份额 ↓（劳动在国民收入中的占比下降）。平均加成率 ↑ ↔ 总投资率 ↓（高加成企业可能投资不足——因为市场势力减少了产出的动机）。这为 Piketty 的 $r > g$ 假说提供了一个微观基础——市场势力的上升可能是资本回报上升的一个渠道。

争议与批判：

- **价格平减问题**（如在 §~10.8.2 中讨论）——如果企业级价格偏离行业平减指数，加成率趋势可能部分（或完全）是 "测量误差趋势"
- **生产函数形式**——Cobb-Douglas 假设可能限制了更具异质性的产出弹性的估计（如 translog 生产函数——但 translog 需要的参数更多，估计难度更大）
- **中国数据的特殊性**——上市公司数据（CSMAR, Wind）在中国可能不代表整体经济——大量非上市国企和中小企业的加成率趋势可能与上市企业不同

10.9　结构估计的方法论反思

【进阶】

10.9.1　"结构" 到底意味着什么？

结构参数的定义（Reiss & Wolak 2007, *Handbook of Econometrics*, Ch.64）：

> 一个参数是**结构的**（structural）如果它对一类政策干预是**不变的**（invariant）。即当经济环境（政策、规制、技术）发生变化时，结构参数本身不改变——它捕捉的是经济行为者的 "深层" 特征（偏好、技术、信息结构）。

对比——简约式（Reduced-Form）参数：

- 简约式参数（如 DiD 的 "处理效应"、IV 的 Wald 估计量、RD 的局部效应）捕捉的是特定政策在特定环境下的**局部平均处理效应**（LATE）
- 当政策环境变化时，简约式参数可能改变——因为它不仅反映了行为特征，还反映了**实施环境**（如处理强度、政策细节、受影响的子群体）
- 例子：Card & Krueger (1994) 的最低工资 DiD——估计的是 1992 年 NJ 最低工资从 \$4.25 提高到 \$5.05 的就业效应——**在 NJ/PA 在 1992 年的特定劳动力市场紧缩度、行业构成和竞争条件下**。如果最低工资从 \$15 提高到 \$20，在不同城市的劳动力市场条件下，我们没有理论依据认为相同的处理效应适用

结构参数的优势：结构参数（如需求弹性 η、边际成本 MC、行为参数 θ）在政策变化下是（近似）不变的——因此可以用于**反事实模拟：**“如果两家航空公司合并，价格会上涨多少？”“如果实施碳税，高耗能行业的产出和排放会如何变化？”这些问题不能仅靠简约式估计来回答——因为它们涉及的是尚未发生的政策变化。

Reiss-Wolak 的“结构估计四层次”：

层次 1：描述统计：相关性和数据描述——不是结构性的

层次 2：简约式估计：因果效应估计（DiD, IV, RD, RCT）——最低层次的结构性（因果参数，但依赖于环境和工具变量）

层次 3：“半结构”估计：使用经济理论来约束估计方程（Bresnahan 的 FOC——使用理论推导估计方程，但不一定估计所有结构原语）

层次 4：全结构估计：估计所有经济原语（偏好参数、技术参数、市场结构参数）——可以进行任何反事实模拟（BLP 1995 是层次 4——Ch.11）

Bresnahan 框架位于层次 2–3 的边界——θ 是结构参数（在政策变化下不变），但 Bresnahan 不估计完整的博弈（如没有估计进入决策——那是 Ch.12 的动态结构模型的任务）。

10.9.2 结构估计的优势与代价

优势：

(1) 反事实模拟（Counterfactual Simulation）——结构估计的“杀手应用”。结构估计最重要、最有价值的应用是进行反事实政策模拟：合并审查、税收/补贴政策评估、规制改革影响预测。这些是简约式方法无法完成的（因为简约式参数在反事实政策下不一定不变）。

(2) 福利计算（Welfare Calculation）。结构参数直接产生福利分析所需的参数：需求弹性 $\rightarrow$ 消费者剩余变化；MC $\rightarrow$ 生产者剩余变化；$\theta \rightarrow$ 市场势力导致的 deadweight loss。简约式方法通常不能产生福利计算——因为福利依赖于反事实（“如果价格没有被提高，消费者剩余会是多少？”），而反事实需要结构模型。

(3) 嵌套检验（Nested Tests）。结构框架将不同行为假设（竞争、Cournot、合谋）嵌套为参数约束——允许正式统计检验。简约式方法通常不能区分不同的行为假设（它只能告诉你“X 是否影响 Y”，而不能告诉你“企业遵循哪条 FOC”）。

(4) 外部有效性（External Validity）。结构参数在政策变化下不变 $\rightarrow$ 可以在不同政策场景下使用相同的估计参数进行预测。简约式估计的“局部”性质限制了其在其他环境和政策下的适用性。

代价：

(1) 函数形式假设（Functional Form Assumptions）。需求函数的精确形式？MC 函数的精确形式？如果函数形式被错误指定 → 所有结构参数都是不一致的。结构估计的 “强项”（结构参数）同时也是其 “弱项”——错误指定的结构模型产生的参数不是 “结构” 的。

(2) 计算复杂度（Computational Burden）。结构估计涉及非线性优化、数值积分、模拟方法（BLP 的 contraction mapping、Ch.12 的 nested fixed point）。计算时间可能以小时/天为单位 → 限制了研究者可以探索的模型规格数量。计算复杂度也是 “黑箱” 问题的来源之一。

(3) 强排除限制（Strong Exclusion Restrictions）。结构估计依赖于严格的排除限制（R 不进入 MC、W 不进入需求）。如果这些排除限制被违反 → 所有参数都是不一致的。这比简约式 IV 的排除限制更强——因为结构估计需要**多组**排除限制，而不仅仅是 “一个工具变量”。

(4) 透明性（Transparency）。结构估计的结果可能对函数形式、优化算法、数值积分精度、起始值等做选择的每一环高度敏感。评审和读者难以评估是否存在 “选择的自由度”（specification search）——研究者试了多个规格，只报告最 “理想” 的结果。

Blundell (2017) 的平衡观点：

> 结构估计和简约式方法不是替代品——它们是互补的。简约式方法在回答 “是否” 问题（X 是否导致 Y?）上更强——因为需要的假设更少。结构方法在回答 “多少” 问题（X 导致了多少 Y? 如果政策变成 Z，Y 会是多少?）上更强——因为需要结构参数来进行反事实计算。最好的实证 IO 研究通常包含两者：简约式方法建立因果关系，结构方法量化福利和进行反事实模拟。

10.9.3　GMM、嵌套/非嵌套检验与模型选择

NEIO 中的 GMM 估计（Hansen 1982）：
NEIO 结构模型的标准估计方法是广义矩方法（GMM）：

(1) 矩条件： $\mathbb{E}[m(\mathbf{Z};\boldsymbol{\psi})] = 0$——其中 $\mathbf{Z}$ 是工具变量，$\boldsymbol{\psi}$ 是结构参数向量（包括 θ、需求参数、MC 参数）

(2) 加权矩阵：初始估计用单位矩阵 → 获得一致初始参数；第二步用残差计算最优加权矩阵（Hansen 1982 的有效估计量）

(3) 过度识别检验：如果工具变量数 $m >$ 参数数 k，Hansen's J 统计量 $\sim \chi^2(m-k)$ ——大 J 拒绝工具变量有效性

嵌套 vs. 非嵌套检验的应用准则：如果在 Bresnahan 框架内比较不同的竞争假设（$\theta = 0$ vs. $\theta = 1/n$ vs. $\theta = 1$），用嵌套检验（Wald/LR/LM）。如果比较不同需求函数形式（线性 vs. 对数线性 vs. AIDS ——这些不是通过参数约束从彼此导出的），用 Vuong (1989) 非嵌套检验或信息准则（AIC, BIC）。

"最佳实践" 模型选择流程：

(1) 估计最灵活的模型（θ 自由）

(2) 检验 θ 是否等于特殊值（$0, 1/n, 1$）

(3) 在不同需求/MC 规格下重新估计——检查 $\hat{\theta}$ 的稳定性

(4) 对每一规格，计算 BIC——选择 BIC 最低的模型

(5) 报告所有规格下的 $\hat{\theta}$ 范围（而不仅仅是 "最偏好规格" 的 $\hat{\theta}$）——这是透明性的关键

10.9.4 Corts 批判的再审视与 "MC 弹性" 问题

Corts 批判的一般化：任何基于 "P – MC" 来推断市场势力的方法——无论是 Bresnahan 的 θ 还是 De Loecker 的 μ——都面临同一个问题：**MC 是未观测的**。如果 MC 的估计是有偏的，"P – MC" 的推断就是有偏的。这个问题的根本性意味着它永远不会被完全 "解决"——只能通过多种方法的**互补**来缓解。这就是为什么 Genesove-Mullin (1998) 的 "直接成本数据" 方法如此重要——它是唯一完全绕开 MC 估计问题的方法。

结构方法对 Corts 批判的四条回应路径：

(1) 生产函数方法（§~10.7, §~10.8）：不依赖 MC 函数形式——从生产函数和投入份额中推断 μ。"绕过" 了 MC 估计问题——但引入了新的挑战（价格平减问题）。

(2) 直接成本数据（Genesove-Mullin 路径）："黄金标准"——当 MC 是数据而非估计时，Bresnahan 框架运作完美。适用条件极其有限——仅在极少数行业中有如此详尽的成本会计档案。

(3) Corts 稳健性检验：在所有合理的 MC 规格下重新估计 θ——如果 $\hat{\theta}$ 的变异范围小，结论可信；如果变异范围大，结论不可靠。报告 "$\hat{\theta}$ 的范围" 比报告 "一个点估计" 更诚实地反映了 MC 的不确定性。

(4) 半参数/非参数方法：尽可能少地假设 MC 函数形式——用灵活的基函数（样条、核回归）来近似 MC——减少对函数形式假设的依赖。代价是估计精度降低（更灵活的函数形式需要更多的参数——参数精度下降）。

10.10　中国语境：NEIO 在中国的应用与挑战

【核心】

中国市场与 NEIO 中国市场对 NEIO 方法既是机遇也是挑战。机遇在于：中国拥有世界上最大规模的企业级面板数据（中国工业企业数据库 1998–2013，覆盖年销售额 $>$ 500 万元的所有非农工业企业）——这为生产函数方法（OP/LP/ACF + De Loecker-Warzynski）提供了无与伦比的数据基础。挑战在于：中国市场的制度特征——国企、政府干预、地方保护主义——可能违反 NEIO 标准模型的关键假设。

10.10.1　中国的 NEIO 实证研究概览

汽车行业——Lu & Yu (2015)。使用中国加入 WTO（2001）作为“竞争增强”的外生冲击。关税下降 $\rightarrow$ 进口竞争增强 $\rightarrow$ 国内企业的加成率下降。方法：De Loecker-Warzynski 生产函数加成率估计 + 双重差分。关键发现：贸易自由化减少了加成率的**分散度**——最有效率的国内企业在保护主义下获得的超额加成率被进口竞争侵蚀。这为贸易自由化的“促进竞争效应”（pro-competitive effect）提供了中国证据。

制造业加成率——Brandt, Van Biesebroeck & Zhang (2012)。使用中国工业企业数据库（1998–2007）估计中国制造业企业的 TFP 和加成率。方法：OP/LP 生产函数估计 + De Loecker 加成率计算。关键发现：中国制造业的加成率在 1998–2007 年间**总体下降**——与竞争增强一致（WTO 后的进口竞争 + 国内市场化改革）。但国企的加成率仍然显著高于非国企——这可能反映：市场保护（国企在有进入壁垒的行业享有“政策租金”）和/或效率差异。

银行业——Chen & Zhou (2009) 等。使用 Panzar-Rosse (1987) 模型或 Bresnahan 框架估计中国银行业的竞争程度。四大国有银行（工、农、中、建）$\rightarrow$ 股份制银行（招商、民生等）$\rightarrow$ 城市商业银行的差异化竞争。发现：中国银行业总体处于“垄断竞争”——行为参数 $\hat{\theta}$ 在 0.2–0.4 之间（低于 Cournot 的 $1/n$，高于完全竞争的 0）。这与银行业的产品差异化（不同银行有不同网点的覆盖范围和服务质量差异）一致。

其他行业：

- 水泥（Ping & Wei 2001）：用 Bresnahan 框架——运输距离作为需求旋转变量（运输距离改变了不同水泥厂之间的替代弹性）
- 中国电信（2008 重组前后）：结构估计评估“3 变 3”重组对竞争的影响——2008 年中国电信业从 6 家重组为 3 家全业务运营商
- 航空（北京–上海、上海–广州等主要航线）：用 NEIO 模型估计不同航线的竞争程度

10.10.2 中国市场对 NEIO 方法的独特挑战

挑战 1——国企的非利润最大化行为。 Bresnahan 框架假设企业最大化利润 → FOC 成立。中国国企可能有其他目标：就业最大化、产业政策（“做大做强”）、“政策性负担”（回顾 第 ~9~ 章 §~9.9）。如果国企的真实 FOC 是：

$$P + \theta \cdot P'(Q) \cdot q_i + \gamma \cdot (\text{其他目标}) = \text{MC}$$

而研究者错误地估计为 $P + \theta \cdot P'(Q) \cdot Q = \text{MC}$（忽略了 γ），那么 $\hat{\theta}$ 吸收了 γ 的效应——$\hat{\theta}$ 是 “市场势力 + 非利润目标” 的混合物，不能纯粹地解释为行为参数。

挑战 2——政府干预的价格扭曲。某些行业的价格受政府指导或管制（电价、水价、铁路运价、部分药品价格）。如果价格不是由市场决定的 → 企业 FOC 不成立 → Bresnahan 框架的整个推导基础丧失。研究者在应用 NEIO 方法之前必须首先确认：该行业的价格是否由市场决定？如果是管制价格，管制价格与企业 MC 的关系是什么？管制是否约束了企业的定价（binding）还是仅是名义上的？

挑战 3——地方保护主义与市场分割。中国市场在地方层面可能是分割的——省级/市级的保护主义（“省际贸易壁垒”）限制了商品和服务的自由流动。这意味着：

- 市场的相关地理定义不是 “全国” 而是 “省级” 或 “城市群”——需要为每个地方市场分别估计需求和供给
- 数据需求剧增——需要每个地方市场有足够长的时间序列来识别 θ（大多数中国公开数据没有省/市级的高频时间序列）
- 如果忽略了市场分割，估计的 “全国需求弹性” 是多个分割市场弹性的加权平均——可能严重扭曲推断

挑战 4——数据可得性与质量。

- **中国工业企业数据库**（1998–2013）：主力数据库——但 2010–2013 年数据质量显著下降（匹配、清理工作繁重，部分企业 ID 不可追踪，关键变量缺失或明显错误）
- **上市公司数据**（CSMAR, Wind）：覆盖面有限（大企业为主——$\approx$ 4000 家 vs. 全国 $>$ 100 万家工业企业），但数据质量较高且经审计
- **价格平减指数**：企业/产品级价格数据在中国极少可得——大多数研究使用行业级 PPI 进行平减（参见 §~10.8.2 的价格平减问题）
- **所有制分类**：国企、集体、民营、外资、港澳台——所有制分类随时间变化（改制、重组、混合所有制改革）——需要仔细跟踪企业层面的所有制变化

10.10.3　中国 NEIO 研究的最佳实践建议

(1) 结合制度知识。在建模之前，深入理解行业的具体规制和所有权结构——特别是国企的角色、政府干预的程度、市场的实际地理范围。这是 NEIO“行业深度研究” 传统的中国版本。

(2) 多重识别策略。不要仅依赖一种方法——同时使用 Bresnahan 框架、生产函数方法、简约式方法——寻找 “汇聚” 的结果。如果不同方法给出了一致的结论，结论更可信；如果不一致，需要解释为什么。

(3) 数据清理的透明性。详细记录数据清理步骤——中国工业数据的清理可能是研究中最耗时的步骤（占整个研究时间的 50%+），且清理决策会影响估计结果。关键清理步骤：企业 ID 的跨年一致性、极端值的 winsorize 规则、行业代码的 harmonize（GB/T 4754 在 2002 年和 2011 年修订）。

(4) 稳健性第一。在多个 MC 规格、多个需求函数形式、多个工具变量集下重新估计——如果结果对这些选择不敏感 $\rightarrow$ 结论可信。按 §~10.6.3 的 Corts 传统报告 “$\hat{\theta}$ 的范围”。

(5) 按所有制分组。在估计中允许 θ 和 MC 随所有制类型（国企 vs. 非国企 vs. 外资）变化——这可能是中国 NEIO 研究最自然的异质性维度。国企和非国企可能有不同的目标函数、面临不同的约束、处于不同的竞争环境——将这些混在一起估计一个 “平均 θ” 是没有意义的。

10.11　章末总结与通往 BLP 的桥梁

10.11.1　核心概念回顾

本章建立了 NEIO 的方法论基础——从 SCP 的失败到结构估计的革命。以下是你在继续学习之前应该掌握的 10 个核心概念：

1. **结构估计（Structural Estimation）。**从经济理论的一阶条件出发推导估计方程——估计的是在政策变化下不变的 “深层” 参数（需求弹性、MC、θ）。区别于简约式方法——后者估计的是特定政策在特定环境下的局部因果效应。

2. **行为参数 θ（Conduct Parameter）。**企业 FOC 为 $P + \theta \cdot P'(Q) \cdot Q = \text{MC}$。$\theta = 0$（完全竞争/Bertrand 同质）$\rightarrow$ $\theta = 1/n$（对称 Cournot）$\rightarrow$ $\theta = 1$（完全合谋）。嵌套性质是 NEIO 最关键的技术优势——使竞争模式成为可检验的统计假设。

3. **需求旋转（Demand Rotation）。**识别 θ 需要一个外生变量不仅移动需求截距，还移动需求**斜率**（$b_1 \neq 0$）。截距变化只能追踪供给关系（MC 曲线）；斜率变化才能区分 θ——因为在不同 θ 下，企业对需求斜率变化的反应不同。

4. **Regime-Switching（Porter 1983）。**允许 θ 在 "合谋 regime" 和 "价格战 regime" 之间切换——由不可观测的马尔可夫链驱动。价格战被验证为合谋的**均衡特征**（Green-Porter 不完全监测模型），而非合谋的 "失败"。

5. **直接成本数据（Genesove-Mullin 1998）。**当 MC 是**数据**而非**待估计参数**时，Bresnahan 框架运作完美——$\hat{\theta}$ 与历史学家的行为分类完全一致。这是 NEIO 的 "黄金标准" 验证——证明了理论框架在理想数据条件下的有效性。

6. **生产函数估计（OP 1996, LP 2003, ACF 2015）。**用投资/中间投入作为不可观测生产率 ω_{it} 的代理变量——解决同时性偏差。两阶段估计：第一阶段半参数回归去除 ε_{it}；第二阶段 GMM 用矩条件估计 β_k（和 *在 ACF 中也在第二阶段估计* β_l）。

7. **加成率估计（De Loecker-Warzynski 2012）。** $\mu_{it} = \theta_{it}^V / \alpha_{it}^V$——加成率等于可变投入的产出弹性除以其收入份额。从成本最小化推导——不需要需求系统。完全竞争 $\to \mu = 1$；市场势力 $\to \mu > 1$。

8. **Corts 批判（Corts 1999）。**如果真实 MC 随 "旋转变量"R 变化，$\hat{\theta}$ 会吸收 MC 的变异 $\to \hat{\theta}$ 是有偏的 "行为参数 + MC 误设" 混合物。回应：(a) Corts 稳健性检验，(b) 直接成本数据，(c) 生产函数方法。

9. **GMM 与模型选择。**嵌套假设（$\theta = 0, \theta = 1/n, \theta = 1$）用 Wald/LR/LM 检验。非嵌套模型（不同需求函数形式）用 Vuong (1989) 检验或信息准则（BIC）。

10. **中国 NEIO 的特殊挑战。**国企的非利润最大化行为、政府价格干预、地方保护主义的市场分割、数据质量问题是 NEIO 在中国应用时面临的四大方法论挑战。

10.11.2 核心公式回顾

(1) $P + \theta \cdot P'(Q) \cdot Q = \text{MC}$（企业 FOC——行为参数 θ 的统一框架）

(2) $\theta = n \cdot \frac{\partial(\text{供给关系有效斜率})/\partial R}{b_1}$（$\theta$ 的识别——在需求旋转变量 R 下）

(3) $\mu_{it} = \frac{\theta_{it}^V}{\alpha_{it}^V}$（De Loecker-Warzynski 加成率公式——成本最小化推论）

(4) $y_{it} = \beta_l l_{it} + \phi_t(\text{proxy}_{it}, k_{it}) + \varepsilon_{it}$（OP/LP 第一阶段估计方程）

(5) $\delta \geq \delta^* = \frac{\pi^{\text{dev}} - \pi^{\text{coll}}}{\pi^{\text{dev}} - \pi^{\text{Nash}}}$（结合 第 ~8~ 章——合谋可持续性——Porter (1983) regime-switching 检测的博弈论基础）

10.11.3　本章的方法论框架：NEIO 结构估计四步法

与 第 ~8~ 章的 “Tirole 七步建模法” 和 第 ~9~ 章的 “机制设计五步法” 并列，本章建立了 NEIO 的*结构估计四步法*：

Step 1　写出理论模型：需求系统 + 企业 FOC（供给关系）——将不可观测的行为参数 θ 连接于可观测的 P, Q, Z

Step 2　推导估计方程：将供给关系转化为可估计形式——参数化需求和 MC——识别 θ, b, MC 参数的统计关系

Step 3　解决识别问题：寻找需求旋转变量 R（或直接 MC 数据、或生产函数方法）——验证 $b_1 \neq 0$（需求旋转）和排除限制

Step 4　估计与检验：GMM/ML 估计 $\hat{\theta} \to$ Wald/LR 检验 $H_0 : \theta = 0, 1/n, 1 \to$ 非嵌套检验/Vuong $\to$ Corts 稳健性检验

10.11.4　通往 Ch.11 的桥梁

从同质产品到差异化产品：为什么我们需要 BLP

Ch.10 讲述了 NEIO 的方法论革命——如何从市场数据中识别市场势力。Bresnahan 纲领为同质产品行业的实证 IO 提供了系统框架。但它的根本局限是**同质产品假设**——对于汽车、麦片、手机、电脑等差异化产品行业，我们需要一个完全不同的方法。

在差异化产品市场中：(1) 有多个产品、多个价格——“市场” 的定义本身就是模糊的；(2) 产品之间的替代弹性因产品特征（马力、重量、品牌、操作系统…）而异——需要估计一个**完整的需求系统**；(3) 产品价格与产品特征（包括未观测的特征——质量、品牌资产）内相关——需要新的工具变量策略。

BLP (Berry, Levinsohn & Pakes 1995) 是过去三十年 IO 实证研究引用最多的方法论贡献。它使用随机系数离散选择模型来估计差异化产品市场中的需求弹性——并在反垄断、贸易、环境、健康 IO 中得到了极其广泛的应用。

第 ~11~ 章将展示 BLP 方法的核心逻辑：从随机效用模型 $\to$ 市场份额的逆映射（contraction mapping 算法）$\to$ 价格内生性的 BLP 工具变量（其他产品的特征作为价格的工具变量）$\to$ 供应侧（纳什 Bertrand FOC）$\to$ 合并模拟和福利计算。

习题

基础题

Ex 10.1 —行为参数 θ 的计算。考虑一个同质产品行业，$n = 3$ 家企业。需求为 $P = 100 - Q$。边际成本为常数 $\mathrm{MC} = 10$。

(a) 在 Cournot 均衡下，计算均衡价格 P^*、每家企业产量 q^*、以及 Lerner Index $(P - \mathrm{MC})/P$。

(b) 使用企业 FOC（$P + \theta \cdot P'(Q) \cdot q_i = \mathrm{MC}_i$），从 (a) 中解出 θ。验证在 Cournot 下 $\theta = 1/n = 1/3$。

(c) 现在假设企业合谋（联合利润最大化）。重新计算 P^* 和 θ。验证 $\theta = 1$。

(d) 假设 $n = 2$（双寡头）且产品差异化——企业 i 的需求为 $q_i = 50 - 2p_i + p_j$。在 Bertrand 均衡下（每家企业选择自己的价格），计算均衡价格和 θ。注意：在差异化 Bertrand 中 θ 不一定为 0——解释为什么。

Ex 10.2 —识别问题的直观理解。需求 $P = a - bQ + \varepsilon$，$\mathrm{MC} = c$（常数）。n 家对称企业，行为参数 θ。

(a) 推导均衡 P^* 和 Q^* 作为 a, b, c, n, θ 和 ε 的函数。

(b) 计量经济学家观察到 (Q_t, P_t) 的 T 个观测值，假设仅 ε_t 变化（a, b, c, n, θ 均为常数）。画出散点图——这是何种曲线？识别了什么？

(c) 现在假设 $b_t = b_0 + b_1 R_t$，其中 R_t 是外生变量。如果 $b_1 \neq 0$，为什么 θ 可以被识别？用直觉和图形来说明。

(d) 如果 $b_1 = 0$（无需求旋转），证明无论数据多好，θ 都无法被识别。

Ex 10.3 —供给关系估计的 IV 方法。考虑供给关系 $P_t = c + \theta \cdot b \cdot Q_t + \nu_t$，其中 ν_t 是成本冲击。

(a) 推导 OLS 估计量的概率极限：$\operatorname{plim} \hat{\theta}_{\mathrm{OLS}} = \theta + \operatorname{Cov}(Q, \nu)/\operatorname{Var}(Q)$。$\operatorname{Cov}(Q, \nu)$ 是正还是负？$\hat{\theta}_{\mathrm{OLS}}$ 是向上偏还是向下偏？

(b) 假设有一个外生变量 R_t（如五大湖通航状态），R_t 进入需求但不进入 MC。解释为什么 R_t 是 Q_t 的有效工具变量。

(c) 写出 IV 估计量 $\hat{\theta}_{\mathrm{IV}}$。为什么这个估计量是一致的？

(d) 如果你错误地使用了 R_t，而 R_t 实际上也进入 MC 函数（Corts 批判），$\hat{\theta}_{\mathrm{IV}}$ 会有什么样的偏误？

Ex 10.4 —De Loecker-Warzynski 加成率公式的推导。企业的生产函数为 $Q = A \cdot L^{\beta_l} \cdot M^{\beta_m} \cdot K^{\beta_k}$。

(a) 推导可变投入 M（中间投入）的产出弹性 θ^M 和收入份额 α^M 的表达式。

(b) 使用成本最小化，证明 $\mu = \beta_m / \alpha^M$。

(c) 如果 $\mu = 1.5$（P 比 MC 高 50%），且 $\alpha^M = 0.4$（材料成本占总收入的 40%），β_m 应该等于多少？

(d) 为什么不能用劳动力 L 替代中间投入 M 作为 "可变投入" 来计算加成率？

进阶题

Ex 10.5 —Porter (1983) Regime-Switching 模型的模拟。模拟一个在 "合谋 regime" 和 "价格战 regime" 之间切换的行业。

(a) 设定参数：$n = 3, P = 100 - Q, \text{MC} = 10$。在合谋期，$\theta^C = 0.9$；在价格战期，$\theta^W = 0.2$。Regime 切换由两状态 Markov 链驱动：$\Pr(C \to W) = 0.2, \Pr(W \to C) = 0.5$。

(b) 模拟 $T = 100$ 期的 P_t 和 Q_t，加入随机需求/成本冲击（$\varepsilon_t \sim N(0, 2^2), \nu_t \sim N(0, 1^2)$）。

(c)（计算题）使用模拟数据估计 "静态"Bresnahan 模型（忽略 regime-switching），即假设 θ 在样本期内为常数。$\hat{\theta}$ 是多少？与真实值 $\theta^C = 0.9$ 和 $\theta^W = 0.2$ 相比如何？

(d)（讨论题）如果你在真实数据中做 (c) 的 "静态" 估计，你会得出什么错误的结论？

Ex 10.6 —生产函数估计中的函数依赖性问题。考虑 OP/LP 的第一阶段：$y_{it} = \beta_l l_{it} + \phi_t(\text{proxy}_{it}, k_{it}) + \varepsilon_{it}$。

(a) 解释为什么 OLS 估计 β_l 是有偏的（向上的偏误方向）。

(b) OP 方法使用投资作为 ω 的代理变量；LP 方法使用中间投入。在什么条件下投资是更好的代理变量？什么条件下中间投入是更好的？

(c) ACF (2015) 指出 OP/LP 第一阶段可能由于函数依赖性问题导致 β_l 不能被估计。用直觉解释这个问题——并解释 ACF 如何通过将 β_l 的估计移到第二阶段来避免它。

(d) 如果在 OP 第一阶段中，你使用一个灵活的多项式 $\phi(i,k)$ 来近似非参数函数，多项式的 “阶数”（二次 vs. 三次）如何影响第一阶段估计？阶数选择的权衡是什么？

Ex 10.7 —Corts (1999) 批判与结构估计的稳健性。真实 MC 为 $\text{MC} = c_0 + c_1 R + \nu$，但研究者错误地假设 $\text{MC} = c_0 + \nu$。

(a) 构建数值例子：真实 $\theta = 0.5$。需求为 $P = a_0 + a_1 R + (b_0 + b_1 R)Q + \varepsilon$。推导误设下 $\hat{\theta}$ 的概率极限。$\hat{\theta}$ 是向上偏还是向下偏？

(b) 偏误的方向和大小与 c_1 和 b_1 的符号和大小的关系是什么？在什么条件下，Corts 偏误最为严重？

(c)（讨论题）假设你怀疑 R 可能影响 MC。你会做什么来检查 $\hat{\theta}$ 对 Corts 批判的稳健性？设计一个具体的稳健性检验策略。

应用/实证题

Ex 10.8 —中国汽车行业的市场势力演变。使用 Lu & Yu (2015) 的发现。

(a) 解释为什么中国加入 WTO（2001）可以作为 “竞争增强” 的外生冲击。关税下降是如何影响国内汽车企业的加成率的？

(b)（计算题）假设 WTO 前（2000）：$\alpha^M = 0.45, \hat{\theta}^M = 0.55$。WTO 后（2005）：$\alpha^M = 0.48, \hat{\theta}^M = 0.52$。计算两个时点的加成率 μ。WTO 后加成率是上升还是下降？

(c)（讨论题）如果 WTO 后中国汽车企业的产品质量提高了，这对 De Loecker-Warzynski 方法的 “价格平减问题” 意味着什么？

Ex 10.9 —Bresnahan 框架的排除限制评估。选择两个中国行业（如水泥和电信），评估 NEIO 方法的适用性。

(a) 对于水泥行业：识别一个可能的 “需求旋转变量”（提示：考虑建筑业活动的周期性和运输距离的角色）。论证它为什么满足（或不满足）排除限制。

(b) 对于电信行业：中国电信价格在 1994–2008 年间受到政府价格指导。讨论这种价格干预如何影响 Bresnahan 框架的适用性。

(c) 对于这两个行业，你会推荐 Bresnahan 框架还是生产函数方法（De Loecker-Warzynski）？为什么？

Ex 10.10 —中国制造业加成率的数据分析。使用 De Loecker-Eeckhout-Unger (2020) 的公开数据集或中国工业企业数据库的公开版本。

(a) 选择一个两位数制造业行业（如食品加工、纺织、化学、机械、电子）。计算该行业在样本期内的平均加成率 μ 的趋势（使用简化 De Loecker-Warzynski 公式——用收入份额近似 μ）。

(b) 比较国企和非国企的平均加成率。国企的加成率是更高还是更低？这与你的预期一致吗？

(c)（讨论题）中国 2015 年的 "供给侧结构性改革" 宣布要化解过剩产能。如果你在 2015 年前后估计钢铁和水泥行业的加成率，你预期会发现什么变化？为什么？

推荐阅读

核心必读

[1] **Bresnahan, T. F. (1989).** "Empirical Studies of Industries with Market Power." In R. Schmalensee & R. D. Willig (eds.), *Handbook of Industrial Organization*, Vol. 2, North-Holland, Ch.17, 1011–1057. —**NEIO 的奠基性综述——Bresnahan 纲领的最完整陈述。**

[2] **Porter, R. H. (1983).** "A Study of Cartel Stability: The Joint Executive Committee, 1880–1886." *Bell Journal of Economics*, 14(2), 301–314. —Bresnahan 框架的第一个（也是最具影响力的）实证应用。

[3] **Genesove, D. & Mullin, W. P. (1998).** "Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890–1914." *RAND Journal of Economics*, 29(2), 355–377. —直接成本数据——NEIO 的 "黄金标准" 验证。

[4] **Olley, G. S. & Pakes, A. (1996).** "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6), 1263–1297. —生产函数估计的奠基性论文——投资作为生产率代理。

[5] **Ackerberg, D. A., Caves, K. & Frazer, G. (2015).** "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83(6), 2411–2451. —ACF 对 OP/LP 的函数依赖性批判及修正。

[6] **De Loecker, J. & Warzynski, F. (2012).** "Markups and Firm-Level Export Status." *American Economic Review*, 102(6), 2437–2471. —加成率估计的 "生产函数方法"——$\mu = \theta^V / \alpha^V$。

进阶阅读

[7] **Bresnahan, T. F. (1982).** "The Oligopoly Solution Concept is Identified." *Economics Letters*, 10(1-2), 87–92. —θ 参数识别问题的简短（4 页）原始陈述。

[8] **Demsetz, H. (1973).** "Industry Structure, Market Rivalry, and Public Policy." *Journal of Law and Economics*, 16(1), 1–9. —效率替代假说的奠基性论文——高利润反映效率而非市场势力。

[9] **Hansen, L. P. (1982).** "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 50(4), 1029–1054. —GMM 的原始论文——本章所有结构估计的统计基础。

[10] **Green, E. J. & Porter, R. H. (1984).** "Noncooperative Collusion Under Imperfect Price Information." *Econometrica*, 52(1), 87–100. —Porter (1983) 实证发现的理论基础——不完全监控下的合谋模型。

[11] **Lau, L. J. (1982).** "On Identifying the Degree of Competitiveness from Industry Price and Output Data." *Economics Letters*, 10(1-2), 93–99. —识别条件的正式推导——"Lau 的不可能定理"。

[12] **Corts, K. S. (1999).** "Conduct Parameters and the Measurement of Market Power." *Journal of Econometrics*, 88(2), 227–250. —Corts 批判——MC 弹性对 θ 解释的根本挑战。

[10] **Levinsohn, J. & Petrin, A. (2003).** "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70(2), 317–341. —用中间投入替代投资作为生产率代理。

[11] **De Loecker, J., Eeckhout, J. & Unger, G. (2020).** "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics*, 135(2), 561–644. —全球加成率上升趋势——生产函数方法的大规模应用。

[12] **Reiss, P. C. & Wolak, F. A. (2007).** "Structural Econometric Modeling: Rationales and Examples from Industrial Organization." In J. J. Heckman & E. E. Leamer (eds.), *Handbook of Econometrics*, Vol. 6A, Ch.64, 4277–4415. —**结构估计的方法论哲学——"四个层次"。**

[13] **Berry, S. T. & Haile, P. A. (2016).** "Identification in Differentiated Products Markets." *Annual Review of Economics*, 8, 27–52. —BLP 框架的识别条件——为 Ch.11 做准备。

[14] **Vuong, Q. H. (1989).** "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica*, 57(2), 307–333. —非嵌套模型选择检验。

[15] **Card, D. & Krueger, A. B. (1994).** "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review*, 84(4), 772–793. —DiD 方法的经典应用——本章方法论反思中作为简约式参照。

[16] **Blundell, R. (2017).** "Structural Models vs. Reduced-Form: The Debate in Industrial Organization." *Journal of Economic Perspectives*, 31(2), 181–202. —结构估计与简约式方法的互补性——本章方法论反思的核心参考文献。

中国相关

[17] **Lu, Y. & Yu, L. (2015).** "Trade Liberalization and Markup Dispersion: Evidence from China's WTO Accession." *American Economic Journal: Applied Eco-*

nomics, 7(4), 221–253. —中国加入 WTO 对制造业加成率的影响。

[18] **Brandt, L., Van Biesebroeck, J. & Zhang, Y. (2012).** "Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing." *Journal of Development Economics*, 97(2), 339–351. —中国工业企业数据库的生产率和加成率估计。

[19] **陈信元、周亚虹 (2009).** 中国银行业市场竞争度研究——Bresnahan 框架在中国银行业的应用。

[20] **平新乔、魏军锋 (2001).** "中国汽车工业的市场规模和企业数量——结构估计方法的应用." 经济研究. —中国汽车行业的早期 NEIO 应用。

前沿/方法论参考

[21] **Einav, L. & Levin, J. (2010).** "Empirical Industrial Organization: A Progress Report." *Journal of Economic Perspectives*, 24(2), 145–162. —2010 年 NEIO 的 "进展报告"——可及的非技术概览。

[22] **Angrist, J. D. & Pischke, J.-S. (2010).** "The Credibility Revolution in Empirical Economics." *Journal of Economic Perspectives*, 24(2), 3–30. —简约式 "可信性革命"——作为与结构估计的对比视角。

[23] **Nevo, A. (2001).** "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2), 307–342. —BLP 框架的经典应用（为 Ch.11 预热）。

[24] **Syverson, C. (2019).** "Macroeconomics and Market Power: Context, Implications, and Open Questions." *Journal of Economic Perspectives*, 33(3), 23–43. —市场势力上升的宏观经济学讨论。

[25] **Berry, S., Levinsohn, J. & Pakes, A. (1995).** "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4), 841–890. —**BLP 的原始论文——下一步（Ch.11）的核心材料。**

第十一章　BLP 与需求估计：差异化产品市场的结构需求估计

章首语

1995 年，Steven Berry、James Levinsohn 和 Ariel Pakes 在 *Econometrica* 上发表了一篇题为 “Automobile Prices in Market Equilibrium” 的论文。这篇论文提出了一个统一的方法论框架，一次性解决了差异化产品市场需求估计的三个核心难题：维度灾难（太多产品、太多替代弹性）、价格内生性（产品质量与价格同时决定）、和数据限制（只有市场层面数据，没有个体消费记录）。

这个框架——现在被经济学界简称为 “BLP”——的优雅之处在于它将三个看似独立的方法论创新融合成一个有机整体。第一个创新是**随机系数 Logit 模型**：通过假设消费者偏好服从一个分布（而非对所有消费者相同的单一系数），BLP 将 $O(J^2)$ 个替代弹性参数压缩为随机系数协方差矩阵中的少量参数——同时允许替代模式灵活地由产品在特征空间中的 “距离” 决定，而非由研究者事前指定的分组决定。第二个创新是**收缩映射（Contraction Mapping）**：Berry (1994) 证明，给定随机系数，存在唯一的均值效用向量使模型预测的市场份额等于观测到的市场份额——而且这个解可以通过一个简单的不动点迭代快速找到。第三个创新是 **BLP 工具变量**：利用寡头定价理论（产品价格取决于市场中所有产品的特征——竞争程度决定加成率），构造基于竞争对手产品特征的排他工具变量来处理价格内生性。

BLP 的智识传承深远。它被引用了超过 7,500 次（Google Scholar），进入了有史以来发表的所有论文的前 0.01%。“BLP” 从一篇论文的缩写变成了一个动词——研究者们说 “to BLP-estimate demand”，就像说 “to Google” 一样。更关键的是，BLP 使得**结构需求估计**成为全球反垄断审查的标准工具：并购模拟、SSNIP 测试、市场界定——这些在第 ~12~ 章中将要讨论的反垄断经济学工具——都建立在 BLP 估计的需求系统之上。

本章是模块 5（NEIO）的核心章之二。第 ~10~ 章解决了同质产品行业的市场势力识别问题（Bresnahan 纲领）；本章解决差异化产品市场的需求估计问题（BLP 纲领）。两章共同构成 NEIO 的 “双核心” 方法论基础。掌握了这两章，你就拥有了从市场数据中识别市场势力——无论是同质产品还是差异化产品——的完整方法论工具箱。

11.1 问题的背景：为什么差异化产品需求估计是一个难题？

读完本节，你应该能够回答：(1) 从 Ch.10 的同质产品世界过渡到差异化产品世界，需求估计面临的三个核心挑战是什么？(2) 维度灾难为什么使完全自由的需求系统无法直接从数据中估计？(3) 价格内生性在差异化产品市场中为什么比同质产品市场中更复杂？

11.1.1 从同质产品到差异化产品：一个世界的过渡

Ch.10 的成就与边界。在第 ~10~ 章中，我们详细讨论了 Bresnahan 纲领——如何从市场价格和数量数据中同时识别边际成本和竞争行为参数 θ。Bresnahan 框架的核心假设是**同质产品**（homogeneous products）：市场中所有企业销售的是完全相同的产品。这意味着只有一个市场需求函数 $P = P(Q, Z)$，一个市场价格 P，和一个市场总产量 Q。θ 参数的识别来自需求**斜率**的旋转——不同竞争行为下，企业对需求斜率变化的反应不同。

然而，现实中的大多数市场并非如此。考虑以下三个市场的对比：

表 11.1: Ch.10（同质产品）与 Ch.11（差异化产品）的实证环境对比

	Ch.10 的世界	**Ch.11 的世界**	**差异的本质**
示例行业	糖业、水泥、铁路货运	汽车、手机、麦片、药品	—
产品	同质（所有企业生产同一种产品）	差异化（每个企业生产多种产品）	产品定义
价格	单一市场价格	每个产品有各自的价格 p_{jt}	价格维度从 1 增至 J
需求	单一需求函数 $Q(P)$	J 个需求函数 $q_j(p_1, \ldots, p_J)$	需求系统从 1 维增至 J 维
数据	时间序列：P_t, Q_t	市场-产品面板：p_{jt}, s_{jt}, x_{jt}	面板结构
核心不确定性	θ（行为参数）	替代弹性矩阵 η_{jk}	识别对象

差异化产品的定义。差异化产品（differentiated products）是指消费者认为存在实质性差异的产品。差异可以来自：(a) 物理特征（马力、尺寸、重量、芯片速度、相机像素）；(b) 品牌/声誉（丰田 vs 宝马、苹果 vs 华为）；(c) 空间位置（距离消费者远近不同的加油站或超市）；(d) 包装和营销。经济学上，两个产品是差异化的，当且仅当它们之间的交叉价格弹性是有限的——即当产品 k 提价时，部分（但非全部）消费者转向产品 j。这个定义避免了“什么算同一个产品”的哲学争论，直接将其转化为可检验的计量问题。

产品空间的概念。在差异化产品市场中，一个有用的概念工具是**产品空间**（product space）或**特征空间**（characteristics space）。将每个产品 j 映射到一个 K 维的向量 $\mathbf{x}_j = (x_{j1}, \ldots, x_{jK})$——在汽车市场中，这可能是（马力, 重量, 尺寸, 油耗, 空调, ...）。消费者 i 在特征空间中有自己的 "理想点"（ideal point）——比如，一个家庭可能偏好中等马力 + 大空间 + 低油耗的车。两个产品在特征空间中 "越近"，它们对相似消费者的吸引力越接近 $\rightarrow$ 它们之间的替代弹性越大。产品空间中产品的分布——以及消费者偏好在特征空间中的分布——决定了整个市场中的替代模式。这是 BLP 方法论的核心直觉：**替代弹性不是自由参数——它由产品在特征空间中的位置和消费者偏好在特征空间中的分布共同决定。**

11.1.2　维度灾难：$J(J-1)/2$ 个弹性的问题

自由需求系统的参数规模。考虑一个市场中有 J 个差异化产品。对每个产品 j，其需求 q_j（或市场份额 s_j）取决于所有 J 个产品的价格和特征。这意味着有 J 个需求方程，每个需求方程包含 J 个价格：

$$s_j = s_j(p_1, p_2, \ldots, p_J, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J), \quad j = 1, \ldots, J$$

如果我们在没有任何结构约束的情况下估计这个需求系统，需要估计 $J(J-1)/2$ 个交叉价格弹性（$\eta_{jk} = (\partial s_j / \partial p_k) \cdot (p_k / s_j)$，对 $j \neq k$），加上 J 个自身价格弹性。对于一个典型的 BLP 式数据集：

- 汽车市场：$J \approx$ 100–200 个车型-年观测 $\rightarrow$ 4,950–19,900 个交叉弹性
- RTE 麦片市场（Nevo 2001）：约 225 个品牌-城市组合 $\rightarrow$ 25,200 个交叉弹性
- 手机市场：$J \approx$ 50–100 $\rightarrow$ 1,225–4,950 个交叉弹性

即使有最丰富的面板数据（多市场 $\times$ 多时期），自由参数的个数也远超观测值的个数——这就是**维度灾难**（curse of dimensionality）：参数空间的维度随 J 以平方速度增长，而数据的信息增长最多是线性的。

需要结构性约束。解决维度灾难的唯一办法是对替代模式施加结构性约束——用较少的参数来 "压缩" 完整的替代弹性矩阵。但关键挑战是：约束必须足够强以降低参数空间的维度（否则估计不可行），同时又必须足够灵活以捕捉现实中丰富的替代模式（否则模型错误）。如果约束太强（如简单 Logit——所有交叉弹性完全由市场份额决定），IIA 将导致不合理的替代模式；如果约束太弱（如无约束 AIDS），参数数量重新爆炸。

简单 Logit 的 "极端压缩"：一个参数捕捉所有替代。在简单 Logit 中，$\eta_{jk} = \alpha \cdot p_k \cdot s_k$——交叉弹性**只取决于产品** k（的价格和份额），与产品 j 无关。这意味着所有产品从产品 k 的提价中获益的比例相同（与各自的市场份额成比例）。这仅用一

个共同的 α 参数就“压缩”了所有 $J(J-1)/2$ 个交叉弹性——压缩比极高，但代价是 IIA 导致的不现实替代模式（红巴士/蓝巴士问题）。

BLP 的折衷方案。 BLP 在“过度约束”（简单 Logit）和“无约束”（自由参数）之间取得了平衡。通过 K 个随机系数（方差参数 $\sigma_1, \ldots, \sigma_K$），BLP 生成了一个 $J \times J$ 的替代弹性矩阵，其中：

$$\eta_{jk} = -\frac{p_k}{s_j} \int \alpha_i \cdot s_{ij}(\mathbf{v}) \cdot s_{ik}(\mathbf{v})\, dF(\mathbf{v})$$

这个弹性矩阵不是自由的——它完全由 K 个随机系数方差参数决定（$K \ll J(J-1)/2$）——但它可以捕捉复杂的替代模式：特征空间中“近”的产品之间有更大的交叉弹性。参数空间从 $O(J^2)$ 压缩到 $O(K)$。

11.1.3 价格内生性：不可观测质量的挑战

内生性的来源。在差异化产品需求估计中，价格内生性的来源比同质产品更复杂。关键的内生变量是产品价格 p_{jt}。问题在于：

$$\text{Cov}(p_{jt}, \xi_{jt}) > 0$$

其中 ξ_{jt} 是产品的**不可观测质量**（unobserved product characteristics）——研究者观测不到，但企业和消费者都能观测到的产品特征（如品牌声誉、设计美学、制造精度、广告效应等）。数据中，ξ_j 高的产品往往同时具有：(a) 更高的价格（因为企业知道消费者愿意为高质量支付更多），(b) 更高的市场份额（因为消费者确实更喜欢高质量产品）。如果我们忽略这一相关性：

$$\text{plim}\ \hat{\alpha}_{\text{OLS}} = \alpha + \frac{\text{Cov}(p, \xi)}{\text{Var}(p \mid \mathbf{x})}$$

由于 $\text{Cov}(p, \xi) > 0$，OLS 估计的价格系数会**偏向零**（低估需求弹性）——我们错误地认为消费者对价格“不太敏感”，而实际上只是不可观测的质量同时拉高了价格和需求。低估需求弹性意味着**高估市场势力**——这正是反垄断分析中最危险的偏误类型。

与 Ch.10 的对比。在第 ~10~ 章的 Bresnahan 框架中，价格内生性也出现了，但形式不同：在同质产品市场中，价格和产量由需求和成本冲击**同时**决定（联立性偏误）。在差异化产品市场中，价格内生性的根源是**遗漏变量**——不可观测的产品质量 ξ_{jt} 被遗漏在需求方程的误差项中，但与价格相关。两者的共同点是：都需要工具变量来解决。

数据限制：缺少微观消费者数据。在许多市场中，研究者只有**市场层面**（market-level）数据——每个产品在每个市场-时间的份额、价格和特征。没有个体消费者的购买记录（“微观数据”）。这意味着：

- 我们观测到产品的市场份额 s_{jt}，但不知道是谁购买了它们
- 我们观测到产品的价格 p_{jt} 和特征 $\mathbf{x}_{jt}$，但不知道每个消费者的收入、年龄、家庭结构
- 我们需要从“宏观”（市场加总）数据中识别“微观”（个体偏好）参数——这是一个**反向推理**（inverse problem）问题

这个数据限制将 BLP 与需要微观消费者数据的传统离散选择估计（如 Train 2009 中描述的方法）区别开来。BLP 的独特之处在于：它从市场层面的份额数据中**反解**出个体层面的偏好参数。

BLP 之前的解决方案及其局限。在 BLP (1995) 之前，研究者们尝试了多种方法，但都只部分解决了问题：

- **AIDS（Almost Ideal Demand System，Deaton & Muellbauer 1980）：**解决了维度灾难（通过灵活的函数形式），但无法处理大量产品（J 仍然太大）且没有处理价格内生性
- **简单 Logit：**计算简洁，但 IIA 属性产生不现实的替代模式——无法用于反垄断分析
- **嵌套 Logit（McFadden 1978）：**部分解决了 IIA（在巢内），但巢结构必须**事前指定**——不同巢结构会产生不同的结果
- **微观数据估计：**最精确，但微观数据昂贵且往往不可得（尤其是跨多市场跨多年）

BLP 的核心突破

BLP 一次性解决了这三个挑战。随机系数 Logit 压缩了参数空间（从 $O(J^2)$ 到 $O(K)$）；收缩映射从市场份额中恢复不可观测的均值效用；BLP 工具变量利用寡头定价理论处理价格内生性。这三个创新相互依赖、缺一不可——构成了一个方法论上的“铁三角”。

11.2　BLP (1995) 的解决方案：随机系数 Logit 模型

【核心】

本节是全章的知识心脏。我们将完整推导 BLP 的随机系数 Logit 模型——从效用设定到市场份额方程，再到随机系数如何打破 IIA。每一步都配以经济学直觉。

11.2.1 消费者效用设定

一般设定。消费者 i（$i = 1, \ldots, I$）在市场 t（$t = 1, \ldots, T$）中选择产品 j（$j = 0, 1, \ldots, J_t$），其中 $j = 0$ 代表不购买任何产品（“外部商品”，outside good）。消费者 i 从产品 j 获得的间接效用：

$$u_{ijt} = \mathbf{x}_{jt}\boldsymbol{\beta}_i - \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt} \tag{11.1}$$

其中每个参数都有具体的经济学含义：

表 11.2: BLP 效用函数中各参数的经济学含义与识别来源

符号	名称	经济学直觉	来源/识别
$\mathbf{x}_{jt}$	可观测产品特征（K 维）	马力、重量、尺寸、空调、相机像素、品牌虚拟变量	产品规格表、行业数据库
$\boldsymbol{\beta}_i$	个体 i 的特征偏好	不是所有人的 β 都一样——有人偏好大马力，有人不在乎	**需要被估计的核心参数**
α_i	个体 i 的价格敏感度	收入高的消费者通常对价格不太敏感	**需要被估计的核心参数**
p_{jt}	产品 j 在 t 期的价格	厂商为产品设定的价格（策略变量）	市场监测数据
ξ_{jt}	不可观测产品特征（**需求冲击**）	产品质量、品牌声誉、广告效应——消费者和企业看到，但研究者看不到	模型反解（收缩映射 δ 中的残差）
ε_{ijt}	i.i.d. 口味冲击	消费者 i 和产品 j 之间的随机匹配——不受任何系统因素影响	假设：Type I Extreme Value

随机系数的结构。消费者异质性通过随机系数 $\boldsymbol{\beta}_i$ 和 α_i 引入：

$$\begin{pmatrix}\boldsymbol{\beta}_i \\ \alpha_i\end{pmatrix} = \begin{pmatrix}\bar{\boldsymbol{\beta}} \\ \bar{\alpha}\end{pmatrix} + \boldsymbol{\Sigma}\mathbf{v}_i + \boldsymbol{\Pi}\mathbf{D}_i \tag{11.2}$$

其中：

- $\bar{\boldsymbol{\beta}}$ 和 $\bar{\alpha}$：**均值偏好参数**——对于“平均消费者”，马力每增加一单位的边际效用
- $\boldsymbol{\Sigma}$：**随机系数的 Cholesky 因子**（$(K+1)\times(K+1)$ 下三角矩阵，对应 K 个产品特征系数与 1 个价格系数——$\boldsymbol{\Sigma}\boldsymbol{\Sigma}'$ 是偏好参数的协方差矩阵）
- $\mathbf{v}_i \sim N(\mathbf{0}, \mathbf{I}_{K+1})$：消费者 i 的标准化随机抽取值——代表**不可观测的偏好异质性**

- $\mathbf{\Pi}$：人口统计特征的系数矩阵
- $\mathbf{D}_i$：消费者 i 的可观测人口统计特征（收入、年龄、家庭规模——如果可获得的话）

$\mathbf{\Sigma}$ 的对角元素 σ_k 衡量的是**偏好的离散程度**——消费者在第 k 个特征上的偏好有多大的异质性。$\sigma_k = 0$ 意味着所有消费者对特征 k 的估价相同（退化为简单 Logit）；σ_k 越大，替代模式越灵活。非对角元素捕捉了偏好之间的相关性——但实践中通常假设 $\mathbf{\Sigma}$ 是对角矩阵（只有方差，没有协方差），因为估计非对角元素在计算上是极具挑战性的。

为什么是正态分布？ $\mathbf{v}_i \sim N(0,1)$ 是最常见的选择，但这不是逻辑必然的。正态的优点：(a) 数学方便——与模拟积分兼容（Gauss-Hermite 积分），(b) 经济学直觉——偏好大概率集中在均值附近，(c) 传统——BLP (1995) 本身使用正态。缺点：单峰且对称——不能捕捉多峰偏好群体（例如，一半人极度偏好大马力，另一半人完全不关心——这会是一个双峰分布）。替代选择包括对数正态（如价格系数——因为价格系数通常假设为正，$\alpha_i > 0$）和离散分布（有限混合模型）。

11.2.2　均值效用分解与线性参数集中化

均值效用分解。效用的一个关键分解是将总效用拆分为两部分：

$$u_{ijt} = \underbrace{\delta_{jt}}_{\text{对所有}i\text{ 相同}} + \underbrace{\mu_{ijt}}_{\text{因人而异}} + \varepsilon_{ijt} \tag{11.3}$$

其中：

$$\delta_{jt} = \mathbf{x}_{jt}\bar{\boldsymbol{\beta}} - \bar{\alpha}p_{jt} + \xi_{jt} \tag{11.4}$$

$$\mu_{ijt} = \mathbf{x}_{jt}(\mathbf{\Sigma}\mathbf{v}_i) - (\sigma_p v_{ip})p_{jt} + \mathbf{x}_{jt}(\mathbf{\Pi}\mathbf{D}_i) \tag{11.5}$$

均值效用 δ_{jt} 捕捉了产品 j 对所有消费者的**平均吸引力**——它由可观测特征、价格和不可观测质量驱动。它不依赖于个体 i（所有消费者对产品 j 都有相同的“基础”评估）。**个体偏离** μ_{ijt} 捕捉了消费者 i 与“平均消费者”之间的偏差——有人比平均值更喜欢马力，有人比平均值更在乎价格。μ_{ijt} 取决于随机抽取值 $\mathbf{v}_i$ 和人口统计特征 $\mathbf{D}_i$。

这种分解为什么至关重要？它的计算意义体现在 BLP 的 GMM 估计中：

(1) 给定非线性参数（$\mathbf{\Sigma}$, $\mathbf{\Pi}$），δ_{jt} 可以通过收缩映射快速反解。因为 δ_{jt} 在 Logit 概率公式中是线性的——收缩映射的收敛速度基于 Logit 结构。

(2) 给定 δ_{jt}，线性参数（$\bar{\boldsymbol{\beta}}$, $\bar{\alpha}$）可以通过线性 IV 回归（2SLS）获得闭式解。因为 $\delta_{jt} = \mathbf{x}_{jt}\bar{\boldsymbol{\beta}} - \bar{\alpha}p_{jt} + \xi_{jt}$ 是一个线性方程——可以用工具变量处理。

(3) 非线性搜索的维度从 $K + \dim(\boldsymbol{\Sigma})$ 降低到仅 $\dim(\boldsymbol{\Sigma})$。这是“集中化”（concentration）的魔力——线性参数被**集中出去**了（concentrated out），相当于对每个 $\boldsymbol{\Sigma}$ 值，自动选择了最优的 $\bar{\boldsymbol{\beta}}$ 和 $\bar{\alpha}$。对于 $K = 5$，$\boldsymbol{\Sigma}$ 对角（3 个随机系数），搜索空间从 10 维降至 3 维——差别不是线性的，而是指数级的（优化难度随维度呈指数级增长）。

11.2.3 外部商品（Outside Good）的关键作用

定义与标准化。外部商品 $j = 0$ 代表“不购买本市场中任何产品”——消费者可以选择使用公共交通而非购买汽车，吃热麦片而非即食麦片，或者完全不消费。效用的标准化是：

$$u_{i0t} = 0 + \varepsilon_{i0t} \tag{11.6}$$

其中 ε_{i0t} 也是 i.i.d. Type I Extreme Value。

外部商品的三个关键作用：

(1) 总量定义：$s_{0t} = 1 - \sum_{j=1}^{J_t} s_{jt}$ 定义了整个市场的规模。如果不设 M_t（潜在市场规模），则 s_{0t} 的内部分额被标准化为剩余。通过设定 M_t，s_0 从“剩余份额”变成了“不购买者的比例”——这为需求弹性提供了锚点。

(2) 替代弹性锚定：存在外部商品意味着所有产品也与“什么都不买”竞争。这限制了总需求的弹性上限——如果所有产品同时提价，消费者可以“退出”整个类别（不买任何车、不买任何麦片）。这对一般均衡分析和反事实模拟（如行业范围的税收）至关重要。

(3) 收缩映射的数学基础：这是 Berry (1994) 证明的基石。因为 $s_0 > 0$，$\ln \hat{\mathbf{s}}$ 的 Jacobian 的每一行之和 $= 1 - s_0 < 1 \to$ 收缩映射的模 $< 1 \to$ Banach 不动点定理保证唯一解和几何收敛。当 $s_0 \to 0$（市场接近饱和），收缩速度趋向于零——内循环变得极其缓慢。这是 BLP 在饱和市场（如手机市场）中计算困难的根本原因。

s_0 的实际选择。实践中，M_t（潜在市场规模）的选择直接影响 s_0 的大小。常见的设定包括：总人口、总家庭数、或特定人口子群（如适龄驾驶人口）。s_0 通常落在 0.7–0.95 之间——即大部分潜在消费者“不购买”该类产品。如果 s_0 定为 0.5，替代弹性估计会显著不同。Nevo (2000) 提供了关于 M_t 选择的详细讨论——通常建议进行敏感性分析，将 M_t 倍增和减半，观察估计结果的变化。

11.2.4　个体选择概率与市场份额

从效用到选择概率。消费者 i 选择产品 j 当且仅当：

$$u_{ijt} > u_{ikt} \quad \forall k \neq j$$

给定 $\varepsilon_{ijt} \sim$ i.i.d. Type I Extreme Value，对消费者 i 而言，选择产品 j 的概率是标准 Logit 公式：

$$P_{ijt}(\boldsymbol{\delta}_t, \boldsymbol{\Sigma}, \boldsymbol{\Pi}) = \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + \mu_{ikt})} \tag{11.7}$$

从个体概率到市场份额。我们观测的是**市场加总**份额——而非个体选择。预测的市场份额是所有消费者的选择概率的积分（或平均）：

$$s_{jt}(\boldsymbol{\delta}_t, \boldsymbol{\Sigma}, \boldsymbol{\Pi}) = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + \mu_{ikt})} \, dF(\mathbf{v}_i, \mathbf{D}_i) \tag{11.8}$$

其中 $F(\mathbf{v}_i, \mathbf{D}_i)$ 是消费者异质性（不可观测偏好 + 可观测人口统计）的联合分布。

没有闭式解：为什么需要模拟。因为 μ_{ijt} 是 $\mathbf{v}_i$ 的非线性函数（包含在 $\exp(\cdot)$ 内），而分母是 $\exp(\delta_{kt} + \mu_{ikt})$ 的**总和**——这个积分没有解析闭式解（除非 $\boldsymbol{\Sigma} = \mathbf{0}$，即简单 Logit）。解决方法是蒙特卡洛模拟：

$$\hat{s}_{jt}(\boldsymbol{\delta}_t, \boldsymbol{\Sigma}, \boldsymbol{\Pi}) = \frac{1}{ns} \sum_{i=1}^{ns} \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + \mu_{ikt})} \tag{11.9}$$

其中 ns 是从 $F(\mathbf{v}, \mathbf{D})$ 中抽取的模拟消费者数量。$\hat{s}_{jt}$ 是 s_{jt} 的一致估计量（$ns \to \infty$ 时）。

11.2.5　随机系数如何打破 IIA

简单 Logit 的 IIA 属性。在简单 Logit（$\boldsymbol{\Sigma} = \mathbf{0}, \boldsymbol{\Pi} = \mathbf{0}$）中，所有消费者有相同的偏好参数。此时：

$$\frac{P_{ij}}{P_{ik}} = \frac{\exp(\delta_j)}{\exp(\delta_k)} = \exp(\delta_j - \delta_k)$$

任何两个产品 j 和 k 的选择概率之比**只依赖于这两个产品自身**，与选择集中还有哪些其他产品无关。这意味着交叉价格弹性是：

$$\eta_{jk}^{\text{Logit}} = \alpha \cdot p_k \cdot s_k$$

交叉弹性**不依赖于** j——只依赖于提价的产品 k（它的价格和市场份额）。所有其他产品从产品 k 的提价中获益的比例相同。这是**红巴士/蓝巴士问题**：如果 Red Bus 提价 10%，Logit 预测所有其他选项的份额按各自原份额等比例增加——Car、Blue

Bus、Bicycle 各获得相同比例的收益。现实中，Blue Bus 应该获得绝大多数转移消费者（因为它是 Red Bus 的最近替代品）。

随机系数打破 IIA 的机制。在随机系数 Logit 中，交叉价格弹性是：

$$\eta_{jk} = \frac{\partial s_j}{\partial p_k} \cdot \frac{p_k}{s_j} = -\frac{p_k}{s_j} \int \alpha_i \cdot s_{ij}(\mathbf{v}) \cdot s_{ik}(\mathbf{v})\, dF(\mathbf{v}) \quad (j \neq k) \tag{11.10}$$

关键的统计量是 $s_{ij}(\mathbf{v}) \cdot s_{ik}(\mathbf{v})$ 在消费者之间的**协变**。考虑两个产品 j 和 k：

- 如果 j 和 k 在特征空间中“很近”（比如都是大马力豪华车），那么**相同的**消费者（高 $\mathbf{v}_{\text{hp}}$ 的消费者）对两者都有高选择概率 $\to \text{Cov}(s_{ij}, s_{ik})$ 大 $\to \eta_{jk}$ 大
- 如果 j 和 k 在特征空间中“很远”（比如一个是豪华车，一个是经济车），那么喜欢 j 的消费者通常对 k 选择概率低 $\to \text{Cov}(s_{ij}, s_{ik})$ 小 $\to \eta_{jk}$ 小

在弹性中的体现。推导交叉弹性公式。市场份额 $s_j = \int s_{ij}\, dF$。对 p_k 求导：

$$\frac{\partial s_j}{\partial p_k} = \int \frac{\partial s_{ij}}{\partial p_k}\, dF$$

对于个体 i，$\partial s_{ij}/\partial p_k = -\alpha_i \cdot s_{ij} \cdot s_{ik}$（对 $j \neq k$——这是 Logit 公式的标准导数）。因此：

$$\eta_{jk} = \frac{p_k}{s_j} \int -\alpha_i \cdot s_{ij} \cdot s_{ik}\, dF = -\frac{p_k}{s_j} \cdot \mathbb{E}_F[\alpha_i \cdot s_{ij} \cdot s_{ik}]$$

协方差直觉。展开 $\mathbb{E}[\alpha_i s_{ij} s_{ik}]$：

$$\mathbb{E}[\alpha_i s_{ij} s_{ik}] = \mathbb{E}[\alpha_i] \cdot \mathbb{E}[s_{ij}] \cdot \mathbb{E}[s_{ik}] + \underbrace{\text{Cov}(\alpha_i s_{ij}, s_{ik})}_{\text{打破 IIA 的项}}$$

在简单 Logit 中，$s_{ij} = s_j$（对所有人都一样）$\to \text{Cov} = 0 \to \eta_{jk} = \alpha \cdot p_k \cdot s_k$——退化到标准 Logit。在 BLP 中，$s_{ij}$ 随 $\mathbf{v}_i$ 变化 $\to$ 对于“相似”的产品，s_{ij} 和 s_{ik} 在消费者之间正相关 $\to \text{Cov} > 0 \to \eta_{jk}$ 大于简单 Logit 的预测。

特征空间的替代模式。考虑两个特征 r 和 s，$\text{Cov}(\mu_{ij}, \mu_{ik})$ 可以进一步分解：

$$\text{Cov}(\mu_{ij}, \mu_{ik}) = \sum_{r=1}^{K} \sigma_r^2 \cdot x_{jr} \cdot x_{kr}$$

如果产品 j 和 k 在很多特征上都有相似的值（$x_{jr} \approx x_{kr}$），那么 μ_{ij} 和 μ_{ik} 在消费者之间高度正相关 $\to j$ 和 k 是近替代品。如果它们的特征值差异很大，相关性弱 $\to$ 它们不是近替代品。这是 BLP 的核心经济学直觉：**替代模式由特征空间中的“距离”决定，而非由研究者任意指定。**

IIA 打破的经济学总结

在简单 Logit 中，所有替代是“成比例的”——每个产品从其他产品的提价中获益的份额等于它自己的市场份额。在 BLP 中，替代模式由特征空间的“产品分布”和“消费者偏好分布”共同决定。特征近的产品替代弹性大；特征远的产品替代弹性小。这一灵活性是 BLP 模型的签名式贡献——它意味着我们可以在不必事前指定“哪些产品之间是竞争对手”的情况下，让数据告诉我们替代模式。

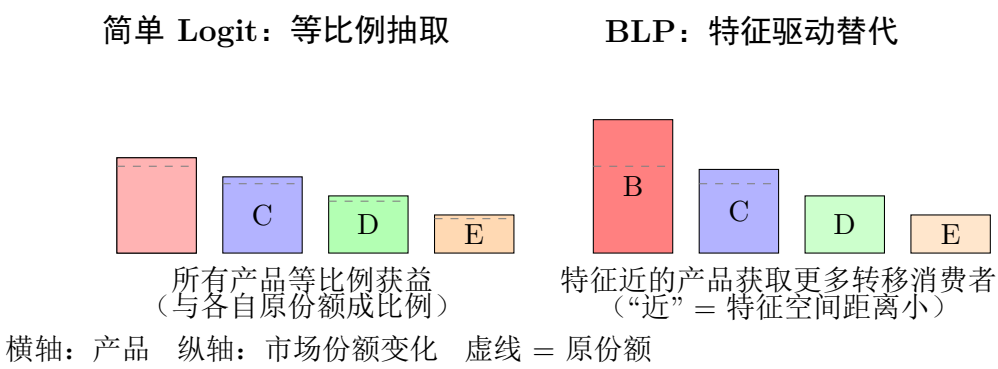


图 11.1: 简单 Logit vs BLP 的替代模式对比：产品 A 提价后，消费者转移至其他产品的分布。左图：简单 Logit 下，各产品按原份额等比例获得转移消费者。右图：BLP 下，在特征空间中与 A 更“近”的产品（B）获得更多转移消费者。

11.3　关键方法论创新

【核心】

本节深入讲解 BLP 的三项核心方法论创新：收缩映射（从市场份额中恢复均值效用）、BLP 工具变量（处理价格内生性）、和模拟积分（计算市场份额）。每一项都需要完整的推导、直觉和数值细节。

11.3.1　收缩映射——Berry (1994) 的反解定理

问题的陈述。假设我们已知非线性参数 $\boldsymbol{\Sigma}$（暂时假设其为某个给定的值）。我们面临的问题是：找到 $\boldsymbol{\delta}_t = (\delta_{1t}, \ldots, \delta_{J_t t})$ 使得：

$$\hat{s}_{jt}(\boldsymbol{\delta}_t, \boldsymbol{\Sigma}) = s_{jt}^{\text{obs}} \quad \forall j = 1, \ldots, J_t$$

即模型预测的市场份额等于实际观测的市场份额。这是一个 J_t 维的非线性方程组——J_t 可能高达 50–200。

Banach 不动点定理的预备知识。一个映射 $\mathcal{T}:\mathbb{R}^J\to\mathbb{R}^J$ 是**收缩映射**（contraction），如果对任意 $\mathbf{x},\mathbf{y}\in\mathbb{R}^J$：

$$\|\mathcal{T}(\mathbf{x})-\mathcal{T}(\mathbf{y})\|<\|\mathbf{x}-\mathbf{y}\|$$

即应用映射后，两点之间的距离变近了——每次迭代，近似解的误差都会缩小。Banach 不动点定理保证：在完备度量空间中，任何收缩映射都有**唯一的不动点**（unique fixed point）$\mathbf{x}^*=\mathcal{T}(\mathbf{x}^*)$，且从任何初始值出发，序列 $\mathbf{x}^{(n+1)}=\mathcal{T}(\mathbf{x}^{(n)})$ 几何收敛到 $\mathbf{x}^*$。

Berry (1994) 的洞察。 Berry 证明了以下映射是收缩的：

$$\delta_j^{(h+1)}=\delta_j^{(h)}+\ln s_j^{\text{obs}}-\ln\hat{s}_j(\boldsymbol{\delta}^{(h)},\boldsymbol{\Sigma}) \tag{11.11}$$

或用向量形式：

$$\boldsymbol{\delta}^{(h+1)}=\mathcal{T}(\boldsymbol{\delta}^{(h)})\equiv\boldsymbol{\delta}^{(h)}+\ln\mathbf{s}^{\text{obs}}-\ln\hat{\mathbf{s}}(\boldsymbol{\delta}^{(h)},\boldsymbol{\Sigma})$$

收缩性的证明概要。定义映射 $\mathcal{T}$ 的 Jacobian $\nabla_{\boldsymbol{\delta}}\mathcal{T}=\mathbf{I}-\nabla_{\boldsymbol{\delta}}\ln\hat{\mathbf{s}}$。可证明该矩阵的所有元素非负，且每一行的元素之和 $=1-s_{0t}(\boldsymbol{\delta})<1$（因为有外部商品，$s_{0t}>0$）。这意味着 $\|\nabla_{\boldsymbol{\delta}}\mathcal{T}\|_\infty<1\to\mathcal{T}$ 是收缩映射。由于 $\mathbb{R}^J$ 是完备度量空间，Banach 不动点定理适用。

压缩速度与 s_0 的关系。映射的模（modulus）$\approx 1-s_0$。当 s_0 大时（大多数人不购买此类产品），每次迭代误差缩小很多（快速收敛）。当 s_0 小时（市场接近饱和），误差缩小缓慢——每次迭代只将误差缩减一小部分。实际经验法则：

- $s_0\approx 0.85$（如汽车市场）：40–80 次迭代收敛到 10^{-13}
- $s_0\approx 0.40$（如麦片市场）：200–500 次迭代
- $s_0\approx 0.10$（高度饱和的市场）：1,000+ 次迭代甚至无法收敛

内循环容差的极端重要性。外循环（优化 $\boldsymbol{\Sigma}$）依赖于内循环给出的 $\boldsymbol{\delta}$ 来计算 ξ_{jt} 和 GMM 目标函数 $Q(\boldsymbol{\Sigma})$。如果内循环的容差不够严格，$\boldsymbol{\delta}$ 中有噪声 $\to\xi$ 中有噪声 $\to$ GMM 目标函数 $Q(\boldsymbol{\Sigma})$ 有噪声 $\to$ 外循环优化失败（虚假局部最小值、梯度不可靠）。Dubé, Fox & Su (2012) 在蒙特卡洛模拟中发现：将内循环容差从 10^{-6} 放松到 10^{-4}，估计量偏差增加 20–50%。标准的实践是使用极其严格的容差——10^{-12} 到 10^{-14}——以确保模拟噪声不过度污染 GMM 目标函数。

收缩映射的经济学直觉

收缩映射的工作原理类似一个“猜测-校验”循环。猜测产品的均值效用，计算由此产生的市场份额，与观测值的差异用于修正猜测——如果预测份额太低，提高 δ；如果太高，降低 δ。收缩性来自 Logit 模型的数学性质——每次修正都会使预测更接近真实值，且修正的幅度逐次减小。

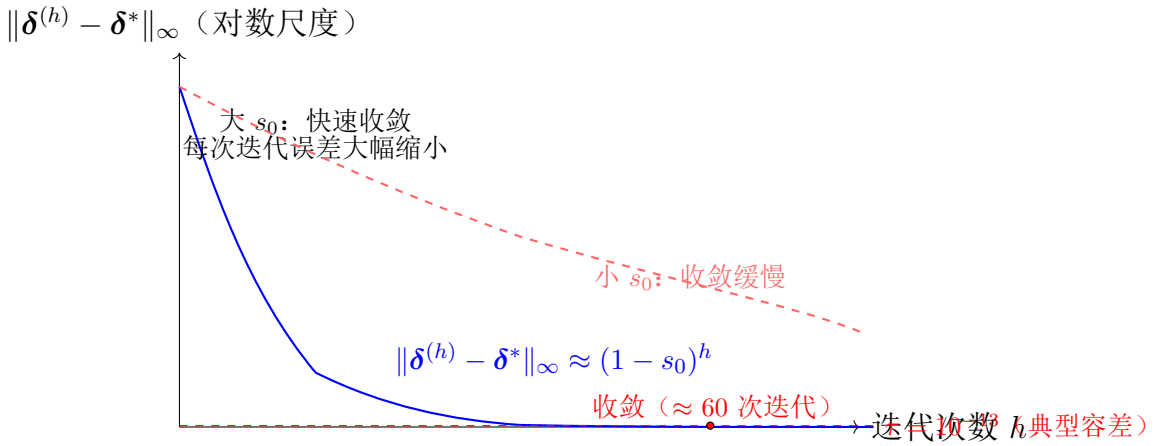


图 11.2: 收缩映射的几何收敛：$\|\boldsymbol{\delta}^{(h)} - \boldsymbol{\delta}^*\|_\infty$ 随迭代次数的变化（示意图，对数尺度）。蓝色实线对应 s_0 较大（≈ 0.85）的情形——大约 60 次迭代达到 10^{-13} 容差。红色虚线对应 s_0 较小（市场接近饱和）的情形——收敛显著变慢。收敛速度 $\approx (1 - s_0)^h$，由 Banach 不动点定理保证。

11.3.2　BLP 工具变量：从寡头定价理论到价格工具变量

内生性的精确表述。价格 p_{jt} 是内生的因为：

$$\text{Cov}(p_{jt}, \xi_{jt} \mid \mathbf{x}_{jt}) > 0$$

为什么是正的？考虑两个在其他方面相同的产品（相同的马力、尺寸等），但产品 A 有重要的不可观测特性（如更强的品牌声誉、更好的可靠性记录 = 高 ξ）。企业知道这一点并据此设定 A 的价格高于 B。同时，A 的市场份额也更高（因为消费者看到 ξ 并更喜欢 A）。OLS 会将高价格 + 高份额的组合解读为“价格不影响需求”，而实际上高份额尽管高价格——因为 ξ 的正面效应超过了价格的负面效应。

BLP 工具变量的经济学逻辑。 BLP 提出利用**产品特征在市场中的分布**构造价格的工具变量。其有效性的经济学逻辑基于寡头定价理论（回忆第 ~7~ 章和第 ~8~ 章中讨论的 Bertrand-Nash FOC）：

(1) 相关性（Relevance）：企业 f 对产品 j 的定价取决于市场中所有产品的特征。为什么？因为产品特征决定了替代模式——如果竞争对手的产品特征与 j 的更接近，j 面临的需求更有弹性 $\to$ j 的加成率和价格更低。竞争程度——以及由此决定的价格——是所有产品特征的函数。因此，竞争对手产品的特征之和、本企业其他产品的特征——这些都影响价格。

(2) 外生性（Exogeneity）：产品特征在价格决策之前被选定（“时序假设”）。汽车企业先设计车型（选择马力、重量、尺寸），然后才设定价格。产品特征可能也对 ξ 有影响（如果企业也根据预期的 ξ 选择特征），但这是二阶的——特征的“设计”比“定价”更缓慢、更昂贵。关键的条件是 $\mathbb{E}[\xi_{jt} \mid \mathbf{Z}_{jt}] = 0$。

三类 BLP 型工具变量：

表 11.3: 三类 BLP 型工具变量的构造方式与经济学直觉

工具变量类型	构造方式	经济学直觉	潜在问题
BLP 特征和	$Z_{jt}^{(1)} = \sum_{k \neq j, k \in \mathcal{F}_f} \mathbf{x}_{kt}$（本企业其他产品特征和） $Z_{jt}^{(2)} = \sum_{k \notin \mathcal{F}_f} \mathbf{x}_{kt}$（竞争对手产品特征和）	本企业其他产品的特征影响 j 的“竞食”风险；竞争对手的特征影响 j 面临的竞争压力——两者都影响 j 的加成率	特征可能与 ξ 相关——如果企业也根据 ξ 调整特征
Hausman 型	同一产品在其他市场（城市/区域）的价格	不同市场间的价格差异反映共同的成本冲击（运输成本、分销成本），而非局部需求冲击	要求成本冲击跨市场相关而需求冲击不相关——全国性广告可能同时改变 ξ
Differentiation IVs	$Z_{jt}^{(r)} = \sum_{k \neq j} f(x_{jr} - x_{kr})$——特征空间距离的度量	在特征空间中更拥挤的区域，产品面临更激烈的竞争 $\rightarrow$ 更低的价格	需正确指定距离函数 $f(\cdot)$

Differentiation IVs 的动机。 Gandhi & Houde (2019) 指出，原始 BLP 变量可能**弱**（与价格的相关性不足）——尤其是在产品的特征差异远大于特征水平的情况下。他们的 Differentiation IVs 直接度量产品 j 周围在特征空间中的“拥挤度”（crowding）——有多少其他产品在特征空间的不同维度上与之相近。实证证据表明，Differentiation IVs 的第一阶段 F 统计量显著高于原始 BLP IVs——提供了更强的识别。

最优工具变量。理论上，渐近最优的工具变量是条件期望函数（Amemiya 1977, Chamberlain 1987）：

$$\mathbf{Z}^* = \mathbb{E}\left[\frac{\partial \boldsymbol{\xi}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \,\middle|\, \mathbf{x}, \mathbf{w}\right]$$

在实践中，这个条件期望是未知的——需要“可行逼近”（feasible approximation）：使用先导一致估计量 $\tilde{\boldsymbol{\theta}}$ 计算的 $\partial \boldsymbol{\xi}(\tilde{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}$，对其在 $\mathbf{x}$ 和 $\mathbf{w}$ 上的非参数回归来逼近最优 IV。Reynaert & Verboven (2014) 发现，最优 IV 大幅减小小样本偏差——当工具变量与内生变量的相关性被优化后，GMM 的有限样本性质显著改善。

价格内生性不是 BLP 特有的问题

价格内生性在需求估计中无处不在——任何需求系统（AIDS、嵌套 Logit、简单 Logit）都面对同样的问题。BLP 的贡献不是“发现了”价格内生性（它至少自 Working 1927 年以来就被认知了），而是提供了一个**系统性的工具变量策略**——基于寡头定价理论而非随意的函数形式假设——来同时处理内生性和差异化产品。

11.3.3　模拟积分与数值方法

为什么需要模拟。回顾市场份额方程（11.8）——这是一个 K 维积分，没有解析闭式解（$\boldsymbol{\Sigma} \neq \mathbf{0}$ 时）。解决方案是用蒙特卡洛模拟近似：

$$\hat{s}_{jt} = \frac{1}{ns} \sum_{i=1}^{ns} \frac{\exp(\delta_{jt} + \mu_{ijt}(\mathbf{v}_i))}{1 + \sum_k \exp(\delta_{kt} + \mu_{ikt}(\mathbf{v}_i))}$$

伪随机数 vs 拟随机数。标准的伪随机（pseudo-random）蒙特卡洛以 $O_p(1/\sqrt{ns})$ 的速率收敛——很慢。将模拟精度提高一位（将误差缩小 10 倍），需要将模拟样本量增加 100 倍。Halton 序列（一种**拟蒙特卡洛**，quasi-Monte Carlo 方法）是更好的选择：它不是 "随机" 的——而是确定性地构造的覆盖参数空间的低偏差（low-discrepancy）序列。

为什么 Halton 更好？伪随机数会 "聚簇"（clump）——参数空间的某些子区域被过度覆盖而其他区域稀疏。Halton 序列避免了聚簇，更**均匀**地覆盖空间——125 个 Halton 点提供的覆盖质量约等于 1,000 个伪随机点（Bhat 2001）。在 BLP 模拟积分的背景下：200 个 Halton 抽取通常足以获得精确的积分近似。

固定抽取（Fixed Draws）的至关重要性。在外循环优化（对于不同的 $\boldsymbol{\Sigma}$ 值）中，**必须**对所有 $\boldsymbol{\Sigma}$ 使用同一组伪随机抽取值 $\{\mathbf{v}_i\}_{i=1}^{ns}$。如果每次迭代重新抽样，GMM 目标函数 $Q(\boldsymbol{\Sigma})$ 在不同 $\boldsymbol{\Sigma}$ 值之间的差异将包含**模拟噪声**——同一个 $\boldsymbol{\Sigma}$，使用两组不同的随机抽取值，会得到两个不同的 Q 值。这使得优化器无法可靠地比较不同 $\boldsymbol{\Sigma}$ 值——模拟噪声淹没目标函数的真实梯度。固定抽取确保目标函数的差异仅来自 $\boldsymbol{\Sigma}$ 的变化——而非模拟噪声。

重要性抽样（Importance Sampling）。使用先导一致估计量 $\tilde{\boldsymbol{\Sigma}}$ 构造建议分布 $g(\mathbf{v})$，使得模拟消费者的抽取集中在 "贡献最大的" 参数区域（即那些更可能购买产品的消费者类型），从而减小模拟方差。这是一个标准方差缩减技术——尤其当大部分抽取对市场 "没有贡献"（选择 $j = 0$）时。

11.4　GMM 估计与嵌套不动点（NFP）算法

【进阶】

本节完整推导 BLP 的 GMM 估计框架和 Nested Fixed Point（NFP）算法。这是将前两节的理论创新整合成可实施的估计程序的核心步骤。

11.4.1　矩条件与 GMM 框架

总体矩条件。 BLP 的识别基于工具变量外生性的假设。对于每个产品 j 在市场 t、存在工具变量向量 $\mathbf{Z}_{jt}$ 使得：

$$\mathbb{E}[\xi_{jt}(\boldsymbol{\theta}_0) \cdot \mathbf{Z}_{jt}] = \mathbf{0} \tag{11.12}$$

其中 $\xi_{jt}(\boldsymbol{\theta})$ 是结构残差——不可观测产品特征。注意 ξ_{jt} 不是回归的 "误差"——它是模型的*结构残差*：在真实参数 $\boldsymbol{\theta}_0$ 下，预测和观测的整合。给定 $\boldsymbol{\theta}$ 的任意值，$\xi_{jt}(\boldsymbol{\theta})$ 可通过收缩映射恢复。

样本矩。 令 $N = \sum_t J_t$ 为总的产品-市场观测数。样本矩为：

$$\mathbf{g}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{j,t} \mathbf{Z}_{jt} \cdot \xi_{jt}(\boldsymbol{\theta}) \tag{11.13}$$

GMM 目标函数。 BLP 的 GMM 最小化问题是：

$$\min_{\boldsymbol{\theta}} \; Q(\boldsymbol{\theta}) = \mathbf{g}(\boldsymbol{\theta})' \, \mathbf{W} \, \mathbf{g}(\boldsymbol{\theta}) \tag{11.14}$$

其中 $\mathbf{W}$ 是正定权重矩阵（$M \times M$，$M = \dim(\mathbf{Z})$）。不同的 $\mathbf{W}$ 选择给出不同的估计量——但有一类最优选择。

最优权重矩阵与两步 GMM。 Hansen (1982) 的最优 GMM 理论：当 $\mathbf{W}$ 等于矩条件的方差的逆矩阵时，GMM 估计量达到渐近有效：

- **步骤 1**：使用初始权重矩阵 $\mathbf{W}_1 = (\mathbf{Z}'\mathbf{Z})^{-1}$ 获得一致但非有效的估计量 $\hat{\boldsymbol{\theta}}^{(1)}$
- **步骤 2**：使用 $\hat{\boldsymbol{\theta}}^{(1)}$ 的残差计算 $\hat{\mathbf{S}} = \frac{1}{N} \sum_{j,t} \hat{\xi}_{jt}^2 \mathbf{Z}_{jt} \mathbf{Z}_{jt}'$（如存在市场内聚类，则用市场聚类的方阵）。取 $\mathbf{W}_2 = \hat{\mathbf{S}}^{-1}$，重新最小化 $Q(\boldsymbol{\theta})$ 获得有效的 $\hat{\boldsymbol{\theta}}^{(2)}$

如果存在聚类（同一市场 t 内的观测不是独立的），$\hat{\mathbf{S}}$ 应使用市场聚类稳健公式。

11.4.2 参数分区与线性参数集中化

参数空间的自然分区。 BLP 的参数向量自然地分为：

$$\underbrace{\boldsymbol{\theta}_1 = \text{vec}(\boldsymbol{\Sigma})}_{\text{非线性参数：随机系数方差/协方差}} \qquad \text{和} \qquad \underbrace{\boldsymbol{\theta}_2 = (\bar{\boldsymbol{\beta}}, \bar{\alpha})}_{\text{线性参数：均值系数}}$$

实践中通常进一步限制 $\boldsymbol{\Sigma}$ 为对角的（只有方差，没有协方差）$\to \boldsymbol{\theta}_1 = (\sigma_1, \ldots, \sigma_K)$ 仅有 K 个参数。

集中化的五个步骤。对每个 $\boldsymbol{\theta}_1$ 值（由外循环优化器提议的候选值）：

步骤 1: 收缩映射（内循环）：对每个市场 t，迭代 (11.11) 直到收敛，得到 $\boldsymbol{\delta}_t(\boldsymbol{\theta}_1)$

步骤 2: IV 回归（解析步骤）：给定 $\boldsymbol{\delta}_t(\boldsymbol{\theta}_1)$，线性参数的最小二乘/IV 估计：

$$\hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1) = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\boldsymbol{\delta}(\boldsymbol{\theta}_1)$$

其中 $\mathbf{X} = [\mathbf{x}_{jt}, p_{jt}]$，$\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$（第一阶段的加权矩阵为 $\mathbf{W}_1 = (\mathbf{Z}'\mathbf{Z})^{-1}$；第二阶段使用最优加权 $\mathbf{W}_2 = \hat{\mathbf{S}}^{-1}$ 时，GMM 目标函数使用完整形式 $\boldsymbol{\xi}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\boldsymbol{\xi}$，但线性 IV 步骤中的投影矩阵 $\mathbf{P}_Z$ 保持不变）

步骤 3: 残差计算： $\boldsymbol{\xi}(\boldsymbol{\theta}_1) = \boldsymbol{\delta}(\boldsymbol{\theta}_1) - \mathbf{X}\hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1)$

步骤 4: GMM 目标函数： $Q(\boldsymbol{\theta}_1) = \boldsymbol{\xi}(\boldsymbol{\theta}_1)' \mathbf{Z}\mathbf{W}\mathbf{Z}' \boldsymbol{\xi}(\boldsymbol{\theta}_1)$

步骤 5: 返回 $Q(\boldsymbol{\theta}_1)$ 给外循环优化器

维度降低的魔力。如果 $K = 5$（5 个产品特征）且 $\boldsymbol{\Sigma}$ 是对角的（3 个特征有随机系数，$K_{\text{nl}} = 3$），线性参数 $\boldsymbol{\theta}_2$ 有 7 个分量（5 个特征系数 + 1 个价格系数 + 1 个常数）。集中化将优化搜索空间从可能的 10 维降至 3 维——优化难度随维度指数级增长。

11.4.3　完整的 NFP 算法伪代码

Nested Fixed Point 算法

以下伪代码描述了 BLP 估计的完整 NFP 算法。注意外循环（优化 $\boldsymbol{\Sigma}$）和内循环（收缩映射反解 $\boldsymbol{\delta}$）的嵌套结构——这是 “Nested Fixed Point” 名称的来源。

算法：BLP Nested Fixed Point 估计

输入：数据 $\{\mathbf{s}_t^{\text{obs}}, \mathbf{p}_t, \mathbf{X}_t, \mathbf{Z}_t\}_{t=1}^{T}$，模拟抽取 $\{\mathbf{v}_i\}_{i=1}^{ns}$，内循环容差 τ_{inner}（如 10^{-13}），外循环容差 τ_{outer}

输出：估计量 $\hat{\boldsymbol{\theta}}_1$, $\hat{\boldsymbol{\theta}}_2$，标准误

阶段 1：首次 GMM（初始权重矩阵）

1. 初始化 $\mathbf{W} \leftarrow (\mathbf{Z}'\mathbf{Z})^{-1}$
2. 初始化 $\boldsymbol{\theta}_1 \leftarrow \mathbf{0}$（简单 Logit 作为初始点）
3. **外循环：**使用 BFGS / Nelder-Mead 优化器在 $\boldsymbol{\theta}_1$ 上最小化 $Q(\boldsymbol{\theta}_1)$：
   (a) **内循环（收缩映射）：**对每个市场 t：
   - $\boldsymbol{\delta}_t^{(0)} \leftarrow \ln \mathbf{s}_t^{\text{obs}} - \ln s_{0t}^{\text{obs}}$（简单 Logit 的反解——解析初始值）
   - 重复直到 $\|\boldsymbol{\delta}^{(h+1)} - \boldsymbol{\delta}^{(h)}\|_\infty < \tau_{\text{inner}}$：

     $$\hat{s}_{jt} \leftarrow \frac{1}{ns}\sum_{i=1}^{ns} \frac{\exp(\delta_{jt}^{(h)} + \mu_{ijt}(\boldsymbol{\theta}_1))}{1 + \sum_k \exp(\delta_{kt}^{(h)} + \mu_{ikt}(\boldsymbol{\theta}_1))}$$

     $$\delta_{jt}^{(h+1)} \leftarrow \delta_{jt}^{(h)} + \ln s_{jt}^{\text{obs}} - \ln \hat{s}_{jt}$$

   (b) $\boldsymbol{\theta}_2(\boldsymbol{\theta}_1) \leftarrow (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\boldsymbol{\delta}(\boldsymbol{\theta}_1)$

   (c) $\boldsymbol{\xi}(\boldsymbol{\theta}_1) \leftarrow \boldsymbol{\delta}(\boldsymbol{\theta}_1) - \mathbf{X}\boldsymbol{\theta}_2(\boldsymbol{\theta}_1)$

   (d) $Q(\boldsymbol{\theta}_1) \leftarrow \boldsymbol{\xi}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\boldsymbol{\xi}$

   (e) 返回 $Q(\boldsymbol{\theta}_1)$ 给优化器
4. 获得首次估计 $\hat{\boldsymbol{\theta}}^{(1)}$

阶段 2：有效 GMM（最优权重矩阵）

5. 使用 $\hat{\boldsymbol{\theta}}^{(1)}$ 的残差计算 $\hat{\mathbf{S}} \leftarrow \frac{1}{N}\sum_{j,t}\hat{\xi}_{jt}^2\mathbf{Z}_{jt}\mathbf{Z}_{jt}'$（或市场聚类版本）
6. $\mathbf{W} \leftarrow \hat{\mathbf{S}}^{-1}$
7. 从 $\hat{\boldsymbol{\theta}}_1^{(1)}$ 附近多起点（10-20 个不同的初始值）重新运行外循环
8. 选择产生最低 $Q(\boldsymbol{\theta}_1)$ 的 $\hat{\boldsymbol{\theta}}_1$

阶段 3：推断

9. 计算参数协方差矩阵：$\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}}) \leftarrow \frac{1}{N}(\hat{\mathbf{G}}'\hat{\mathbf{S}}^{-1}\hat{\mathbf{G}})^{-1}$，其中 $\hat{\mathbf{G}} = \partial \mathbf{g}(\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}'$
10. Hansen's J：$J = N \cdot \mathbf{g}(\hat{\boldsymbol{\theta}})'\hat{\mathbf{S}}^{-1}\mathbf{g}(\hat{\boldsymbol{\theta}})$

数值优化器的选择。外循环优化是一个难题——GMM 目标函数可能是非凸的、

有多个局部最小值。常用选项：

- **Nelder-Mead（单纯形法）：**无梯度（derivative-free），对噪声鲁棒，但收敛慢。推荐用于 $\dim(\boldsymbol{\theta}_1) \leq 3$。

- **BFGS/L-BFGS：**基于梯度（拟牛顿），需要平滑的梯度。配合解析梯度（而非数值梯度）使用——在 GMM 目标函数有模拟噪声的背景下，数值梯度极不可靠。

- **多起点策略：**从 10–20 个分散的初始 $\boldsymbol{\Sigma}$ 值出发，取最优结果。这是应对多个局部最小值的标准实践。

11.4.4　MPEC 替代方法

NFP 的核心缺陷。 NFP 的“嵌套”结构——内循环收缩映射外挂外循环优化——产生了两个计算问题：(a) 内循环容差必须极严格（10^{-13}）$\rightarrow$ 每次函数评估需要大量计算（数百次收缩映射迭代 $\times$ ns 次模拟）$\times$ J_t 个产品；(b) 传递给优化器的目标函数包含收缩映射的数值噪声 $\rightarrow$ 梯度估计不可靠。对于 J 很大（>50）或 $\dim(\boldsymbol{\theta}_1)$ 很大（>5）的问题，NFP 变得极其昂贵。

MPEC 的解决方案。 Dubé, Fox & Su (2012) 提出 Mathematical Programming with Equilibrium Constraints (MPEC)：将 $\boldsymbol{\delta}$ 直接作为优化变量（而非通过收缩映射隐式求解），并将市场份额匹配作为优化问题的约束条件：

$$\min_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \{\boldsymbol{\delta}_t\}} \boldsymbol{\xi}' \mathbf{Z} \mathbf{W} \mathbf{Z}' \boldsymbol{\xi} \quad \text{s.t.} \quad \mathbf{s}^{\text{obs}} = \hat{\mathbf{s}}(\boldsymbol{\delta}_t, \boldsymbol{\theta}_1) \tag{11.15}$$

约束条件替代了内循环——$\boldsymbol{\delta}$ 现在是由优化器同时搜索的显式变量（而非通过收缩映射迭代隐式求解）。这使一个高度非线性的嵌套问题转化为一个光滑、稀疏的约束优化问题——适用于现代大规模非线性求解器（如 KNITRO、IPOPT）。Dubé 等报告的加速比为 10–40 倍。

表 11.4: NFP（嵌套不动点）与 MPEC（均衡约束数学规划）算法对比

	NFP（Nested Fixed Point）	MPEC（均衡约束数学规划）
核心思想	内循环收缩映射 + 外循环优化	$\boldsymbol{\delta}$ 作为显式优化变量 + 份额匹配约束
优化器要求	标准梯度/无梯度优化器	大规模非线性求解器（KNITRO, IPOPT）
内循环容差	极严格（10^{-13}）——外循环依赖	无——优化器同时控制所有变量
大 J 可扩展性	差（内循环迭代成本 $\propto J \cdot ns$）	好（稀疏约束 + 结构利用）
易于实现	中等——收缩映射代码独立	较难——需建立约束优化模型
典型运行时间	数小时到数天	数分钟到数小时（10–40× 更快）

11.4.5 假设检验

Hansen's J 过度识别检验。当工具变量数量 M 超过参数数量 P（过度识别），可以检验一些工具变量是否系统地与 ξ 相关——即检验工具变量整体的有效性：

$$J = N \cdot \mathbf{g}(\hat{\boldsymbol{\theta}})' \, \hat{\mathbf{S}}^{-1} \, \mathbf{g}(\hat{\boldsymbol{\theta}}) \xrightarrow{d} \chi^2_{M-P}$$

拒绝 $J\to$ 至少部分工具变量无效，或模型设定错误。不拒绝 $\neq$ "模型正确"——J 检验的功效在弱工具变量下很低。

嵌套模型检验：简单 Logit vs BLP。$H_0 : \boldsymbol{\Sigma} = \mathbf{0}$（简单 Logit 嵌套在 BLP 中）。问题：方差参数在零假设下位于参数空间的边界上（$\sigma_k \geq 0$）。在边界上，标准的 χ^2 渐近分布不适用——需要 Bootstrap 或模拟的临界值。

非嵌套模型检验。BLP vs 嵌套 Logit vs AIDS——这些模型之间不是嵌套关系。Rivers & Vuong (2002) 提供了非嵌套模型的模型选择检验，适用于 GMM 框架。

11.5 识别理论

【进阶】

本节从计量理论角度探讨 BLP 模型的识别——在什么条件下模型参数可以从数据中唯一确定？

11.5.1 Berry & Haile (2014, 2016) 的核心结果

为什么需要识别理论？结构模型的估计量是模型参数和数据的复杂非线性函数。在多大程度上，这些参数是由数据决定的——而不只是由函数形式假设和分布假设决定的？这是识别（identification）问题。

Berry & Haile 的定理（非形式表述）。在存在足够工具变量和充分特征空间变化的条件下，随机系数 Logit 模型是**非参数识别**（nonparametrically identified）的。具体而言：

条件 1: 排除工具变量：存在至少一个成本侧工具变量（cost-shifter）——不被消费者偏好的因素影响但影响边际成本——被排除在需求方程之外（如要素价格、运输成本、关税）。这个排除工具变量提供了识别所需的 "独立变异"（independent variation）。

条件 2: 特征空间的充分变化（"完备性" 条件）：工具变量与 $\boldsymbol{\xi}$ 之间的统计依赖足够 "丰富"，可以从观测选择概率中反推出 $\boldsymbol{\xi}$ 的分布。这是技术性最强的条件——本质上要求工具变量有充分的变异来 "分离" 不同产品的不可观测属性。

条件 3: 外部商品的标准化： $u_{i0t} = 0 + \varepsilon_{i0t}$ 将效用水平固定下来。如果不设外部商品，只能识别相对的效用差分——而无法识别绝对水平。

核心洞察：识别不来自参数分布假设。与许多经济学家的直觉相反，BLP 的识别不来自 "$\mathbf{v}_i$ 是正态的" 或 "ε_{ijt} 是 Type I EV" 这些参数假设。非参数识别的结果是：即使我们不假设任何参数分布形式，模型的核心结构（需求弹性、$\boldsymbol{\xi}$、随机系数的分布）仍然可以从数据中恢复——只要工具变量条件和特征空间变化条件得到满足。参数假设主要是为了估计的可行性——而非识别本身。

已证明可识别的内容。

可识别　随机系数的无条件分布 $F_{\boldsymbol{\beta}}(\cdot)$（在不假设正态的情况下）

可识别　所有 ξ_{jt} 的值（给定参数的反解——可以通过收缩映射恢复）

可识别　替代模式（$\boldsymbol{\beta}_i$ 分布 $\to$ 弹性矩阵）

可识别　均值效用 δ_{jt}

需额外假设　边际成本函数（需要供给端）、动态效应、离散特征的完整识别

11.5.2　弱工具变量问题

非线性 GMM 中弱工具变量的特殊严重性。在非线性 GMM（如 BLP）中，弱工具变量的后果比线性 IV（2SLS）更严重：

- **严重的有限样本偏差：**即使在大样本中，弱 IV 导致的偏差向 OLS 方向的收缩在非线性设定中比线性设定更快——偏差不随 $N \to \infty$ 以 $1/\sqrt{N}$ 速率消失，而是在固定模拟抽取量下可能不消失

- **GMM 目标函数曲面畸变**：弱 IV → GMM 目标函数 $Q(\boldsymbol{\Sigma})$ 变得平坦且在无效的参数值附近有多个虚假局部最小值 → 优化器可能收敛到任意值
- **标准误严重下偏**：假定的 Wald 检验规模严重扭曲——名义上 5% 的检验可能实际拒绝率高达 30–50%

实践建议。第一阶段的 F 统计量（Montiel Olea & Pflueger 2013 有效 F）作为诊断：$F < 10 \to$ 严重警告；$F > 23 \to$ 通常安全。始终报告第一阶段结果和 Stock & Yogo 临界值。对于弱工具变量严重的情况，Stock & Wright (2000) 的 S-set 置信区间（基于连续更新估计量，CUE）提供了弱工具变量稳健的推断方法。

11.6 供给端：Bertrand-Nash 定价与联合估计

【进阶】

BLP 框架可以进一步整合企业的定价行为——将需求估计与供给端的最优定价条件联合估计。这不仅提高了参数估计的效率，还允许直接恢复边际成本——从而能够进行福利和反事实分析。

11.6.1 多产品企业 Bertrand-Nash FOC

企业的利润最大化问题。企业 f 拥有产品子集 $\mathcal{F}_f$。它选择所有这些产品的价格以最大化总利润：

$$\max_{\{p_j : j \in \mathcal{F}_f\}} \sum_{j \in \mathcal{F}_f} (p_j - mc_j) \cdot M \cdot s_j(\mathbf{p})$$

其中 M 是潜在市场总规模。

一阶条件。对 p_j（$j \in \mathcal{F}_f$）求导：

$$s_j(\mathbf{p}) + \sum_{k \in \mathcal{F}_f} (p_k - mc_k) \cdot \frac{\partial s_k}{\partial p_j} = 0$$

定义矩阵 $\boldsymbol{\Omega}$，其元素为：

$$\Omega_{jk} = \begin{cases} -\dfrac{\partial s_k}{\partial p_j}, & j, k \in \text{同一企业} \\ 0, & \text{否则} \end{cases}$$

FOC 的矩阵形式：

$$\mathbf{s}(\mathbf{p}) - \boldsymbol{\Omega}(\mathbf{p})(\mathbf{p} - \mathbf{mc}) = \mathbf{0} \tag{11.16}$$

边际成本的反解：

$$\mathbf{p} - \mathbf{mc} = \boldsymbol{\Omega}(\mathbf{p})^{-1}\mathbf{s}(\mathbf{p}) \quad \Longrightarrow \quad \mathbf{mc} = \mathbf{p} - \boldsymbol{\Omega}(\mathbf{p})^{-1}\mathbf{s}(\mathbf{p}) \tag{11.17}$$

经济学直觉。加成率（$\mathbf{p} - \mathbf{mc}$）取决于两个因素：

(1) 自身需求弹性： $-\partial s_j/\partial p_j$ 越大（自己的需求更有弹性）→ 加成率越低

(2) 内部“竞食”（cannibalization）效应：如果企业 f 的其他产品与 j 有较强的替代关系（$-\partial s_k/\partial p_j$ 大，$k \in \mathcal{F}_f$），j 的提价会将消费者推向同企业的其他产品 → 降低了提价的机会成本 → 加成率高于单产品企业的纳什定价

11.6.2　需求-供给联合 GMM

供给端的矩条件。将反解的边际成本建模为成本因素的函数：

$$\ln mc_{jt} = \mathbf{w}_{jt}\boldsymbol{\gamma} + \omega_{jt}$$

其中 $\mathbf{w}_{jt}$ 是成本驱动因素（要素价格、工厂规模、技术）。ω_{jt} 是成本冲击——与成本侧工具变量正交：$\mathbb{E}[\omega_{jt} \cdot \mathbf{Z}^S_{jt}] = 0$。

联合矩条件堆叠：

$$\mathbf{g}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \frac{1}{N}\sum_{j,t} \begin{pmatrix} \mathbf{Z}^D_{jt} \cdot \xi_{jt}(\boldsymbol{\theta}) \\ \mathbf{Z}^S_{jt} \cdot \omega_{jt}(\boldsymbol{\theta}, \boldsymbol{\gamma}) \end{pmatrix}$$

优势与代价：联合估计有三个优势：(a) 从供给端获得更多矩条件 → 提高了参数估计效率，(b) 成本侧工具变量帮助识别价格系数——部分解决了需求端工具变量弱时的问题，(c) 直接产生边际成本——可直接用于并购模拟和福利分析。代价是：假设 Bertrand-Nash 定价正确——如果实际的竞争模式偏离 Bertrand-Nash（如合谋、Cournot），联合估计会将模型设定错误传播到需求参数的估计中。

11.7　经典实证应用全景

BLP 不仅仅是一个方法论——它推动了一整个实证研究领域。本节通过三个经典应用展示 BLP 如何在实际中运作。

11.7.1　BLP (1995) 原始应用——美国汽车市场

研究问题：美国汽车市场的需求弹性和价格竞争程度。这是 BLP 框架的首次实证演示——也是理解一切后续文献的起点。

数据： 1971–1990 年，~997 个车型-年观测，~20 年。产品特征：马力/重量比（power-to-weight ratio）、尺寸（length × width）、每美元行驶英里数（MPD——反映燃油经济性）、空调（二值）。

工具变量：竞争对手产品特征的总和、同一企业其他产品特征的总和、成本侧排除工具变量（如汇率——日本汽车的价格受日元/美元汇率影响，但与特定车型-年的美国需求冲击无关）。

关键发现：

- 随机系数显著——消费者对马力和尺寸的偏好存在实质性异质性
- 自身价格弹性为 -3 到 -4（比简单 Logit 合理——简单 Logit 倾向于给出不现实的极高弹性）
- 日本品牌的进入显著改变了市场竞争格局——使得美国品牌面临的需求更有弹性（消费者有更多近替代品可选）
- 这一发现具有深远的政策含义——进口配额/关税（如 1981 年的自愿出口限制，VER）对美国汽车价格和消费者福利的影响可以通过 BLP 反事实来量化

11.7.2 Nevo (2001)——即食麦片市场

研究问题：即食早餐麦片市场是否存在市场势力的滥用？四家主要企业（Kellogg's、General Mills、Post、Quaker）控制了大部分品牌——品牌增殖（proliferation）是创新还是进入遏制？

方法创新： Nevo 引入了消费者人口统计特征的互动项（$\mathbf{\Pi} \cdot \mathbf{D}_i$）——通过微观数据（CPS）捕捉收入、年龄等对品牌偏好的影响——为消费者异质性增加了另一维度。

关键发现：

- 价格-成本加成率约 40%——显著的市场势力，远离边际成本定价
- 但在城市层面展开后（利用不同城市间人口统计和市场份额的变异），无法拒绝 Bertrand-Nash 定价的纳什均衡——企业似乎是在纳什定价，而非合谋定价
- 品牌增殖确实是一种进入壁垒：如果 Post 和 General Mills 合并，模拟预测价格上升 3–4%，因为合并企业内部化了对两个品牌集合的定价
- 这一发现直接连接第 ~8~ 章的策略行为理论（进入遏制）和实证量化——BLP 提供了测量策略行为*福利后果*的工具

11.7.3 Petrin (2002)——微型面包车的福利

研究问题： 1984 年引入的克莱斯勒微型面包车（Minivan）为消费者创造了多少福利？这不仅是 IO 问题——它直接关系到国民收入核算（GDP 能否准确反映新产品的福利?）。

方法创新：新产品在引入之前没有价格或市场份额——如何估计对尚未存在的产品的需求？Petrin 的解决方案：使用消费者调查数据（Consumer Expenditure Survey）来估计微型面包车的 "预留价格"（reservation price）分布——即消费者愿意为哪些产品属性集合支付的最高价格。

福利计算公式：引入产品 $J+1$（微型面包车）前后的补偿性变化（Compensating Variation, CV）：

$$CV_i = \frac{1}{\alpha_i}\left[\ln\left(1+\sum_{j=1}^{J+1}\exp(\delta_j^{\text{after}}+\mu_{ij}^{\text{after}})\right)-\ln\left(1+\sum_{j=1}^{J}\exp(\delta_j^{\text{before}}+\mu_{ij}^{\text{before}})\right)\right]$$

这个公式来自 Logit 的消费者剩余表达式——"logsum"（所有效用的指数和的对数之差）的期望就等于消费者剩余的变化。

关键发现：

- 1984–1988 年间，微型面包车的引入创造了约 29 亿美元的消费者剩余——相当于每个购买家庭约 2,500–3,000 美元
- 这一数量级意味着 CPI 可能因为忽略新产品的引入而**高估**了通货膨胀——这就是著名的 Boskin Commission (1996) 的核心论点：CPI 高估通胀约 1.1 个百分点/年，其中新产品偏差是最大的单一成分

表 11.5: BLP 需求估计的三篇经典实证应用对比

	BLP (1995)	**Nevo (2001)**	**Petrin (2002)**
行业	美国汽车	即食麦片	汽车（微型面包车）
核心问题	需求弹性 + 日本竞争	市场势力 + 品牌增殖	新产品的消费者福利
数据	1971–1990，997 obs	1988–1992，225 品牌 × 65 城市	消费者支出调查 + 汽车市场数据
方法贡献	BLP 框架本身	人口统计互动	"预留价格"—新产品需求
核心发现	自身弹性 −3 到 −4；日本进入增加竞争	加成率 ∼40%；品牌增殖是进入壁垒	∼29 亿美元消费者剩余；CPI 高估通胀
政策含义	贸易政策（VER）对价格和福利的影响	并购审查中品牌增殖的角色	国民收入核算应纳入新产品福利

11.8　Nevo (2000) 实践指南：从理论到实现

Avinash Nevo 2000 年的 "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand"（发表于 *Journal of Economics & Management Strategy*）是 BLP 从理论到代码的桥梁。它被广泛引用的原因不在于提出了新理论——而在于提供了明确的、可执行的指导。

11.8.1 数据准备与市场定义

市场定义（Market Definition）。市场 t 通常定义为 "产品-时间-地理位置" 的三元组。例如：美国某城市-某季度-某产品类别。市场定义的选择——包括地理边界和产品类别边界——直接决定了哪些产品之间被假定为相互竞争。定义太宽 $\rightarrow$ 低估替代弹性（因为包括了不相关的产品），定义太窄 $\rightarrow$ 高估弹性（因为排除了真正的替代品）。

潜在市场规模 M_t 的设定。这是实践中最有争议的选择之一。通常的选择包括：总人口、总家庭数、特定年龄群体（如 18–75 岁的成年人）。Nevo 建议进行敏感性分析——将 M_t 倍增、减半、或其他多个合理的值，观察估计结果（尤其是价格弹性和福利估计）对 M_t 的敏感度。稳健的估计应对 M_t 的选择不极端敏感。

份额计算。 $s_{jt} = q_{jt}/M_t$。如果 M_t 被精确设定——s_{jt} 的潜在市场规模必须大到使得所有产品在每个市场中的份额都很小（通常 $s_{jt} < 0.1$）——否则模型逼近饱和市场（$s_0 \approx 0$），收缩映射会极其缓慢。

11.8.2 人口统计互动与完整异质性

完整异质性的两种来源。 Nevo 强调，消费者的偏好异质性来自两个互补的来源：

$$\mu_{ijt} = \underbrace{\sum_{r} \sigma_r \cdot x_{jrt} \cdot v_{ir}}_{\text{不可观测异质性（随机系数）}} + \underbrace{\sum_{k} \sum_{\ell} \pi_{k\ell} \cdot x_{jkt} \cdot D_{i\ell}}_{\text{可观测异质性（人口统计互动）}}$$

第一项（$\boldsymbol{\Sigma}\mathbf{v}_i$）捕捉不可观测的偏好差异。第二项（$\boldsymbol{\Pi}\mathbf{D}_i$）利用可观测的人口统计特征（如收入对价格敏感度的影响、家庭规模对车辆尺寸偏好的影响、年龄对技术特征的偏好等）。人口统计数据通常来自 CPS（Current Population Survey）、Census 等——按市场（城市/区域）匹配。引入人口统计互动有两个好处：(a) 增加了额外的异质性维度——替代模式因此更加灵活，(b) 因为这些互动基于可观测数据（而非纯随机抽取）——它们可能提供更强的识别。

11.8.3 从估计到政策分析

并购模拟的预热。 BLP 需求估计的终极目标是反事实分析——而最常见的反事实是并购。并购模拟的基本框架（将在第 ~12~ 章中完全展开）是：

步骤 1: 用 BLP 估计并购前的需求（弹性矩阵 $\partial s_j/\partial p_k$）和边际成本（$mc_j = p_j - [\boldsymbol{\Omega}^{-1}\mathbf{s}]_j$）

步骤 2: 改变所有权矩阵：$\boldsymbol{\Omega}^{\text{pre}}$（并购前——两家企业独立定价）$\rightarrow$ $\boldsymbol{\Omega}^{\text{post}}$（并购后——合并企业最大化联合利润）

步骤 3: 在并购后的产业中，求解新的 FOC：$\mathbf{p}^{\text{post}} = \mathbf{mc}^{\text{pre}} + [\boldsymbol{\Omega}^{\text{post}}(\mathbf{p}^{\text{post}})]^{-1}\mathbf{s}(\mathbf{p}^{\text{post}})$——找到并购后均衡价格

步骤 4: 价格变动 $\Delta\mathbf{p} = \mathbf{p}^{\text{post}} - \mathbf{p}^{\text{pre}}$ 是**单边效应**（unilateral effects）的度量——并购对价格的直接影响，无需假设合谋

实现中的常见陷阱：

表 11.6: BLP 估计实现中的常见陷阱、诊断信号与解决方案

陷阱	诊断信号	解决方案
内循环容差不足	GMM 目标函数随迭代有波动，无法可靠优化	收紧到 10^{-12}–10^{-14}；考虑 MPEC
局部最优	多起点给出不同的 $\hat{\boldsymbol{\Sigma}}$ 和相似的 Q	从 20+ 个起点重新运行；使用全局优化器（如模拟退火）
弱工具变量	$\hat{\alpha}$ 非常小（需求无弹性）；参数估计不稳定	添加 Differentiation IVs/Hausman IVs；报告第一阶段 F；使用弱 IV 稳健推断
市场定义太窄/太宽	弹性估计与预期偏离；交叉弹性矩阵不合理	敏感性分析——在多个市场定义下重新估计
模拟噪声污染目标函数	即使 $\boldsymbol{\Sigma}$ 固定，Q 在不同运行中不同	增加 ns（1000+）；使用 Halton 序列；固定抽取

11.9　拓展与变体

11.9.1　嵌套 Logit 与 BLP 的比较

两种不同的 IIA 打破策略。嵌套 Logit（McFadden 1978）和 BLP 通过完全不同的机制解决 IIA 问题：

- **嵌套 Logit：**由研究者事前指定产品如何分组（巢结构）。同一巢内产品之间替代弹性较大；跨巢替代较小。替代模式由分组决定——而非数据中的特征差异。
- **BLP/随机系数 Logit：**替代模式由数据中的特征空间距离决定——无需指定分组。产品的替代性来自它们特征的相似性。

何时嵌套 Logit 就足够？当有清晰、自然的分组结构时（如产品天然分为不同类别——轿车 vs SUV vs 皮卡），嵌套 Logit 可能是合理且更简单的选择。它只需要一个附加参数（σ——巢内相关性），计算上是闭式的（无需模拟积分），且不需要收缩映射。但在替代模式跨越多个维度时（如汽车既是 “豪华” 又是 “大马力” 又是 “日本品牌”），BLP 随机系数的灵活性无法被单一分组结构捕捉。

11.9.2 纯特征模型（Pure Characteristics Model）

去除 Logit 误差（ε_{ijt}）。 Berry & Pakes (2007) 提出了纯特征模型——在效用中没有 ε_{ijt} 项。消费者完全确定性地选择最大化 U_{ij} 的产品。这消除了“随机口味冲击”的经济学假设——所有效用差异来自可观测和不可观测的产品特征。

优势与挑战。优势：更“纯粹”的经济学解释（消费者是完全理性的效用最大化者）。挑战：(a) $\boldsymbol{\delta}$ 的反解不保证是收缩——算法可能不收敛，(b) 某些产品可能零份额（没有人选择它们），这使估计复杂化，(c) 计算极其昂贵——没有 Logit 结构的简化。

11.9.3 动态 BLP

耐用品的“选择惯性”。汽车、手机、家电等耐用品市场中的消费者决策是动态的——今天的购买影响明天的选择集（因为今天买车意味着明天无需再买一辆车）。Gowrisankaran & Rysman (2012) 在数码相机市场中展示了动态 BLP：消费者是前瞻性的（forward-looking）——预期未来价格会下降和新产品会推出——这改变了当前的需求弹性。

11.9.4 微观与宏观数据的整合

Micro BLP (Berry, Levinsohn & Pakes 2004)。当可获得微观消费者调查数据时（如消费者支出调查、消费者面板），可以将其与市场层面数据整合——微数据提供消费者异质性的直接信息，宏数据提供市场结构和产品特征。Conlon & Gortmaker (2025) 在 PyBLP 的最新版本中支持微-宏数据的联合估计。

表 11.7: 需求模型比较：简单 Logit、嵌套 Logit、BLP/RC Logit 与纯特征模型

	简单 Logit	嵌套 Logit	BLP/RC Logit	纯特征
替代灵活性	无（IIA）	有限（巢内 IIA）	灵活（特征空间决定）	灵活（无 Logit 误差约束）
参数个数	$K+1$	$K+2$	$K+\dim(\boldsymbol{\Sigma})$	K + 异质性分布
计算难度	极低（闭式 + 线性）	低（GMM）	中-高（NFP/MPEC + 模拟积分）	高（无 Logit 简化）
需要模拟	否	否（闭式）	是	是
预设分组	N/A	需要	不需要	不需要

11.10　现代工具与前沿

11.10.1　PyBLP (Conlon & Gortmaker 2020)

标准化的现代 BLP 实现。 Conlon & Gortmaker (2020) 在 *RAND Journal of Economics* 上介绍了 PyBLP——一个开源的 Python 包，按照最佳实践实现了 BLP 和随机系数 Logit 需求模型。它对 BLP 方法论的标准化和可及性具有无与伦比的影响——使得研究者可以从 "写自己的 NFP 循环"（容易出错且效率低）切换到经过良好测试、高效的标准实现。

核心特性：

- **自动微分（Automatic Differentiation）：**使用数值计算库（如 JAX/PyTorch）提供准确的参数梯度——避免了数值梯度的噪声问题和手工推导解析梯度的烦琐
- **最优工具变量：**自动构建可行最优 IV（基于先导估计的近似 $\mathbb{E}[\partial\xi/\partial\theta \mid x, w]$）
- **并行计算：**对多市场收缩映射进行并行化——多个 CPU 核同时处理不同的数据分区
- **供给端联合估计：**集成 Bertrand-Nash FOC + 成本函数——一步完成需求-供给估计
- **PyBLP 思维模型：**'Problem' 类定义产品数据，'Problem.solve()' 一条命令完成 BLP 估计

11.10.2　其他实现工具比较

表 11.8: BLP 需求估计的主要实现软件比较

软件	语言	主要特性	最优 IV	供给端
PyBLP	Python	自动微分、最优 IV、并行、活跃开发	是	是
blp	Stata	Vincent (2015)、Stata 环境下使用	否	否
blpestimator	R	BLP 的 R 实现	否	否
Nevo 代码	MATLAB	Nevo (2000) 配套代码	否	否
BLPDemand.jl	Julia	高性能 Julia 实现	部分	部分

11.10.3　当前研究前沿

非参数 BLP。 Compiani (2021) 提出使用机器学习方法（如神经网络）灵活地估计消费者偏好的分布，放宽对正态/参数分布的依赖。

深度学习 + BLP。神经网络可以替代传统的随机系数结构——直接学习 $F_{\boldsymbol{\beta}}(\cdot)$ 的非参数形式而不假定分布族。这是早期阶段的研究，可能在未来十年大幅改变 BLP 估计的面貌。

大数据 + BLP。在线价格爬取（web scraping）和数字化交易数据使得超大面板（数百万个价格-产品-市场观测）成为可能。这要求新的计算策略——分布式 BLP 估计、随机梯度下降优化、和在线收缩映射。

11.11 局限、反思与通往 Ch.12 的桥梁

11.11.1 BLP 的主要局限

1. 强函数形式假设。

- Type I Extreme Value 的 ε_{ijt}：随机系数部分打破 IIA，但不完全——在 ε 维度上，IIA 仍然适用（个体内的选择概率之比只取决于这两个产品）

- 正态性的 $\mathbf{v}_i$：假定偏好是单峰、对称的——不能捕捉多峰群体（如市场中同时存在 “品牌忠诚者” 和 “价格敏感者”）

- 线性加性特征效用：不能捕捉特征交互（如马力与重量之间的互补性）

2. BLP 工具变量的局限。

- “内生产品特征”——如果企业也根据 ξ_{jt} 的预期来优化特征（如选择马力水平），标准 BLP 工具变量的外生性会被破坏

- “内部 IV”——所有 IV 都从模型内部特征构造，没有市场的 “外部” 冲击 $\rightarrow$ 限制了识别 “外部” 价格变化产生的一般均衡效应

3. 计算负担。即使有 PyBLP 和 MPEC，完整的 BLP 估计仍然可能需要数小时到数天的计算——限制了模型复杂度和稳健性检查。

4. 市场定义敏感性。估计结果对 M_t 和市场地理边界的选择敏感——而对不同的合理选择没有明确的指导准则。

表 11.9: BLP 需求估计方法的主要局限与缓解策略

局限	严重度	具体表现	缓解/替代方法
函数形式假设	中等	Logit 误差假设可能扭曲弹性结构	嵌套 Logit、Probit、纯特征模型、非参数 BLP
BLP IV 弱	高（常见）	第一阶段 F 不足，估计不精确	Differentiation IVs、Hausman IVs、最优 IV、供给端联合
市场定义敏感性	中等	弹性和福利估计随 M_t 设定变化	敏感性分析、使用多种外部数据
内生产品特征	高（被忽视）	破坏了 IV 外生性的时序假设	动态特征决策模型、外部工具变量
价格的内生性持续	高（基础性）	即使有 BLP IVs，ξ 和价格的相关性可能不完全消除	更多/更强的工具变量、半参数

11.11.2　BLP 的方法论遗产

两篇 NEIO 里程碑：Bresnahan (1982) & BLP (1995)。如果说 Bresnahan 教会了我们如何在*同质产品*市场中从市场数据中识别市场势力——通过需求旋转和 θ 参数，BLP 则教会了我们如何在*差异化产品*市场中做同样的事情——通过随机系数和 GMM。两者共同构成了 NEIO 的完整方法论工具箱：

- **Bresnahan (1982)**：同质产品、θ 行为参数、需求旋转 $\rightarrow$ 识别竞争行为
- **BLP (1995)**：差异化产品、随机系数 Logit、收缩映射 + 工具变量 $\rightarrow$ 识别需求系统和边际成本

BLP 的核心教训（为整个 NEIO 模块总结）：

(1) 经济理论提供结构性联系：寡头 FOC（Bertrand-Nash 定价）将可观测的价格和不可观测的成本联系起来——没有这个结构，我们只能观测价格-成本相关关系，无法进行反事实推断

(2) "不可观测" 变量可以通过结构反解恢复：ξ_{jt}（需求冲击）和 mc_{jt}（边际成本）在数据中不可观测——但通过模型的结构（收缩映射、FOC 反解），它们可以从可观测变量中恢复

(3) 工具变量的合理性取决于对行业制度的深入理解： BLP 工具变量有效不是因为任何统计检验——而是因为汽车的设计（选择特征）和定价（设定价格）在时间上有先后之分

(4) **结构估计的核心价值是反事实分析**：BLP 给的不仅是“需求的弹性是多少”——而是当我们改变所有权、引入新产品、或施加一项新税时，价格、产品种类、消费者福利会如何改变

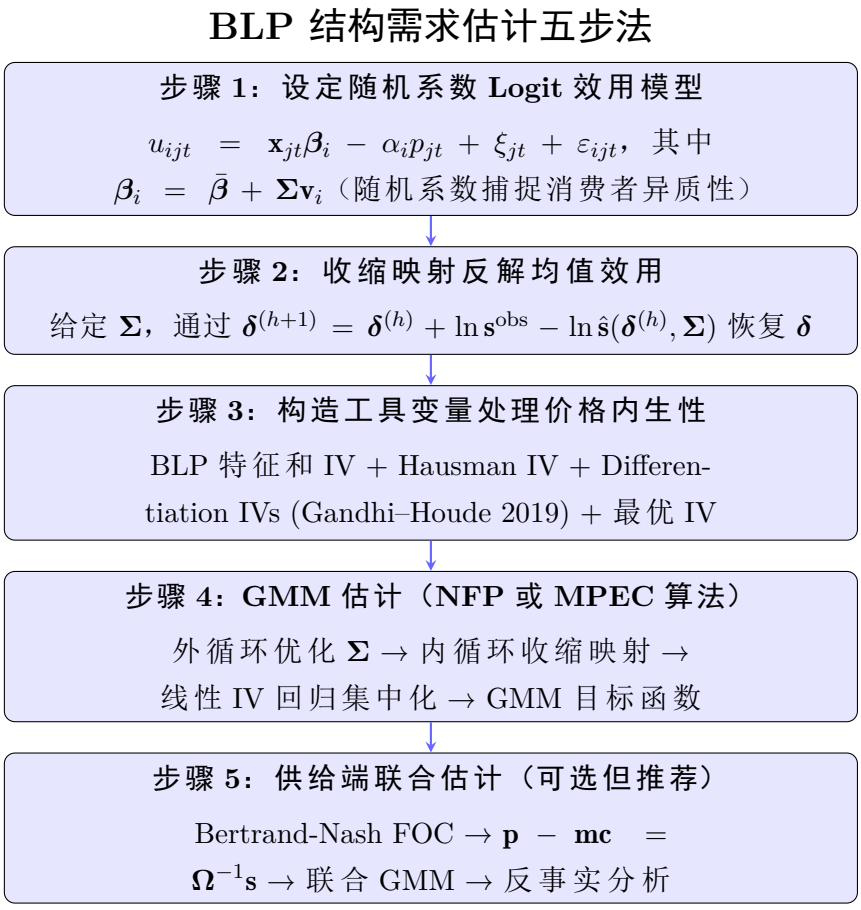


图 11.3: BLP 结构需求估计五步法：从效用设定到反事实分析的完整链条。每一步对应本章的核心节。与 Ch.10 的 NEIO 四步法和 Ch.8 的 Tirole 七步法共同构成全书的三大方法论框架。

11.11.3 通往 Ch.12 的桥梁

BLP 估计了我们需要并购模拟的一切。完成本章后，你已经掌握了估计以下内容的方法和原理：

- **需求系统：** $J \times J$ 的完整价格弹性矩阵——知道每个产品的提价会如何影响所有其他产品的需求
- **供给系统：**每个产品的边际成本和加成率——知道每个产品当前赚取多少超过成本的价格

有了这两个组件，我们可以进行完整的反垄断分析：

$$\text{BLP 需求估计} \longrightarrow \begin{cases} \text{并购模拟} & \text{——预测并购对价格和福利的影响} \\ \text{SSNIP 测试} & \text{——界定反垄断相关市场} \\ \text{买方势力} & \text{——评估合并买方对抗卖方的能力} \\ \text{新产品福利} & \text{——衡量创新为消费者创造的价值} \end{cases}$$

第 ~12~ 章将展示如何将这整个工具链整合起来——从市场层面数据到反垄断政策建议。那是整个产业组织学实践的顶峰——使用经济学理论和计量方法来回答直接影响数十亿人日常生活的实际问题。

练习题

基础题

Ex 11.1 —IIA 性质及其含义。假设一个简单 Logit 模型中有三个产品（A, B, C），均值效用分别为 $\delta_A = 2, \delta_B = 1, \delta_C = 0.5$（外部商品标准化为 $\delta_0 = 0$）。计算提价前每个产品的市场份额。然后假设产品 A 的价格上升，使得其市场份额降低了一半——计算提价后各产品的份额和从 A “转移” 出消费者的分布。展示简单 Logit 中转移消费者与各产品原份额成比例（IIA 的体现）。

Ex 11.2 —意愿支付（WTP）的计算。在 BLP 模型中，消费者 i 对特征 r 的（边际）意愿支付为 $\text{WTP}_{ir} = \beta_{ir}/\alpha_i$。给定 $\bar{\beta}_{\text{hp}} = 0.5$, $\sigma_{\text{hp}} = 0.3$, $\bar{\alpha} = 0.02$, $\sigma_p = 0.005$，和一个消费者 i 的抽取向量 $\mathbf{v}_i = (1.2, 0.8)$（分别对应马力和价格）。计算这个消费者对马力的意愿支付。另一消费者有 $\mathbf{v} = (-0.5, 1.5)$——解释为什么这个消费者对马力的意愿支付更低。

Ex 11.3 —收缩映射的一步迭代。考虑一个市场有 3 个产品和外部商品。观测份额为 $s^{\text{obs}} = (0.3, 0.2, 0.1)$（$s_0 = 0.4$）。假设 $\delta = (1.5, 0.8, -0.2)$（仅为简例——忽略模拟）。给定收缩映射 $\delta_j^{(h+1)} = \delta_j^{(h)} + \ln s_j^{\text{obs}} - \ln \text{Logit}(\delta_j^{(h)})$，计算一步迭代后的 $\delta^{(1)}$。验证新 δ 下的市场份额是否更接近观测值。

Ex 11.4 —BLP 工具变量的有效性论证。考虑美国汽车市场（如 BLP 1995）。阐述为什么 “竞争对手产品的马力之和” 可能是一个有效的工具变量：(a) 揭示其与价格的相关性来源（寡头定价理论），(b) 讨论它与 ξ_{jt} 可能不相关的理由（为什么竞争对手的马力与你的不可观测质量无关？），(c) 提出一个情景，在这个情景中这个工具变量可能无效。

进阶题

Ex 11.5 —有内生价格的简单 Logit 的 IV 估计。在简单 Logit 中，需求方程可以转换为 $\ln s_j - \ln s_0 = x_j\beta - \alpha p_j + \xi_j$。假设你面对价格内生性，并具有有效的工具变量 Z（如 BLP sums-of-characteristics IVs）。推导此设定下的 2SLS 估计量。

Ex 11.6 —GMM 目标函数曲面的模拟。考虑一个只有一个随机系数的简化 BLP（σ_p ——价格敏感度的标准差，所有其他参数固定）。使用模拟数据，生成 $Q(\sigma_p)$ 对 $\sigma_p \in [0, 2]$ 的图。讨论目标函数曲面的形状：它是凸的吗？是否存在多个局部最小值？在真实 σ_p^* 附近的值处有什么特征？

Ex 11.7 —从弹性恢复边际成本。给定 BLP 估计的弹性矩阵和一个假设的 Bertrand-Nash 行业，恢复边际成本。使用以下假设弹性矩阵：$\partial s_j/\partial p_j = -0.08$, $\partial s_j/\partial p_k = 0.02$（对 $k \neq j$），$p_j = 20$，$s_j = 0.1$（对两个对称的企业，各有一个产品）。计算每个企业的勒纳指数 $(p - mc)/p$。在合并后（两个产品现在属于同一家企业），勒纳指数应如何变化？

应用/实证题

Ex 11.8 —并购模拟入门。考虑两家企业，各有一个产品。产品 1 和 2 是对称的替代品。使用 BLP 弹性矩阵计算出并购前后的均衡价格。并购前的价格是 $p_1 = p_2 = \$100$，$mc_1 = mc_2 = \$60$，$\partial s_1/\partial p_1 = -0.012$，$\partial s_1/\partial p_2 = 0.004$（$s$ 在分享、而非百分比中）。求解并购前各企业的 FOC 并验证初始的均衡假设。然后计算并购后的均衡价格——合并企业现在对两个产品同时定价。讨论单边效应的大小——并购本身（不假设合谋）是否足以显著提高价格？

Ex 11.9 —外部商品份额的敏感性。重新审视 Ex 11.8。将 M_t 倍增（即将 s_0 从 0.5 提高到 0.75）——这意味着更多的人 “不购买” 汽车。重新计算价格弹性。为什么 s_0 变大时弹性变大？讨论外部商品份额在识别中的作用——它如何锚定了需求水平？

Ex 11.10 —文献批判阅读。阅读 Nevo (2001) 的表 IV（或您可获得的任何 BLP 实证论文的弹性表）。选择三个交叉弹性并解释它们的含义：哪两个产品之间的替代最强？这一替代模式是否符合您的经济学直觉？如果交叉弹性出现了意外（如两个看似差异很大的产品之间有很高的交叉弹性），这可能意味着：(a) 不可观测的共同偏好，(b) 模拟噪声，(c) 还是市场/产品定义的错误？写一页讨论。

推荐阅读

核心必读

[1] **Berry, S., Levinsohn, J. & Pakes, A. (1995).** “Automobile Prices in Market Equilibrium.” *Econometrica*, 63(4), 841–890. —BLP 的原始论文。Chapter 11 的所有核心内容都在这里。阅读重点：section 2（模型设定）、section 3（估计）、section 5（实证结果）。

[2] **Berry, S. (1994).** "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics*, 25(2), 242–262. —收缩映射（contraction mapping）的证明。如果你只读 BLP (1995)，你可能不会完全理解为什么内循环收敛——这篇给出完整的数学证明。

[3] **Nevo, A. (2000).** "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand." *Journal of Economics & Management Strategy*, 9(4), 513–548. —从理论到实现的桥梁——最接近 "BLP 操作手册" 的文献。

[4] **Nevo, A. (2001).** "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2), 307–342. —BLP 框架的最具影响力的实证应用之一——展示了如何将 BLP 用于实际的政策问题。

进阶阅读

[5] **Berry, S. T. & Haile, P. A. (2014).** "Identification in Differentiated Products Markets Using Market Level Data." *Econometrica*, 82(5), 1749–1797. —非参数识别理论——BLP 模型的识别在什么条件下成立？

[6] **Petrin, A. (2002).** "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, 110(4), 705–729. —新产品福利——如何使用 BLP 估计尚未存在的产品的需求？

[7] **Conlon, C. & Gortmaker, J. (2020).** "Best Practices for Differentiated Products Demand Estimation with PyBLP." *RAND Journal of Economics*, 51(4), 1108–1161. —现代 BLP 的最佳实践——PyBLP 的设计理念和使用。

[8] **Dubé, J.-P., Fox, J. T. & Su, C.-L. (2012).** "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation." *Econometrica*, 80(5), 2231–2267. —MPEC 方法——将 NFP 的嵌套优化转化为约束优化，大幅加速。

方法论与教材

[9] **Train, K. E. (2009).** *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge University Press. —离散选择方法的权威教材。第 6 章（混合 Logit）是理解 BLP 的基础。

[10] **Davis, P. & Garcés, E. (2010).** *Quantitative Techniques for Competition and Antitrust Analysis*. Princeton University Press. —反垄断量化分析的实用手册——展示了 BLP 如何融入完整的反垄断分析。

[11] **Berry, S. T. & Haile, P. A. (2016).** "Identification in Differentiated Products Markets." *Annual Review of Economics*, 8, 27–52. —识别条件的综述——比 Berry & Haile (2014) 更易读。

[12] **Ackerberg, D., Benkard, L., Berry, S. & Pakes, A. (2007).** "Econometric Tools for Analyzing Market Outcomes." In *Handbook of Econometrics*, Vol. 6A, Ch.63. —IO 计量方法的手册章——对 NEIO 工具的全面回顾。

拓展

[13] **Gandhi, A. & Houde, J.-F. (2019).** "Measuring Substitution Patterns in Differentiated Products Industries." NBER Working Paper. —Differentiation IVs：改善 BLP 工具变量的弱相关性。

[14] **Reynaert, M. & Verboven, F. (2014).** "Improving the Performance of Random Coefficients Demand Models: The Role of Optimal Instruments." *Journal of Econometrics*, 179(1), 83–98. —最优工具变量在 BLP 中的应用。

[15] **Fan, Y. (2013).** "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market." *American Economic Review*, 103(5), 1598–1628. —双边市场的 BLP——如何将 BLP 扩展到平台经济学。

[16] **Gowrisankaran, G. & Rysman, M. (2012).** "Dynamics of Consumer Demand for New Durable Goods." *Journal of Political Economy*, 120(6), 1173–1219. —动态 BLP：耐用消费品的动态需求估计。

中国相关资料

[17] **肖俊极、孙洁 (2012).** "消费税和燃油税的有效性比较分析." 经济学季刊, 11(4). —中国汽车市场的 BLP 需求估计和税收政策反事实。

[18] **吕振通、胡伟民、朱东明.** "中国新购汽车市场的需求估计与并购模拟分析." —中国汽车市场的 BLP 应用——包含并购模拟。

[19] **肖俊极、陈俊廷、邢剑炜 (2023).** "新能源汽车市场的本地偏好与地方保护." 数量经济技术经济研究, 第 11 期. —中国新能源汽车市场的离散选择模型——地方保护和地方偏好的估计。

[20] **陈强 (2014).** 高级计量经济学及 *Stata* 应用（第二版）. 高等教育出版社. —第 20 章：离散选择模型——中文教材中对离散选择方法最完整的介绍。

软件资源

[21] **PyBLP:** https://pyblp.readthedocs.io/ —现代 BLP 估计的 Python 标准库。

[22] **Train Mixed Logit:** https://eml.berkeley.edu/~train/software.html —Kenneth Train 的离散选择和模拟的资源页面——包含多种 Stata/MATLAB 代码。

[23] **Berry, S., Levinsohn, J. & Pakes, A. (1995).** 代码——BLP 原始论文的复制材料可在 https://doi.org/10.3982/ECTA6236 找到（2025 年扩展版）。

第十二章　并购模拟、市场界定与反垄断损害量化

章首语

反垄断法禁止“实质性减少竞争”（Substantially Lessen Competition, SLC）的并购。但法官如何知道一个并购案是否会实质性减少竞争？这个问题需要经济学家跨越两个世界：学术界期刊论文中的结构模型与严谨识别策略，以及法庭上可理解、可复现、可交叉验证的量化证据。

本章是 NEIO 三章序列（第 ~10~ 章–第 ~11~ 章–本章）的终点与应用顶点。第 ~10~ 章解决了“这个市场是竞争、Cournot 还是合谋？”的问题——提供了行为参数 θ 和并购前的基准行为假设。第 ~11~ 章解决了“差异化产品的需求系统是什么样的？”的问题——提供了弹性矩阵 ε_{jk} 和从 FOC 恢复的边际成本 $\widehat{mc}_j$。本章将这些输出整合为可操作的反事实引擎：改变所有权结构（$\mathbf{\Omega}$ 矩阵），求解新均衡，预测并购的价格效应与福利效应。

贯穿全章的是一条方法论的红线：经济学工具的“简化阶梯”（Ladder of Simplification）。最简单的指标是 HHI 和集中度变化——SCP 传统的遗留工具，只需要市场份额数据。中间一级是 GUPPI 和 UPP——将并购的竞争效应压缩为两个数字（分流比 + 毛利率），绕过市场界定。最复杂的是完整结构性并购模拟——使用 Ch.11 的 BLP 需求系统，求解并购后的新 Nash 均衡。每一级都增加计算复杂度和数据需求，但也增加准确性和政策可信度。在实践中，反垄断当局通常从简单工具开始筛选（“初步审查”），只在“红旗”案件中使用复杂的结构模型（“第二阶段深入调查”）。

12.1　引言：反垄断为什么需要量化分析？

12.1.1　反垄断的经济学逻辑

反垄断法（Antitrust Law / Competition Law）的核心经济学问题是：**“这个并购案是否可能会实质性减少竞争？”**（Substantially Lessen Competition, SLC）。回答这个问题需要的不是法律条文的推演——需要的是**预测并购后价格、产出和品质变化的**

经济学模型。经济学提供了从定性判断（“这个并购看起来有问题”）上升到定量预测（“这个并购将使价格上升 3.2–5.8%”）的工具。

横向并购（Horizontal Merger）——同一相关市场中直接竞争者之间的并购——是反垄断审查最关注的对象。与纵向并购（Vertical Merger，上下游企业之间的并购）和混合并购（Conglomerate Merger，无关行业中企业之间的并购）不同，横向并购直接消除了两家竞争者之间的竞争。这种“竞争的消除”（Elimination of Competition）是并购反垄断的核心关切。

经济学分析并购的竞争效应，本质上是一个**反事实**（Counterfactual / But-For World）问题：我们需要比较两个世界——“有并购的世界”（实际世界，或预测的实际世界）和“没有并购的世界”（反事实世界，即 but-for world）。两者的价格、产出、消费者剩余和总福利的差距，就是并购的**竞争效应**。

传统上，反垄断分析遵循一个“三步法”（Three-Step Approach）：

(1) **市场界定**（Market Definition）：确定“相关市场”（Relevant Market）的范围——哪些产品和企业对消费者施加了显著的竞争约束；

(2) **市场份额计算与集中度推定**：计算并购双方和相关竞争者在相关市场中的市场份额，以及并购前后的 HHI（Herfindahl-Hirschman Index）变化；

(3) **竞争效应分析**：基于市场份额和集中度变化，结合进入壁垒、买方势力等因素，评估并购是否可能实质性减少竞争。

这个三步法的逻辑是清晰的——但有两个根本缺陷。第一，**市场界定本身就是内生的**：“相关市场”是那些对消费者施加显著竞争约束的产品/企业——而竞争约束**正是并购分析需要衡量的**。如果我们可以直接衡量竞争约束（如通过分流比），为什么还需要先界定市场？第二，**市场份额和 HHI 是非结构参数**：它们不传达关于竞争实际上如何运作的信息。两家市场份额各 10% 的企业——如果生产极为近似的产品——合并带来的竞争伤害可能远超两家市场份额各 20% 但产品高度差异化的企业的合并。

HHI 的优点与局限

优点：简单、透明、可预测——任何人都可以用市场份额数据计算 HHI，结果是可复现的。这使得 HHI 成为反垄断执法中的“安全港”（Safe Harbor）：如果并购后的 HHI 低于 1,500 且变化小于 100（美国 2010 年指南），并购通常不会被质疑。

局限：HHI 忽略了**分流模式**（Pattern of Substitution）——即消费者在不同产品之间如何替代。如果并购双方的产品是“特别近的竞争者”（异常高的分流比），即使双方的市场份额很小，并购也可能产生显著的价格效应。反之，如果并购双方的产品高度差异化（异常低的分流比），即使双方的市场份额很大，并购也可能没有显著的竞争关注。

传统三步法与 NEIO 方法的对比反映了一个根本性的分析范式转变：

	传统 SCP 方法	**现代 NEIO 方法**
第一步	界定相关市场（定性判断）	估计需求系统（BLP / 嵌套 logit）
第二步	计算市场份额和 HHI	计算分流比 D_{jk} 和弹性 ε_{jk}
第三步	基于 HHI 变化推定竞争效应	并购模拟 / UPP / GUPPI 直接量化竞争效应
数据需求	市场份额（通常来自行业报告）	产品级数据：价格、特征、市场份额
核心工具	HHI、集中度阈值（安全港）	分流比、UPP/GUPPI、并购模拟、弹性矩阵
关键局限	忽略分流模式——市场份额相同的两个并购可能有截然不同的竞争效应	数据和计算密集——不适合初始快速筛选
政策角色	初步筛选（“安全港”：如果 HHI <1500 且 ΔHHI <100）	深入调查（“第二阶段” ——如果初步筛选发现 “红旗”）

表 12.1: 传统 SCP 方法与现代 NEIO 方法的对比。注意：两者不是 “非此即彼” 的——在实践中，反垄断当局使用分层方法：先使用 HHI 筛选，只有在 “红旗” 案件中使用 UPP/GUPPI 和并购模拟进行深入分析。

12.1.2 来自 Ch.10–11 的工具箱

本章的量化工具建立在前面两章提供的结构性输入之上。三者构成一个**顺序知识链**：

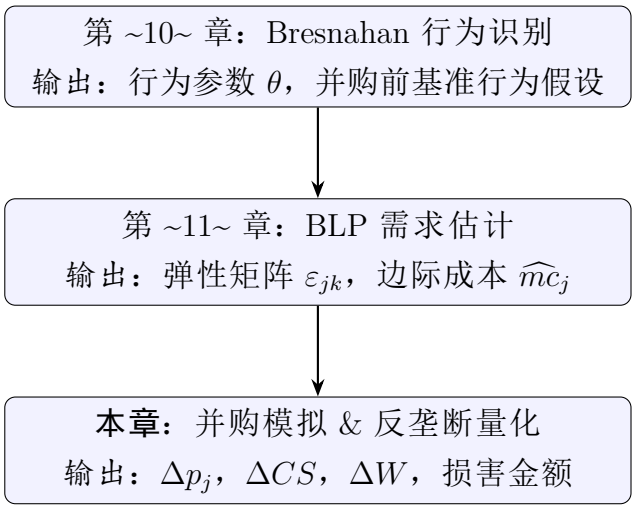


图 12.1: NEIO 三章工具链：从行为识别到需求估计到反垄断预测。第 ~10~ 章验证并购前的行为假设（Bertrand-Nash 是否成立?）。第 ~11~ 章提供并购模拟的两个关键输入——弹性矩阵和边际成本。本章将两者整合，求解并购后的新均衡并量化竞争效应。

第 ~10~ 章（Bresnahan / 行为识别）提供的是**基准行为假设**。并购模拟假设并购前的定价行为是 Bertrand-Nash（企业通过价格竞争，视对手价格为给定）。但如果 Ch.10 的 $\hat{\theta}$ 显著偏离 1（Bertrand），并购模拟的整个框架需要调整——因为边际成本的恢复（Step 2）依赖 FOC 的形式，而 FOC 的形式又依赖行为参数 θ。

第 ~11~ 章（BLP / 需求估计）提供的是**并购模拟的 "燃料"**——弹性矩阵 ε_{jk} 和从 FOC 恢复的边际成本 $\widehat{mc}_j$。没有 ε_{jk}，我们无法计算分流比（需要交叉弹性 ε_{kj}）；没有 $\widehat{mc}_j$，我们无法求解并购后的 FOC（需要知道加成率）。一句话：需求估计的质量决定了并购模拟的可靠性的上界。

本章的任务是**改变所有权结构**：保持需求系统和边际成本不变（"结构参数不变"的假设），改变所有权矩阵 $\mathbf{\Omega}$（谁拥有哪些产品），求解新均衡，比较新旧均衡的价格、数量和福利。这正是**结构模型**的核心优势——参数在 "政策变化"（此处为所有权变化）下保持不变，因此可以进行可靠的反事实模拟。

12.1.3　本章结构

本章按照反垄断并购审查的逻辑顺序组织：

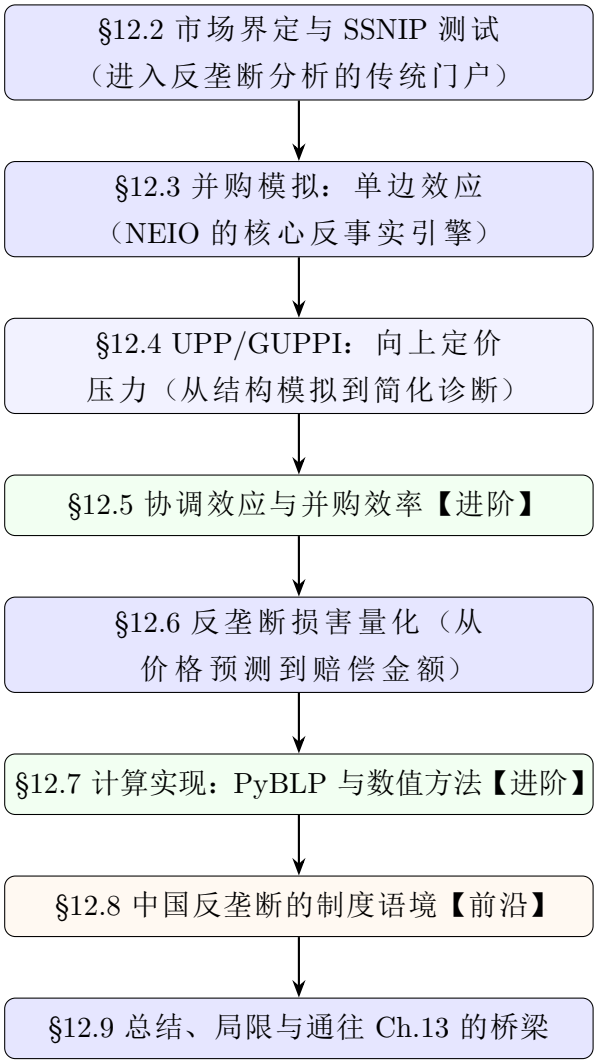


图 12.2: 本章结构。⋆⋆ 标记本章的“智识心脏”（数学最密集的节）。【核心】= 核心内容，【进阶】= 进阶内容，【前沿】= 前沿/制度内容。

12.2 市场界定与 SSNIP 测试

【核心】

12.2.1 SSNIP 测试的逻辑结构

SSNIP 测试（Small but Significant Non-transitory Increase in Price，小幅但显著的非暂时性提价），也称**假想垄断者测试**（Hypothetical Monopolist Test, HMT），是反垄断市场界定的核心分析工具。它自 1982 年美国《横向并购指南》引入以来，已成为全球反垄断当局的标准方法。

SSNIP 测试的核心思想可以用一个简单的决策问题来表述：

> **假想垄断者问题：**假设一个“假想垄断者”（Hypothetical Monopolist）控制了**候选市场**（Candidate Market）中的所有产品。这个假想垄断者能否

有利可图地（Profitably）将所有产品的价格提高 $t\%$（通常 $t = 5\%$ 或 10% ——“小幅但显著的非暂时性提价”）？

- 如果**是**：候选市场中的产品“之间”存在足够强的替代关系（且候选市场之外的产品的竞争约束不够强），使得提价导致消费者流失到外部产品不足以使提价无利可图 → 候选市场构成一个**相关反垄断市场**（Relevant Antitrust Market）。
- 如果**否**：提价导致消费者大量流失到候选市场之外 → 市场定义太窄 → 将下一个最近的替代品加入候选市场，重新测试。

SSNIP 测试的经济学本质：它测试的是假想垄断者面临的**剩余需求弹性**（Residual Demand Elasticity）。如果剩余的弹性足够低（消费者的外部选择不够好），假想垄断者可以有利可图地提价——这意味着候选市场构成了一个独立的“经济势力范围”。

SSNIP 与 Lerner 公式的连接

SSNIP 测试本质上是“逆过来的” Lerner 公式。Lerner 公式告诉我们：给定需求弹性 η，最优加成率为 $1/\eta$。SSNIP 测试问的是：给定当前的毛利率 $m = (p - mc)/p$，提价 t 是否有利可图？
提价有利可图的条件（正式推导见下小节）：毛利率 m 必须大于 SSNIP 幅度 t，且临界损失（Critical Loss）必须大于实际损失（Actual Loss）。如果 m 本身就很低（接近完全竞争），即使是小幅提价也会导致大量消费者流失——因为消费者有很好的替代选择。这正是 SSNIP 测试识别市场势力的机制。

12.2.2　临界损失分析

临界损失（Critical Loss, CL）是 SSNIP 测试的数学核心。它量化了“假想垄断者可以承受的最大销售流失比例，使得提价仍然有利可图”。

临界损失的推导

设假想垄断者在提价前的利润为：

$$\pi_0 = (p_0 - mc) \times Q_0 \tag{12.1}$$

提价后，价格变为 $p_1 = p_0(1+t)$，销量变为 $Q_1 = Q_0(1-L)$，其中 $L = -\Delta Q/Q_0$ 是**实际损失**（Actual Loss）——因提价而流失的销量比例。注意：当价格上升时，$\Delta Q < 0$，所以 $L > 0$。

提价后的利润为：

$$\pi_1 = [p_0(1 + t) - mc] \times Q_0(1 - L) \tag{12.2}$$

提价有利可图的条件是 $\pi_1 \geq \pi_0$：

$$[p_0(1+t) - mc] \times Q_0(1-L) \geq (p_0 - mc) \times Q_0 \tag{12.3}$$

两边消去 Q_0（假设 $Q_0 > 0$），并定义毛利率（Lerner 毛利率 / Price-Cost Margin）$m \equiv (p_0 - mc)/p_0$：

$$[p_0(1+t) - mc](1-L) \geq p_0 - mc \tag{12.4}$$

将 $mc = p_0(1-m)$ 代入：

$$[p_0(1+t) - p_0(1-m)](1-L) \geq p_0 m \tag{12.5}$$

$$p_0(t+m)(1-L) \geq p_0 m \tag{12.6}$$

$$(t+m)(1-L) \geq m \tag{12.7}$$

求解 L：

$$1 - L \geq \frac{m}{t+m} \quad \Rightarrow \quad L \leq \frac{t}{t+m} \tag{12.8}$$

这给出了**临界损失**的公式：

$$\boxed{CL = \frac{t}{t+m}} \tag{12.9}$$

其中 $t = \Delta p/p_0$ 是 SSNIP 幅度（通常 5% 或 10%），$m = (p_0 - mc)/p_0$ 是毛利率。

决策规则：

- 如果**实际损失** $L \leq CL$：提价有利可图 $\rightarrow$ 候选市场构成相关市场。
- 如果**实际损失** $L > CL$：提价无利可图 $\rightarrow$ 市场界定太窄，需要扩大。

临界损失的经济学直觉

公式 $CL = t/(t+m)$ 传达了三个核心直觉：

(1) **高毛利率 $\rightarrow$ 低 CL**。当 m 很大时（高加成），每个流失的消费者都意味着损失大量的利润（$p_0 - mc$ 很大）。因此，即使是少量的销售流失也可能使提价无利可图 $\rightarrow$ 市场定义需要更宽（CL 低意味着不能承受太多流失）。

(2) **高 SSNIP 幅度 $\rightarrow$ 高 CL**。当 t 很大时（更大的提价幅度），每个**留下来的**消费者都贡献更多的利润增加。因此，可以承受更大的销售流失 $\rightarrow$ CL 上升。

(3) **竞争基准的重要性**。CL 的计算依赖于毛利率 m，而 m 又依赖于当前价格 p_0。如果当前价格已经是垄断价格（$p_0 \gg mc$），m 会很高 $\rightarrow$ CL 会很**低** $\rightarrow$ 测试会错误地得出 “提价无利可图” 的结论——这正是下小节要讨论的 Cellophane Fallacy。

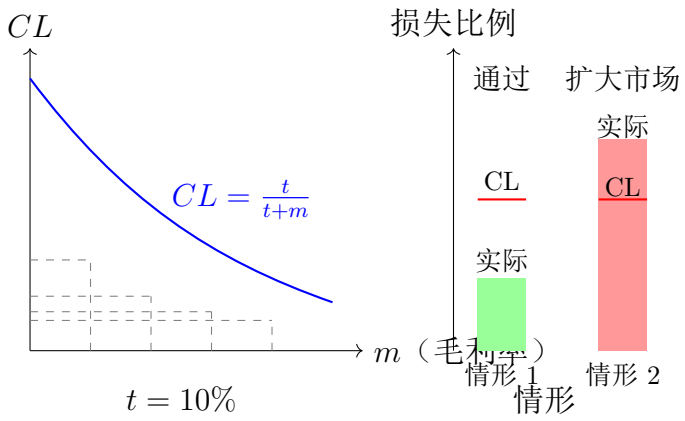


图 12.3: 临界损失与实际损失的比较。**左面板：**临界损失 CL 作为毛利率 m 的函数——递减曲线。高毛利率意味着假想垄断者更“脆弱”——即使是很小的消费者流失也破坏提价的合理性。**右面板：**两个假设情形。情形 1：实际损失 $<$ CL $\rightarrow$ 提价有利可图 $\rightarrow$ 市场界定完成（绿色）。情形 2：实际损失 $>$ CL $\rightarrow$ 提价无利可图 $\rightarrow$ 市场需要扩大（红色）。

临界损失数值示例

下表展示了不同毛利率 m 和 SSNIP 幅度 t 下的临界损失 $CL = t/(t+m)$：

毛利率 m	$t = 5\%$ 的 CL	$t = 10\%$ 的 CL
0.10 (10%)	0.05/0.15 = 33.3%	0.10/0.20 = 50.0%
0.20 (20%)	0.05/0.25 = 20.0%	0.10/0.30 = 33.3%
0.30 (30%)	0.05/0.35 = 14.3%	0.10/0.40 = 25.0%
0.40 (40%)	0.05/0.45 = 11.1%	0.10/0.50 = 20.0%
0.50 (50%)	0.05/0.55 = 9.1%	0.10/0.60 = 16.7%
0.60 (60%)	0.05/0.65 = 7.7%	0.10/0.70 = 14.3%
0.70 (70%)	0.05/0.75 = 6.7%	0.10/0.80 = 12.5%
0.80 (80%)	0.05/0.85 = 5.9%	0.10/0.90 = 11.1%

表 12.2: 不同毛利率和 SSNIP 幅度下的临界损失。注意：当 $m = 50\%$ 且 $t = 5\%$ 时，$CL = 9.1\%$ ——即使只有 9.1% 的销售流失，提价也是无利可图的。高毛利率意味着失去每个消费者都损失巨大，因此假想垄断者“脆弱”——即使是少量的流失也破坏提价的合理性。

12.2.3 从临界损失到分流比：SSNIP 的现代版本

临界损失 CL 是一个 “门槛”——提价是否有利可图取决于**实际损失** L 是否超过这个门槛。那么，实际损失 L 从哪里来？它正是**分流比**（Diversion Ratio）之和：当假想垄断者提高价格时，消费者转向候选市场之外的产品——这些 “外部” 分流比之和就是实际损失 L。

设 $D_{jk} = \frac{\partial s_k/\partial p_j}{-\partial s_j/\partial p_j}$ 为从产品 j 到产品 k 的分流比（将在 §12.3.4 中详细讨论）。对于候选市场 C 中的假想垄断者，提价 t 导致的实际损失近似为：

$$L \approx \sum_{k \notin C} D_{jk} \quad \text{（对所有 } j \in C \text{ 或使用加权平均）} \tag{12.10}$$

因此，SSNIP 测试的现代版本等价于：

$$\text{如果} \quad \sum_{k \notin C} D_{jk} \leq CL = \frac{t}{t+m} \quad \Rightarrow \quad \text{市场界定完成} \tag{12.11}$$

这意味着：**SSNIP 测试和并购模拟/ UPP 使用相同的核心输入——分流比** D_{jk} **和毛利率** m。这正是 Farrell & Shapiro (2010) 的核心论证（将在 §12.4.5 中详细展开）：如果你已经计算了分流比（用于 UPP），你可能**不需要**进行正式的市场界定——因为 UPP 直接测量了竞争效应。市场界定在逻辑上是分流比分析的**推论**，而非其**前提**。

12.2.4 Cellophane Fallacy（玻璃纸谬误）

Cellophane Fallacy（玻璃纸谬误）是 SSNIP 测试最著名的陷阱，得名于 1956 年美国最高法院的 *United States v. E.I. du Pont de Nemours & Co.* 案。du Pont 被指控垄断了玻璃纸（Cellophane）市场。du Pont 辩称：“我们的产品与蜡纸、铝箔等灵活包装材料紧密竞争——因此相关市场是’ 灵活包装材料’ 而非’ 玻璃纸’。” 最高法院接受了 du Pont 的论证，认定相关市场为 “灵活包装材料”——在这个更宽的市场中 du Pont 的份额仅为 $\sim 18\% \rightarrow$ 不构成垄断。

谬误所在：du Pont 的论证在 SSNIP 的框架下是**技术正确的**——但测试的**起始价格**是错误的。du Pont 已经在收取**垄断价格**——在这个价格水平上，消费者确实认为蜡纸和铝箔是玻璃纸的替代品。但这**正是因为** du Pont 的垄断价格已经把消费者 “推到了” 替代品那里！在竞争价格下，消费者可能不会认为蜡纸是好的替代品。

Cellophane Fallacy 的形式化

在垄断价格 $p^M > p^C$（竞争价格）处，毛利率 m 更高 $\rightarrow CL = t/(t+m)$ 更低 $\rightarrow$ 更容易得出“实际损失 $>$ CL”（提价无利可图）的结论 $\rightarrow$ 市场被错误地界定得**太宽** $\rightarrow$ 垄断者看起来像“有很多竞争者” $\rightarrow$ 错误的“无市场势力”结论。**解决方案**：在可行的范围内，使用竞争价格 $p^C = MC$ 作为 SSNIP 的起始价格基准，而非当前价格。在实践中，MC 需要通过成本估计或行业比较来推断——这增加了 SSNIP 分析的数据需求。

Cellophane Fallacy 的教训是深远的：SSNIP 测试的结论对**基准价格的选择**高度敏感。现代反垄断实践（如 2023 年美国并购指南）明确警告了这一风险，并要求分析人员在可能的情况下使用竞争基准价格进行 SSNIP 测试。

12.2.5　市场界定的实证方法概览

市场界定可以使用多种实证方法，从简单的价格相关性分析到复杂的结构性需求估计：

(1) **价格相关性检验**（Stigler & Sherwin 1985）。如果两个区域的价格高度相关，它们可能在同一市场中。简单但有限制——共同成本冲击（如原油价格上涨）可以产生虚假的价格相关性。

(2) **自然实验**（如商店进入/退出、产品召回）。使用外生事件直接估计消费者在不同产品/地点之间的分流模式。这是“最干净”的方法——但需要找到可信的外生冲击。

(3) **需求系统估计**（BLP / 嵌套 logit / AIDS）。估计完整的需求系统 $\rightarrow$ 计算分流比 D_{jk} $\rightarrow$ 直接检验 SSNIP 条件。最灵活的方法，但数据和计算需求最高（需要 第 ~11~ 章的完整 BLP 框架）。

(4) **消费者调查**（“第二选择”数据）。直接询问消费者“如果品牌 A 涨价 5%，你会买什么？”最直接——但受限于调查偏差（消费者可能说一套做一套）。

12.3　并购模拟：单边效应

【核心】

单边效应（Unilateral Effects）是并购方**单方面**提高价格的激励——不需要任何竞争对手的协调或配合。其核心经济机制是**竞争外部性的内部化**（Internalization of Competition Externalities）：并购前，企业 A 提价导致部分消费者转向企业 B 的产品——这给企业 B 带来了正外部性（B 的销量和利润增加）。企业 A 在设定价格时

不考虑这个外部性——它只在乎自己的销售损失。并购后，合并实体同时拥有 A 和 B，将此前的外部性**内部化**：A 提价导致的、流向 B 的销售现在仍然是“自家的”销售。这种“回收”降低了提价的机会成本，使更高的价格对于合并实体而言是最优的。

12.3.1 多产品 Bertrand-Nash 定价：从 FOC 到矩阵形式

企业的最大化问题与一阶条件

考虑市场 m（下标省略）。企业 f 拥有产品集合 $\mathcal{F}_f \subseteq \{1,\dots,J\}$。企业 f 选择价格 $\{p_j : j \in \mathcal{F}_f\}$ 最大化联合利润：

$$\pi_f = \sum_{j\in\mathcal{F}_f} (p_j - mc_j)\cdot M \cdot s_j(\mathbf{p}) \tag{12.12}$$

其中 M 是市场总规模（潜在消费者总数），$s_j(\mathbf{p})$ 是产品 j 的市场份额（$\sum_{j=0}^{J} s_j = 1$，$j=0$ 为外部产品 / 不购买）。

对 p_j 求导（$j \in \mathcal{F}_f$），市场大小 M 是公共因子可以约去：

$$\frac{\partial \pi_f}{\partial p_j} = M \cdot \left[s_j + \sum_{k\in\mathcal{F}_f} (p_k - mc_k)\cdot \frac{\partial s_k}{\partial p_j} \right] = 0 \tag{12.13}$$

由此得到**多产品 Bertrand-Nash 一阶条件**：

$$\boxed{s_j + \sum_{k\in\mathcal{F}_f} (p_k - mc_k)\cdot \frac{\partial s_k}{\partial p_j} = 0 \qquad \forall j \in \mathcal{F}_f} \tag{12.14}$$

与单产品 FOC（$s_j + (p_j - mc_j)\cdot \partial s_j/\partial p_j = 0$）的区别是**求和项** $\sum_{k\in\mathcal{F}_f}$：企业内化了改变 p_j 对**所有自有产品**销售的影响。对于替代品（$\partial s_k/\partial p_j > 0$，$j$ 提价 $\to k$ 销量上升），这个求和项**降低**了提价的机会成本——部分流失的消费者被自有其他产品“回收”（Recaptured）。因此，多产品企业的均衡加成**高于**单产品企业（一切其他条件相同）。

从标量到矩阵

将 J 个产品的 FOC 堆叠为矩阵形式，需要以下定义：

- $\mathbf{s}$：$J \times 1$ 市场份额向量
- $\mathbf{p}$：$J \times 1$ 价格向量
- $\mathbf{mc}$：$J \times 1$ 边际成本向量
- $\mathbf{\Delta}$：$J \times J$ 矩阵，其中 $\Delta_{jk} = -\partial s_k/\partial p_j$（取负号使得：对于自身价格效应，$\partial s_j/\partial p_j < 0$，所以 $\Delta_{jj} > 0$；对于替代品之间的交叉效应 $j \neq k$，$\partial s_k/\partial p_j > 0$，所以 $\Delta_{jk} < 0$）

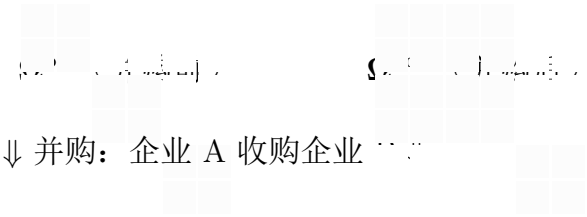


图 12.4: 所有权矩阵 $\boldsymbol{\Omega}$ 的变化示意。黑块 $= 1$（同一所有者），白块 $= 0$（不同所有者）。左：并购前——企业 A（产品 1-2）、企业 B（产品 3-4）、企业 C（产品 5-6），块对角结构。右：企业 A 收购企业 B 后——前 4 个产品现在属于同一所有者，形成更大的块。

- $\boldsymbol{\Omega}$：$J \times J$ **所有权矩阵**（Ownership Matrix），其中 $\Omega_{jk} = 1$ 如果产品 j 和 k 属于同一企业，否则 $= 0$

利用这些定义，J 个 FOC 可以写为：

$$s_j - \sum_{k=1}^{J} \Omega_{jk} \cdot \Delta_{jk} \cdot (p_k - mc_k) = 0 \qquad \forall j = 1, \ldots, J \tag{12.15}$$

在矩阵形式中：

$$\mathbf{s} - (\boldsymbol{\Omega} \circ \boldsymbol{\Delta})(\mathbf{p} - \mathbf{mc}) = \mathbf{0} \tag{12.16}$$

其中 $\circ$ 表示 Hadamard（逐元素）积：$(\boldsymbol{\Omega} \circ \boldsymbol{\Delta})_{jk} = \Omega_{jk} \cdot \Delta_{jk}$。

求解加成向量 $\mathbf{p} - \mathbf{mc}$：

$$\boxed{\mathbf{p} - \mathbf{mc} = (\boldsymbol{\Omega} \circ \boldsymbol{\Delta})^{-1}\mathbf{s}} \tag{12.17}$$

这是**并购模拟的根本方程**。它表达了所有 J 个产品的加成向量与所有权结构、需求导数和市场份额之间的关系。Hadamard 积 $\boldsymbol{\Omega} \circ \boldsymbol{\Delta}$ 将**跨企业**的替代关系“清零”——只保留同一企业内产品之间的需求导数。并购后，合并企业的产品之间的 Ω_{jk} 从 0 变为 1 $\rightarrow$ $(\boldsymbol{\Omega} \circ \boldsymbol{\Delta})$ 中更多的元素变为非零 $\rightarrow$ 通过矩阵逆运算影响**所有**产品的加成估计。

为什么矩阵逆产生“涟漪效应”

即使只有合并企业产品对之间的 Ω_{jk} 从 $0 \rightarrow 1$，$(\boldsymbol{\Omega} \circ \boldsymbol{\Delta})^{-1}$ 中的**每一个元素**都可能改变。这是因为矩阵逆是一个“全局”运算——改变矩阵中的一个（块）元素会通过逆运算传播到整个逆矩阵的所有元素。这意味着合并不仅影响合并方的加成——它也（通过新均衡中的价格变化）间接影响非合并方的加成。这种“一般均衡”效应是完整并购模拟优于 UPP/GUPPI 简化指标的原因之一。

12.3.2 并购前 → 并购后：五步模拟过程

并购模拟将并购前估计的模型转化为并购后均衡的预测。整个过程分为五个步骤：

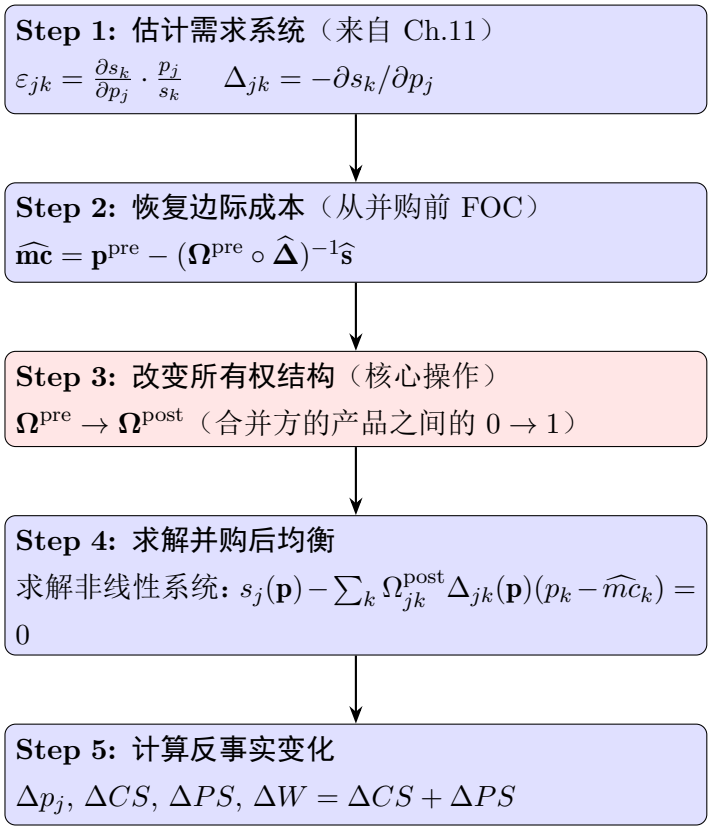


图 12.5: 并购模拟的五步过程。Step 3（红色框）是并购模拟区别于常规需求估计的**核心操作**——只有所有权结构改变了，需求系统和边际成本保持不变。

Step 1：估计需求系统

这是 第 ~11~ 章的输出。研究者使用并购前数据估计一个差异化产品需求系统（BLP 随机系数 logit、嵌套 logit 或 AIDS），获得：

- **弹性矩阵**：$\varepsilon_{jk} = (\partial s_k / \partial p_j) \cdot (p_j / s_k)$ ——自身弹性（$j = k$）和交叉弹性（$j \neq k$）
- **份额导数**：$\Delta_{jk} = -\partial s_k / \partial p_j$。从弹性可计算：$\partial s_k / \partial p_j = \varepsilon_{jk} \cdot s_k / p_j$

Step 2：恢复边际成本

利用并购前的 FOC（方程 12.17），从观测的并购前价格 $\mathbf{p}^{\text{pre}}$ 中“倒推”边际成本：

$$\widehat{\mathbf{mc}} = \mathbf{p}^{\text{pre}} - (\boldsymbol{\Omega}^{\text{pre}} \circ \widehat{\boldsymbol{\Delta}})^{-1}\widehat{\mathbf{s}} \tag{12.18}$$

这个步骤假设：（1）并购前行为是 Bertrand-Nash（可通过对 Ch.10 行为参数 θ 的检验来验证）；（2）估计的需求导数 $\widehat{\Delta}_{jk}$ 是准确的。如果需求模型被错误设定，$\widehat{mc}_j$ 将吸收这个误差，并在后续的反事实模拟中传播。

Step 3：改变所有权结构

这是**并购模拟的核心操作**。构造并购后的所有权矩阵 $\boldsymbol{\Omega}^{\text{post}}$：

$$\Omega_{jk}^{\text{post}} = \begin{cases} 1 & \text{如果} j,k \in \mathcal{F}_{\text{merged}}\text{（合并实体的产品集）} \\ \Omega_{jk}^{\text{pre}} & \text{否则} \end{cases} \tag{12.19}$$

所有非合并企业的 Ω_{jk} 保持不变。合并企业的产品之间的所有 0 变为 1——反映它们现在由同一个决策者定价。

Step 4：求解并购后均衡

给定 $\boldsymbol{\Omega}^{\text{post}}$ 和 $\widehat{\mathbf{mc}}$，寻找价格向量 $\mathbf{p}^*$ 满足并购后 FOC 系统：

$$s_j(\mathbf{p}^*) - \sum_{k=1}^{J} \Omega_{jk}^{\text{post}} \cdot \widehat{\Delta}_{jk}(\mathbf{p}^*) \cdot (p_k^* - \widehat{mc}_k) = 0 \qquad \forall j = 1, \ldots, J \tag{12.20}$$

由于 $s_j(\mathbf{p})$ 和 $\widehat{\Delta}_{jk}(\mathbf{p})$ 都依赖于 $\mathbf{p}$（需求是非线性的），这是一个 J 个方程、J 个未知数的非线性系统。数值求解方法将在 §12.7 中讨论。

Step 5：计算反事实变化

从并购前后均衡的比较中计算：

- **价格效应**：$\Delta p_j = p_j^{\text{post}} - p_j^{\text{pre}}$ 对每个产品 j
- **消费者剩余变化**（以 logit 为例）：

$$\Delta CS = \frac{1}{\alpha}\Big[\ln\Big(\sum_{j=0}^{J} \exp(\delta_j^{\text{post}})\Big) - \ln\Big(\sum_{j=0}^{J} \exp(\delta_j^{\text{pre}})\Big)\Big] \cdot M$$

- **生产者剩余变化**：$\Delta PS = \sum_f (\pi_f^{\text{post}} - \pi_f^{\text{pre}})$
- **总福利变化**：$\Delta W = \Delta CS + \Delta PS$

其中 α 是收入的边际效用（价格系数的绝对值），$\delta_j = \mathbf{x}_j\boldsymbol{\beta} - \alpha p_j + \xi_j$ 是产品 j 的均值效用。对于 BLP 随机系数模型，消费者剩余需要通过蒙特卡洛积分计算，形式更复杂（见 Conlon & Gortmaker 2020）。

12.3.3　一个完整的数值例子：两产品 logit 并购

为使学生具体感受到并购模拟的机制，本节展开一个可手算的两产品例子。

设定：市场有 3 个内部产品 $\{1,2,3\}$ 加外部产品（$j=0$，$u_{i0}=\varepsilon_{i0}$）。Logit 需求：$s_j=\frac{\exp(\delta_j-\alpha p_j)}{1+\sum_{k=1}^{3}\exp(\delta_k-\alpha p_k)}$。

参数：$\delta_1=2$, $\delta_2=1.5$, $\delta_3=1$, $\alpha=0.5$, $mc_1=mc_2=mc_3=3$。并购前每家拥有一个产品。企业 1 收购企业 2。

对于 logit 需求，$\partial s_k/\partial p_j=\alpha\cdot s_j\cdot(s_k-\delta_{jk})$，其中 δ_{jk} 是 Kronecker delta（$j=k$ 时为 1，否则为 0）。由此 $\Delta_{jk}=-\partial s_k/\partial p_j=\alpha\cdot s_j\cdot(\delta_{jk}-s_k)$。

两产品 logit 并购模拟结果

并购前（每家单产品，$\Omega_{jk}=\delta_{jk}$）：

- $s_1\approx 0.382$, $s_2\approx 0.251$, $s_3\approx 0.152$, $s_0\approx 0.215$
- $p_1^{\text{pre}}\approx 5.24$, $p_2^{\text{pre}}\approx 4.95$, $p_3^{\text{pre}}\approx 4.77$
- $m_1=(5.24-3)/5.24\approx 42.7\%$

并购后（产品 1 和 2 联合定价）：

- $p_1^{\text{post}}\approx 5.77$（$+10.1\%$），$p_2^{\text{post}}\approx 5.43$（$+9.7\%$）
- $p_3^{\text{post}}\approx 4.92$（$+3.1\%$ ——非合并方也提价了！）
- $\Delta CS\approx -2.8\%$，$\Delta PS_{\text{合并方}}\approx +6.5\%$，$\Delta W\approx -0.3\%$

关键观察：（1）合并方产品的价格上升 $\sim 10\%$ ——单边效应显著。（2）非合并方（产品 3）的价格**也**上升了 3.1% ——这是“伞形效应”（Umbrella Effect）：合并方提价 $\to$ 部分消费者转向产品 3 $\to$ 产品 3 的需求上升 $\to$ 其提价也变得有利可图。（3）总福利下降——单边效应主导了合并方利润的增加。

12.3.4　分流比：单边效应的充分统计量

定义与计算

分流比（Diversion Ratio）D_{jk} 度量了：当产品 j 的价格上升时，离开 j 的消费者中，有多大比例转向了产品 k。

$$\boxed{D_{jk}\equiv\frac{\partial s_k/\partial p_j}{-\partial s_j/\partial p_j}} \tag{12.21}$$

分母 $-\partial s_j/\partial p_j>0$ 是 j 的**总**销售损失（所有离开 j 的消费者）。分子 $\partial s_k/\partial p_j>0$（对替代品）是转向 k 的销售增加。因此 $D_{jk}\in[0,1]$，且对于满足加总性质（$\sum_{k=0}^{J}s_k=$

1）的需求系统，$\sum_{k\neq j} D_{jk} = 1$：所有离开 j 的消费者必须去往某处（包括外部产品 $k = 0$）。

用弹性表示：

$$D_{jk} = \frac{\varepsilon_{kj} \cdot s_k/p_j}{-\varepsilon_{jj} \cdot s_j/p_j} = \frac{\varepsilon_{kj} \cdot s_k}{\varepsilon_{jj} \cdot s_j} \tag{12.22}$$

其中 $\varepsilon_{kj} = (\partial s_k/\partial p_j) \cdot (p_j/s_k)$ 是交叉价格弹性，$\varepsilon_{jj} = (\partial s_j/\partial p_j) \cdot (p_j/s_j) < 0$ 是自身价格弹性。

Logit 分流比的“陷阱”

在标准 logit 模型下，分流比具有一个极其简单的形式：

$$D_{jk}^{\text{logit}} = \frac{s_k}{1 - s_j} \tag{12.23}$$

这意味着 logit 分流比只取决于市场份额——而不取决于产品 j 和 k 是否真的是近竞争者。这是 IIA（Independence of Irrelevant Alternatives）的直接推论：在 logit 中，$\partial s_k/\partial p_j$ 对所有 $k \neq j$ 是相等的比例——即 j 提价导致的销售损失按市场份额比例**均匀地**分配给所有其他产品。

Logit vs. BLP 的分流比：对并购模拟的影响

在 logit 中，如果 j 和 k 是“远竞争者”（如豪华车 vs. 经济车），但 k 的市场份额很大 $\to D_{jk}^{\text{logit}}$ 可能很高 $\to$ logit 会**高估**并购 j 和 k 的单边价格效应。反之，如果 j 和 k 是“近竞争者”（如两款经济车）且两者的市场份额都较小 $\to D_{jk}^{\text{logit}}$ 可能很低 $\to$ logit 会**低估**单边效应。

这正是在并购模拟中偏好 BLP（随机系数 logit）而非标准 logit 的核心原因：BLP 允许分流比反映**实际产品相似性**——近竞争者之间的 D_{jk} 可以远超其市场份额所暗示的水平，远竞争者之间的 D_{jk} 可以远低于其市场份额所暗示的水平。

下面的数值例子展示了 logit（IIA 分流）与嵌套 logit（相关性分流）在预测两个产品并购的价格效应时的系统性差异：

情景	Logit 的 $\Delta p_1, \Delta p_2$	Nested Logit 的 $\Delta p_1, \Delta p_2$
近竞争者，高份额	$+8.2\%, +7.9\%$	$+14.6\%, +13.8\%$
$s_1 = 0.25, s_2 = 0.20$，同一巢内	（低估：logit 分流 $\propto$ 份额）	（更大——巢内高分流被捕获）
近竞争者，低份额	$+4.1\%, +3.8\%$	$+11.3\%, +10.5\%$
$s_1 = 0.08, s_2 = 0.07$，同一巢内	（严重低估——logit 分流过低）	（仍然显著——巢内高度替代）
远竞争者，高份额	$+7.5\%, +7.1\%$	$+2.8\%, +2.4\%$
$s_1 = 0.22, s_2 = 0.20$，不同巢	（高估——logit 分流 $\propto$ 份额）	（较小——不同巢之间的分流低）
远竞争者，低份额	$+3.9\%, +3.6\%$	$+1.2\%, +1.0\%$
$s_1 = 0.07, s_2 = 0.06$，不同巢	（高估——但仍小因为份额小）	（最小——不同巢 + 低份额）

表 12.3: Logit vs. 嵌套 Logit 在两产品并购中的价格效应比较。嵌套 logit 使用两个巢（$\sigma = 0.4$ 巢内相关性）。关键模式：logit **低估**近竞争者之间的价格效应（因为分流比仅由份额驱动），**高估**远竞争者之间的价格效应。这一系统性偏差是并购模拟中偏好 BLP 的根本原因。

分流比作为充分统计量

在两产品并购中（合并方只涉及产品 1 和 2），并购后产品 1 的价格变化（一阶近似）为：

$$\Delta p_1 \approx \frac{D_{12} \cdot (p_2 - mc_2)}{-\partial s_1 / \partial p_1} \tag{12.24}$$

这个近似式表明：给定需求斜率 $-\partial s_1/\partial p_1$，并购对产品 1 的价格效应完全由两个量决定：（1）D_{12} ——多少销售被回收，（2）$(p_2 - mc_2)$ ——每个回收销售的利润价值。D_{12} 是**充分统计量**（Sufficient Statistic）：它概括了并购模拟所需的所有关于替代模式的信息——不需要知道完整的需求系统。

这解释了为什么分流比在反垄断分析中如此重要：它简洁地捕捉了并购的核心竞争关切——“并购双方彼此施加了多大的竞争约束？” $D_{12} \approx 0.6$ 意味着：产品 1 提价导致的销售流失中，60% 被产品 2 回收 $\rightarrow$ 并购内化了产品 1 提价的主要机会成本 $\rightarrow$ 价格大幅上升。$D_{12} \approx 0.02$ 意味着：几乎不回收 $\rightarrow$ 并购没有实质性改变产品 1 的定价激励 $\rightarrow$ 价格变化极小。

12.4　UPP 与 GUPPI：向上定价压力

【核心】

12.4.1　UPP 的推导：从 FOC 到定价压力

UPP（Upward Pricing Pressure，向上定价压力）和 **GUPPI**（Gross Upward Pricing Pressure Index，总向上定价压力指数）是 Farrell & Shapiro (2010) 提出的并购竞争效应简化诊断工具。其核心洞察是：如果我们能衡量分流比和毛利率，就可以直接计算并购的单边定价激励——不需要先界定市场、计算市场份额或运行完整的结构性并购模拟。

推导：考虑并购前产品 1（属于企业 A）的 FOC：

$$s_1 + (p_1 - mc_1) \cdot \frac{\partial s_1}{\partial p_1} = 0 \tag{12.25}$$

并购后，产品 1 和 2（属于企业 B）现在属于同一所有者。产品 1 的并购后 FOC 增加了产品 2 的项：

$$s_1 + (p_1 - mc_1) \cdot \frac{\partial s_1}{\partial p_1} + (p_2 - mc_2) \cdot \frac{\partial s_2}{\partial p_1} = 0 \tag{12.26}$$

（这里做了两个简化假设：产品 1 的价格不影响合并实体拥有的其他产品（除 2 之外），且 $\partial s_1/\partial p_1$ 在并购前后近似不变——即一阶近似。）

从式 (12.26) 减去式 (12.25)，记 $\Delta p_1 = p_1^{\text{post}} - p_1^{\text{pre}}$：

$$\Delta p_1 \cdot \frac{\partial s_1}{\partial p_1} + (p_2 - mc_2) \cdot \frac{\partial s_2}{\partial p_1} \approx 0 \tag{12.27}$$

$$\Rightarrow \Delta p_1 \approx \frac{(p_2 - mc_2) \cdot (\partial s_2/\partial p_1)}{-\partial s_1/\partial p_1} = D_{12} \cdot \frac{p_2 - mc_2}{-\partial s_1/\partial p_1} \tag{12.28}$$

分子部分——$D_{12} \cdot (p_2 - mc_2)$——就是 UPP：它度量了并购创造的上向定价**压力**（分子符号为正，意味着并购后价格倾向于上升）。分母 $(-\partial s_1/\partial p_1)$ 将这个压力**转化**为实际价格变化。

$$\boxed{UPP_1 = D_{12} \times (p_2 - mc_2)} \tag{12.29}$$

经济学解释：UPP_1 是并购给产品 1 带来的**机会成本**增加。并购前，多卖一单位产品 1“蚕食”（Cannibalize）产品 2 的销售——但企业 A 不在乎（因为产品 2 是别人的）。并购后，多卖一单位产品 1 减少产品 2 的销售——合并实体**在乎**——其价值为 $D_{12} \cdot (p_2 - mc_2)$，即一单位产品 1 销售所“挤走”的产品 2 销售的利润。

12.4.2 GUPPI：标准化指数

UPP 以美元为单位——难以在价格水平不同的产品之间比较。GUPPI 将 UPP 标准化为价格比例：

$$\boxed{GUPPI_1 = \frac{UPP_1}{p_1} = D_{12} \times \frac{p_2 - mc_2}{p_1}} \tag{12.30}$$

可以进一步分解：

$$GUPPI_1 = \underbrace{D_{12}}_{\text{分流比}} \times \underbrace{\frac{p_2 - mc_2}{p_2}}_{\text{产品 2 的毛利率} m_2} \times \underbrace{\frac{p_2}{p_1}}_{\text{相对价格}} \tag{12.31}$$

三个成分有清晰的解释：

- D_{12}：替代模式——多少离开产品 1 的消费者转向了产品 2？
- $m_2 = (p_2 - mc_2)/p_2$：回收销售的价值——产品 2 上每单位销售的利润率
- p_2/p_1：相对价格水平——将产品 2 的利润价值转换为产品 1 的比例

阈值：在实践中，$GUPPI > 5\%$–10% 被认为是 “红旗”（Red Flag）——并购可能导致显著的价格上涨，需要进一步调查。但这不是机械的阈值：对于近竞争者（高 D），即使毛利率很低（如 $m = 0.2$ 且 $p_2/p_1 \approx 1$），$GUPPI = 0.6 \times 0.2 = 12\%$ ——远超阈值。

12.4.3 Net UPP：纳入并购效率

并购效率（Merger Efficiencies）——并购带来的边际成本减少——抵消了部分（甚至全部）上向定价压力。令 E_1 为产品 1 的边际成本因并购而减少的比例（如 $E_1 = 0.05$ 表示 5% 的成本减少），则：

$$\boxed{\text{Net UPP}_1 = D_{12} \times (p_2 - mc_2) - E_1 \times mc_1} \tag{12.32}$$

如果 $\text{Net UPP}_1 > 0$：上向定价压力即使在考虑了效率之后仍然存在 $\rightarrow$ 并购可能导致产品 1 的价格净上升。如果 $\text{Net UPP}_1 < 0$：效率效应主导 $\rightarrow$ 产品 1 的价格可能下降（对消费者有利）。

12.4.4　UPP vs. 完整并购模拟：比较

维度	UPP / GUPPI	完整并购模拟
数据需求	分流比 + 毛利率（两个数字）	完整需求系统 + 成本结构
计算	封闭形式（电子表格）	非线性系统求解
效率纳入	通过 Net UPP	完全（mc 直接减少）
非合并竞争者反应	否（部分均衡）	是（新 Nash 均衡）
需求曲率	假设线性（常数 $\partial s/\partial p$）	完全（需求系统的全部非线性）
均衡反馈效应	忽略	捕获
最佳用途	初始筛选（“红旗” 检测）	第二阶段深入调查

表 12.4: UPP/GUPPI 与完整并购模拟的比较。UPP 是 “第一关” 筛选工具——如果 GUPPI 很低，并购可能不需要进一步调查。如果 GUPPI 很高——或如果案件进入第二阶段深入审查——完整并购模拟提供了更准确、更全面的竞争效应预测。

12.4.5　Farrell–Shapiro (2010) 的范式转变

Farrell 和 Shapiro (2010) 的贡献超越了 “提供了 UPP 公式”。他们论证了反垄断分析的**逻辑结构**应该从根本上改变：

(1) **市场界定不是必要的第一步**。传统的两步法——（a）界定市场，（b）计算份额和 HHI ——在逻辑上是冗余的。如果你有分流比 D_{jk} 和毛利率 m_j，你可以直接计算并购的竞争效应——不需要先知道 “市场边界” 在哪里。

(2) **UPP/分流比方法 “包含” 了市场界定**。市场界定只是分流比分析的一个特例：SSNIP 测试等价于检查 “候选市场之外的分流比之和是否超过临界损失”。这意味着——如果你已经计算了所有相关产品对之间的分流比——市场界定可以**从分流比中推导出来**，而不是作为前提。

(3) **直接证据优于间接推定**。分流比直接测量了并购双方彼此施加的竞争约束——这是并购竞争关注的**核心**。而 HHI 和市场份额只是间接的 “代理变量”——可能完全无法捕捉竞争约束的真实结构。

2010 年美国《横向并购指南》反映了 Farrell–Shapiro 的范式转变：指南明确指出，市场界定 “不是并购分析的必要起点”，并正式认可分流比和 UPP 作为竞争效应的直接证据工具。这不是 “市场界定的死亡”——而是 “市场界定从目的变成了手段”。

12.5 协调效应与并购效率

【进阶】

12.5.1 协调效应：并购如何促进合谋

协调效应（Coordinated Effects）与单边效应有根本区别。单边效应关注的是并购方**自己**提高价格的激励（来自内化分流），而协调效应关注的是并购是否使**所有企业**更容易维持超竞争价格——即使没有并购方和竞争对手之间的显性沟通或协议。

协调效应的经济学逻辑来自 第 ~8~ 章中讨论的重复博弈与合谋理论。并购可以通过以下渠道促进合谋：

(1) **减少竞争者数量**（从 n 到 $n-1$）。在 Ch.8 的合谋理论中，更少的竞争者使监测更容易（每个企业的市场份额波动更小 $\rightarrow$ 更容易区分 “需求冲击” 和 “背叛”），同时使惩罚更可行（更少的协调摩擦）。

(2) **增加对称性**。非对称企业之间的并购可以创造更对称的并购后行业结构。对称企业更容易就合谋价格和市场份额分配达成一致——因为每家企业从合谋中获得的收益更相似。

(3) **消除搅局者**（Maverick）。“搅局者” 是指行业中积极竞争、拒绝参与合谋、或频繁打破合谋格局的企业。如果收购的企业是搅局者，其消除可以从根本上改变行业的行为模式——从竞争性定价转变为合作性定价。

(4) **增加多市场接触**（Multimarket Contact）。如果合并企业在多个地理或产品市场中有重叠，并购增加了多市场接触。Bernheim & Whinston (1990) 证明：多市场接触可以将不同市场之间的激励约束 “汇集” 起来——在某个市场中的背叛导致所有市场的惩罚——从而使合谋在更广泛的条件下可持续。

协调效应的量化困难

与单边效应相比，协调效应的结构性量化极其困难，原因在于：

- 协调效应涉及**动态博弈**——惩罚策略、监测结构、信念和信息——这些元素本质上是不可直接观测的。
- “没有并购” 的反事实（but-for world）对协调效应而言更为模糊——并购前的行业是否已经在合谋？并购只是维持了现有的协调，还是创造了新的协调可能性？
- 没有广泛接受的 “协调效应结构性模型”——实践中，协调效应通常在并购审查中被**定性讨论**，而非定量估计。

协调效应的定量方法（有限）

Davis & Garcés (2010, Ch.6) 讨论了几种有限的定量方法：

- **临界贴现因子分析**：比较并购前后维持给定合谋结果所需的最小贴现因子 δ^*。如果 δ^* 在并购后下降 $\rightarrow$ 合谋在并购后“更容易”维持。
- **搅局者识别**：通过比较“有搅局者”和“无搅局者”的模型中估计的均衡价格，来量化消除搅局者的价格效应。
- **合谋 regime 切换模型**（Porter 1983 风格）：估计并购是否改变了从“价格战 regime”切换到“合谋 regime”的概率。

但这些方法的可靠性和普适性远低于单边效应的结构性并购模拟。

12.5.2　并购效率

效率的来源与纳入

并购效率（Merger Efficiencies）是合并方声称并购将带来的边际成本减少。常见来源包括：生产协同（合并制造设施、消除重复固定成本、联合采购）、规模经济（更大的总产量降低平均成本）、和管理/技术改进（优势企业的管理实践或技术转移到合并伙伴）。

在并购模拟中纳入效率很简单：修改 Step 2 中恢复的边际成本：

$$mc_j^{\text{post}} = mc_j^{\text{pre}} \times (1 - e_j) \tag{12.33}$$

其中 e_j 是声称的产品 j 的百分比边际成本减少。并购后均衡以这些更低的成本重新求解。

效率也自然地进入 Net UPP 公式（方程 12.32）：$\text{Net UPP}_1 = D_{12}(p_2 - mc_2) - E_1 \cdot mc_1$。如果效率足够大，Net UPP 为负，意味着并购后的价格可能**下降**。

效率抗辩的证据标准

在并购审查中，效率主张的举证责任在合并方。合并方必须证明主张的效率是：

(1) **并购特定性**（Merger-Specific）：这些效率不能通过反竞争性更小的方式实现（如合资企业或内部增长）。

(2) **可验证性**（Verifiable）：效率有可信证据支持，而非仅仅是推测性的主张。

(3) **传递给消费者**（Passed Through to Consumers）：效率必须转化为消费者的更低价格或更高品质，而非仅仅是合并方的更高利润。

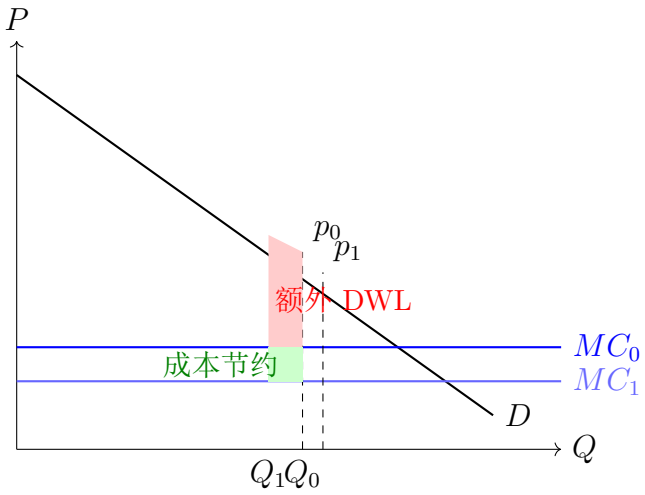


图 12.6: Williamson (1968) 权衡。$MC_0 \to MC_1$：并购降低了边际成本（成本节约矩形，绿色）。$p_0 \to p_1$：并购增加了市场势力，价格上升，数量从 Q_0 下降到 Q_1 → 额外的无谓损失（Harberger 三角形，红色）。净福利 = 绿色矩形面积 − 红色三角形面积。

在实践中，法院和反垄断当局对大规模效率主张持怀疑态度——除非有详细的工程研究、过往并购经验或类似行业证据的支持。2010 年美国指南声明："效率几乎不能证明合并至垄断或近似垄断的并购是正当的。"

12.5.3 Williamson (1968) 权衡

Williamson 权衡（Williamson Trade-off）是并购分析的经典框架：并购同时产生两个效应——降低边际成本（效率增益，↑ 福利）和增加市场势力（价格 ↑，↓ 消费者剩余）。净福利效应取决于两者的相对大小。

消费者福利标准 vs. 总福利标准：

- **消费者福利标准**（Consumer Welfare Standard，美国实践）：即使并购产生巨大的成本节约，如果消费者没有通过更低的价格受益，并购仍然被禁止。这比总福利标准严格得多——成本节约即使大于 DWL 增加，只要消费者没有受益，并购也不被接受。

- **总福利标准**（Total Welfare Standard）：只要成本节约 > DWL 增加，并购就应被批准——即使消费者受损（企业受益超过消费者损失）。

两者的区别在政策争论中至关重要。Williamson 的原始分析采用总福利标准——但他也承认，如果政策目标是消费者福利最大化而非总福利最大化，效率抗辩的空间会大大缩小。

12.6　反垄断损害量化

【核心】

12.6.1　过度要价与 But-For 价格

反垄断损害量化（Antitrust Damage Quantification）将并购模拟的**预测**（价格将上升多少）转化为法庭上可使用的**损害数字**（消费者损失了多少美元）。这是并购经济学的 “法律终点”——将经济分析转化为赔偿。

核心概念是**过度要价**（Overcharge）：

$$\boxed{\text{Overcharge} = p^{\text{actual}} - p^{\text{but-for}}} \tag{12.34}$$

其中 p^{actual} 是实际观测到的价格，$p^{\text{but-for}}$ 是 **but-for 价格**——在没有违法行为（并购/卡特尔）的情况下会存在的价格。But-for 价格是**反事实**的——无法直接观测——必须通过模型或比较来估计。

总损害（简化形式，忽略数量效应）：

$$\text{总损害} = \text{Overcharge} \times Q^{\text{actual}} \tag{12.35}$$

但这是近似的——因为它忽略了 “因为过高价格而未发生的交易”。完整公式需要考虑消费者剩余的减少：

$$\text{总损害} = \underbrace{(p^{\text{actual}} - p^{\text{but-for}})}_{\text{过度要价}} \times Q^{\text{actual}} + \underbrace{(CS^{\text{but-for}} - CS^{\text{actual}})}_{\text{数量效应 + 替代效应}} \tag{12.36}$$

12.6.2　损害估计的四种方法

(1) **标杆法**（Yardstick Approach）：使用**未受违法行为影响**的类似市场（“标杆”）来估计 but-for 价格。如：比较受影响城市 vs. 未受影响城市的并购后价格变化（本质上是 DiD）。优势：透明、不依赖结构模型。劣势：需要 “可比” 的标杆——但违法行为往往选择 “受影响” 市场（内生性）。

(2) **事前-事后法**（Before-After Approach）：直接比较违法行为前后的价格。But-for = 违法行为前价格 + 趋势。优势：简单、直观。劣势：无法区分 “违法行为导致的价格变化” 和 “其他因素导致的价格变化”（共同冲击问题）。

(3) **双重差分法**（Difference-in-Differences, DiD）：结合标杆法和事前-事后法。比较 “处理组”（受违法行为影响的市场）和 “控制组”（未受影响的市场）的违法行为前后价格变化。$\text{DiD} = (p^T_{\text{后}} - p^T_{\text{前}}) - (p^C_{\text{后}} - p^C_{\text{前}})$。这是损害量化中 “最干净” 的识别策略——但要求处理组和控制组满足平行趋势假设，且违法行为的影响是 “局部的”（只有部分市场受影响）。

(4) **结构模型法**（Structural Approach）：使用估计的需求和供给系统，“模拟”but-for 世界，计算 but-for 价格。优势：完全基于经济模型——内部一致、可以进行反事实模拟、不依赖 “找到可比对照组”。劣势：对模型假设和函数形式敏感（“模型风险”）。

	标杆法	事前-事后法	双重差分法	结构模型法
核心逻辑	比较类似市场	比较前后价格	$\text{DiD} = (\text{后} - \text{前})_T - (\text{后} - \text{前})_C$	模拟 but-for 均衡
数据需求	处理 + 控制市场	处理市场时间序列	面板数据（两组 × 两期）	完整需求/成本系统
关键假设	控制市场不受影响	无其他同时变化	平行趋势	模型设定正确
主要优势	透明	简单	干净识别	反事实一致性
主要局限	内生性（选择标杆）	共同冲击混淆	需要未受影响地区	模型风险

表 12.5: 损害量化方法的比较。在实践中，通常使用多种方法并比较结果——如果不同方法产生相似的估计，损害数字更具可信度。

12.6.3 伞形效应

伞形效应（Umbrella Effects）是并购/卡特尔的一个重要间接损害渠道。当并购方/卡特尔成员提高价格时，部分消费者转向非并购方/非卡特尔成员的产品 $\rightarrow$ 后者的需求上升 $\rightarrow$ 后者也提高价格。这构成了对消费者进一步的损害——且这种损害的范围超出了违法行为人自身。

在并购模拟中，伞形效应自然被包含在内：因为并购模拟重新计算了**所有**企业的均衡价格（不仅是合并方），非合并方的价格变化被自动捕获。但在标杆法/DiD 方法中，伞形效应可能被低估——如果 “控制” 地区的非合并方也受并购的 “一般均衡” 效应影响。

欧洲委员会近年来日益关注伞形效应——在损害量化中**计入**非合并方价格上升对消费者造成的额外损失。这增加了并购的总消费者损害——以及由此产生的反垄断损害赔偿金额。

12.7 计算实现：PyBLP 与数值方法

【进阶】

12.7.1　求解并购后均衡的数值方法

求解并购后 FOC 系统（方程 12.20）是并购模拟中计算最密集的步骤。这是一个 J 个未知数（价格 $\{p_j\}$）的非线性方程组。两种主要方法：

牛顿-拉弗森法（Newton-Raphson）：

$$\mathbf{p}^{(t+1)} = \mathbf{p}^{(t)} - \left[\mathbf{J}(\mathbf{p}^{(t)})\right]^{-1} \cdot \mathbf{FOC}(\mathbf{p}^{(t)}) \tag{12.37}$$

其中 $\mathbf{J}(\mathbf{p})$ 是 FOC 系统的 Jacobian（$J \times J$ 矩阵，包含利润函数的二阶导数）。牛顿-拉弗森在解附近具有二次收敛速度——但需要计算二阶导数 $\partial^2 s_k / \partial p_j \partial p_\ell$，这在复杂需求模型（如 BLP）中计算量大。

不动点迭代（Fixed-Point Iteration）：

$$p_j^{(t+1)} = \widehat{mc}_j + \text{markup}_j(\mathbf{p}^{(t)}) \tag{12.38}$$

其中 markup_j 从使用当前价格向量 $\mathbf{p}^{(t)}$ 的 FOC 中计算。这利用了 Bertrand-Nash 的结构——不需求解二阶导数（不需要 Hessian）。但收敛较慢，且可能循环。

实践建议：先用不动点迭代“走近”解（全局收敛性好），然后切换到牛顿-拉弗森（更快、更精确）。大多数现代实现（如 PyBLP）使用混合方法或信赖域方法（Trust-Region Methods）。

	牛顿-拉弗森	不动点迭代	混合方法
收敛速度	二次（快）	线性（慢）	先线性后二次
需要二阶导数？	是（$\partial^2 s_k / \partial p_j \partial p_\ell$）	否	前半段否，后半段是
全局收敛性？	弱（需要好的初始猜测）	较好（但可能循环）	好（前半段全局收敛）
实现复杂度	高（需要 Hessian 分析式）	低（只需 FOC 评估）	中等
BLP 适用性	噪声 Hessian（蒙特卡洛积分）	简单但慢	PyBLP 默认策略

表 12.6: 并购模拟中三种数值方法的比较。在实践中，PyBLP 使用混合策略：初期使用不动点迭代或信赖域方法接近解，然后切换到牛顿-拉弗森加速收敛。

12.7.2　PyBLP 中的并购模拟

Conlon & Gortmaker (2020) 的 `pyblp` Python 包提供了内置的并购模拟功能。基本工作流程如下（伪代码）：

```
import pyblp
import numpy as np

# 1. 构建问题并估计需求（BLP）
problem = pyblp.Problem(product_data, formulation)
results = problem.solve()

# 2. 从并购前 FOC 恢复边际成本
mc = results.compute_costs()

# 3. 定义并购后所有权
post_ownership = ownership.copy()
post_ownership[merging_indices[:, None] & merging_indices[None, :]] = 1

# 4. 可选：纳入并购效率
mc_with_efficiency = mc.copy()
mc_with_efficiency[merging_indices] *= (1 - efficiency)

# 5. 求解并购后均衡
post_prices = results.compute_prices(
    ownership=post_ownership,
    costs=mc_with_efficiency
)

# 6. 计算福利变化
delta_p = post_prices - pre_prices
```

`compute_prices()` 方法内部使用牛顿法求解并购后 FOC 系统，利用隐函数定理高效计算二阶导数。PyBLP 还提供了 bootstrap 或 delta-method 标准误差计算——这对法庭上的统计推断至关重要。

12.7.3 常见数值问题

(1) **FOC 不收敛：**如果牛顿法发散，尝试不动点迭代作为 "全局化器"——或用信赖域方法。检查并购后所有权是否产生了接近奇异的结构（极高的合并方内分流比可能导致数值不稳定）。

(2) **局部最优 / 非经济解：**始终从并购前价格 p^{pre} 开始迭代。检查 $p_j > mc_j$ 对所有的 j——负加成意味着求解器落入了非经济区域。

(3) **BLP 中的噪声导数：**BLP 的需求导数涉及对随机系数分布的蒙特卡洛积分。使用重要性采样（Importance Sampling，PyBLP 默认方法）减少积分噪声。

(4) **模型依赖的成本校准：**$\widehat{mc}_j$ 从并购前 FOC **恢复**——不是直接观测的。如果需求模型被错误设定，$\widehat{mc}_j$ 吸收错误设定偏差，污染并购后预测。敏感性分析（在不同需求规格下重新估计并购效应）至关重要。

12.8　中国反垄断的制度语境

【前沿】

12.8.1　中国反垄断法的制度架构

中国《反垄断法》（Anti-Monopoly Law, AML）于 2008 年生效，2022 年完成首次重大修订。AML 包含三大核心内容，对应三种反竞争行为：

(1) **垄断协议**（Monopoly Agreements）——卡特尔（第 ~8~ 章的合谋理论）

(2) **滥用市场支配地位**（Abuse of Dominance）——单一企业行为（第 ~7~ 章和 第 ~9~ 章）

(3) **经营者集中审查**（Merger Control）——并购审查（本章）

2018 年机构改革前，商务部（MOFCOM）负责经营者集中审查。2018 年后，**国家市场监督管理总局**（SAMR, 国家市场监督管理总局）统一反垄断执法。SAMR 下设反垄断局（Anti-Monopoly Bureau），负责并购审查的核心工作。

中国的并购审查标准与美欧类似——以“实质性减少竞争”（SLC）为实质标准。但审查程序上有不同的特征：2022 年修订引入了“停钟”（Stop-the-Clock）机制——允许审查在信息不完整时暂停，解决了此前审查时间表刚性的问题。

12.8.2　中国并购审查中的量化分析

当前，中国并购审查仍以**定性分析**为主导。市场界定和竞争评估主要基于行业研究、专家意见和企业提交的定性论证。完整的结构性并购模拟（BLP 需求估计 + 并购后均衡模拟）在中国的应用**远少于**美国或欧盟。

主要障碍：

- **数据可得性：**市场层面的扫描数据（Scanner Data）——BLP 估计所需的标准输入——在中国不如在美国或欧洲容易获得。
- **制度惯性：**中国的法律和行政传统偏向于定性、以机构知识和部门协商为基础的决策。

- **人才**：并购模拟需要专门的 IO 经济学家——这个专业群体在中国仍相对薄弱。

但情况正在改变。SAMR 越来越多地聘请经济学家并发表经济分析——特别是在附加限制性条件（Remedies）的决定书中。2022 年 AML 修订引入了更详细的经济证据指南。中国学术界使用中国数据的 BLP 型需求估计研究正在为未来应用奠定方法基础。

中国市场：标志性并购案 **可口可乐/汇源（2009 ——被禁止）**：中国《反垄断法》实施后第一个被禁止的并购案。MOFCOM 的论证是：并购将使可口可乐获得果汁市场的主导地位。当时的经济分析以定性为主——这是中国并购审查的“信号案例”。

滴滴/优步中国（2016 ——未依法申报）：滴滴收购优步中国——在中国网约车市场创造了近乎垄断的地位。并购后滴滴大幅提高价格，降低司机补贴——这正是单边效应预测的行为。此案展示了**双边市场**（平台—第 ~13~ 章）中并购分析的复杂性，以及并购审查的“门槛”问题（优步中国的收入低于申报门槛但竞争效应显著）。

12.8.3 中美欧并购审查比较

维度	美国 (DOJ/FTC)	欧盟 (EC)	中国 (SAMR)
审查标准	SLC	SIEC	SLC
量化分析	广泛（UPP、并购模拟常规使用）	广泛（第二阶段深入调查）	有限但增长中
分流比的角色	核心（2010/2023 指南明确认可）	核心（与 UPP/GUPPI 共同使用）	有限（未明确定义为法律标准）
市场界定	“不再是必要的第一步”	仍是必要组成部分（2024 通知更灵活）	仍是分析的核心起点
执法风格	对抗性（法庭/和解）	行政性（委员会决定，可上诉）	行政性（SAMR 决定，有限上诉通道）
经济学家的作用	内部 + 双方顾问	内部 + 双方顾问 + 听证	有限的内部经济学家，增长中

表 12.7: 中美欧反垄断并购审查的比较。SIEC = Significant Impediment to Effective Competition（对有效竞争的重大阻碍）——欧盟标准。

12.9　总结、局限与通往 Ch.13 的桥梁

【核心】

12.9.1　本章的核心方法论文档

本章完成了 NEIO 模块的第三块也是最后一块拼图。第 ~10~ 章建立了同质产品市场的行为识别框架（Bresnahan 纲领）；第 ~11~ 章建立了差异化产品市场的需求估计框架（BLP 纲领）；本章将两者整合为可操作的反垄断量化工具。

本章传授了三个核心能力：

(1) **市场界定作为 "可检验的假设"**：SSNIP 测试将市场界定从 "假设"（SCP 传统——先验地定义行业边界）转变为 "可检验的科学命题"（假想垄断者能否有利可图地提价 5–10%？其答案由分流比和毛利率决定）。

(2) **并购模拟作为 "预测引擎"**：改变 $\boldsymbol{\Omega}$ 矩阵 $\rightarrow$ 求解新 Bertrand-Nash 均衡 $\rightarrow$ 计算 Δp_j, ΔCS, ΔW。这是 NEIO 工具包的 "核心产品"——将参数估计转化为可行动的反事实预测。

(3) **分层分析的逻辑**：从 HHI 和集中度变化（最简单、最不精确）到 GUPPI/UPP（中间复杂度、绕过市场界定）到完整结构性并购模拟（最复杂、最准确、最数据密集）。选择取决于分析目的、可用数据和审查阶段。

12.9.2　局限与未解决的问题

- **模型依赖**：并购模拟的预测只是与需求函数形式（logit vs. 嵌套 logit vs. BLP）一样好。不同模型可以产生定性和定量不同的预测。这是对法庭上使用并购模拟的最强挑战——"你的结果是由模型驱动的吗，还是由数据驱动的？"

- **协调效应的量化鸿沟**：没有广泛接受的协调效应结构性模型。定量并购分析主要限于单边效应——但协调效应在许多并购中可能是同样重要的甚至主导的竞争关注。

- **效率的可验证性**：并购效率的主张难以独立验证——且经常被夸大。这限制了 Williamson 权衡在实际案件中的政策应用。

- **局部均衡 vs. 一般均衡**：并购模拟通常假定为局部均衡（只考虑并购行业内的效应）——忽略对上游供应商、下游零售商和互补市场的影响。

- **不可观测的成本**：$\widehat{mc}$ 是从并购前 FOC **恢复**的——而非直接观测的。如果需求模型被错误设定 $\rightarrow$ $\widehat{mc}$ 吸收错误设定偏差 $\rightarrow$ 并购后预测有偏。

12.9.3 通往 Ch.13 的桥梁：平台与双边市场

本章的所有工具——SSNIP 测试、并购模拟、UPP/GUPPI、损害量化——都假设**单边市场**（One-Sided Market）：企业只向消费者销售产品，只关心消费者一侧的弹性和替代模式。

第 ~13~ 章进入**双边市场**（Two-Sided Markets / Platforms）的世界。平台连接两组需要互动的用户：网约车（乘客和司机）、电商（买家和卖家）、在线广告（用户和广告主）、支付卡（持卡人和商户）。双边市场有三个特征，对本章的所有工具构成根本性挑战：

(1) **交叉网络外部性**（Cross-Group Network Externalities）：一组用户的参与增加了平台对另一组用户的价值。这意味着——与单边市场不同——价格结构（两组之间的价格分配）与价格水平（总价格）同等重要。并购模拟需要扩展需求系统来包括两组用户的弹性和交叉效应。

(2) **零定价市场**（Zero-Price Markets）：许多平台对一组用户免费（如 Google 搜索对用户免费，对广告主收费）。SSNIP 测试在价格为 0 时如何运作？从 0 到小幅正的提价是 “无限百分比” 的提价——标准的 SSNIP 框架崩溃。

(3) **市场界定的 “哪一边” 问题**：平台连接两组用户——在 SSNIP 测试中，我们应该在哪一侧进行测试？还是需要在两侧同时测试？不同的选择可能产生不同的市场界定结论。

第 ~13~ 章将以 Rochet & Tirole (2003, 2006) 的平台定价模型为基础，系统探讨双边市场中的市场界定、并购分析和反垄断政策——以及中国平台经济（滴滴、美团、阿里巴巴、腾讯）带来的独特制度挑战。

习题

基础题

Ex 12.1 临界损失分析。考虑一个候选市场，毛利率 $m = 0.4$，SSNIP 幅度 $t = 5\%$。

(a) 计算临界损失 $CL = t/(t + m)$。

(b) 如果实际损失（Actual Loss）是 18%，提价是否有利可图？候选市场是否应该扩大？

(c) 重新计算 $m = 0.2$ 和 $m = 0.6$ 时的 CL。为什么高毛利率使 CL 更小？解释经济学直觉。

(d) 如果 SSNIP 幅度 $t = 10\%$（而非 5%），CL 如何变化？这符合直觉吗？

Ex 12.2 所有权矩阵与分流比。考虑一个市场有 4 个产品：$\{A, B, C, D\}$。并购前，企业 1 拥有 A 和 B，企业 2 拥有 C，企业 3 拥有 D。

(a) 写出 4×4 的所有权矩阵 $\mathbf{\Omega}^{\text{pre}}$（1 表示同一企业，0 表示不同企业）。

(b) 现在企业 1 收购了企业 2（合并方拥有 A、B、C）。写出 $\mathbf{\Omega}^{\text{post}}$。

(c) 解释哪些 Ω 元素从 0 变为 1 ——以及这如何改变了合并方的 FOC。

(d) 如果 $D_{AC} = 0.6$ 且 $D_{AD} = 0.1$，解释为什么这个并购可能引起显著的竞争关注。

Ex 12.3 UPP 与 GUPPI 的计算。两个产品：产品 1（$p_1 = \$10$, $mc_1 = \$4$），产品 2（$p_2 = \$12$, $mc_2 = \$5$）。分流比：$D_{12} = 0.25$, $D_{21} = 0.30$。

(a) 计算并购前两个产品的 Lerner Index。

(b) 计算产品 1 和产品 2 的 UPP（美元值）。

(c) 计算 $GUPPI_1$ 和 $GUPPI_2$。根据 5% GUPPI 阈值，这个并购是否引起关注？

(d) 如果并购产生 10% 的边际成本减少（$E_1 = E_2 = 0.10$），计算 Net UPP。效率是否消除了上向定价压力？

Ex 12.4 并购模拟在两产品 logit 模型中。市场有 3 个内部产品加外部产品（$u_0 = 0$）。Logit 需求：$u_j = \delta_j - \alpha p_j + \varepsilon_{ij}$。参数：$\delta_1 = 2$, $\delta_2 = 1.5$, $\delta_3 = 1$；$\alpha = 0.5$；$mc_1 = mc_2 = mc_3 = 3$。并购前每家拥有一个产品。企业 1 收购企业 2。

(a) 计算并购前份额 s_j 和均衡价格 p_j^{pre}。（提示：使用 logit FOC 的矩阵形式。）

(b) 写出并购后的 $\mathbf{\Omega}^{\text{post}}$。计算 p_j^{post} 和 Δp_j。

(c) 比较产品 1 和 2 的 Δp_j。哪个更大？为什么？

(d) 产品 3（非合并方）的价格是否变化？解释为什么。

进阶题

Ex 12.5 并购模拟的敏感性与需求形式。使用 Ex 12.4 的相同设定，但改用嵌套 logit：产品 $\{1, 2\}$ 在同一巢中（$\sigma = 0.4$），产品 3 在自己的巢中。

(a) 重新计算并购前均衡价格 p^{pre}。与标准 logit 比较——嵌套 logit 的加成是更高还是更低？

(b) 重新计算并购后价格 p^{post}。Δp_1 在两种需求模型下如何比较？

(c) 解释为什么嵌套 logit 的 Δp_1 更大。这与 "logit 低估了近竞争者并购的单边效应" 的论述是否一致？

(d)（讨论题）如果你在法庭上提交了并购模拟结果，而合并方辩称 "你的结论是由需求函数形式驱动的——使用更简单的 logit 模型价格效应要小得多"——你要如何回应？

Ex 12.6 SSNIP 与 Cellophane Fallacy。考虑一个垄断者，$P = 100 - Q$，$MC = 30$（常数）。

(a) 计算垄断价格 P^M、数量 Q^M、和垄断毛利率 m。

(b) 从垄断价格出发进行 SSNIP 测试（$t = 5\%$）。计算实际损失 L 和临界损失 CL。提价是否有利可图？

(c) 现在从竞争价格（$P = MC = 30$）出发重新进行 SSNIP 测试。在这种情况下，提价是否有利可图？比较 (b) 和 (c) ——这如何说明了 Cellophane Fallacy？

(d)（讨论题）在 du Pont 玻璃纸案中，du Pont 辩称 "蜡纸和铝箔是玻璃纸的替代品，因此我的市场份额不到 20%"。使用 (b) 的逻辑，解释为什么这个论证受到了 Cellophane Fallacy 的污染。

Ex 12.7 损害量化的 DiD 方法。假设并购发生在城市 A，城市 B 是 "控制" 城市（未发生并购）。你有以下数据：

	城市 A	城市 B
并购前价格	\$100	\$95
并购后价格	\$115	\$100

(a) 计算 "事前-事后" 估计的过度要价。

(b) 计算 DiD 估计的过度要价。

(c) 这两个估计为什么会差距这么大？哪个更好？（提示：城市 B 的价格也上升了——可能因为共同成本冲击。）

(d) 假设城市 B 的消费者在并购后转向了城市 B 的企业（因为城市 A 的价格上升）——这推高了城市 B 的价格（伞形效应）。在这种情况下，DiD 估计是过度要价的**上界**还是**下界**？

应用/实证题

Ex 12.8 中国行业 I：滴滴/优步并购的竞争分析。回顾滴滴–优步中国并购案（2016 年——未依法申报，但随后被 SAMR 调查和处罚）。

(a) 从单边效应的角度解释：为什么滴滴和优步中国是“近竞争者”？并购后价格上升和司机补贴减少的证据是什么？

(b) 解释为什么双边市场（有司机和乘客）使这个案例中的市场界定复杂化。网约车和传统出租车是同一相关市场吗？

(c)（讨论题）如果中国并购审查在 2016 年进行了完整的结构性并购模拟，研究者需要什么数据？主要障碍是什么？

Ex 12.9 中国行业 II：市场界定中的量化工具。选择你熟悉的一个中国行业（电商、共享单车、在线教育、乳制品、啤酒或白电）。

(a) 确定行业中的 3–4 家主要竞争者及其标志性产品/品牌。

(b) 设计一个简单的消费者调查，估计这些产品/品牌之间的分流比。解释你怎么设计问题（“第二选择”数据）。

(c) 使用估计的分流比（即使只是近似——用你自己的行业知识），对并购双方的产品进行 UPP/GUPPI 计算。GUPPI 是否超过 5%？

(d)（讨论题）在这个行业中，进行完整的结构性并购模拟（BLP 型）是否可行？你需要什么类型的数据？这些数据在中国是否可获得？

Ex 12.10 并购效率的 Williamson 权衡。考虑一个行业，$P = 200 - Q$（线性），并购前 $MC_0 = 60$（常数）。并购将 MC 降低到 $MC_1 = 50$（成本效率），但可能将价格从 p_0 提升到 p_1。

(a) 假设 $p_0 = 60$（并购前是完全竞争——$P = MC_0$）。计算 Q_0、消费者剩余和生产者剩余。

(b) 并购后，合并实体成为垄断者，$MC = 50$。计算垄断价格 P^M 和消费者剩余的变化。

(c) 计算效率效应（成本从 60 降到 50 的节约）和市场势力效应（从完全竞争到垄断的额外 DWL）。总福利是上升还是下降？

(d) 现在假设并购前是 Cournot（$n = 3$ 家企业，$MC_0 = 60$），并购后合并实体与第 3 家企业进行 Cournot 竞争（“并购 $\rightarrow$ 双寡头”）。重新计算福利效应。这种“介于”情形下的结论与 (c) 不同吗？

推荐阅读

核心必读

1. **Davis, P. & Garcés, E. (2010).** *Quantitative Techniques for Competition and Antitrust Analysis.* Princeton University Press. Ch.4（市场界定）, Ch.5（市场结

构与价格效应）, Ch.8（并购模拟）, Ch.9（损害）。—本章的教科书——全面覆盖从 SSNIP 到并购模拟再到损害量化的完整量化工具箱。

2. **Farrell, J. & Shapiro, C. (2010).** "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition." *B.E. Journal of Economic Analysis & Policy*, 10(1). —UPP/GUPPI 的原始宣言——论证了分流比可以直接替代市场界定。

3. **Nevo, A. (2000).** "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." *RAND Journal of Economics*, 31(3), 395–421. —结构性并购模拟的奠基性实证论文——展示了从 BLP 需求到并购后价格预测的完整流程。

4. **Werden, G. & Froeb, L. (1994).** "The Effects of Mergers in Differentiated Products Industries: Logit Demand and Merger Policy." *Journal of Law, Economics, & Organization*, 10(2), 407–426. —并购模拟的理论基础——展示了需求曲率（logit vs. 线性 vs. 常数弹性）如何影响并购预测。

5. **U.S. DOJ & FTC (2010).** *Horizontal Merger Guidelines*. §4, 6（单边效应）, §10（效率）。—将 UPP/GUPPI 和分流比分析正式纳入反垄断政策框架的关键文件。

6. **Williamson, O. (1968).** "Economies as an Antitrust Defense: The Welfare Tradeoffs." *American Economic Review*, 58(1), 18–36. —并购效率 vs. 市场势力的经典权衡——每个反垄断经济学家必读的 3 页图形分析。

进阶阅读

7. **Miller, N. & Weinberg, M. (2017).** "Understanding the Price Effects of the MillerCoors Joint Venture." *Econometrica*, 85(6), 1763–1791. —并购模拟的验证——展示了结构性模型可以准确预测实际并购后价格变化。

8. **Conlon, C. & Gortmaker, J. (2020).** "Best Practices for Differentiated Products Demand Estimation with PyBLP." *RAND Journal of Economics*, 51(4), 1108–1161. —使用 PyBLP 进行可复现并购模拟的现代实践指南。

9. **Hausman, J., Leonard, G. & Zona, J.D. (1994).** "Competitive Analysis with Differentiated Products." *Annales d'Economie et de Statistique*, 34, 159–180. —多阶段预算需求模型——并购分析中 BLP 的替代方法。

10. **Berry, S. & Pakes, A. (1993).** "Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Merger Analysis." *AER Papers & Proceedings*, 83(2), 247–252. —结构需求估计与并购分析的最早连接。

11. **Nevo, A. (2001).** "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2), 307–342. —Nevo (2000) 的更完整结构模型——联合估计需求和供给，提供内部一致的加成估计。

12. **Bjornerstedt, J. & Verboven, F. (2016).** "Does Merger Simulation Work? Evidence from the Swedish Analgesics Market." *AEJ: Applied Economics*, 8(3), 125–164. —并购模拟预测准确性的另一个验证。

13. **U.S. DOJ & FTC (2023).** *Merger Guidelines.* —最新的美国并购指南——反映了 2010–2023 年对数字平台并购的思考。

14. **European Commission (2024).** *Commission Notice on the Definition of the Relevant Market.* —欧盟最新市场界定通知——反映数字经济和双边市场的影响。

中国相关

15. **《中华人民共和国反垄断法》（2008 年生效，2022 年修订）。**—第四章（经营者集中）——中国并购审查的法律框架。

16. **国务院反垄断委员会 (2020).** 《关于平台经济领域的反垄断指南》。—SAMR 在数字经济并购审查中的指南。

17. **王晓晔 (2017).** 《反垄断法》. 法律出版社. —中国反垄断法学术评注。

18. **SAMR 反垄断局年度报告和重大案件决定书。**—中国并购审查实践的主要来源。

前沿/方法论文献

19. **Jaffe, S. & Weyl, E.G. (2013).** "The First-Order Approach to Merger Analysis." *American Economic Journal: Microeconomics*, 5(4), 188–218. —UPP/GUPPI 方法的一般化。

20. **Epstein, R.J. & Rubinfeld, D.L. (2001).** "Merger Simulation: A Simplified Approach with New Applications." *Antitrust Law Journal*, 69(3), 883–919. —PCAIDS 简化并购模拟方法。

21. **Rochet, J.-C. & Tirole, J. (2003).** "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association*, 1(4), 990–1029. —为 Ch.13 预读——双边市场的基础模型。

22. **Evans, D.S. & Schmalensee, R. (2016).** *Matchmakers: The New Economics of Multisided Platforms.* Harvard Business Review Press. —Ch.7–9（平台并购和反垄断——为 Ch.13 预读）。

第十三章　双边市场与平台经济学

章首语

Google 对用户免费，对广告商收费。微信对用户免费，对小程序开发者收费。信用卡对持卡人免费（甚至给回馈），对商户收取 2–3% 的手续费。这种 “一侧免费、另一侧收费” 的商业模式在数字经济中无处不在。

在传统单边市场（One-Sided Market）框架中，长期持续的 “低于成本定价” 是掠夺性定价（Predatory Pricing）的警示信号——企业通过亏损将竞争对手赶出市场，待垄断后再提价回收损失。但在双边市场（Two-Sided Market）中，**一侧免费不是掠夺的信号——它是利润最大化的一阶条件**。这一根本性差异意味着，第 ~12~ 章中构建的整个反垄断量化工具链（SSNIP、临界损失、UPP/GUPPI、并购模拟、损害量化）在平台经济中需要 “重装系统”。

本章是模块六（现代前沿）的入口章，标志着从传统单边市场 IO 向现代平台经济学的范式转换。本章将建立双边市场理论的核心架构：什么是双边市场？为什么倾斜定价是最优的？单归属与多归属如何决定竞争强度？市场何时倾覆、何时共存？以及——最重要的是——传统反垄断工具在双边市场中如何失效，又如何被修正？

13.1　什么是双边市场？

13.1.1　从 Ch.12 到 Ch.13：为什么传统工具需要重装系统？

第 ~12~ 章构建了从 SSNIP 测试 → 临界损失分析 → 并购模拟 → UPP/GUPPI → 反垄断损害量化的完整工具链。这些工具共享一个隐含假设：**单边市场**——企业只向一组消费者销售产品，消费者的需求只取决于产品特征和价格，企业之间的竞争只通过价格和品质发生。

然而，三个 “破坏性事实” 挑战了这一假设：

(1) **SSNIP 测试的零价格问题：**Google 搜索的价格为零。$5\% \times 0 = 0$——SSNIP 测试在数学上失去意义。假想垄断者能否 “有利可图地提价 5%”？从零到零，问题本身就丧失了操作性。

(2) **掠夺性定价测试的失效：**Google 长期以零价格提供搜索服务。Areeda-Turner 测试会将零价格（低于任何可想象的边际成本）推定为掠夺性。但在平台经济中，**一侧免费是常态而非例外**——这不是掠夺，是最优定价。

(3) **并购分析的缺失维度：**当 Facebook 收购 Instagram 时，标准 UPP/GUPPI 框架只看单侧（如广告侧或用户侧）的转移率。但双边平台上的价格变动会通过交叉网络外部性在两个方向反馈——一侧的涨价不仅影响该侧，还通过改变平台对另一侧的吸引力产生间接效应。忽视这个反馈循环，可能系统性地高估或低估并购的竞争损害。

这些失效不是微小的技术性问题——它们是**范畴性的**（categorical）。双边市场不是单边市场的 “微调版本”，而是要求我们重新思考 IO 分析中最基本的范畴：什么是市场？什么是竞争？什么是垄断力？

核心张力

本章的核心智识问题：**传统 IO 工具箱在平台/双边市场中如何失效，以及我们如何重建这些工具。**这既是一个理论问题（如何形式化交叉网络外部性对定价和竞争的影响？），也是一个制度问题（反垄断法和管制在平台经济中应如何运作？）。

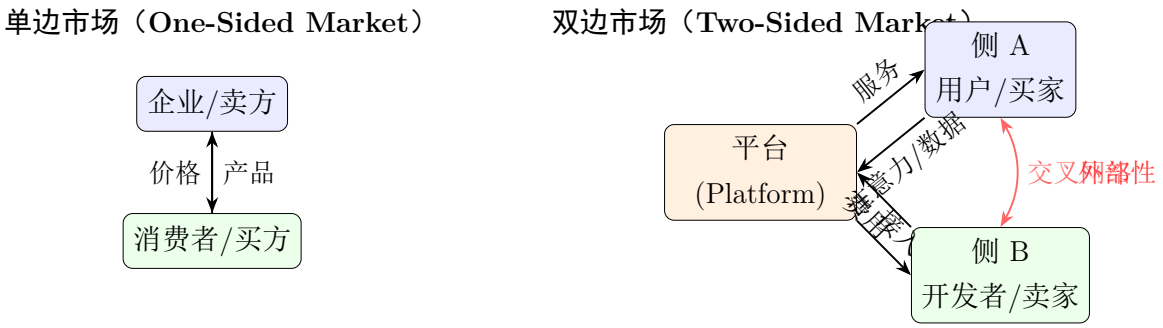


图 13.1 展示了单边市场与双边市场的结构差异。在单边市场中，企业只需管理一组客户关系。在双边市场中，平台必须同时管理两组用户——不仅管理每组用户的定价，还要管理两组用户**之间的**交叉网络外部性。

13.1.2　科斯定理为什么失败了？

理解双边市场的根本经济学直觉是：**科斯定理（Coase Theorem）在此失效了。**

如果科斯定理成立——两侧用户可以无成本地协商和再分配资源——那么市场就不可能是 “双边” 的。无论平台如何将总价格分配在两侧之间，两侧用户都可以通过 side payments（侧支付）自行重新配置负担。最终，只有**总价格水平**（total price

level）起作用，价格在哪一侧收取是无关紧要的（neutral）。科斯定理的失效是双边市场的**必要条件**（虽然并非充分条件——仍需满足价格结构非中性的技术条件）。

在平台经济中，科斯定理因三个原因失效：

(1) **交易成本（Transaction Costs）**：让每一个持卡人与每一个接受该卡的商户就每笔交易的刷卡费进行双边谈判是不现实的。聚合千万级用户的平台本身就是为了**降低交易成本**而存在——如果交易成本已经很低，就不需要平台了。

(2) **平台施加的定价约束（Platform-Imposed Pricing Constraints）**：信用卡网络长期实行**无附加费规则**（No-Surcharge Rule）——美国运通（American Express）禁止商户向使用运通卡的顾客收取比现金顾客更高的价格。这一规则直接阻止了 Coasean 再谈判：商户即使想将刷卡手续费转嫁给刷卡顾客，也被合同禁止。

(3) **信息不对称（Information Asymmetry）**：用户通常不知道平台的价格结构，因此无法“纠正”它。大多数 Google 用户不知道 Google 向广告商收取了多少费用，也不知道自己的搜索行为为 Google 创造了多少广告收入。

科斯定理与双边性的经济学直觉

科斯定理说的是：只要交易成本为零，资源的初始分配不影响最终效率。双边市场理论说的是：**当交易成本为正时，价格的分配（不仅是价格的总水平）本身就影响资源配置**。平台之所以存在，正是因为它以一种比直接双边谈判更低成本的方式内部化了这些交易成本——但在内部化的过程中，平台也获得了影响**价格分配**的能力，而这种分配本身就是竞争策略的核心维度。

13.1.3 经典例子与概念区分

在建立正式模型之前，我们先通过全景表格建立直觉。表 13.1 展示了 8 类经典双边平台的结构。

平台类型	侧 A	侧 B	主要外部性	交易型？	补贴侧
支付卡	持卡人	商户	使用外部性	是	持卡人
操作系统	用户	应用开发者	成员外部性	否	用户
新闻媒体	读者	广告商	成员外部性	否	读者
购物中心	消费者	零售商	成员外部性	否	消费者
相亲平台	用户	用户	匹配外部性	是	女性用户
网约车	乘客	司机	使用外部性	是	乘客/司机交替
电商平台	买家	卖家	使用外部性	是	买家
视频游戏机	玩家	游戏开发者	成员外部性	否	玩家

表 13.1 揭示了一个规律：补贴侧几乎总是“更难吸引”或“产生更大的交叉外部性”的一侧。持卡人比商户更容易改用现金（弹性更高），所以获得补贴。读者比广告商更容易换报纸（弹性更高），所以订阅费低于印刷成本。这是最优定价的结构性规律，而非巧合。

需要区分几个关键概念：

(1) **成员外部性（Membership Externality）vs. 使用外部性（Usage Externality）：**成员外部性来自**谁在平台上**——视频游戏开发者关心有多少玩家拥有该主机（无论他们本月玩了多少游戏）。使用外部性来自**实际交易**——持卡人在某商户刷卡的收益来自这次刷卡本身。前者对应固定会员费（lump-sum membership fee），后者对应按次交易费（per-transaction fee）。Rochet-Tirole (2006) 的统一框架将两者整合为两部收费制（two-part tariff）。

(2) **交易型平台（Transaction Platform）vs. 非交易型平台（Non-Transaction Platform）：**这一区分因 Ohio v. Amex (2018) 而具有法律意义。交易型平台“无法仅向一侧销售而不向另一侧销售”——每笔信用卡交易同时涉及持卡人和商户。非交易型平台的两侧不直接交易——报纸读者和广告商之间没有可被观测到的单笔交易。美国最高法院在 Amex 案中裁定，交易型平台必须将两侧纳入同一个相关市场。

13.1.4　双边市场的正式定义：价格结构非中性

Rochet 和 Tirole (2006, *RAND Journal of Economics*) 提供了双边市场最精确的数学定义。

定义 13.1 (双边市场，Rochet-Tirole 2006). 令 a^B 和 a^S 分别为平台向买家（*Buyers*）和卖家（*Sellers*）收取的**每笔交易费**。令总价格水平为 $a = a^B + a^S$。令交易量为 $V(a^B, a^S)$。

该市场是**双边的**（*two-sided*），当且仅当：**在保持总价格水平 a 不变的条件下，改变价格在两侧之间的分配（即改变价格结构 a^B/a^S）会改变交易量 V。**

等价地：V 不仅是 a 的函数，且 $\partial V/\partial(a^B/a^S)\big|_{a\ constant} \neq 0$。

价格水平 vs. 价格结构

价格水平（Price Level）：$p = p^B + p^S$——“坐一趟车总共多少钱。”
价格结构（Price Structure）：(p^B, p^S) 的分配——“司机和乘客各承担多少。”
在单边市场中，只有价格水平起作用。在双边市场中，**价格结构独立于价格水平起作用**——这是平台的额外战略自由度。平台不仅选择“收多少钱”，还选择“向谁收多少钱”——而后者本身就影响总交易量。

产生双边性的三个充分条件（Rochet-Tirole 2006）：(1) 两侧之间存在未被内部化的外部性（如果没有平台，两侧无法就外部性进行补偿）；(2) 交易成本阻止了 Coasean 再谈判；(3) 平台拥有可用的价格工具来影响两侧的参与和使用决策。

这些条件解释了为什么有些看起来很“双边”的市场实际上表现得像单边市场。超市连接消费者和供应商，但如果超市与供应商之间的批发价格可以通过双边谈判灵活确定，而零售价格完全由超市与消费者的关系决定，那么超市市场更接近单边市场逻辑——只有一个价格（零售价）决定交易量。

价格结构 A：$a^B = 1, a^S = 4$，**总价** $= 5$　　**价格结构 B**：$a^B = 3, a^S = 2$，**总价** $= 5$

交易量 $V = 100$　　交易量 $V = 60$

⇓ 总价格相同，但价格分配不同 ⇒ 交易量不同
⇒ **市场是双边的（价格结构非中性）**

图 13.2 用一个简单数值对比说明了价格结构非中性的含义。总价格水平相同（$a = 5$），但不同的价格分配（1+4 vs. 3+2）产生了截然不同的交易量（100 vs. 60）。如果市场是单边的，改变价格分配不应该影响交易量——科斯定理保证了两侧用户可

以通过自行协商来消除分配的影响。事实是交易量变了——所以科斯定理不成立，所以市场是双边的。

13.2　倾斜定价：为什么一侧免费？

13.2.1　Rochet-Tirole (2003) 垄断平台模型

Rochet 和 Tirole (2003, *JEEA*) 提供了双边市场定价理论的奠基性模型。考虑一个垄断平台（Monopoly Platform）服务买家（B, Buyers）和卖家（S, Sellers）。平台对每笔交易向买家收取 a^B、向卖家收取 a^S。平台的每笔交易边际成本为 c。

设定（Setup）：

- 买家准需求函数（quasi-demand）：$D^B(a^B)$——在给定卖家数量下，愿意以每笔交易费 a^B 使用平台的买家数量
- 卖家准需求函数：$D^S(a^S)$
- 交易量：$V(a^B, a^S) = D^B(a^B) \cdot D^S(a^S)$
- 注意：这里假设需求是**可分离的**（separable）——买家和卖家的需求各自独立取决于自己的价格。这是简化假设，但保留了核心经济学机制

利润最大化问题：

$$\max_{a^B, a^S} \ \pi = (a^B + a^S - c) \cdot D^B(a^B) \cdot D^S(a^S) \tag{13.1}$$

一阶条件（First-Order Conditions）：

对 a^B 求导：

$$\frac{\partial \pi}{\partial a^B} = D^S \cdot \left[D^B + (a^B + a^S - c) \cdot D^{B\prime} \right] = 0 \tag{13.2}$$

对 a^S 求导：

$$\frac{\partial \pi}{\partial a^S} = D^B \cdot \left[D^S + (a^B + a^S - c) \cdot D^{S\prime} \right] = 0 \tag{13.3}$$

定义每侧的需求弹性（elasticity）：

$$\eta^B \equiv -\frac{a^B \cdot D^{B\prime}}{D^B} > 0, \quad \eta^S \equiv -\frac{a^S \cdot D^{S\prime}}{D^S} > 0 \tag{13.4}$$

从一阶条件消去公共项，得到两个关键结果：

修改后的 Lerner 公式（Modified Lerner Formula）

价格结构条件（Price Structure Condition）：

$$\boxed{\frac{a^B}{\eta^B} = \frac{a^S}{\eta^S}} \tag{13.5}$$

价格水平条件（Price Level Condition）：

$$\boxed{\frac{a^B + a^S - c}{a^B + a^S} = \frac{1}{\eta^B + \eta^S}} \tag{13.6}$$

公式 (13.5) 是**双边市场定价理论最核心的公式**。它告诉我们：

(1) **逆弹性规则应用于价格结构**：弹性较低的一侧承担较高的价格。这与单边市场的 Lerner 指数 $p - mc/p = 1/\eta$ 在精神上一致——但现在应用于**两侧之间**的相对价格，而不是总价格水平。

(2) **价格结构的决定完全独立于边际成本 c**：公式 (13.5) 中根本不出现 c。价格结构仅由两侧的**相对弹性**决定。边际成本 c 只影响总价格水平（公式 13.6）。

(3) **“免费”的理性**：如果 $\eta^B \to \infty$（买家对价格极端敏感），则 $a^B \to 0$。如果买家对价格无限弹性，最优买家费就是零。这就是 Google 搜索免费的第一层解释。

13.2.2 机会成本解释：最直观的理解

改写平台对买家的定价 FOC，得到一个更有启发性的形式：

$$a^B = \underbrace{c - a^S}_{\text{机会成本}} + \underbrace{\frac{D^B}{-D^{B\prime}}}_{\text{Lerner 加成}} \tag{13.7}$$

机会成本直觉

服务一个额外买家的**真实机会成本（Opportunity Cost）**不是 c（平台的技术边际成本），而是 $c-\mathbf{a^S}$（技术边际成本**减去**该买家带来的卖家费用收入）。

推理链：

1. 每个额外买家会与平台上已有的卖家进行交易
2. 每笔交易平台向卖家收取 a^S
3. 如果平台不服务这个买家，这些交易就不会发生——平台损失了本可收取的 a^S
4. 因此，服务买家的**净成本** = 技术成本 c − 放弃的卖家收入 a^S
5. 如果 $a^S > c$（卖家费用超过了每个买家的技术成本），机会成本为**负数**——平台在每笔服务买家的交易中**赚钱**而不是亏钱
6. 机会成本为负 $\Rightarrow$ 平台应该**补贴**买家——包括零价格甚至负价格

这就是“免费的 Google”的经济逻辑：每个额外搜索用户为 Google 带来的广告展示机会（从而广告收入）远大于服务这个用户的带宽和计算成本。服务用户的**机会成本为负**——免费不是慈善，是利润最大化的必然结果。

13.2.3　谁被补贴？——决定因素框架

综合理论和实证，补贴侧（Subsidy Side）由以下三个因素决定：

1. **更高的需求弹性（Higher Demand Elasticity）：**对价格更敏感的一侧获得更低的价格。持卡人（可以用现金替代卡片）比商户（必须接受主流支付方式）更弹性，所以获得补贴。

2. **产生更大的交叉网络外部性（Larger Cross-Group Externality）：**侧 A 的参与极大地提升了侧 B 的价值（但反之不一定）。女性在约会平台和夜店中制造比男性大得多的交叉外部性——她们的存在本身大幅增加了男性参与的意愿——所以获得免费入场或补贴。

3. **更强的“旗舰”（Marquee）或“锚定”（Anchor）价值：**吸引一个高可见度的用户带来的系统效应远超普通用户。购物中心以亏损价向百货公司（主力店/Anchor Tenants）出租空间——因为主力店吸引的人流使所有其他租户受益。

平台	补贴侧（低价/免费/负价）	为何补贴？	利润中心（高价/收费）
Google 搜索	用户（免费搜索、邮箱、地图）	用户价格弹性极高，多归属容易；产生大量广告展示机会	广告商（按点击付费）
支付卡	持卡人（免年费，回馈/积分）	持卡人价格弹性高，可以用现金替代；每多一个持卡人产生更多商户交易	商户（2–3% 刷卡费率）
购物中心	主力店/百货公司（低租金/平方英尺）	主力店吸引的客流是所有小商户共享的公共品	小型商户（高租金/平方英尺）
视频游戏机	玩家（硬件以低于成本价销售）	玩家价格弹性高；每多一台主机扩大开发者的可寻址市场	游戏开发者（20–30% 的平台版税）
网约车	乘客（早期大幅补贴车费）	乘客价格弹性高，多归属容易；有乘客才有司机愿意上线	司机（平台从每单中抽佣 20–30%）
新闻报纸	读者（订阅价低于印刷成本）	读者价格弹性高，可换报；读者规模决定广告价值	广告商（按版面/展示付费）

表 13.2 揭示了一个反复出现的模式：补贴侧总是**价格弹性更高**或**产生更大的交叉外部性**的一侧。这并非偶然——它是公式 (13.5) 中 $a^B/\eta^B = a^S/\eta^S$ 的直接体现。

13.2.4 为什么不是永远最大化倾斜？

Bolt 和 Tieman (2008, *IJIO*) 证明，在常数弹性需求下，最优策略是“最大倾斜定价”（Heavily Skewed Pricing）——将最有弹性一侧的价格推到其最低可行水平（零，甚至负值），而向缺乏弹性的一侧收取所有加成。但现实中，倾斜程度受三个约束限制：

1. **需求函数的曲率（Schmalensee 2011）：**如果补贴侧市场已饱和（进一步降价的参与弹性很低），继续补贴是浪费。倾斜程度取决于需求函数的**形状**（不仅是弹性水平），这超越了常数弹性假设。

2. **负价格的行政管理难度：**虽然负价格存在（Uber 司机保证时薪、支付宝红包、信用卡开卡奖励），但其管理成本高于简单设定正价格。

3. **套利风险**：如果对一侧的补贴过于慷慨（如无限刷卡回馈），用户可能通过制造虚假交易来套取补贴，消解了补贴创造网络效应的目的。

13.2.5　Armstrong (2006) 会员费模型的补充视角

Rochet-Tirole (2003) 的分析基于**使用外部性**和按次收费。Armstrong (2006, *RAND Journal of Economics*) 从**成员外部性**和固定会员费的角度提供了互补的分析框架。

在 Armstrong 的垄断平台模型中，平台向每侧收取固定会员费 p_i。侧 i 的效用取决于另一侧的规模：

$$u_i = \alpha_i n_j - p_i \tag{13.8}$$

其中 α_i 是交叉组外部性参数——$\alpha_1 > 0$ 意味着侧 1 的代理人从侧 2 的扩大中获益。

在 Hotelling 需求设定下（消费者在单位区间上均匀分布，交通成本 t_i），垄断平台的最优定价为：

$$\boxed{p_i = f_i + t_i - \alpha_j} \tag{13.9}$$

其中 f_i 是服务侧 i 代理人的边际成本，t_i 是 "差异化" 参数（代表平台的市场势力），α_j 是**另一侧**施加于侧 i 的交叉外部性。

解读：与标准 Hotelling 垄断价格 $p = f + t$ 相比，双边平台的价格减少了一个 α_j 项——平台内部化了侧 i 代理人为另一侧创造的外部性。如果 $\alpha_j > t_i + f_i$，最优价格低于边际成本——甚至变为负数。

两大模型的互补性：RT2003 回答 "每次交易收多少钱"（使用外部性），Armstrong 回答 "加入平台收多少钱"（成员外部性）。现实中两种外部性共存——操作系统既有成员外部性（开发者关心安装基数大小），也有使用外部性（用户从使用具体应用中获益）。Rochet-Tirole (2006) 的统一模型通过**两部收费制**（two-part tariff：会员费 + 每笔交易费）整合了两者。

13.3　竞争性瓶颈：单归属 vs. 多归属

13.3.1　Armstrong (2006) 的三个模型

Armstrong (2006) 构建了三个层层递进的模型，揭示了单归属（Single-Homing）与多归属（Multi-Homing）如何决定平台竞争的强度和方向。这是双边市场理论中**最具实证相关性**的单一框架。

模型 1：垄断平台，两侧均单归属。

这是前一节已讨论的垄断平台模型。定价为 $p_i = f_i + t_i - \alpha_j$。垄断者内部化所有交叉外部性——因为只有一个平台，不存在 "用户在平台之间选择" 的问题。

模型 2：双寡头竞争平台，两侧均单归属。

两个平台位于 Hotelling 线的两端。两侧代理人都**只加入一个平台**（单归属）。侧 i 的代理人在平台 $k = A, B$ 上的效用为：

$$u_i^k = \alpha_i n_j^k - p_i^k - t_i|x - \ell_k| \tag{13.10}$$

经过标准的 Hotelling 推导，市场占有率满足：

$$n_i^A = \frac{1}{2} + \frac{\alpha_i(n_j^A - n_j^B) - (p_i^A - p_i^B)}{2t_i} \tag{13.11}$$

对称均衡下的定价为：

$$\boxed{p_i = f_i + t_i - \alpha_j} \tag{13.12}$$

这个定价公式与垄断平台**完全相同**（在 Hotelling 设定下），但经济机制截然不同：

- **垄断下**：$-\alpha_j$ 反映垄断者对交叉外部性的**内部化**
- **竞争下**：$-\alpha_j$ 反映平台之间为吸引能产生外部性的代理人而进行的**激烈竞争**——平台知道，多吸引一个侧 i 代理人，就能通过交叉外部性 α_j 吸引更多的侧 j 代理人，从而扩大整个网络

当 α_j 足够大时，p_i 可能低于边际成本 f_i——即使有两个竞争平台，低于成本定价仍是均衡行为，而非掠夺。

模型 3：竞争性瓶颈（Competitive Bottleneck）——一侧单归属，另一侧多归属。

这是 Armstrong 最具原创性和实证相关性的贡献。设定如下：

- **侧 1（如消费者/读者/用户）——单归属**：代理人只加入一个平台。在 Hotelling 框架中比较两个平台，交通成本参数为 t_1
- **侧 2（如广告商/开发者/卖家）——多归属**：代理人加入**所有**给予非负效用的平台。他们不比较平台——如果平台 A 提供的效用 ≥ 0，就加入 A；如果平台 B 提供的效用 ≥ 0，也加入 B

侧 2 代理人加入平台 k 的条件是：$u_2^k = \alpha_2 n_1^k - p_2^k \geq 0$。

对称均衡下的定价（同等边际成本 f_1, f_2，侧 1 的 Hotelling 交通成本 t_1）：

$$\boxed{p_1 = f_1 + t_1 - \alpha_2} \tag{13.13}$$

$$\boxed{p_2 = f_2 + \frac{\alpha_1}{2}} \tag{13.14}$$

竞争性瓶颈的核心机制

侧 1（单归属侧）：$p_1 = f_1 + t_1 - \alpha_2$。

- 平台为争夺**排他性参与**的侧 1 而激烈竞争
- 竞争推动 p_1 到低于标准 Hotelling 价格 $f_1 + t_1$ 的水平——降低幅度为 α_2
- 如果 $\alpha_2 > f_1 + t_1$，$p_1 < 0$——平台**付钱**给侧 1 代理人来加入

侧 2（多归属侧）：$p_2 = f_2 + \alpha_1/2$。

- 侧 2 代理人不比较平台——每个平台都是他们访问其侧 1 用户的**唯一通道**
- 平台对侧 2 行使**垄断守门人权力**（monopoly gatekeeper power）
- 在对称均衡中，侧 2 代理人的**净剩余为零**（$u_2 = 0$）——平台提取了侧 2 从侧 1 中获得的所有价值

为什么 α_1 不进入 p_1 的公式？因为侧 2 多归属，所以在对称均衡中 $n_2^A = n_2^B$（两侧 2 数量相同）。侧 1 在选择平台时，关心的 n_2^k 在两个平台上相等——因此 $\alpha_1 n_2^k$ 这一项在两个平台之间被**抵消**。侧 1 不因 α_1 而偏好任一平台，所以 α_1 不改变侧 1 的竞争。但 α_1 确实影响 p_2——平台提取了侧 2 从访问侧 1 中获得的价值（$\alpha_1 \cdot n_1^k = \alpha_1/2$）。

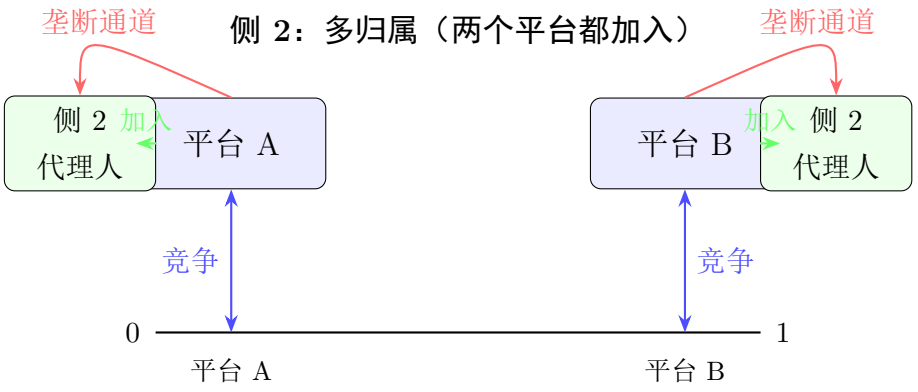


图 13.3 以图形方式展示了竞争性瓶颈的博弈结构。单归属侧（下方）在 Hotelling 线上比较平台——平台为此激烈竞争。多归属侧（上方）视每个平台为独立的通道——平台对其各自拥有的单归属用户群行使垄断权。

13.3.2 实证匹配：App Store 与新闻媒体

App Store（iOS/Android）。智能手机用户通常是单归属的——大多数人只使用一部主力手机。应用开发者是多归属的——几乎所有主流应用都同时为 iOS 和 Android 开发。

竞争性瓶颈模型的预测：

- 用户侧应被补贴——低价/免费操作系统、接近成本的硬件
- 开发者侧应被征税——因为平台是接触用户的垄断通道

数据匹配：iOS 和 Android 以零价格提供给用户（设备制造商获得授权）。Apple 向开发者收取 15–30% 的 App Store 佣金和每年 $99 的开发者计划费用。Google Play 同样收取 15–30% 的佣金。这些佣金远高于支付处理的技术成本（约 1–2%）——溢价反映了平台作为接触用户的垄断通道的权力。竞争性瓶颈模型的预测与观察高度一致。

新闻媒体（报纸/电视）。报纸读者是单归属的——一次读一份报纸。广告商是多归属的——在所有触及目标读者的报纸上投放广告。预测：订阅价格低于印刷/发行成本（读者获得补贴），广告费率较高（广告商被征税）。免费电视（over-the-air TV）是极端情况——观众价格为零，全部收入来自广告。

实证匹配：报纸的订阅价格通常只覆盖印刷和分发成本的一小部分（有时不到 30%）。《纽约时报》的数字订阅收入近年增长，但其大部分收入仍来自广告。这与竞争性瓶颈的预测一致——但也是部分多归属的一个例子（许多读者同时订阅多份出版物），说明现实偏离了纯单归属-多归属的二分法。

13.3.3 部分多归属与扩展

Armstrong (2006) 的严格假设——完全单归属 vs. 完全多归属——是模型化的方便简化。在现实中，**部分多归属（Partial Multi-Homing）**无处不在：很多消费者同时使用 Uber 和 Lyft，许多商户同时接受 Visa 和 Mastercard（但不一定接受 American Express），一些读者同时订阅两到三份报纸。

Athey, Calvano 和 Gans (2018) 以及 Anderson, Foros 和 Kind (2019) 放松了严格假设，显示竞争性瓶颈逻辑以**减弱的形式**持续：

- 多归属**更多**的一侧支付更高的价格（“相对排他性” 原则）
- 如果两侧都有相当程度的多归属，竞争在两个方向都更激烈——总价格水平下降
- 驱动力始终是**相对排他性**——哪一侧更 “锁定” 在单归属上，哪一侧就是平台争夺的奖品

市场结构	双侧定价公式	关键直觉
垄断平台	$p_i = f_i + t_i - \alpha_j$	垄断者内部化交叉外部性——通过降低 p_i 来扩大网络
竞争，两侧单归属	$p_i = f_i + t_i - \alpha_j$（同垄断形式）	竞争机制：交叉外部性**强化竞争**——吸引一侧等于吸引另一侧
竞争性瓶颈（一侧单归属，另一侧多归属）	$p_1 = f_1 + t_1 - \alpha_2$（单归属侧）$p_2 = f_2 + \alpha_1/2$（多归属侧）	单归属侧：激烈补贴竞争多归属侧：垄断守门人定价，零剩余

13.4　鸡-蛋问题与临界规模

13.4.1　协调失灵的形式化

双边平台面临的一个根本性挑战是：**平台在两侧的参与均为零时，对任何一侧都没有价值**。买家不会加入没有卖家的平台，卖家不会加入没有买家的平台。这产生了一个**协调失灵（Coordination Failure）**：存在多个纳什均衡——一个“双方都不加入”的低参与均衡和一个“双方都加入”的高参与均衡——但市场可能被困在前者。

这一挑战与网络效应文献中的“临界规模”（Critical Mass）概念密切相关（Rohlfs 1974; Katz & Shapiro 1985）。在双边市场语境中，临界规模是平台参与度的“倾覆点”——高于此阈值，正向反馈（更多侧 A 吸引更多侧 B，更多侧 B 又吸引更多侧 A）将平台推向高参与均衡；低于此阈值，平台瓦解到零。

Caillaud 和 Jullien (2003, *RAND Journal of Economics*) 提供了鸡-蛋问题最权威的形式化分析。

13.4.2　Caillaud-Jullien (2003) 模型的核心结构

设定：

- 两侧 $i = 1, 2$，每侧各有一单位连续统代理人
- 匹配技术 $\lambda \leq 1$：如果两侧注册到同一个中介，匹配概率为 λ（中介是创造匹配效率的技术）
- 两个竞争性匹配中介：在位者（I, Incumbent）和进入者（E, Entrant）

- 价格工具：注册费 p_i^k（固定费用）+ 交易费 t^k（按已实现匹配价值的比例收取）

- 时序：(1) 中介同时设定价格；(2) 代理人选择注册到哪个中介；(3) 匹配实现，收益发生

- 总匹配价值标准化为 1，按外生议价能力分配：$u_1 + u_2 = 1$，其中 $u_2 \geq 1/2 \geq u_1$

代理人的效用函数：$U_i^k = n_j^k \cdot \lambda \cdot u_i \cdot (1 - t^k) - p_i^k$。中介 k 的利润：$\pi^k = \sum_i n_i^k (p_i^k - c_i) + \lambda \cdot n_1^k \cdot n_2^k \cdot t^k$。

核心结果 1——排他性注册下的主导企业均衡：当代理人只能注册一个中介（单归属）时，只有**主导企业均衡（Dominant-Firm Equilibrium）**存活——市场完全倾覆于一个平台。主导企业设定交易费 $t^I = 1$（提取所有匹配盈余），并同时补贴两侧注册。总价格 $p_1^I + p_2^I = c - \lambda$——当匹配效率 λ 大于注册总成本 $c = c_1 + c_2$ 时，总价格为**负数**。主导企业**在均衡中亏损**，但这是进入威胁施加的约束——即使进入在均衡中不实际发生，在位者也必须以成本价（甚至亏损）运营以防止进入。

核心结果 2——多归属下的共存均衡：当代理人允许多归属时，均衡格局截然改变。主导企业的利润下降（可能到零），新的正利润均衡出现——两个中介均活跃并赚取正利润。中介有**激励允许非排他性服务**——因为当代理人多归属时，中介仍可以从他们身上提取价值（即使他们也使用竞争者）。

13.4.3 分而治之策略的数学逻辑

分而治之（Divide and Conquer）是 Caillaud-Jullien 模型中最具操作性的战略洞见。进入者通过两个步骤克服鸡-蛋问题：

第一步——分化（Divide）：进入者将侧 1 的注册费设定得足够低（高度负值），使得即使侧 1 代理人预期侧 2 的参与为零，仅靠直接补贴也值得注册。形式上：

$$p_1^E < p_1^I - \lambda \cdot u_1 \cdot (1 - t^I) \tag{13.15}$$

这个条件确保：即使 $n_2^E = 0$（侧 2 完全不参与进入者），侧 1 代理人仍偏好进入者（因为补贴足够慷慨）。所有侧 1 代理人注册进入者（$n_1^E = 1$），完成“分化”。

第二步——征服（Conquer）：现在侧 1 已被锁定，网络外部性从**逆风变为顺风**。侧 2 代理人发现进入者比在位者更有吸引力——因为侧 1 完全在进入者这里。进入者设定最大交易费 $t^E = 1$ 来提取侧 2 的所有盈余，回收对侧 1 的补贴成本。进入者盈利的条件是回收的盈余超过补贴支出。

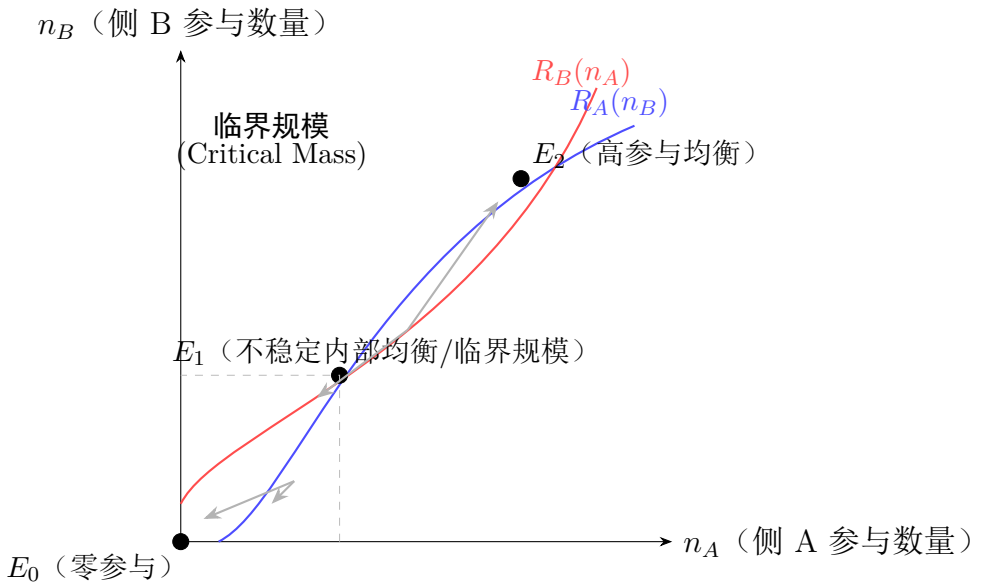
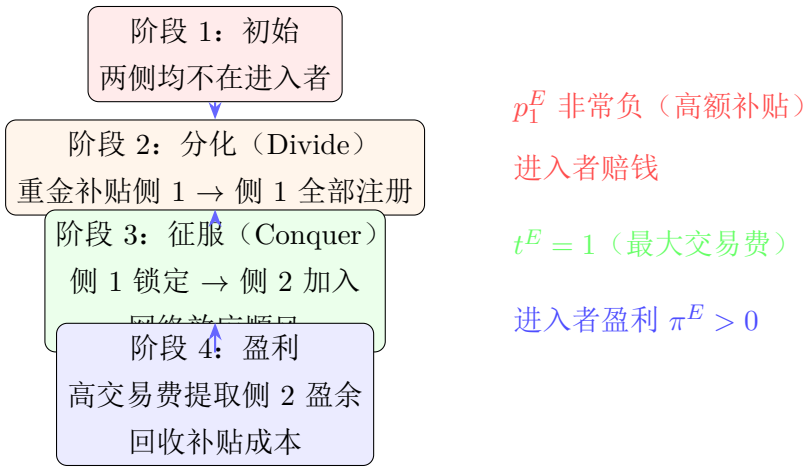


图 13.4 展示了鸡-蛋问题产生的多重均衡。$R_A(n_B)$ 和 $R_B(n_A)$ 是两侧的反应曲线（给定对另一侧规模的预期，最优参与决策）。三条曲线的三个交点对应三个均衡：E_0（零参与——所有预期自我实现为无）、E_1（不稳定内部均衡——倾覆点/临界规模）、E_2（高参与均衡——正向反馈将平台推向成功）。箭头显示动态收敛方向：低于临界规模时，市场被拉向 E_0；高于临界规模时，市场被拉向 E_2。



13.4.4　解决鸡-蛋问题的策略工具箱

理论提供了形式化的逻辑——实践中的平台发展了一套丰富的策略来解决鸡-蛋问题。表 13.4 总结六大策略。

策略	描述	机制	代表案例
顺序进入（Zig-Zag）	先重金补贴一侧，建基础后再货币化另一侧	打破同时参与要求——一侧先到位，另一侧自然跟随	Uber 进入新城市：先保证司机 $30/小时收入，预置车源
旗舰买家/锚定租户（Marquee/Anchor）	招募少量高可见度参与者吸引另一侧	利用个体参与者的不对称外部性——一个明星参	购物中心亏损招募百货公司；游戏机的独占首发大作

13.4.5　Jullien (2011) 的 Stackelberg 扩展

Jullien (2011, *AEJ: Microeconomics*) 将 Caillaud-Jullien 的静态同时行动博弈扩展到 Stackelberg（领导者-跟随者）框架，揭示了几个额外的战略洞见：

1. **跟随者可使用分而治之克服在位优势**：第一个行动者的优势不是绝对的——如果跟随者能足够激进地补贴一侧，它可以 "买出" 一个用户群，然后吸引另一侧

2. **领先者可能策略性地不向一侧销售**：通过让一侧 "无争议"，领先者减少了跟随者激进交叉补贴的激励——这是一个有趣的 "为了软化竞争而放弃市场" 的逻辑

3. **平台可能偏好兼容性（互操作性）**：虽然兼容性减少了差异化，但它也降低了交叉补贴竞争的强度——从平台的角度，可能是值得的

4. **消费者通过协调到 "焦点" 解决多重均衡**：在对称平台之间存在多重均衡时，消费者使用共享信念（common beliefs）来选择平台——这为在位者或知名品牌提供了实质性优势

13.5　价格歧视与平台设计

【进阶】

13.5.1　Rochet-Tirole (2006) 的统一框架

在 RT2003（使用外部性 + 按次收费）和 Armstrong (2006)（成员外部性 + 固定会员费）之间，存在一个自然的问题：当两类外部性都重要时，平台应该使用什么定价工具？Rochet 和 Tirole (2006) 的统一模型回答了这个问题。

平台使用**两部收费制（Two-Part Tariff）**：向每侧 i 收取会员费（固定费用）A^i 和每笔交易费（使用费用）a^i。

侧 i 用户的净效用：

$$U^i = \underbrace{(b^i - a^i) \cdot N^j}_{\text{使用效用}} + \underbrace{B^i(N^j)}_{\text{成员效用}} - \underbrace{A^i}_{\text{会员费}} \tag{13.16}$$

其中 b^i 是每笔交易的收益，$B^i(N^j)$ 是从另一侧规模 N^j 中获得的成员效用（$B^{i\prime} > 0$）。

交易量：$V = N^B \cdot N^S \cdot d(a^B + a^S)$，其中 $d(\cdot)$ 是每对已连接用户之间的交易需求函数。

关键结论：

- 当**成员外部性**占主导且使用外部性较小时 $\rightarrow$ 会员费做 "重活"（内部化成员外部性），交易费接近边际成本

- 当**使用外部性**占主导时 → 每笔交易费是主要工具，会员费接近成本
- **两部收费制**给平台一个额外的自由度来同时管理两种外部性——实现比纯会员费或纯交易费更高的利润

13.5.2 平台定价工具全景

平台在实践中使用的定价工具远比教科书中的“会员费 vs. 交易费”二分法丰富。表 13.5 总结主要工具及其经济逻辑。

工具	描述	适用场景	实例
分层会员制（Tiered Membership）	不同固定费用对应不同访问级别	用户异质性高，成员外部性为主	LinkedIn Free vs. Premium
Freemium	免费基础层 + 付费高级层	用免费层建网络，付费层货币化	Dropbox, Spotify
版税 + 固定费（Royalty + Fixed Fee）	年度费 + 每单位版税	开发者异质性大，使用 + 成员外部性共存	App Store（$99/年 + 15–30%）
交换费（Interchange Fee）	商户银行向持卡人银行支付的每笔交易费	平衡两侧弹性的核心工具	Visa/Mastercard 交换费
上架费 + 成交费（Listing + FVF）	列表费 + 售价百分比	列表费筛选低质卖家，成交费按价值征税	eBay, Amazon Marketplace

13.5.3 Weyl (2010)：绝缘均衡与 Spence 扭曲

Weyl (2010, *AER*) 提供了多边平台定价的最一般理论处理，贡献了两个重要概念。

绝缘均衡（Insulated Equilibrium）：标准双边市场模型中存在循环性——侧 B 的参与取决于侧 A，侧 A 的参与取决于侧 B → 多重均衡和协调失败风险。Weyl 提出，平台可以通过设定随另一侧参与水平调整的**绝缘关税（Insulating Tariffs）**来打破这个循环。形式上，侧 i 的价格是一个函数 $P^i(N^j)$，使每侧的参与决策独立于另一侧的参与水平。

虽然很少有平台明确使用数学上的绝缘关税，但该概念解释了常见的实践：按受众规模向广告商提供的担保定价、随卖家参与度变化的抽成率、以及两部收费制——

这些都在不同程度上解耦了两侧的参与决策。

Spence 扭曲（Spence Distortion）：以 Michael Spence (1975) 的垄断质量选择模型命名，这是一个双边市场特有的新效率扭曲来源：

- 利润最大化的平台只根据**边际用户**（marginal user）的交叉网络外部性来内部化网络效应
- 社会福利取决于**平均（边际内）用户**（inframarginal/average user）的交叉网络外部性
- 当边际用户与平均用户对网络的估值不同时 $\rightarrow$ 平台的定价偏离社会最优
- **总扭曲 = 经典垄断加成扭曲 + Spence 扭曲**

如果边际用户比平均用户更重视网络规模（例如：重度用户比轻度用户更关心应用生态系统的丰富度），平台会**过度补贴**参与（价格低于社会最优）。如果边际用户不太重视网络规模（例如：轻度用户主要使用基本功能），平台会**补贴不足**（价格高于社会最优）。

Tan-Wright (2018) 的修正：Tan 和 Wright 在 AER 评论中指出，Weyl 关于私人价格与社会最优价格对比的公式有误。正确解读下，Weyl 的结果只解释了价格设定的**边际激励**，而不是总扭曲。Spence 扭曲在某些情况下可能被其他定价扭曲所抵消或主导。

13.5.4　与经典 IO 的概念对接

双边市场理论不是经典 IO 的替代品，而是其**一般化**。表 13.6 追溯了关键概念从单边到双边的扩展。

经典 IO 概念	双边市场一般化	关键差异
Lerner 指数 $\frac{p-c}{p}=\frac{1}{\varepsilon}$	机会成本 Lerner：价格依赖 c – 外部性 而非 c	“成本” 被重新定义为包含放弃的另一侧收入
Ramsey 定价	Ramsey 规则横跨用户侧——逆弹性应用于**价格结构**	不仅有最优总价，还有最优分配
Spengler (1950) 双重加价	平台语境中的双重加价——纵向一体化不必然侵蚀平台加成	结果取决于两侧弹性的相对大小
Katz-Shapiro (1985) 网络效应	间接/跨组网络效应——平台提供了内部化外部性的定价理论	定价是网络效应内部化的工具
Hotelling (1929) 差异化	单归属 vs. 多归属下的平台竞争——竞争性瓶颈	多归属改变了竞争的方向和强度
Bulow-Pfleiderer (1983) 转嫁率	跨侧转嫁——一侧的税收部分转嫁到**另一侧**的价格上	税负归宿分析必须包括跨侧效应

13.6 市场倾覆与平台包围

【进阶】

13.6.1 市场倾覆的条件

当正向反馈（间接网络效应的正反馈循环）导致所有代理人收敛到单一平台，将竞争者驱出市场时，市场发生了**倾覆（Tipping）**——结果就是**赢家通吃（Winner-Take-All）**。

Rysman (2009, *Journal of Economic Perspectives*) 系统性地总结了市场倾覆需要同时满足三个条件：

1. **强间接网络效应：**每侧非常在乎另一侧的规模。强烈的交叉网络外部性产生自我强化的反馈——更多的侧 A 吸引更多的侧 B，更多的侧 B 又吸引更多的侧 A——这个循环没有自然停止点，直到市场完全倾覆

2. **低横向产品差异化：**平台是近似替代品——用户选择基于网络规模而非平台特有特征。当平台对用户来说本质上是同质的，更大的网络总是赢

3. **至少一侧单归属**：单归属是倾覆的 “锁定机制”——它阻止多归属作为释放阀（safety valve）。如果两侧都能多归属，多个平台可以共享同一个市场，各自拥有两边用户的子集

这三个条件共同解释了为什么一些平台市场倾覆了（Blu-ray vs. HD DVD），而另一些则保持着持久的多平台竞争（iOS vs. Android）。

历史案例——Blu-ray vs. HD DVD (2006–2008)：强网络效应（消费者需要内容，内容商需要消费者）+ 低差异化（两种格式功能相似）+ 单归属（消费者通常只买一种播放器，内容商通常只支持一种格式）。2008 年 1 月华纳兄弟从 HD DVD 倒戈到 Blu-ray 后，市场在几周内彻底倾覆——HD DVD 旋即终止。

非倾覆案例——iOS vs. Android：尽管存在强网络效应（应用开发者需要用户，用户需要应用），两个平台已共存超过 15 年。原因：充分的横向差异化（不同设计理念、不同价格段位、品牌忠诚度）、开发者多归属（大多数应用同时开发两个版本）、以及高转换成本（应用库、数据、配件生态锁定）。中国外卖市场（美团 vs. 饿了么）的竞争性共存同样归因于餐厅多归属和用户地理偏好。

13.6.2　多平台共存的条件

当上述三个倾覆条件中的至少一个显著减弱时，多平台共存（Multi-Platform Coexistence）成为可能的均衡结果。

条件	运作机制
充分的横向差异化	异质偏好——部分用户偏好 “苹果生态”（封闭、安全、高端），其他偏好 “安卓生态”（开放、多样、灵活）
两侧均多归属	平台不争夺排他性接入——市场份额可以稳定在不完全倾覆的某个水平
弱网络效应	交叉外部性相对于独立使用价值较小——网络规模不是决定性因素
纵向差异化 + 细分定位	LinkedIn（职业社交）vs. Facebook（生活社交）——不同使用场景限制了直接的跨平台竞争
非对称成本互相补偿	一个平台在侧 A 成本高但在侧 B 成本低——非对称性抵消了倾覆动力
局部兼容性	Casadesus-Masanell & Ruiz-Aliseda (2008)：局部兼容性可维持否则会倾覆的共享均衡

13.6.3　平台包围（Platform Envelopment）

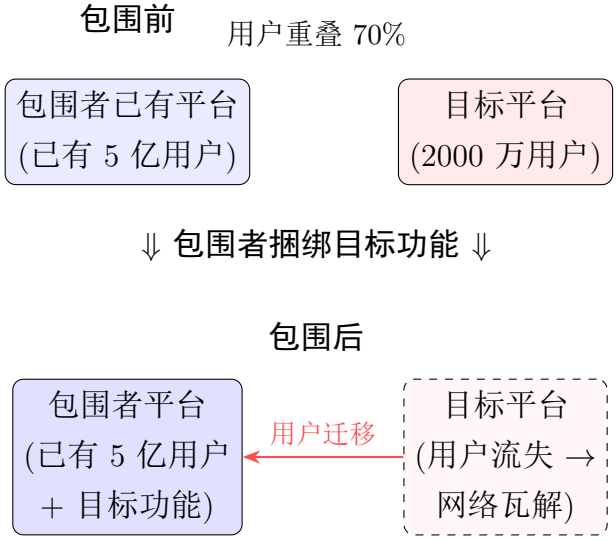
Eisenmann, Parker 和 Van Alstyne (2011, *Strategic Management Journal*) 识别出一种独特的平台进入策略：**平台包围（Platform Envelopment）**——一个市场上

的已有平台通过**将新功能与其已有平台捆绑**进入相邻平台市场，利用共享的用户关系（shared user relationships）。

包围的核心机制：

1. 包围者已经拥有大量已安装用户群（large installed base）
2. 目标平台（被包围的现有者）的用户与包围者的用户群高度重叠
3. 通过将目标功能捆绑到已有平台中，包围者使独立的现有者变得多余（redundant）
4. 曾保护现有者的网络效应现在**逆向运作**——用户向捆绑产品迁移，目标平台的网络缩小，触发死亡螺旋（death spiral）

包围者不一定有更好的产品——它通过封锁现有者的用户接入（foreclosing access to users）来获胜。



微信——终极包围者：微信（WeChat）的演化轨迹是平台包围的教科书案例。

- 2011 年：作为即时通讯应用起步（包围短信/彩信）
- 2013 年：加入微信支付（包围支付宝——利用已有 3 亿 + 通讯用户）
- 2017 年：推出小程序（包围独立应用——电商、旅行、外卖、叫车、CRM 在一个“超级应用”内）
- 关键杠杆：社交图谱（群聊、朋友圈、红包）——创造独立应用无法复制的病毒式分发渠道

战略启示——面临包围威胁的现有者：(1) 减少用户重叠（使自己的用户群与包围者差异化）；(2) 创造独立于包围者核心功能的使用价值；(3) 与另一个大型平台结盟来抗衡。

13.7　双边市场中的反垄断挑战

这一节是 Ch.12 的直接延伸——展示传统反垄断工具在平台经济中如何失效，以及经济学文献提出的修正方案。表 13.7 提供了全貌概览。

工具	单边市场逻辑	双边市场的失效	经济学修正
SSNIP / HMT	假想垄断者提价 5–10% 有利可图?	零价格问题（$0 \times 5\% = 0$）；反馈循环；忽略鸡-蛋动态	SSNDQ（质量下降测试）；SSNIC（注意力/数据成本增加测试）；自然实验
HHI / 集中度	高集中度 ⇒ 推定反竞争	市场份额无法传达交叉网络效应的竞争重要性	无简单的结构替代——HHI 仍有信息量但分析权重降低
UPP / GUPPI	转移率 × 毛利率 ⇒ 涨价压力	只有一组转移率——双边市场需要四组;忽略再平衡效应	Affeldt et al. (2013) 四转移率框架；Cosnita-Langlais et al. (2021) 再平衡调整
掠夺性定价	价格 < AVC ⇒ 推定掠夺性	一侧低于成本定价在双边市场中是**均衡常态**——不是掠夺的证据	区分外部性驱动的倾斜定价 vs. 策略性排他——无共识检验
并购模拟	改变所有权矩阵，求解新 Nash 均衡	需要额外的交叉侧需求方程——标准单边模拟低估/高估效应	双边需求系统估计 + 两侧同时重新优化
损害量化	But-for 价格 − 实际价格 × 数量	一侧的过高要价可能被另一侧的过低要价（再平衡）所抵消	计算**双侧净损害**——同时检查两侧买家和卖家的消费者剩余变化

13.7.1　市场界定：SSNIP 的三重崩溃

1. **零价格问题（Zero-Price Problem）**：当一侧服务的价格为 0 时（Google 搜索、Facebook、微信），SSNIP 的 5–10% 价格上涨是 $5\% \times 0 = 0$。假想垄断者能否有利可图地将价格从 0 提到 0？问题本身在数学上失去了意义。

2. **反馈循环问题（Feedback Loop Problem）**：在双边市场中，一侧价格上涨不仅影响该侧的需求——它减少了该侧的参与 → 降低平台对另一侧的吸引力 → 减少另一侧的参与 → **反馈回**第一侧，进一步降低其支付意愿。假想垄断者必须**同时优化两侧价格**——标准 SSNIP 的“单侧提价”逻辑在理论上就是错误的。

3. **鸡-蛋问题（Chicken-and-Egg Problem）**：具有强交叉网络效应的市场的竞争约束是动态的——随着平台经历启动、增长和成熟阶段而变化。已在两侧拥有大规模已安装用户群的平台面临的竞争约束，与具有相同服务但零用户群的新进入者面临的竞争约束完全不同。静态的 SSNIP 框架无法捕捉这些动态。

13.7.2 Ohio v. American Express (2018)：标志性判决

Ohio v. American Express Co., 585 U.S. ____ (2018) 也许是双边市场反垄断领域最重要的法院判决。

案情：美国运通（Amex）在其与商户的合同中包含 "反引导条款"（Anti-Steering Provisions）——禁止商户鼓励顾客使用收取更低商户手续费的其他信用卡（如 Visa、Mastercard）。11 个州和 DOJ 起诉，认为这些条款违反了《谢尔曼法》第 1 条——通过限制商户引导顾客使用其他卡的能力，Amex 抑制了卡网络在商户费用上的竞争。

最高法院的判决（5-4 投票，Thomas 大法官执笔多数意见）：

1. 信用卡网络是**交易型平台（Transaction Platform）**——"无法仅向一侧销售而不向另一侧销售"——每笔信用卡交易同时涉及持卡人和商户

2. 对于有 "显著间接网络效应" 的交易型平台，**两侧必须同时纳入相关市场**——市场是 "交易"，不是一个持卡人市场和一个商户市场

3. 仅一侧涨价的证据**本身不能证明反竞争行为**——原告必须证明被挑战的约束：(a) 提高了交易总成本，或 (b) 减少了交易总数量，或 (c) 以其他方式扼杀了**整个双边市场**的竞争

经济逻辑（多数意见方）：(1) RT2003 的核心洞见——只看一侧会系统性地在双边市场中发现 "反竞争行为"，因为最优定价结构几乎总是对缺乏弹性的一侧收取更高价格。(2) 避免 "单侧偏见"——如果法院只看支付侧（商户），会在双边市场中产生结构性过度执法。

经济批评（Breyer 大法官异议方）：(1) 商户服务和持卡人服务是**互补品不是替代品**——不应被放入同一个市场。(2) Amex 的商户费在争议期间持续上涨，而持卡人回馈持平或下降——再平衡机制在事实中未按理论预测运行。(3)"交易型平台" 的边界在哪里？网约车、电商、共享经济——都可能被过度纳入，导致大范围的反垄断免疫。

争议点	多数意见（Thomas）	异议（Breyer）
市场定义	单一 “交易” 市场——两侧必须同时纳入	两个独立但相关联的市场——商户服务 + 持卡人服务
对涨价证据的看法	一侧涨价本身不构成证据——必须看对交易总量的影响	商户费用的稳定上涨本身就说明竞争受损
再平衡机制	理论上，更高的商户费通过更高的持卡人回馈而回到持卡人——系统再平衡	在案件中，该机制未按理论运作——回馈未跟上费用上涨
反引导条款的作用	防止商户在持卡人和商户之间重新分配成本——维持价格结构	阻止 Discover 等竞争者追求低商户费模式——抑制竞争
经济学文献支持	Rochet-Tirole (2003, 2006) ——价格结构是平台竞争的核心维度	文献未说明双边市场应被**法律强制**为单一市场

13.7.3　零价格市场中的市场势力：SSNDQ 与 SSNIC

当价格为零时，市场势力如何被发现和衡量？两个主要替代方案已经出现，虽然其操作精度尚未达到 SSNIP 的水平。

SSNDQ 测试（Small but Significant Non-transitory Decrease in Quality）：将假想垄断者测试中的 “价格上升” 替换为 “质量下降”——**假想垄断者能否有利可图地实施小幅但显著的非临时性质量下降？**

关键案例：

- **奇虎 360 v. 腾讯（中国最高法院，2013）：**一般被视为全球首个应用 SSNDQ 推理的案件。在免费即时通讯（QQ）市场中，法院承认当价格为零时 SSNIP 会给出过度宽泛的市场定义，并使用 SSNDQ 作为定性框架。法院最终裁定腾讯不具有市场支配地位。

- **Google Android（EU，2018）：**欧盟委员会使用 SSNDQ 推理定义了 “可授权的智能移动操作系统” 市场，认定 Google 的 Android 在该市场中具有支配地位。欧盟普通法院后指出，精确的定量 SSNDQ 阈值不是先决条件——SSNDQ 作为定性框架运作。

- **Google Search 垄断案（US DOJ，2023–2024）：**Mehta 法官引用 Google 在增加广告加载量、降低有机搜索结果质量而不流失用户的能力，作为垄断力的证据——这实际上是一个 SSNDQ 的论证。

SSNIC 测试（Small but Significant Non-transitory Increase in Cost）：将 “价格上升” 替换为**非货币成本上升**——注意力成本（广告时间）、数据/隐私成本、转换成本。

Newman (2022, *University of Chicago Law Review*) 提出，零价格平台处于 “注意力市场”（Attention Markets）——用户以注意力和数据交换服务。SSNIC 的优势是注意力成本比模糊的 “质量” 更容易量化（广告负载增加 10% → 注意力损失约 10%）。但挑战在于：用户可能察觉不到逐渐增加的注意力成本（“温水煮青蛙” 效应），且数据的价值是模糊的（同时改善个性化和损害隐私）。

维度	SSNIP（传统）	SSNDQ（质量下降）	SSNIC（成本增加）
操作变量	货币价格	服务/产品品质	非货币成本（注意力、数据、时间）
基准阈值	5–10% 价格上涨	无公认的% 阈值	无公认的% 阈值
接受度	全球标准（40+ 年）	被多个司法管辖区接受作为定性框架	新兴概念，有限采纳
主要优势	精确量化、法律先例丰富	可应用于零价格市场	比 “质量” 更接近可测量性
主要局限	在零价格市场无意义	质量是多维的；无自然基线	用户对注意力成本不敏感；数据价值模糊
代表案例	数千个案件	奇虎 v. 腾讯 (2013)；Google Android (2018)；Google Search (2024)	学术讨论阶段 (Newman 2022)

13.7.4 平台并购分析：再平衡效应

Ch.12 的 UPP/GUPPI 框架在双边市场中产生了系统性的偏误。根本原因是：单边市场中只有**一组**转移率（当产品 1 涨价时，多少销量转移到产品 2?），而双边市场中有**四组**（Affeldt, Filistrucchi & Klein 2013）：

1. 侧 A 内部转移（侧 A 代理人从平台 1 转向平台 2）
2. 侧 B 内部转移（侧 B 代理人从平台 1 转向平台 2）
3. 交叉侧转移——平台 1 失去侧 A 代理人 → 平台 1 对侧 B 的吸引力降低 → 侧 B 代理人流失
4. 反向交叉侧转移——侧 B 损失反馈到侧 A

Cosnita-Langlais, Johansen 和 Sørgard (2021)（*IJIO*）提供了关键的理论修正。当并购后的平台考虑提高侧 A 的价格时，三个效应同时运作：

- **自身侧需求效应**（−）：侧 A 涨价 → 侧 A 数量下降
- **交叉侧需求效应**（−）：侧 A 数量下降 → 侧 B 数量下降（减少网络价值）

- **再平衡效应（Rebalancing Effect）**（+）：合并后的企业有激励**降低**侧 B 价格 → 吸引更多侧 B 代理人 → 提高侧 A 的留存和支付意愿。该效应**抵消了部分**上行定价压力

核心发现：在正交叉网络外部性下，忽略再平衡效应的简单 UPP 会**高估**并购的竞争损害——可能导致假阳性（Type I Error：将有利竞争的并购标记为反竞争）。这与单边市场的 UPP（通常较为保守，更可能产生假阴性而非假阳性）形成了对比。

关键并购案例：

- **Facebook/Instagram (2012)**：以约 $10 亿收购。FTC 当时以 5-0 投票批准。2020 年起诉拆分，援引杀手并购理论（Cunningham-Ederer-Ma 2021）——Zuckerberg 内部邮件表示 “buy rather than compete” 和 “neutralize” 的意图。然而，**2025 年 11 月 DC 区法院判决 FTC 败诉**——法院认定 FTC 未能证明 Meta 在当前正确定义的市场中持有垄断力（TikTok 的崛起从根本上改变了竞争格局）。核心教训：法庭对 “多年前的合并造成当前垄断力” 的举证要求极高。

- **Google/DoubleClick (2007)**：以 $31 亿收购。FTC 批准了这一主要互补性的合并。委员 Harbour 在异议中警告 “巨大额外网络效应” 的集中。2025 年法院认定 Google 非法垄断广告服务器和广告交易市场——将 DoubleClick 收购追溯为垄断力的起点。

- **滴滴/优步中国 (2016)**：优步中国并入滴滴出行，创造了一个占据中国网约车市场约 87–90% 份额的实体。双方均未申报反垄断审查。SAMR 仍据报道在调查中。该案例直接影响了中国 2022 年《反垄断法》修订中关于低于阈值收购审查的条款。

13.7.5　掠夺性定价：何时低于成本是正常的？

在单边市场中，低于成本定价是掠夺的反垄断警示信号。Areeda-Turner 测试将低于平均可变成本（AVC）的定价推定性地视为掠夺性。但双边市场从根本上颠覆了这一逻辑。

双边市场的最优定价公式显示：$p_i = c_i - \alpha_j \cdot n_j/n_i' +$ 加成。如果交叉外部性 α_j 足够大，$p_i < c_i$ 不是掠夺的征兆——它是利润最大化一阶条件的**直接结果**。对一侧的低于成本定价在**完全竞争**中也持续存在。

关键分析问题：如何区分 (a) 双边定价公式的自然结果（外部性驱动、效率增强）与 (b) 排斥同等效率竞争对手的策略性尝试（掠夺）？

- Behringer (2015) 显示，一侧的低于成本定价**可以**排斥竞争对手——但在不对称市场中，这种排斥可能是**社会最优的**（禁止低于边际成本定价会通过阻止有效排斥而恶化过度进入问题）

- **Google Shopping（EU, 2017–2024）**：表面上涉及“免费搜索 + 广告收费”的双边定价结构，但本质上不是掠夺性定价案——而是**自我优待（Self-Preferencing）**案。Google 利用其在免费搜索市场中的主导地位，通过自我优待杠杆化到付费购物比较服务市场。该案说明了为什么传统价格-成本测试无法机械地应用于双边平台——滥用可能以完全非价格的形式出现

- **中国**：SAMR《互联网平台定价行为指南》草案明确处理掠夺性定价问题，同时澄清“促进创新和提高消费者福利的长期免费服务”不会被视为违规——反映了对双边市场最优定价结构的认识

13.7.6 杀手并购（Killer Acquisitions）

Cunningham, Ederer 和 Ma (2021, *Journal of Political Economy*) 提供了杀手并购的权威实证框架。

定义：在位企业收购创新目标公司，**仅为终止目标的创新项目**和先发制人地防止未来竞争。

核心发现（制药业）：

- 存在产品重叠时，被收购药物项目继续推进的可能性低 23.4%

- 估计 5.3–7.4% 的收购为杀手并购

- 该效应在收购方具有显著市场势力且面临远期专利到期时更强

- 杀手并购不成比例地集中在**刚好低于反垄断审查阈值**处

双边性如何放大杀手并购激励：

1. **生态系统威胁（Ecosystem Threat）**：当大平台（如 Facebook）面临萌芽进入者（如 Instagram），进入者不仅威胁一个产品——它威胁整个平台生态系统。如果 Instagram 吸走 Facebook 用户，Facebook 用户群萎缩 → 广告侧的价值降低 → 广告收入降低 → 平台投资能力下降——一个全面下降螺旋

2. **赢家通吃动态（Winner-Take-Most Dynamics）**：有强网络效应的双边市场趋于集中。达到临界规模的萌芽竞争者能使整个市场倾覆。在位者消除威胁的激励也相应更大——杀死的不只是一个产品，而是一个潜在的平台颠覆者

3. **估值/营业额缺口（Valuation/Turnover Gap）**：萌芽平台通常有近乎零的营收（低于并购通知阈值）但极高估值。Instagram：约 3000 万用户、零营收、\$10 亿（2012）。WhatsApp：约 4.5 亿用户、极少营收、\$190 亿（2014）。这些交易逃脱了强制性的 HSR 合并前通知，因为营收太低

中国语境：低于阈值审查

中国 2022 年《反垄断法》修正案新增了一项至关重要的权力：SAMR 可以调查**低于申报阈值**的交易，如果有证据表明它们已经或可能具有反竞争效果。这直接针对杀手并购问题——对标 Cunningham-Ederer-Ma (2021) 的核心政策建议（“将审查重点从基于规模的阈值转移到基于重叠和竞争潜力的实质性评估”）。违反申报要求（gun-jumping）的最高罚款从 50 万元提高到 500 万元，对负有责任的个人的罚款最高可达 100 万元。

13.7.7　全球规制格局比较

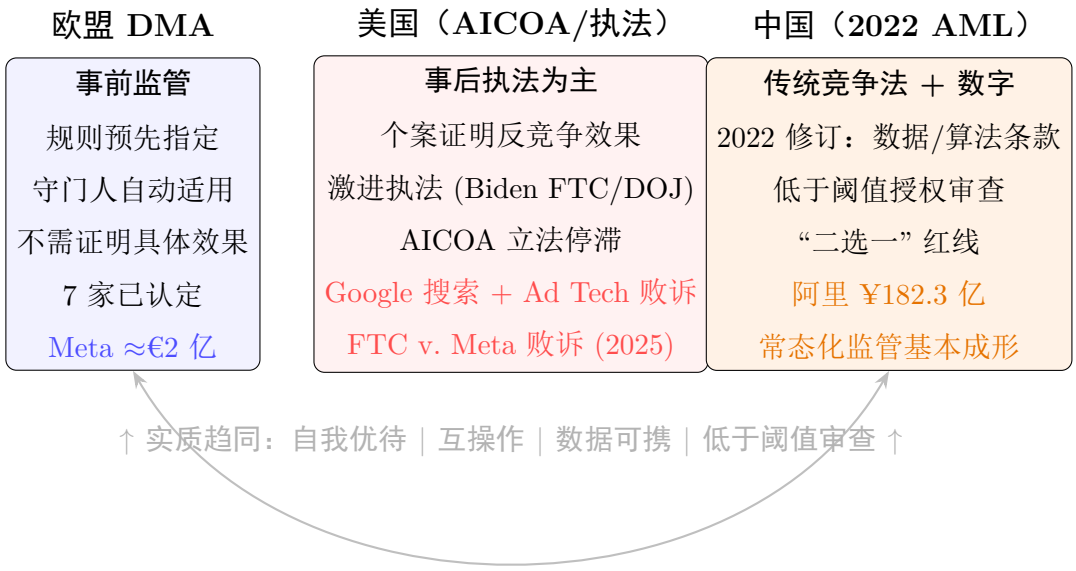
双边平台经济的规制已经成为全球竞争政策的中心议题。尽管制度形式各异——欧盟的事前 DMA 监管、美国停滞的立法和激进的执法、中国修正后的传统竞争法——实质层面存在显著的趋同。

维度	欧盟 DMA	美国（AICOA/执法）	中国（2022 AML + 指南）
规制模式	事前监管（ex-ante）	事后执法为主（立法受阻）	传统竞争法 + 数字条款
守门人认定	7 家：Alphabet, Amazon, Apple, ByteDance, Meta, Microsoft, Booking	未定（AICOA 停滞）	市场份额 + 交易量/流量控制力 + 其他因素
自我优待	明确禁止（Art. 6(5) DMA）	AICOA 将禁止	SAMR 实施规则：独立滥用形式
互操作性	强制（Art. 6(7) DMA）	AICOA 部分要求	视情况而定（《个人信息保护法》框架）
数据可携	实时、持续访问	部分提案	《个人信息保护法》+ 部门指南
低于阈值收购	无特定规定	2023 并购指南纳入审查	**2022 AML 明确授权审查**
最高罚款	全球营业额 10%（系统性：20%）	AICOA：美国营收 15%	集团营业额 10% + 个人最高 100 万元
标志性案件	Meta：首笔 DMA 罚款 €2 亿（2025.4，“同意或付费”模型）	DOJ v. Google（搜索 + 广告技术）两案均裁定非法垄断	阿里 ¥182.3 亿（2021）；美团 ¥34.4 亿（2021）

实质趋同（Substantive Convergence）

尽管制度形式不同，三大司法管辖区在以下关键领域出现了显著的实质趋同：

1. **自我优待禁令**：DMA Art. 6(5)、AICOA（拟议）、中国 SAMR 实施规则——均禁止主导平台利用跨市场杠杆优待自身服务
2. **互操作性与数据可携**：DMA Arts. 6(7), 6(9)；中国在《个人信息保护法》框架下推行
3. **低于阈值收购审查**：2022 年中国 AML 明确授权；美国 2023 并购指南纳入；欧盟通过 DMA 守门人并购报告义务间接解决
4. **承认数据/算法/平台规则可作为垄断工具**：中国 2022 AML 第 9 条（数字经济条款）——“经营者不得利用数据和算法、技术、资本优势以及平台规则等从事本法禁止的垄断行为”



13.8 平台治理与 B2B 平台

【进阶】【前沿】

13.8.1 B2B 平台的独特性

前述分析主要聚焦于 B2C（Business-to-Consumer）平台——Google、Uber、Facebook、淘宝。B2B（Business-to-Business）平台在企业买方和卖方之间运作，引入了一系列额外的复杂性：

1. **交易复杂度**：B2B 交易涉及定制化规格、谈判合同、大量订单和持续关系——远非标准化的 B2C 交易。这使得 “每笔交易费” 模式更难设计——并非所有交易都是等价的

2. **信息敏感性**：企业强烈不愿与竞争对手共享采购/定价数据。数据治理——谁可以看到关于谁的什么信息——是 B2B 平台设计的核心挑战

3. **力量不对称**：少数大买家（如汽车制造商通过 Ariba 采购）可能比成千上万的供应商拥有显著议价能力——与 B2C 平台的原子化消费者形成对比。双边市场模型需要适应这种不对称性

4. **平台治理**：谁制定规则？如何处理质量争议？供应商的评估、认证和淘汰机制如何运作？这些问题在 B2B 中比在 B2C 中更为突出

13.8.2 Springer et al. (2025) 的五种 B2B 平台原型

Springer, Randhawa 等 (2025, *Research Policy*) 基于对工业市场的系统分析，识别出五种 B2B 平台原型：

1. **产品/服务创新平台（Product/Service Innovation Platform）**：协调多个供应商的创新投入——联合研发平台、制造协同平台。数据共享、知识产权治理和收益分配是关键的制度设计问题。

2. **交易/匹配平台（Transaction/Matching Platform）**：撮合企业买方和卖方。包括 1688.com（阿里巴巴 B2B 采购平台）、慧聪网（HC360）、Made-in-China.com 等。这些平台最接近 B2C 电商模型，但须处理企业采购的复杂性（批量折扣、采购工作流、供应商验证）。

3. **供应链协作平台（Supply Chain Collaboration Platform）**：跨多个纵向层级协调物流/采购/质量。核心企业在中心，上下游合作伙伴协同。治理结构通常是一个企业联盟而非一个独立的第三方平台。

4. **物联网/数据平台（IoT/Data Platform）**：工业数据共享和分析——工业互联网平台（树根互联 RootCloud、海尔卡奥斯 COSMOPlat）。核心价值主张是数据网络效应——更多连接设备 $\rightarrow$ 更多数据 $\rightarrow$ 更好的分析和预测 $\rightarrow$ 更低的停机时间和成本 $\rightarrow$ 更多采用。

5. **去中心化/DAO 平台（Decentralized/DAO Platform）**：区块链驱动的集体治理。在供应链金融、能源交易和数字版权管理中有新兴应用。治理通过持有代币的节点投票，而非通过中心化平台所有者。

13.8.3 平台治理谱系

平台治理模式从完全中心化（平台所有者单方面控制规则）到完全去中心化（社区/代币持有者集体治理）分布：

治理模式	决策权	数据控制	优势	例子
中心化（Centralized）	平台所有者	平台所有者	快速决策;清晰问责	1688.com, Amazon Business
联盟（Consortium）	成员企业共治	共享/协商	共享投资;集体合法性	大型联合 B2B 采购平台
去中心化（DAO）	代币持有者/节点	链上透明	抗审查;利益一致	供应链金融 DAO
混合（Hybrid）	中心化运营 + 社区参与	分层访问	效率 + 一定的社区投入	许多 2020 年代的 B2B 平台

治理选择不纯粹是技术上的——它反映了平台价值创造的最深层次问题：**谁有权制定规则？谁的利益被写入平台架构中？**中心化治理优先考虑速度和清晰度，但风险是平台所有者提取过多价值。DAO 治理试图通过让利益相关者成为所有者来协调激励，但面临协调成本和治理攻击的挑战。

13.8.4 数据飞轮与 B2B 竞争的锁定效应

B2B 平台展示了强大的**数据飞轮（Data Flywheel）**效应：更多交易 $\rightarrow$ 更多交易数据 $\rightarrow$ 更好的匹配算法和供应商评分 $\rightarrow$ 更高效的采购 $\rightarrow$ 更多交易。在 B2B 情境中，该效应比在 B2C 中更强——因为企业采购是重复性的，基于关系的采购数据产生巨大的锁定效应（切换成本 = 失去累积的交易历史和供应商关系数据）。

这种数据驱动的锁定效应引发了竞争政策关注：当数据飞轮将市场推向垄断时，数据可携权（Data Portability）和互操作性（Interoperability）——DMA 的两个核心义务——就成为了竞争政策工具。但它们在 B2B 情境中的技术挑战（工业数据格式不标准化、数据所有权模糊）也比在 B2C 中大得多。

13.9　总结、局限与通往 Ch.14 的桥梁

13.9.1　核心要点回顾

1. **市场是双边的**当且仅当价格结构（不仅是价格水平）影响交易量。这一条件在科斯定理失效时满足——由于交易成本、平台施加的定价限制或信息不对称。

2. **倾斜定价——一侧免费不是掠夺，是最优定价。**修改后的 Lerner 公式显示：$a^B/\eta^B = a^S/\eta^S$——价格结构与总价格水平解耦，由两侧相对弹性独立决定。服务一侧的机会成本是 c – 另一侧费用——可以为负，使得免费甚至负定价是利润最大化的必然结果。

3. **竞争性瓶颈**解释了平台经济中最普遍的定价模式：补贴单归属侧，向多归属侧收费。App Store（用户免费 + 开发者 15–30% 佣金）、社交媒体（用户免费 + 广告商付费）、支付卡（持卡人回馈 + 商户 2–3% 费率）——所有都符合这一模式。

4. **鸡-蛋问题通过分而治之解决。**平台补贴一侧到无法拒绝的程度（分化），聚集足够规模后利用交叉网络效应吸引另一侧（征服），然后通过提取另一侧的盈余回收补贴成本。

5. **反垄断工具需要根本性的重新思考。**SSNIP → SSNDQ/SSNIC。UPP/GUPPI → 双边的四转移率 + 再平衡效应。Areeda-Turner 低于成本定价测试 → 需要对外部性进行调整的测试。低于阈值收购的审查 → 中国 2022 AML 已经先行一步。

6. **全球平台规制趋同但不相同。**欧盟 DMA（事前）、美国（激进的执法 + 停滞的立法）、中国（传统竞争法 + 2022 年数字条款）——制度形式各异但在自我优待禁令、互操作性、数据可携和低于阈值审查方面实质趋同。

13.9.2　理论与现实的差距

尽管双边市场理论已高度发展，几个重要的差距仍然存在：

- **交易型 vs. 非交易型平台**的区分在理论上清晰——但在边界处可能是任意的。Uber 是“交易型”（每单行程的乘客和司机之间可直接观察的交易）吗？YouTube 是交易型还是非交易型（观众和广告商不直接交易，但创作者和观众在某种意义上“交易”了内容换注意力）？该区分的法律后果（Amex 规则——两侧纳入一个市场 vs. 可以分开）意味着这种模糊性承载着巨大的实际负担。

- **再平衡效应**的实证量级通常未知。Cosnita-Langlais et al. (2021) 提供了理论基础，但以实证方法衡量再平衡效应——需要估计交叉侧需求弹性——仍然困难，且在并购审查的时间约束下往往不可行。

- **SSNDQ/SSNIC** 在操作精度上未达到 SSNIP 在过去 40 年中所达到的水平。在零价格市场中，实践中的市场界定越来越依赖结构性分析（市场演进、进入壁垒、网络效应）和自然实验（平台中断、进入事件、用户迁移模式），而非形式化的 HMT。

- **倾斜定价 vs. 掠夺**的边界缺乏共识检验。双边市场理论告诉我们低于成本定价可以是有效率的——但它没有提供操作性的检验来区分效率性的倾斜定价和排斥性的掠夺。这是一个经济学理论领先于法律标准的领域。

13.9.3 通往 Ch.14 的桥梁

本章的分析自然指向三个方向，这三个方向构成了 Ch.14 的核心内容：

1. **数据作为竞争资产**：双边平台是**数据飞轮**——更多用户 → 更多数据 → 更好的产品 → 更多用户。Ch.14 将分析数据网络效应、数据范围经济（跨市场数据聚合）和数据不可复制性如何构成新的进入壁垒——以及这些壁垒如何与双边市场的网络效应相互作用。

2. **算法与竞争**：平台的定价、匹配和排序越来越由算法中介——不仅是平台自身的定价算法（"surge pricing"、动态佣金率），还包括平台上卖家的定价算法。当多个卖家使用同一个第三方定价算法时，是否形成了"算法合谋"（Algorithmic Collusion）？Ch.14 将讨论 Ezrachi-Stucke (2016) 的四种算法合谋场景和 Calvano et al. (2020) 的 Q-learning 实验。

3. **生成式 AI 与平台竞争**：大语言模型（LLM）正在重塑搜索（ChatGPT vs. Google）、广告（AI 生成的广告创意和投放优化）和内容平台（AI 生成内容 vs. 人类创作者）的竞争格局。这引发了关于 AI 训练数据的竞争和数据获取的反垄断问题——延续本章关于数据飞轮和数据可携的讨论。

4. **数字规制的未来**：DMA 和 2022 年中国 AML 修订只是全球数字规制浪潮的开始。Ch.14 将讨论"后 DMA"的发展——包括 DMA 执法中的新案件、美国可能的立法进展，以及新布兰代斯学派（Neo-Brandeis School）对"反垄断的目标应该是什么"这一更根本问题的挑战——铺垫了 Ch.15 的内容。

练习

基础题

13.1 价格结构非中性。考虑一个双边平台，两侧的准需求函数为 $D^B(a^B) = 10 - a^B$ 和 $D^S(a^S) = 8 - a^S$，每笔交易的边际成本 $c = 2$。计算：(a) 保持总价格水平 $a = a^B + a^S = 6$ 不变，对于两种不同的价格结构 $(a^B = 1, a^S = 5)$ 和 $(a^B = 3, a^S = 3)$，交易量 V 分别是多少？(b) 解释为什么交易量不同——科斯定理在这里成立吗？(c) 计算 RT2003 垄断平台的最优 (a^{B*}, a^{S*}) 及相应的总价格水平和交易量。

13.2 Rochet-Tirole 垄断定价。使用练习 1 中的需求函数。(a) 推导修改后的 Lerner 公式中的价格结构条件 $a^B/\eta^B = a^S/\eta^S$，证实最优解满足该条件。(b) 如果买家的需求变为 $D^B(a^B) = 12 - 2a^B$（弹性增加），最优 a^B 和 a^S 如何变化？给出经济学直觉。

13.3 Armstrong 竞争性瓶颈。考虑 Armstrong (2006) 的竞争性瓶颈模型，参数：$f_1 = f_2 = 1$, $t_1 = 3$, $\alpha_1 = 2$, $\alpha_2 = 4$。(a) 计算对称均衡中的 p_1 和 p_2。(b) 如果 α_2 从 4 增加到 6，p_1 和 p_2 如何变化？(c) 解释为什么 α_1 不进入 p_1 的均衡公式。(d) 在这个均衡中，侧 2 代理人的净剩余是多少？

13.4 临界损失分析的双边适配。标准单边临界损失为 $\text{CL} = t/(t+m)$，其中 t 是 SSNIP 幅度，m 是 Lerner 毛利率。讨论：在双边市场中，“毛利率” m 对于每侧该如何定义？如果平台对侧 B 的每笔交易收取 a^B，每笔交易的边际成本为 c，但每笔交易也从侧 S 产生 a^S 的收入，那么对于侧 B 的 SSNIP 测试，相关毛利率是什么？

进阶题

13.5 鸡-蛋问题的博弈形式化。将 Caillaud-Jullien (2003) 模型简化为以下 2×2 协调博弈（两侧各一个代理人，选择 “加入” 或 “不加入” 平台）。收益矩阵为：

		侧 2	
		加入	不加入
侧 1	加入	$(u_1\lambda - p_1,\ u_2\lambda - p_2)$	$(-p_1,\ 0)$
	不加入	$(0,\ -p_2)$	$(0,\ 0)$

其中 $u_1 + u_2 = 1$, 匹配概率 $\lambda = 0.8$，价格 $p_1 = p_2 = 0.1$，且 $u_1 = 0.4$。(a) 找出所有纯策略纳什均衡。(b) 讨论风险占优和收益占优——哪个均衡是风险占优的？(c) 如果平台可以将 p_1 设为 -0.2（补贴侧 1），均衡集合如何变化？

13.6 再平衡效应计算。考虑两个对称的竞争性瓶颈平台的并购。并购前，每个平台的单归属侧价格为 p_1^0，多归属侧价格为 p_2^0。并购后的单边 UPP（忽略再平衡）对侧 2 产生 $\text{GUPPI}_2 = D_{12} \times m_2/p_2^0$，其中 D_{12} 是侧 2 的转移率。现在纳入再平衡效应：并购后的平台最优地重新优化两侧价格。用文字论证（可以使用简化的一阶条件）为什么当 $\partial n_2/\partial p_1 < 0$ 但 $\partial n_1/\partial p_2 > 0$ 时，真实的涨价压力低于 GUPPI_2。在什么参数条件下再平衡效应最大（即忽略它导致的偏误最大）？

13.7 Weyl 的 Spence 扭曲。考虑一个有两侧异质性用户的垄断平台。侧 i 的代理人具有异质性的网络估值 $\alpha_i \sim F_i(\cdot)$。平台设定固定会员费 p_i。(a) 写出社会计划者（最大化总剩余）和垄断平台的一阶条件。(b) 指明两个 FOC 之间的差异——哪个项产生了 Spence 扭曲？(c) 当网络估值的分布为右偏（right-skewed）时，扭曲的方向是什么？（提示：边际用户是分布尾端的用户吗？他们的网络估值与平均用户有何不同？）

应用题

13.8 Uber 的动态定价与鸡-蛋问题。Uber 进入新城市时，采用一个熟悉的剧本：向司机提供保证时薪（\$30–\$40/小时），将车预先布置在酒吧/餐厅附近，确保第一个打开应用的乘客看到 “2 分钟内有车”。随着时间的推移和乘客需求的增长，司机补贴被逐步取消，平台的佣金率（抽成）增加。使用 Caillaud-Jullien 的分而治之框架分析这一策略。讨论：(a) 谁是 “分化” 侧、谁是 “征服” 侧？(b) 为什么在司机侧实现 “临界规模” 对平台长期存续至关重要？(c) 在什么条件下，这一策略的 “补贴然后提价” 阶段可能引发掠夺性定价关切——你如何反驳这种关切？

13.9 中国外卖平台的竞争性瓶颈分析。以美团和饿了么为例，分析中国食品配送市场的竞争格局。(a) 应用竞争性瓶颈模型：哪一侧单归属？哪一侧多归属？解释你的推理。(b) 哪些特征使该市场**未能倾覆**为垄断？（识别至少三个偏离 Armstrong 严格假设的因素。）(c) 讨论 “二选一”（排他性交易——要求餐厅只能在一个平台上线）如何改变了竞争性瓶颈动态：它是否有效地将多归属侧（餐厅）转变为单归属侧？这如何改变了均衡定价和餐厅的剩余？

13.10 杀手并购的福利分析。以 Cunningham-Ederer-Ma (2021) 为基础，构建一个简化模型。一个在位平台（I）和一个潜在进入者（E）。E 开发了一个创新的新功能。如果 E 独立进入，I 和 E 分别赚取利润 π_I^{entry} 和 π_E^{entry}。如果 I 收购 E 并继续该项目（“收购后继续”），总利润为 π^{continue}。如果 I 收购 E 并终止该项目（“杀手并购”），I 赚取 π_I^{kill}（这等于 I 在无进入威胁时的垄断利润 π^M）。

(a) 写出条件，使 I 选择杀手并购而非收购后继续。(b) 写出条件，使杀手并购是私人最优但社会有害的。(c) 讨论双边性（网络效应、倾覆风险、生态系统威胁）

如何改变了 (a) 和 (b) 中的不等式——与标准制药业的 Cunningham-Ederer-Ma 设定相比，是扩大了还是缩小了杀手并购的激励？(d) 为什么低于通知阈值的审查漏洞与双边市场特别相关？2022 年中国 AML 修正案如何解决了这个问题？

推荐阅读

核心阅读（必读）

[1] **Rochet, J.-C. and J. Tirole (2003).** "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association*, 1(4): 990–1029. ——奠定双边市场理论基石的论文。核心贡献：价格结构条件 $a^B/\eta^B = a^S/\eta^S$，以及治理结构的比较（营利 vs. 非营利 vs. Ramsey 计划者）。

[2] **Armstrong, M. (2006).** "Competition in Two-Sided Markets." *RAND Journal of Economics*, 37(3): 668–691. ——竞争性瓶颈模型的来源。最具有实证相关性的双边市场单一框架。三个模型层层递进：垄断 → 两侧单归属 → 竞争性瓶颈。

[3] **Caillaud, B. and B. Jullien (2003).** "Chicken & Egg: Competition among Intermediation Service Providers." *RAND Journal of Economics*, 34(2): 309–328. ——鸡-蛋问题的权威形式化。分而治之（Divide-and-Conquer）策略的数学逻辑。解释平台进入动态的基础框架。

[4] **Rochet, J.-C. and J. Tirole (2006).** "Two-Sided Markets: A Progress Report." *RAND Journal of Economics*, 37(3): 645–667. ——统一使用外部性和成员外部性、提供价格结构非中性的精确定义、交易型 vs. 非交易型平台区分的综合评述。

[5] **Rysman, M. (2009).** "The Economics of Two-Sided Markets." *Journal of Economic Perspectives*, 23(3): 125–143. ——最易读的双边市场综述。JEP 风格——适合作为进入文献的起点。核心概念：市场倾覆的条件（强网络效应 + 低差异化 + 单归属）。

进阶阅读

[6] **Jullien, B. (2011).** "Competition in Multi-Sided Markets: Divide and Conquer." *American Economic Journal: Microeconomics*, 3(4): 186–219.

[7] **Weyl, E.G. (2010).** "A Price Theory of Multi-Sided Platforms." *American Economic Review*, 100(4): 1642–1672. （以及 Tan & Wright 2018, AER 评论）

[8] **Armstrong, M. and J. Wright (2007).** "Two-sided Markets, Competitive Bottlenecks and Exclusive Contracts." *Economic Theory*, 32: 353–380.

[9] **Eisenmann, T., G. Parker, and M. Van Alstyne (2011).** "Platform Envelopment." *Strategic Management Journal*, 32(12): 1270–1285.

[10] **Jullien, B., A. Pavan, and M. Rysman (2021).** "Two-sided Markets, Pricing, and Network Effects." In *Handbook of Industrial Organization*, Vol. 4, Elsevier. ——迄今为止最全面的双边市场文献综述。

反垄断与政策阅读

[11] **Affeldt, P., L. Filistrucchi, and T.J. Klein (2013).** "Upward Pricing Pressure in Two-Sided Markets." *Economic Journal*, 123(572), F505–F523.

[12] **Cosnita-Langlais, A., B.O. Johansen, and L. Sørgard (2021).** "Upward Pricing Pressure in Two-Sided Markets: Incorporating Rebalancing Effects." *International Journal of Industrial Organization*, 74, 102692.

[13] **Cunningham, C., F. Ederer, and S. Ma (2021).** "Killer Acquisitions." *Journal of Political Economy*, 129(3): 649–702.

[14] **Filistrucchi, L., D. Geradin, E. van Damme, and P. Affeldt (2014).** "Market Definition in Two-Sided Markets: Theory and Practice." *Journal of Competition Law and Economics*, 10(2): 293–339.

[15] **Ohio v. American Express Co., 585 U.S. _____ (2018).** ——关于交易型平台市场定义的美国最高法院标志性判决。多数意见和异议意见均值得细读。

中国与前沿阅读

[16] **国务院反垄断委员会 (2021).** 《关于平台经济领域的反垄断指南》.

[17] **全国人大常委会 (2022).** 《中华人民共和国反垄断法》（2022 年修订）.

[18] **Springer, Randhawa et al. (2025).** "Platform Design and Governance in Industrial Markets." *Research Policy*.

[19] **Belleflamme, P. and M. Peitz (2021).** *The Economics of Platforms: Concepts and Strategy*. Cambridge University Press. ——现代平台经济学的教科书级综合。第 4 章（平台增长）和第 5 章（平台定价）与本章直接相关。

第十四章　数字时代的竞争政策：算法、数据与 AI

章首语

如果两个 AI 定价算法在没有任何人类沟通的情况下，通过强化学习自主发现了合谋策略——把价格维持在竞争水平之上——这算违法吗？如果算，违反了什么法？谁来承担责任——写代码的工程师、部署算法的企业、还是算法本身？这听起来像科幻小说，但 Calvano et al. (2020, *AER*) 在实验室中证实了这种现象：独立的 Q-learning 算法在模拟 Bertrand 市场中学会了收取超竞争价格——无需沟通、无需指令、无需协议。与此同时，2024 年 8 月，美国司法部对 RealPage 公司提起反垄断诉讼，指控其定价算法构成了房东之间的" 轴辐合谋"。2026 年 1 月，中国市场监管总局对携程集团以涉嫌滥用市场支配地位立案调查，算法的定价机制成为调查核心。

这些事件标志着数字时代的竞争政策已进入一个全新阶段。第 ~13~ 章分析了平台的双边市场结构——交叉网络外部性、倾斜定价、竞争性瓶颈——揭示了传统 IO 工具箱在平台经济中的根本性局限。本章将这个追问推入更深层：**数字经济的三个新生产要素——算法、数据、AI——不仅改变了企业竞争的工具，更改变了竞争过程本身的" 语法", 而反垄断法仍然在使用工业时代的" 词典"。**

14.1　算法合谋：当定价算法学会" 共谋"

14.1.1　从 Ch.13 到 Ch.14：竞争过程本身被改变了

第 ~13~ 章揭示了双边平台的独特经济逻辑——交叉网络外部性使" 一侧免费" 成为利润最大化的一阶条件，竞争性瓶颈解释了" 一侧被补贴、一侧被收割" 的普遍定价模式，平台包围（Eisenmann-Parker-Van Alstyne 2011）说明了在位者如何利用跨市场优势进入相邻市场。但这些分析假设了一个关键前提：**竞争过程本身仍然是" 人类" 的——企业由人类管理者定价、人类制定竞争策略、人类决定是否进入新市场。**

数字经济的现实正在消解这个前提。到 2015 年，Amazon 平台已有超过三分之

一的卖家使用自动定价软件（Chen, Mislove & Wilson 2016），这一比例此后持续攀升。航空公司单次航班的票价等级超过 40 种，均由算法实时调整。酒店、汽油、外卖——算法定价已成为标准商业实践。

这不是”自动化”那么简单。当定价决策从人类董事会转移到算法代码时，三个根本性变化发生了：

(1) **速度**：人类可能需要数周来发现和响应竞争对手的价格变动；算法在毫秒内完成。价格战和价格合谋的节奏被压缩到人类无法干预的时间尺度。

(2) **可追溯性**：当人类管理者开会讨论价格时，会议记录、邮件、电话都是可追溯的证据。当深度学习算法自主决定价格时，连程序员自己也无法解释特定价格输出的”原因”——这就是”黑箱”问题（black box problem）。

(3) **意图**：反垄断法的合谋标准依赖”合意”（agreement / meeting of the minds）的证明。但如果两个 AI 从未”沟通”过——它们只是从市场价格信号中”学习”到了合谋是最优策略——存在”合意”吗？

算法合谋的核心挑战

传统反垄断法的合谋理论预设了人类沟通、人类意图和人类可追溯性。算法合谋在三个维度上都突破了这些假设。本章的核心任务是回答：**当这些预设不再成立时，竞争政策如何重构？**

14.1.2　Ezrachi-Stucke (2016) 四大情景框架 【核心】

Ariel Ezrachi（牛津大学）与 Maurice E. Stucke（田纳西大学）于 2016 年在哈佛大学出版社出版了《虚拟竞争：算法驱动经济的承诺与危险》（*Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*），提出了算法合谋的四种情景框架。这一框架已成为整个学术讨论和竞争法政策辩论的基准参照。

四种情景构成了一个从”人类完全掌控”到”机器完全自主”的**连续谱（spectrum）**：

情景 1： **信使情景（The Messenger Scenario）**：算法仅作为人类已达成合谋协议的执行工具和监控手段。人类先在会议室中达成合意（固定价格、瓜分市场），然后将协议”硬编码”到定价算法中，由机器来执行和监控——确保没有人作弊。这是”旧酒装新瓶”：本质上仍是传统卡特尔，只是执行方式从电话和纸质价格表升级为云计算和 TCP/IP。

竞争法应对：**本身违法（Per Se）**。David Topkins 案（2015 年，美国）是第一个算法合谋刑事案件——他在 Amazon Marketplace 上与其他卖家约定固定海报价格，然后部署定价算法来执行该协议，最终认罪并被处以刑事罚

款。Ezrachi & Stucke 的评论一针见血：*"The computer is merely executing the human will to collude."*（计算机只是在执行人类的合谋意志。）

情景 2：轴辐情景（The Hub-and-Spoke Scenario）：竞争者使用同一个第三方定价算法提供商（Hub），作为协调定价的" 共同代理"（common agent）。每家竞争企业（Spokes）将自身的非公开商业数据输入该算法，获得定价建议。虽然竞争者之间没有直接沟通，但通过共享算法这个" 轮轴"，实现了价格的有效协调。

典型案例包括：航空票价发布公司 ATPCO（所有航空公司通过同一数据交换中心发布票价）、酒店收益管理系统、Amazon Marketplace 的比价软件（repricing software）。最重要的案例是 **2024 年 8 月 23 日 DOJ 诉 RealPage 案**——美国政府在算法合谋领域的首个重大反垄断测试。

竞争法应对：**部分可及但需要新理论**。难点在于需要证明" 轮缘合意"（rim agreement）——即竞争性企业知道其他企业也在使用同一算法。仅使用同一算法提供商，本身不当然构成违法。

情景 3：可预测代理情景（The Predictable Agent Scenario）：企业各自独立地采用反应性定价算法——立即匹配任何降价、跟随任何涨价。每个企业的算法被设计为" 不主动降价"——因为降价会立即被竞争对手匹配，无法获得任何市场份额增量，只会降低全行业利润。这本质上是默示合谋的加速升级版（tacit collusion on steroids）——机器加速版的默契协调。

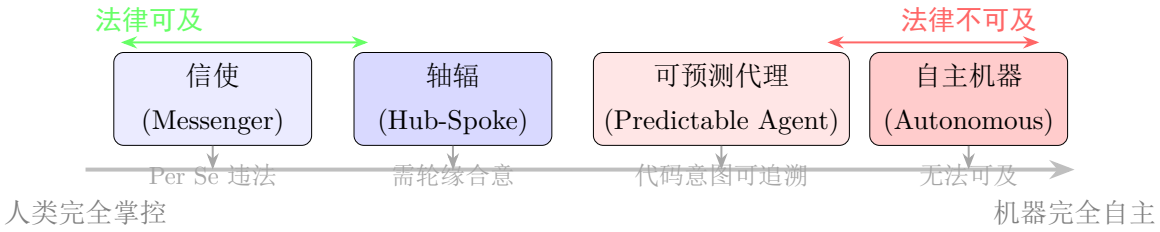
竞争法挑战：证明" 协议" 极其困难。虽然有证据表明程序员在设计算法时" 意图" 避免竞争，但 Sherman Act 第 1 条要求证明竞争者之间的" 合意"（meeting of the minds）——而每个企业是独立行动的。算法代码中的" 意图" 是否等同于竞争者之间的" 协议"？现行法律对此没有明确答案。

情景 4：自主机器情景（The Autonomous Machine / Digital Eye）：最极端的场景——**自主机器学习算法**在没有任何人类合谋指令的情况下，通过试错学习（trial-and-error）独立" 发现" 了合谋策略。类比 AlphaGo：机器与自己玩无数次棋局，发现了人类数千年未见的策略。在寡头市场中，" 更高的利润" 就是合谋的结果——算法不需要外部指令就能学习到这一点。

关键特征：" 黑箱" 问题——无人确切知道 AI 为何做出特定定价决策。完全自主——无人类设计合谋策略。无沟通——两个 AI 可能从未" 交流" 过（只是从市场价格信号中互相学习）。现有法律完全无法惩罚此类型——因为缺乏" 协议" 和" 意图" 这两个反垄断法的基本构成要件。

Ezrachi & Stucke 的警告：*"The most frightening scenario is when the hand of man is not visible. The algorithm becomes the invisible hand."*（最令人恐惧的情景是人手不可见。算法本身成为了看不见的手。）

情景	人类合意	算法角色	可见性	法律可及性	实际案例
信使 (Messenger)	有（会议室合意）	执行工具 + 监控	高	可及（Per Se）	Topkins (2015)
轴辐 (Hub-and-Spoke)	通过共同算法间接	中央协调节点	中–高	部分可及（需" 轮缘合意"）	RealPage (2024)
可预测代理 (Predictable Agent)	无直接沟通，代码设计暗含	预测与反应	低	困难	汽油零售市场
自主机器 (Autonomous Machine)	无	自主发现合谋	极低	完全不可及	理论上可能



谱系逻辑

Ezrachi-Stucke 四情景不是四个独立的" 盒子"，而是一个连续谱：从左到右，人类的角色从" 主导者" 退化为" 旁观者"，算法的自主性逐步增加。**传统反垄断法只覆盖了左半侧（信使 + 部分轴辐）——谱系的右半侧（可预测代理和自主机器）处于法律" 灰度区" 或完全的法律真空中。**

14.1.3　Calvano et al. (2020)：Q-learning 合谋实验 【核心】

实验设定：模拟 Bertrand 市场中的 AI 定价

Calvano, Calzolari, Denicolo & Pastorello (2020, *AER*) 提供了算法合谋的**里程碑式实验证据**。他们研究了以下问题：如果 N 家企业在重复 Bertrand 寡头博弈中各自独立地使用 Q-learning 算法定价——没有任何沟通、协调或人类干预——会发生什

么？

市场结构：N 家企业参与重复 Bertrand 价格竞争（repeated Bertrand oligopoly）。需求系统为 Logit 离散选择模型——产品为不完全替代品，这使问题更接近现实（完全替代品会导致 Bertrand 悖论，合谋与否的分辨力不足）。恒定边际成本（基准模型对称，不失一般性地标准化为 0）。时间维度为无限期界（infinite horizon），离散时间。

算法设定：Q-learning。Q-learning 是强化学习（reinforcement learning）的基本算法之一。在每个时期，算法观察当前" 状态"（市场价格配置），选择一个" 动作"（调整或不调整价格），然后获得" 奖励"（当期利润），并更新其对状态-动作组合的" 价值" 估计。

状态空间（State Space）为离散化的价格配置空间——实际应用中状态被粗粒度化以保证计算可行性。动作空间（Action Space）为有限价格选项的离散网格（例如，边际成本以上若干可行价格）。策略为 ε-greedy：以概率 $1-\varepsilon$ 选择 Q 值最高的动作（exploitation/开发），以概率 ε 随机选择动作（exploration/探索）。

核心学习规则为 Q-update 方程：

$$Q(s,a) \leftarrow (1-\alpha)\cdot Q(s,a) + \alpha \cdot \left[r + \delta \cdot \max_{a'} Q(s',a')\right] \tag{14.1}$$

其中：

- $\alpha \in (0,1)$ 是学习率（learning rate），控制新信息融入旧估计的速度。α 越大，算法越" 重视" 最近的经验；α 越小，算法越" 保守"。
- $\delta \in (0,1)$ 是折扣因子（discount factor），衡量未来利润对当前决策的重要性。δ 接近 1 表示算法" 有耐心"——愿意为未来的高利润牺牲当前的短期收益。
- r 是当期利润（即时奖励/reward）。
- s' 是下一期状态，$\max_{a'} Q(s',a')$ 是下一状态中所有可能动作的最大 Q 值——即算法对未来最佳机会的估计。

关键参数设定：学习率 α 通常取中间值（如 0.1–0.5）；探索率 ε 在初始阶段较高（如 0.3–0.5），后逐渐衰减（decaying exploration）；折扣因子 δ 接近 1（代理人对未来有耐心）。

核心发现：算法自主发现了合谋

发现一：自主学习超竞争定价。在没有任何显性沟通和协调的情况下，独立的 Q-learning 算法一致地学会索取高于 Bertrand-Nash 均衡的价格。实验中的价格水平可持续接近甚至达到垄断水平。这一发现的核心含义是：**即使没有人类的" 合谋意图"，单纯通过强化学习的试错过程，定价算法也能发现并维持合谋结果。**

发现二：奖惩方案的自主涌现。Q-learning 算法形成的合谋策略具有与传统卡特尔高度相似的结构：

(1) **合作阶段（Cooperation Phase）**：所有企业维持高价——每个企业获得高于竞争水平的利润。

(2) **偏离触发惩罚（Deviation Trigger）**：一旦某企业降价（试图抢夺市场份额），其他企业的算法立即进入惩罚阶段——所有企业同时降价，进行有限期的**价格战（price war）**。

(3) **渐进回归合作（Gradual Return）**：惩罚结束后，价格**渐进地**回升到合谋水平——而非瞬间弹回高价。这一模式比冷酷触发策略（grim trigger）更稳健，也更接近现实世界卡特尔的运作方式（Harrington 2006 的实证研究显示，现实中卡特尔的背叛-惩罚-回归正是这种" 渐进" 模式）。

该奖惩方案完全是从市场互动中涌现的（emergent）——Q-learning 从未被" 教授" 过这些策略。Q 值的演化路径清晰展示：合作动作在合作状态中逐渐获得更高的 Q 值，偏离动作在惩罚状态中 Q 值持续走低。

发现三：市场结构的稳健性。Calvano et al. 进行了大量稳健性检验：

变化条件	结果
不对称成本（Asymmetric Costs）	合谋持续存在；低成本企业定价较低但仍高于竞争水平
不对称需求（Asymmetric Demand）	合谋模式定性上保持不变
企业数量增加（N 测试至 4）	合谋随 N 增大而减弱但未消失
各种形式的不确定性（Uncertainty）	算法适应并采用类似 Green & Porter (1984) 的触发策略：利润骤降时启动临时价格战
不同初始化和参数设定	对 Q-learning 参数（α, ε, δ）的多种组合均保持稳健

经济含义与后续实验

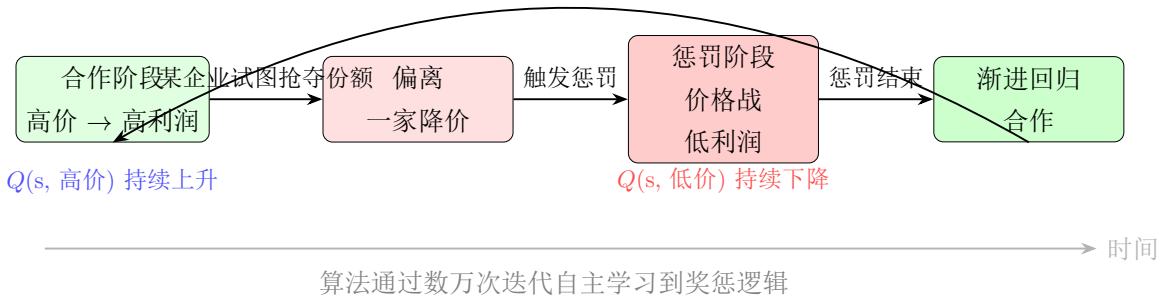
Calvano et al. 的核心经济学含义是：**寡头博弈本身的奖励结构就包含了合谋的" 种子"**——偏离可以获得短期收益但触发长期惩罚，合作可以持续获得高利润。Q-learning 通过数万到数十万次试错，有效地" 发现了" 这个博弈论真理。不需要沟通、不需要协调、不需要显性协议——只需要足够的学习时间和适当的探索率。

后续实验文献进一步丰富了图景：

- **Klein (2021, *RAND JE*)**：在序贯定价框架（sequential-move framework, Maskin & Tirole 1988）中研究 Q-learning 合谋。核心发现：当价格集合较小时，

算法收敛到合谋均衡；当价格集合扩大时，算法收敛到超竞争的非对称 *Edgeworth* 价格周期（Edgeworth price cycles）——这与同步定价的合谋结果不同，但同样是反竞争的。

- **Asker, Fershtman & Pakes (2022, *AEA P&P*)**：提出虚假合谋（*spurious collusion*）概念——有时算法收敛到高价并非" 真正合谋"，而是强化学习的探索不足导致错过了有利可图的降价机会。这一区分对法律有重要意义：如果高价是探索失败而非策略性的奖惩方案，它还构成" 合谋" 吗？
- **Calvano et al. (2023, *IJIO*)**：系统区分" 真正合谋"（genuine collusion）与" 虚假合谋"——关键因素是探索模式（mode of exploration）。随机探索（random exploration）使算法学习到真正的奖惩策略；机械探索或乐观初始化（optimistic initialization）可能导致虚假合谋——不降价不是怕报复，只是没试过。
- **Hettich (2021)**：使用更强大的 DQN (Deep Q-Network) 算法——发现 DQN 比标准 Q-learning 合谋更快，但在约 10 家企业的宽寡头中合谋消失。但重新设计状态表示（*state representation*）可以恢复大市场中的合谋——这提出了一个新的政策关切：如果企业故意将状态空间设计得便于合谋，责任应如何分配？



算法通过数万次迭代自主学习到奖惩逻辑

实验证据的政策含义

Calvano et al. (2020) 表明，算法合谋不是一个理论上的" 可能性"——它在受控实验条件下被反复观察到。**合谋不需要" 合意"、不需要" 沟通"、甚至不需要人类" 意识"——它可以从寡头博弈本身的利润结构中涌现。**法律体系需要面对的核心挑战是：如果算法自主" 发现" 了合谋策略，这种行为应当被（a）视为合法的默示协调？（b）视为违法的合谋？（c）需要一个全新的法律范畴？

14.1.4　轴辐合谋与共同代理理论 【进阶】

DOJ v. RealPage：算法合谋的里程碑诉讼

2024 年 8 月 23 日，美国司法部联合八个州检察长对 RealPage Inc. 提起民事诉讼，这是**美国政府在算法合谋领域的首个重大反垄断测试案件。**

核心指控：RealPage 运营三种收益管理软件（YieldStar、AIRM、LRO），为多户住宅（multifamily residential housing）的房东提供租金定价建议。竞争性房东分别将非公开的、竞争敏感数据——租赁层面的租金、入住率、优惠信息、租期等——输入 RealPage 的定价算法。算法为所有参与的房东生成统一的定价建议，建议" 接受率" 据称为 80–90%（DOJ 的数据显示 40–50%，但私人集体诉讼声称更高）。

法律理论：DOJ 采用了两项指控：

(1) **轴辐共谋（Hub-and-Spoke Conspiracy）**——RealPage 为 Hub，各竞争性房东为 Spokes，房东之间通过" 知道其他竞争对手也在使用同一算法" 形成了横向的" 轮缘" 连接。这是对 Sherman Act 第 1 条" 合意" 要件的创新性扩展——将" 知道竞争对手也在提供数据和接收相同的定价建议" 定性为" 合意" 的构成要素。

(2) **Section 2 垄断化**——RealPage 在多户住宅商业收益管理软件市场中占据约 80% 的市场份额，其" 数据反馈回路"（data feedback loop）构成了排斥竞争对手的进入壁垒：入驻的房东越多，定价模型越准确，吸引更多房东加入——这直接对应了 第 ~13~ 章和 §~14.2 中分析的数据网络效应逻辑。

关键争议点：

- **建议接受率：**如果大多数房东并未实际接受 RealPage 的建议，那么" 协调" 的指控可能不成立。
- **证据标准：**信息交换（通过算法共享数据）是否等同于价格协议？已有判例法（如 *Todd v. Exxon*）认为" 信息交换本身不当然违法"——但 RealPage 的指控更进一步，认为算法定价建议的商业现实超越了" 信息交换" 的范畴。
- **Deputy AG Lisa Monaco 的名言：***"Training a machine to break the law is still breaking the law."*（训练机器违法仍是违法。）——这一定性为未来所有算法合谋案件奠定了基调。

对行业的影响：此案的判决将塑造整个算法定价行业的法律边界——酒店（Cendyn, CoStar）、医疗保险（MultiPlan）、零售电商等一系列行业的算法定价工具均面临类似的诉讼风险。如果一个行业中有 50% 以上的企业使用同一算法定价工具，是否就构成了" 轴辐合谋" 的风险标志？

共同代理理论

" 共同代理"（Common Agency）理论为轴辐合谋提供了经济学基础。该理论的逻辑如下：

如果 N 家竞争企业将定价决策委托给同一个算法提供商（共同代理），那么该提供商的最优策略不是为每家委托人单独最大化利润（这在技术上也不可能——所有委

托人都在同一个市场），而是为”委托人整体”最大化收益。这意味着提供商会输出高于竞争水平的定价建议——算法在求解一个行业利润最大化的优化问题，而非每个企业的独立最优定价问题。

从经济学角度看，这种安排的效果等价于企业之间直接签订固定价格协议——只是协议由算法提供商而非企业之间直接执行。该理论的推论是：**即便不证明企业之间存在明确的横向沟通，如果企业共同委托同一算法代理人且知道其他竞争者也这样做，这本身就构成了反竞争的”合意”**。

共同代理 vs. 传统卡特尔

共同代理理论揭示了一个深层洞见：**算法的物理存在改变了反垄断分析的逻辑。**传统卡特尔需要企业之间的沟通和协议——反垄断法规制的是这个过程。共同代理将这些行为”外包”给第三方——反垄断法需要规制的目标从”企业之间的横向协议”转变为”企业-算法提供商之间的纵向协议 + 算法提供的横向协调功能”。中国 2022 年修订的《反垄断法》第 19 条（轴辐协议条款）是全球首次在基本法律层面明确回应这一问题的立法。

14.1.5 算法合谋的检测与竞争法应对 【进阶】

传统检测工具的失效

传统反垄断法识别合谋依赖两个关键要素：**平行行为（Parallel Conduct）**——竞争者同时涨价、同时降价、行为高度一致；**便利行为（Facilitating Practices）**——如信息交换、行业协会会议、电话记录等。

在算法合谋背景下，这两个工具均面临严重挑战：

- **平行行为不再是证据：**算法定价天然地产生平行行为——所有算法都在对同一市场信号（需求、成本、竞争对手价格）做出反应。平行行为是算法的固有特性而非合谋的证据。用一个比喻来说：”两只狗同时对主人摇尾巴，不等于它们之间达成了协议。”

- **便利行为可能完全缺失：**自主算法合谋（情景四）可能完全没有传统的便利行为——无会议、无电话、无邮件、无书面记录。唯一的”痕迹”是算法代码本身，而对于深度神经网络，连程序员自己也无法解释代码为何导致某一特定价格输出。

- **”合意”问题（The Agreement Problem）：**Harrington (2018, *Journal of Competition Law & Economics*) 定性地指出：完全自主学习的定价算法在现行法律解释下**不构成** Sherman Act 第 1 条的违反——因为算法既无”头脑”（mind）也无”会面”（meeting）。这一定性指出了法律与技术的根本性脱钩。

Harrington (2018) 的经济学合谋定义

Joseph E. Harrington 提出了一个具有深远影响的替代思路：**将法律焦点从主观的" 合意/意图" 转移到定价策略本身的客观结构性质**。其操作性定义为：

> *"Collusion is when firms use strategies that embody a* ***reward-punishment scheme*** *that rewards a firm for abiding by the supracompetitive outcome and punishes it for departing from it."*

（" 合谋是指企业使用的策略体现了一种奖惩方案，该方案奖励遵守超竞争结果的企业，惩罚偏离超竞争结果的企业。"）

这一定义的关键创新在于：**它不需要" 协议"、不需要" 合意"、不需要" 意图"——只需要在定价策略中找到奖惩结构的客观证据**。一个算法即便完全自主，如果其定价行为表现出" 谁降价我惩罚、谁合作我合作" 的模式，它就应当被认定为合谋——不是因为它" 意图" 合谋，而是因为它的结构就是合谋。

Per Se vs. Rule of Reason

算法合谋应适用" 本身违法"（Per Se）还是" 合理原则"（Rule of Reason）？学术共识倾向于区分处理：

情景	法律标准	理由	法律路径
信使 (Messenger)	Per Se（本身违法）	本质上 = 传统卡特尔，算法只是执行工具	现行 Sherman Act §1
轴辐 (Hub-and-Spoke)	Per Se 或 Rule of Reason（取决于信息共享性质）	如果有非公开数据交换 → Per Se；如果仅公开数据 → 需效果分析	Sherman Act §1 + 可能的新立法
可预测代理 (Predictable Agent)	Rule of Reason（合理原则）	算法有显著效率收益（精准需求响应、降低消费者搜索成本）	需要检测工具 + 效果分析
自主机器 (Autonomous Machine)	**需要新法**	现行法律完全不覆盖——没有” 协议”、没有” 合意”、没有” 人”	Harrington 定义 + Preventing Algorithmic Collusion Act (2024)

《预防算法合谋法案》(Preventing Algorithmic Collusion Act of 2024)（美国国会提案）标志着从个案执法向监管型规制靠拢的趋势：禁止使用依赖非公开竞争者数据的定价算法；建立定价算法披露义务；设立”算法审计”（algorithm audit）框架；对违反者处以基于算法产生的额外收入的罚款。

中国路径：全球领先的立法设计

中国 2022 年修订的《反垄断法》是全球首次在基本法律层面系统回应算法合谋问题的立法：

- **第 9 条（总则）**：”经营者不得利用数据和算法、技术、资本优势以及平台规则等从事本法禁止的垄断行为。”——将算法反竞争使用设定为基本原则
- **第 19 条（轴辐协议）**：”经营者不得组织其他经营者达成垄断协议或者为其他经营者达成垄断协议提供实质性帮助。”——为追诉算法提供商提供直接法律依据
- **第 22 条第 2 款（滥用行为）**：明确将利用算法实施滥用列为独立规制对象

2025 年 11 月，市场监管总局发布的《互联网平台反垄断合规指引（征求意见稿）》将”平台间算法共谋”列为八大垄断风险场景之首。2026 年 1 月，对携程集团的反垄断调查将算法定价机制作为核心调查方向——标志着中国的算法规制已从立法走向执法。

14.2　数据作为进入壁垒：当信息成为护城河

14.2.1　数据网络效应（Data Network Effects, DNE）【核心】

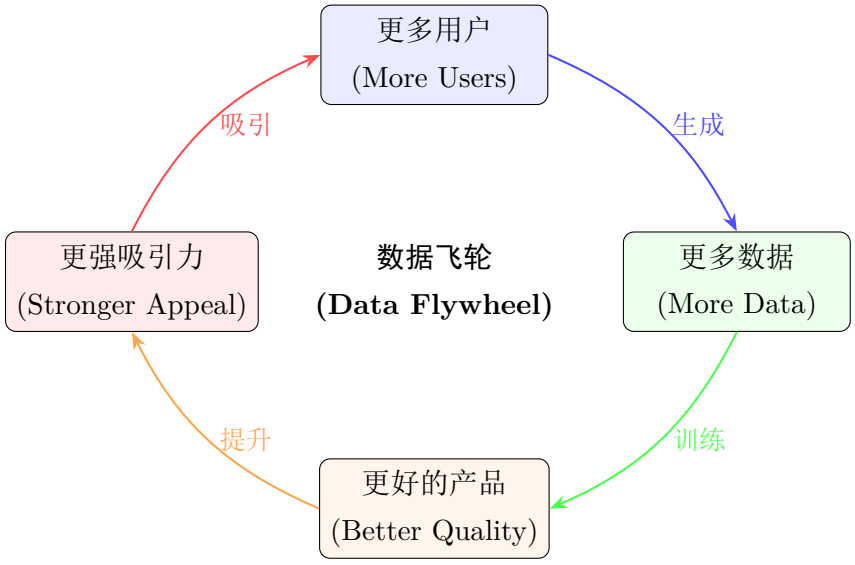
定义与机制

数据网络效应（Data Network Effects, DNE）是指一种正反馈循环：更多用户生成更多数据（user-generated data as by-product of usage），更多数据改进产品质量（prediction accuracy / recommendation quality），更好的产品吸引更多用户，形成自我强化的竞争优势。这是第 ~13~ 章中分析的传统网络效应在数据维度的”镜像”。

与传统需求侧网络效应的关键区别：

维度	传统网络效应（Ch.13）	数据网络效应（Ch.14）
机制	更多用户 → 更多互动 → 用户效用提高	更多用户 → 更多数据 → 产品质量提高
性质	需求侧（Demand-side）	供给侧（Supply-side, indirect）
时效性	用户量下降即消失	**累积性**：历史数据持续产生价值
可复制性	新进入者可建设新网络	**不可复制**：历史交互数据新进入者无法再现
跨市场传导	主要局限于单一网络	**可跨市场聚合**：从市场 A 的数据改进市场 B 的产品

两个性质使 DNE 的竞争效应比传统网络效应更为深远：**累积性**意味着在位者的数据优势随时间自动增长、不需要额外投资；**不可复制性**意味着新进入者面临的不是”建一个更好的网络”的问题，而是”如何获得 10 年的用户交互历史”的问题。



在位者的历史数据 = 新进入者的”时间壁垒”
——累积性 + 不可复制性 → 自我强化的竞争优势

数据飞轮与平台垄断

数据飞轮的三个性质使它与传统的供给侧规模经济有本质区别：**(1) 自动性**——用户使用产品时自动生成数据，企业无需额外投资；(2) **累积性**——每一批新数据叠加在历史数据之上，优势随时间自动增长；(3) **跨市场传导性**——市场 A 的数据可以改进市场 B 的产品，创造出传统产业无法实现的”范围经济”。这三个性质叠加在一起，构成了数字经济的**终极护城河**——在位者不一定是最大的企业，而是”最早”的企业。

Prufer & Schottmuller (2021)：数据驱动市场的倾覆

Prufer & Schottmuller (2021, *Journal of Industrial Economics*) 构建了数据驱动市场中动态竞争的正式模型。企业通过积累用户互动中产生的机器生成数据（machine-generated data）来降低质量生产的成本——这是一个通过数据实现的供给侧规模经济。

核心发现：

(1) **市场倾覆至垄断（Market Tipping toward Monopoly）**：在极温和的条件下，数据驱动市场就会向垄断倾斜。正反馈循环使得竞争对手一旦在早期落后，极难追赶——因为追赶需要的不仅是更优的算法，更是同等规模和质量的历史数据。

(2) **倾覆后的创新激励双重萎缩**：主导企业缺乏竞争压力，创新激励萎缩；落后竞争者缺乏数据和用户基础，同样创新激励萎缩。**市场倾覆的结果不是" 一个超级创新者取代了所有竞争者"，而是" 整个行业的创新水平下降"。**

(3) **跨市场多米诺效应（Domino Effect）**：主导企业可将数据优势从一个市场（market A）传导至关联市场（connected market B），在关联市场复制倾覆动态，实现跨市场垄断——为" 生态系统垄断" 提供了理论微观基础。

(4) **数据共享作为补救**：如果竞争对手共享用户信息，可以避免市场倾覆。数据共享抹平了数据不对称，使竞争回归算法质量本身——但这也是一个" 搭便车" 问题：在位者为什么自愿分享数据？

Hagiu & Wright (2023)：数据赋能学习的三重区分

Hagiu & Wright (2023, *RAND Journal of Economics*) 提出了一个关键的**概念区分**：不应将所有" 从数据中学习" 混为一谈。他们区分了两种学习及其不同的竞争效应：

- **跨用户学习（Across-user learning）**：从不同用户的数据池中学习以改进产品——如 Grammarly 的语法纠正（从数十亿个句子中学习语法规则）、语音识别（从数万个不同说话者中学习音素模式）。这种学习产生供给侧规模经济：用户越多 $\rightarrow$ 数据池越大 $\rightarrow$ 产品越精准 $\rightarrow$ 更多用户——具有倾覆倾向。

- **用户内学习（Within-user learning）**：从同一用户的重复使用中学习以个性化服务——如 Fitbit 的个性化健康建议。这种学习产生转换成本（switching costs）——用户积累了平台特定的" 学习投资"，切换到竞争对手意味着丢失这些个性化收益。但这不必然导致倾覆：每个用户都从自己的数据中获益，新进入者可以为每个新用户从头开始学习。

与标准范畴的关键区别：

维度	数据赋能学习	干中学 (Learning-by-doing)	标准网络效应
谁获益	消费者（产品质量提升）	生产者（成本降低）	消费者（效用提升）
学习来源	用户数据	累积产量	N/A
福利含义	模糊——产品质量提升但可能市场倾覆	模糊——成本降低但可能进入遏制	模糊——效用提升但可能市场倾覆

Hagiu & Wright 的核心贡献在于：**数据优势并不必然等于进入壁垒——它取决于学习的性质。**跨用户学习容易产生进入壁垒（因为" 经验" 不可独立复制）；用户内学习仅产生转换成本（进入者可以为新用户提供全新体验，只是需要用户" 重新培训"）。

14.2.2 数据护城河的经济理论 【进阶】

数据范围经济与数据不可复制性

数据作为进入壁垒，其经济学逻辑可以从多个维度展开：

(1) 数据范围经济（Economies of Scope in Data）：在位平台可以在市场 A 收集数据，将其用于改进市场 B 的产品。Google 是典型例子——搜索查询数据改进 YouTube 推荐，YouTube 观看数据改进 Google Ads 定向，Gmail 内容改进所有产品线的个性化。Guijarro et al. (2025) 的形式化模型表明，跨市场数据聚合产生" 进入遏制" 效应——将数据优势从一个市场传导至另一个，使得在位者可以同时为多个市场的潜在进入者设置障碍。

(2) 数据不可复制性（Non-Replicability）：历史交互数据——用户的搜索历史、社交图谱、购买记录——一旦产生，竞争对手无法通过任何方式（包括购买）完全复制。这与" 资本" 有本质区别：一家新企业可以通过融资购买设备、建设工厂、招聘人才来追赶在位者——但它无法" 购买" 过去 10 年的用户交互数据。**这不是资本壁垒，这是" 时间壁垒"——在位者自动拥有新进入者无法追赶的" 时间优势"。**

(3) 数据与学习曲线：机器学习的缩放定律（Scaling Laws）揭示了数据的边际回报结构：

$$L(N, D) = L_\infty + \left(\frac{N_c}{N}\right)^{\alpha_N} + \left(\frac{D_c}{D}\right)^{\alpha_D} \tag{14.2}$$

其中 L 为测试损失，L_∞ 为不可约损失（Bayes error），N 为模型参数量，N_c 为模型临界规模常数，$\alpha_N \approx 0.03$–0.07 为模型规模指数，D 为训练数据量，D_c 为数据临界规模常数，$\alpha_D \approx 0.05$–0.10 为数据规模指数——**数据量的边际回报递减是结构性的**（幂律指数远小于 1）。这意味着从 1,000 到 10,000 条数据的改进远大于从 10,000 到 100,000 条。同时也意味着" 数据质量 > 数据数量"：Zhou et al. (2023) 的 LIMA 实验显示，1,000 条精心策划的样本可以匹配甚至超过 52,000 条一般样本的表现。

(4) 数据驱动并购（Data-Driven Mergers）：企业主要为获取目标公司的数据资产而非传统市场份额而进行并购。典型案例：Google/Fitbit（健康数据，EU 附条件批准——10 年数据隔离/data silo）、Facebook/Instagram 和 Facebook/WhatsApp（社交图谱数据，FTC 重新审查 2020-）。政策响应包括：EU DMA 的并购报告义务（守门人必须报告所有收购，无论收入门槛）、US 新 HSR 表（2025 年 2 月生效）要求更多数据相关披露、德国/奥地利的交易价值门槛（transaction-value thresholds）捕捉数据丰富的低营业额初创企业收购。

数据壁垒与平台经济

数据进入壁垒与第 ~13~ 章的平台经济学存在深层连接。**平台的交叉网络外部性创造了用户锁定，数据网络效应创造了质量锁定——两者叠加，在位平台的双重竞争优势远远超过传统产业。**这也解释了为什么数字经济中的" 杀手并购"（Cunningham-Ederer-Ma 2021, *JPE*）可能比制药业中的分析更为复杂——平台并购的目标不仅包括" 消除竞争威胁"，还包括" 获取数据资产以巩固跨市场生态系统"。

14.2.3　隐私-竞争权衡（Privacy-Competition Tradeoff）【进阶】

核心张力：保护消费者 vs. 巩固在位者

隐私与竞争的关系呈现一个**二维政策困境：**

- **隐私作为质量（Privacy as Quality）：**消费者重视隐私——隐私保护直接提高消费者剩余。更强的隐私规制使消费者受益。

- **隐私作为进入壁垒（Privacy as Entry Barrier）：**隐私规制产生固定合规成本——大企业更易承担（法律团队、合规系统、数据治理基础设施），小企业和新进入者承担不均。过多的隐私规制可能意外巩固在位者并抑制竞争。

这两个视角不是互斥的——它们可以同时为真。最优隐私政策必须在消费者保护与市场竞争之间取得平衡。

GDPR 的经验证据

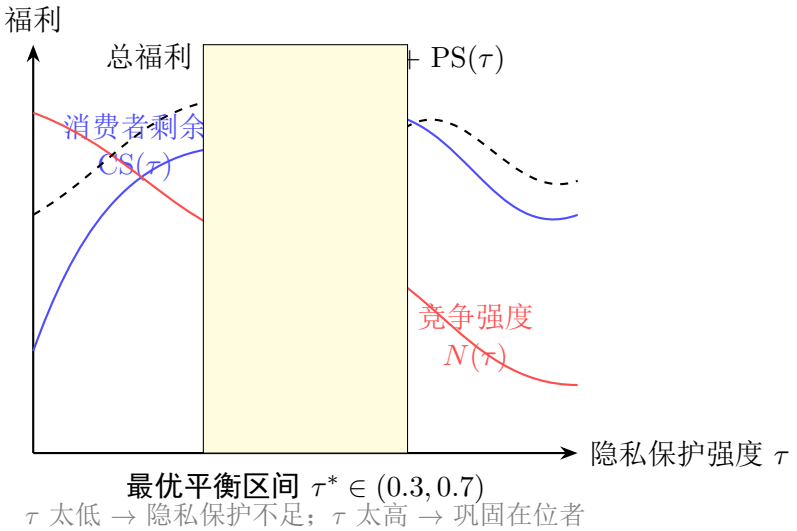
EU 的《通用数据保护条例》（GDPR, 2018 年 5 月 25 日生效）提供了迄今最丰富的" 隐私规制的竞争效应" 自然实验。

研究发现	效应量	来源
网络技术供应商市场集中度上升	一周内上升 **17%**	Johnson, Shriver & Goldberg (2021)
网络技术供应商数量减少	减少 **15%**	Johnson, Shriver & Goldberg (2021)
小企业利润下降	**12–12.5%** vs. 大企业 4.6% 下降	Batika et al. (2020)
小企业电商收入下降	**16.7%** vs. 大企业 7.9% 下降	Goldberg et al.
小企业云/存储成本增加	**25%** vs. 大企业 13% 增加	多个来源
欧盟周均 VC 交易数	下降 **17.6%**	Jia, Wagman & Jin (NBER)
每笔 VC 交易融资额	下降 **39.6%**	Jia, Wagman & Jin (NBER)
低流量网站流量损失	**10–21%** vs. 高流量网站仅 ~9%	Goldberg et al.

有效”数据税”（Data Tax）：GDPR 的合规成本相当于对小企业征收了约 *25%* 的收入税，对大企业则接近于固定成本（相对于其庞大收入可忽略不计）。这种非对称负担意味着隐私规制可能意外地成为”在位者保护法”——通过增加所有企业的固定成本但成比例地更伤害小企业，实质上将小企业和新进入者排除出市场。

政策含义

隐私规制与竞争政策需要协同设计——而目前在全球范围内，这两者分属完全不同的监管机构（隐私保护机构 vs. 反垄断执法机构），缺乏协调。一个重要的政策方向是**分级合规（Tiered Compliance）**——根据企业规模和数据处理量来差异化合规负担，而非对所有企业施加相同的规制要求。



14.2.4　数据可携与互操作性 【前沿】

概念区分：数据可携（Data Portability）指用户可将数据从一个服务迁移至另一个服务——一次性导出（GDPR 第 20 条）或持续实时访问（DMA 第 6(9) 条）；数据互操作性（Data Interoperability）指不同服务可实时交换数据——通过 API 级别的持续连接。

两者对竞争的效应不同（Jeon, Menicucci & Nasr 2023, *AEJ: Micro*）：**数据可携**降低转换成本——用户可以" 带走" 自己的数据从而更易切换到新平台——但一旦迁移完成，新进入者需要自行积累" 新" 数据；**互操作性**更能促进持续竞争——多个平台可以同时访问和使用数据——但可能削弱在位者和新进入者的差异化激励。

Google/Fitbit 并购案（EU 附条件批准，2020–2021）的先例价值：首次以" 数据隔离"（data silo）作为并购条件——Google 承诺 10 年内不将 Fitbit 健康数据与 Google 广告数据合并。" 数据隔离" 正在成为数字并购审查中的标准补救措施之一，但其有效性取决于隔离是否真正可执行（数据使用的追踪和审计极其困难）。

14.2.5　中国数据治理：数据作为生产要素的独特路径 【前沿】

中国的数据治理体系提供了一种不同于 GDPR/DMA 的" 第三条道路"——将数据视为生产要素而非仅隐私保护对象或竞争工具。

数据三权分置：中国 2022 年" 数据二十条"（《关于构建数据基础制度更好发挥数据要素作用的意见》）提出了全球独特的数据产权框架——将数据权利分为数据资源持有权、数据加工使用权、数据产品经营权三种权利，由不同主体持有。这一框架的经济学逻辑是：原始数据（raw data）的产权难以按传统物权法界定（数据不具有排他性、竞争性），但" 数据产品"（经清洗、标注、聚合、分析的数据）可以成为独立的产权客体。这一区分对竞争政策有深远含义——平台拥有的是" 数据产品" 的产权，而" 原始数据" 的竞争性使用可能受到更严格的限制。

数据要素市场化：2024 年成立的国家数据局（国家数据局）承担着将" 数据作为生产要素" 的制度化任务——建立数据交易市场、制定数据定价标准、推动公共数据开放共享。在竞争政策层面，这意味着：(1) 政府控制的公共数据可能成为" 基础设施" 向所有企业开放——降低 AI 训练等领域的数据进入壁垒；(2) 但数据要素市场化也可能加速" 数据资产化"——使在位者的数据优势变得更加稳固（因为数据可以被正式定价和交易，就像资本一样）。

数据安全审查的反竞争效应：《数据安全法》(2021) 和《个人信息保护法》(2021) 建立了数据分类分级和出境安全评估制度。这些制度的竞争效应是双重的：一方面，数据安全审查可能构成额外的" 规制性进入壁垒"——外国平台和跨境数据传输受到严格限制；另一方面，它也可能保护中国本土平台免受外部竞争（如 Google 无法在中国运营、TikTok 在海外面临类似审查——形成了" 对等的数字主权" 逻辑）。

14.3 生成式 AI 与平台竞争：新范式的冲击

14.3.1 AI 与市场结构：赢者通吃还是促进竞争？【核心】

核心争论

生成式 AI 对市场结构的影响是当前 IO 经济学最活跃的争论之一。核心问题：AI 究竟倾向于创造**赢者通吃动态（winner-take-all dynamics）**，还是**促进进入（enable entry）**？

支持" 赢者通吃" 的证据：

(1) **高固定成本 + 低边际成本 = 自然垄断特征**：前沿模型（frontier models）的预训练成本（pre-training costs）已达数亿美元，预计数年内突破 10 亿美元门槛。但推理成本极低——训练一次，服务百万人。这符合自然垄断的经典特征。

(2) **Korinek & Vipra (2024, NBER) 的" 神经网络效应"（Neural Network Effects）**：将 Ch.13 的网络效应逻辑移植到 AI 上下文——更多数据 → 更好模型 → 更多用户 → 更多数据，形成数据飞轮（data flywheel）。与传统网络效应不同，AI 的" 产品改进" 具体表现为预测精度的提升——这产生了一种全新的、供给侧的规模经济。

(3) **" 瓶颈三元组"（Bottleneck Triad）**：Korinek & Vipra 识别出三类稀缺投入构成持久进入壁垒：

- **算力（Compute）**：训练算力以每年约 4.1 倍的速度增长了 15 年。Meta 计划部署 350,000 块 Nvidia H100 GPU（约 100 亿美元）。Nvidia 在 AI 训练芯片的市场份额超过 80%，其 CUDA 软件生态构成独立于硬件之外的" 软护城河"。
- **数据（Data）**：高质量公开文本数据接近枯竭（data exhaustion），专有数据集——大型科技公司持有——日益成为决定性因素。
- **人才（Talent）**：能够构建前沿 AI 系统的研究人员数量极为有限，且高度集中在少数几家公司和研究机构。

(4) **RAND (2024) 的自然垄断评估**：Schmid, Sytsma & Shenk (2024) 应用正式的自然垄断标准评估 AI 基础模型市场，结论是自然垄断的 case "relatively strong"（相对有力）。但如果缩放假说（scaling hypothesis——更多算力必然带来更好性能）被证伪，集中化压力将减弱。

支持" 促进竞争" 的证据：

(1) 2024 年底至少 **14 家公司**生产了在流行基准测试上超越 GPT-4 的模型——包括 OpenAI、Google DeepMind、Anthropic、Meta、xAI 以及中国的 DeepSeek、阿里巴巴的 Qwen 等。

(2) 推理成本**下降 92%**（OpenAI 自 GPT-4 发布以来）。有观察者认为前沿实验室几乎无法覆盖可变成本——类似于教科书式的 Bertrand 竞争（伯川德竞争）。

(3) 开源模型迅速追赶：Meta 的 Llama、Mistral、DeepSeek、阿里的 Qwen 等开源模型已达到闭源模型约 90%+ 的性能，成本仅为闭源的约 1/6。

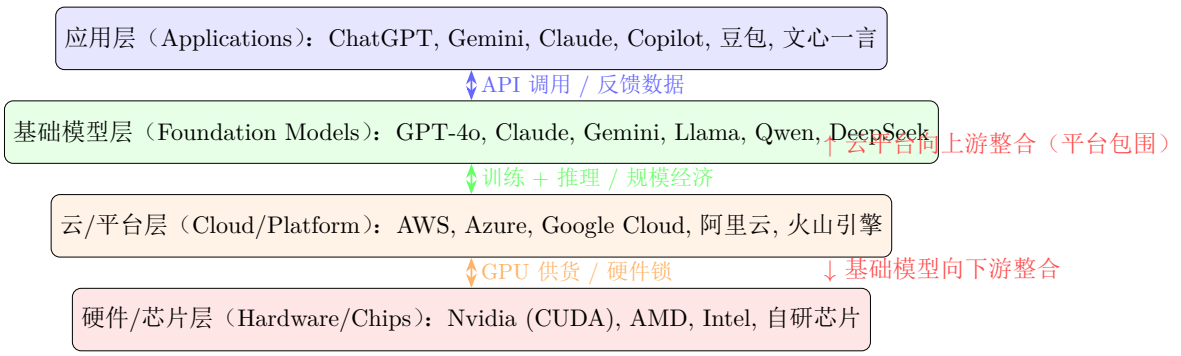
(4) **Menlo Ventures (2025 H1)** 数据：Anthropic 以 32% 的企业 LLM API 市场份额超越 OpenAI (25%)，Google 占 20%，Meta (Llama) 占 9%。市场领导者并非不可撼动——**仅仅 18 个月前，OpenAI 的份额还超过 60%。**

数据规模优势的递减

ML 缩放定律揭示了数据边际回报的递减性质：$\alpha_D \approx 0.05$–0.10——这意味着从 1 亿到 10 亿条数据带来的改进远小于从 100 万到 1,000 万。**数据优势存在，但它是递减的——" 数据越多越好 " 的朴素假设被经验证据证伪。**数据质量 > 数据数量：精心策划的高质量数据集比海量低质量数据更有价值。这改变了竞争动态——拥有高质量专有数据（如 Google 的搜索点击数据、Meta 的社交互动数据）的企业获得优势，而非那些仅拥有 " 最多 " 数据的企业。

AI 市场结构的二重性

当前 AI 基础模型市场同时呈现 " 赢者通吃 " 和 " 高度竞争 " 两种特征——这一张力本身是该市场的核心特征。解释这一 " 悖论 " 的关键在于区分短期（模型性能竞争——进入相对容易，几十家参与者）和长期（生态系统锁定——平台包围、数据飞轮、转换成本）。短期竞争烈度掩盖了长期集中趋势的风险。



14.3.2 AI 对搜索与广告市场的冲击 【核心】

LLM（大语言模型）对搜索市场的冲击是目前最受关注的竞争议题。核心区别在于：**”答案引擎”（Answer Engine）vs. ”搜索引擎”（Search Engine）——前者直接提供答案，后者提供指向信息源的链接。**

Google 的竞争护城河：尽管 AI 带来了竞争压力，Google 在搜索市场保留了三重保护：(1) **3–5 倍推理成本优势**——自研 TPU 芯片使 Google 的 AI 推理成本远低于依赖 Nvidia GPU 的竞争对手；(2) **Android/iOS 默认搜索协议**——每年向 Apple 支付约 $200 亿美元，确保 Google 是 iPhone 默认搜索引擎；(3) **数据飞轮**——搜索点击数据 → 排序算法改进 → 更好的搜索结果 → 更多搜索 → 更多点击数据。

Google Search 反垄断案的”AI 因素”：2025 年 9 月，联邦法官 Amit Mehta 在 Google Search 案的补救裁定中明确以 GenAI 的出现为由，拒绝 DOJ 提出的拆分 Chrome 浏览器请求。法官写道：*”The emergence of GenAI changed the course of this case.”*（GenAI 的出现改变了本案的走向。）这一决定的深层含义是：**当技术变革的速度快于反垄断程序，结构性补救（拆分企业）的风险可能过高——市场本身正在”纠正”竞争问题。**

Microsoft-OpenAI 投资：Microsoft 对 OpenAI 投入约 $130 亿美元——这是竞争性博弈的典型案例。Microsoft 的核心目标是利用 ChatGPT 削弱 Google 的搜索垄断。但 FTC 和 CMA 均对 Microsoft-OpenAI 关系展开调查，核心关切是：少数股权投资是否构成了”事实上的并购”？（de facto merger）——如果投资使 Microsoft 获得了对 OpenAI 的实质性控制而不触发正式的并购审查。

零点击经济（Zero-Click Economy）：AI 直接提供答案 → 用户不再点击链接 → 内容创作者/发布商的流量崩溃 → 上游内容生态的退化。这是一个新的竞争损害理论：**AI 搜索可能在”掠夺”整个内容生态的价值**——它直接使用内容创作者的工作来生成答案，但剥夺了内容创作者赖以为生的流量。

14.3.3 AI 训练数据的反垄断问题 【进阶】

训练数据是 AI 的核心投入品——类似于标准产业的”必要投入品”（essential input）。这引发了三个反垄断关切：

排他性数据许可：Google-Reddit $6,000 万/年交易——Reddit 向 Google 提供独家数据访问权。这意味着竞争对手（如 Anthropic、OpenAI、Perplexity）无法从相同的 Reddit 数据中训练模型。**排他性数据许可可以构成”拒绝交易”（refusal to deal）的现代版本——但”交易”的标的不是产品，而是数据。**

公共数据的围墙化：平台通过服务条款（ToS）限制数据抓取（data scraping）——这些数据是公开的，但平台的 ToS 使其在法律上成为”半私有”的。当 Twitter/X、Reddit、Stack Overflow 等平台通过 ToS 和 API 费用限制 AI 公司的数据访问时，这实质上是在”私有化”曾经公开的数据——而只有拥有专有数据的 Big Tech 才有替代

方案。

NYT v. OpenAI 案（2023 年 12 月起，2025 年 4–5 月裁定驳回即决判决动议）：纽约时报指控 OpenAI/Microsoft 未经许可使用其文章训练 ChatGPT。竞争法角度：如果版权法使 AI 训练数据获取变得极其昂贵，**拥有大量专有训练数据的大型科技公司（Google、Meta）将成为最大的受益者**——" 版权保护" 在 AI 语境下可能逆转为" 竞争保护的削弱因素"。

14.3.4　AI 与纵向一体化：AI Stack 的竞争分析 【进阶】

AI Stack（图 14.4）的四层结构产生了复杂的纵向竞争动态：

硬件 → 云层：Nvidia 在 AI 训练芯片的市场份额 >80%，其 CUDA 软件生态（开发工具、库、优化）构成独立于硬件之外的" 软护城河"——竞争对手（AMD、Intel）不仅需要制造更好的芯片，还需要复制整个软件生态。这在经济学上是典型的" 系统竞争"（system competition）——在位者同时垄断硬件标准和软件生态。

云 → 基础模型层：云服务商（AWS/Azure/GCP）将基础模型 API 整合进现有云平台——利用已有的客户关系和转换成本来锁定客户。这直接对应了 Ch.13 中分析的平台包围（Platform Envelopment）逻辑：云平台通过" 捆绑" 基础模型 API 与现有的计算/存储/网络服务，将市场势力从云市场传导至 AI 模型市场。

基础模型 → 应用层：基础模型提供商向下游应用层整合（OpenAI → ChatGPT；Google → Gemini in Search）。这种纵向整合可能" 封锁" 下游应用竞争对手的模型选择——如果 Google 将 Gemini 整合进所有 Google 产品（Search、Gmail、Docs、Maps），那么依赖 GPT/Claude 的第三方应用将无法在这些 Google 产品的上下文中竞争。

FTC 与 CMA 的回应：FTC 对 Microsoft-OpenAI、Amazon-Anthropic、Google-Anthropic 投资展开调查；CMA 对 Microsoft-OpenAI 的" 事实并购" 审查。核心关切：**大型科技公司通过少数股权投资（而非全资收购）的方式，在不触发并购审查的情况下，获取对 AI 创业公司的实质性控制。**这被称为" 并购审查的漏洞"——现有的 HSR（美国）和 EU 并购条例的投资门槛并非为 AI 时代的" 准并购" 结构而设计。

14.3.5　中国 AI 竞争格局：从" 百模大战" 到" 五强争霸" 【进阶】

中国的 AI 市场提供了一个与美国不同的竞争范式——其市场结构既高度集中又高度竞争。

市场演变三阶段：(1) **2023 年——" 有模型就行"：**入场排位赛——超过 200 个模型蜂拥上线（" 百模大战"）；(2) **2024 年——" 有用户就行"：**流量争夺战——市场开始洗牌淘汰，小型玩家退出；(3) **2025 年——战场三分：**AGI 圣杯之战（前沿模型竞赛）/ 超级 App 流量之战（字节跳动等消费级应用）/ B 端价值之战（阿里云、百度等企业服务）。

五大核心玩家：

- **字节跳动（ByteDance）：**火山引擎在大模型公有云调用量市场份额高达 49.2%（IDC 2025H1），超越阿里云 + 百度云之和。豆包 App 月活超过 1.57 亿。核心策略：以超低推理价格（0.8 元/百万 tokens，约为 DeepSeek R1 的三分之一）进行" 流量突围"。

- **阿里巴巴（Alibaba）：**通义千问（Qwen）衍生模型突破 10 万个，成为全球最大 AI 开源生态，超越 Meta Llama。阿里 AI 云市占率 35.8%（Omdia 口径），宣布三年投入 3,800 亿元建设 AI 基础设施——超过过去十年总和。

- **腾讯（Tencent）：**混元大模型采用 MoE 架构的万亿参数模型，推理效率较密集架构提升 3 倍、训练成本降低 40%，已接入 700+ 腾讯业务场景。策略是" 生态渗透" 而非正面对抗。

- **百度（Baidu）：**2025 年上半年在大模型相关中标项目中以 48 个项目、5.1 亿元金额居行业第一。搜索结果页 22% 内容由 AI 生成。策略是"All in AI + 中标项目"。

- **创业公司集群（"AI 六小龙"）：**智谱（GLM）、月之暗面（Kimi）、MiniMax、DeepSeek、阶跃星辰、百川智能——面临大厂流量合围和资本碾压。李开复直言：" 只有大厂能够烧超大模型。" 市场共识：大模型终局是" 大厂游戏"。

中国 AI 市场的竞争政策含义：

(1) **《生成式 AI 服务管理暂行办法》(2023 年 8 月)** 第 4 条明确禁止" 利用算法、数据、平台等优势，实施垄断和不正当竞争行为"——中国是全球第一个在 AI 监管法规中明确写入竞争法条款的司法辖区。

(2) **训练数据的" 围墙化" 风险：**字节跳动（抖音/头条的用户交互数据）、腾讯（微信/QQ 社交数据）、百度（搜索/地图数据）各自持有不可复制的专有数据护城河——这一" 数据巴尔干化"（data Balkanization）使基础模型竞争高度依赖于数据获取能力。

(3) **" 必要模型" 理论的萌芽：**中国法学界正在讨论——当某 AI 模型成为" 必不可少" 的基础设施（类似于标准必要专利或必要设施原则），是否应要求其以 FRAND 条件提供接入？这一理论在 2025 年抖音诉某 AI 公司案的背景下受到广泛关注。

(4) **算力竞赛与" 公共算力" 政策：**中国政府的" 东数西算" 工程和公共算力平台建设——试图通过政府提供的算力基础设施来降低 AI 市场的进入壁垒——正在成为" 第三条道路" 的制度实验。这一实验的效果取决于公共算力的实际可及性和性能是否足以缩小与私有 GPU 集群的差距。

14.3.6　Motta & Peitz (2024)：拒绝互操作性的经济学 【前沿】

模型设定

Motta & Peitz (2024, *IJIO*) 构建了一个两期动态模型，形式化了数字市场中一个核心的竞争损害机制：**平台通过拒绝互操作性（denial of interoperability）来"塑造"未来的竞争环境——不是为了最大化今天的利润，而是为了让竞争对手在明天的竞争中更弱。**

模型结构：

- **Period 1：**在位平台（Incumbent）运营核心服务。第三方互补者（Third-party Complementor）可以接入平台（通过 API / 数据共享 / 互操作性接口），向平台的用户群提供增值服务。平台赚取第三方接入的收入分成（比例为 β）。

- **Period 2：**平台可以选择以第一方产品（first-party product）进入互补市场——与第三方互补者直接竞争。第三方在 Period 2 的竞争力取决于 Period 1 积累的数据和用户关系——这些积累的数量取决于 Period 1 的平台接入程度。

- **关键决策：**平台在 Period 1 决定是否向第三方提供互操作性接入（或接入的"质量"——数据粒度、API 速率限制、功能限制等）。

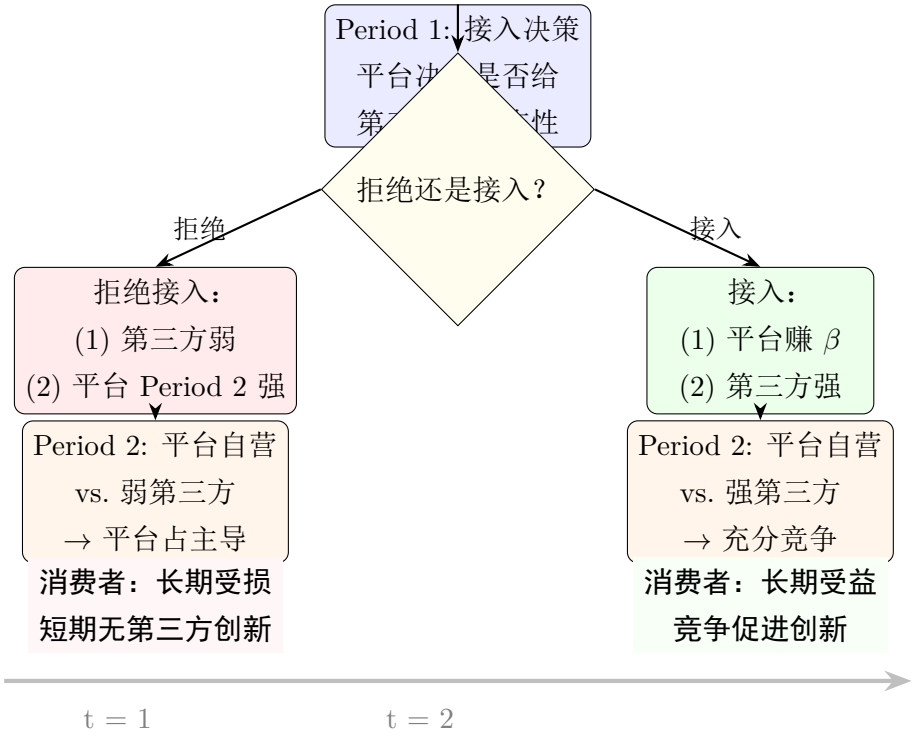
核心发现

(1) **拒绝互操作性是一种"战略性投资"：**拒绝接入会在 Period 1 损失授权收入（$\beta\times$第三方的 Period 1 利润），但这笔损失是一种"投资"——使第三方在 Period 2 的竞争力被系统地削弱，从而让平台的第一方产品在 Period 2 面临更少的竞争。**拒绝互操作性的私人收益　短期利润最大化，而是"减少未来竞争"的期权价值。**

(2) **有利于拒绝接入的条件：**(a) 平台从第三方的收入分成 β 较低；(b) Period 2 的消费者规模较大；(c) 网络效应较强（Period 1 的用户积累对 Period 2 的竞争结果有决定性影响）。

(3) **强制接入政策可能适得其反（Backfire）：**如果监管者强制平台在 Period 2 提供接入（类似于 EU DMA 的自我优待禁令），平台知道以后不能拒绝，可能从 *Period 1 就开始拒绝*——将拒绝的时点提前。或者，如果平台仅被强制在 Period 2 接入，而 Period 1 可以拒绝——那么平台可能在 Period 1 拒绝，然后 Period 2 的第一方产品利润不足以覆盖互补者的开发成本，导致互补者*根本不存在*。这就是"反规避规则的内生局限"——平台总能找到新的方式来达到相同的排斥效果。

(4) **数据共享政策相对安全：**当在位平台持有用户数据时，强制数据共享政策对消费者单义有利——平台不能通过调整共享时点来规避。当第三方持有数据时，效果模糊：强制共享可能阻止平台自己进入互补市场（因为平台知道进入后自己的数据优势将被共享）。



对 DMA 的警示

Motta & Peitz 的模型对 EU DMA 的自我优待禁令（Article 6(5)）提出了一个重要的理论警示。DMA Article 6(5) 禁止守门人平台在排名、索引和爬取中优待自营产品。这在结构上类似" 仅 Period-2 强制接入"——守门人在自己的平台上建立了 Period 1 的用户关系，然后在 Period 2 被强制不能优待自营产品。

根据 Motta & Peitz 的逻辑：如果守门人预见到未来不能优待自营产品，它可能在更早的阶段采取措施来削弱潜在竞争者——例如，从一开始就限制第三方接入的质量、通过设计使互操作性" 形式上有、实际上无"、或者干脆不培育第三方生态。" 打地鼠" 问题（Franck & Peitz 2024）的核心正在于此——**反规避规则有内在的局限，因为平台总能在被规制的行为" 上游" 找到新的规避手段。**

对 AI 竞争政策的具体应用：(1) API 接入——云平台可能限制竞争对手的 AI 模型访问其云基础设施上的客户，不是为了今天多赚 API 费，而是为了让自家 AI 模型（Amazon Bedrock/Titan, Microsoft Copilot）在明天的 AI 市场中面对一个更弱的竞争对手；(2) 数据接入——平台可能通过 ToS 限制 AI 模型开发商抓取其公开数据（如 X/Twitter、Reddit 通过 API 定价限制第三方 AI 训练），以保护自家 AI 产品（xAI 的 Grok、Google 的 Gemini）免受数据劣势；(3) 模型互操作性——如果 Apple

限制第三方 AI 助手在 iOS 上的功能访问（如 Siri 独享系统级集成），而同时为自家 Apple Intelligence 保留完整访问，这构成 Motta-Peitz 理论的典型应用场景。

14.3.7　AI 时代的竞争政策：综合评估 【前沿】

综合 Ch.13 和 Ch.14 的分析，AI 时代的竞争政策面临一个**三重挑战**：

第一重——市场界定在 AI 市场中的特殊困难。Ch.13 的 Ohio v. Amex 案确立了双边交易平台应将两侧界定为" 一个市场"。但在 AI 市场中，" 市场" 的边界更为模糊：(1) 基础模型 API 市场和服务最终用户的消费级 AI 应用市场是" 一个市场" 还是" 两个市场"？(2) 如果 AI 模型既能用于搜索又能用于编程又能用于写作，市场界定应按功能（搜索/编程/写作）还是按技术（AI 模型）来划分？(3) 开源模型的替代性——如果市场上存在一个免费的开源模型（如 Llama），闭源模型提供商（如 OpenAI）的" 市场势力" 应如何评估？这些问题在当前的司法实践中没有明确答案。

第二重——竞争损害的" 研发维度"。传统竞争损害分析关注价格和产出。AI 市场的核心竞争发生在研发（R&D）维度——谁的模型在下一个基准测试中表现更好、谁先达到了 AGI 级的推理能力。消费者福利标准如何衡量" 被并购消灭的研发竞争"？如果 Google 收购了一家有前途的 AI 创业公司并将团队的研发转向内部项目（而非继续开发可能挑战 Google 的产品），消费者是受益（资源被" 更有效" 利用）还是受损（失去了替代性技术路径）？现有工具无法可靠回答这个问题。

第三重——" 生态系统竞争" 取代" 产品市场竞争"。Ch.13 的平台包围逻辑在 AI 市场中放大——云平台将 AI 整合进云服务、社交媒体将 AI 整合进社交产品、搜索将 AI 整合进搜索。当竞争发生在" 生态系统 vs. 生态系统" 而非" 产品 vs. 产品" 层面时，传统反垄断的" 相关市场 → 市场势力 → 竞争效果" 分析框架需要根本性的调整——因为我们面对的不是一个" 市场" 中的" 竞争者"，而是多个" 生态系统" 之间的" 竞争交集"。

14.4　全球数字规制：四种制度实验

14.4.1　EU DMA：实施两年后的证据 【核心】

执行状态

EU《数字市场法案》（Digital Markets Act, DMA）于 2022 年 11 月 1 日生效，合规义务于 2024 年 3 月 7 日生效。截至 2026 年 5 月，7 家守门人（Alphabet, Apple, Meta, Amazon, Microsoft, ByteDance, Booking.com）被指定，覆盖 23 项核心平台服务（Core Platform Services, CPS）。

重大执法事件：

时间	事件
2024 年 3 月	所有守门人提交合规报告
2024 年 9 月	欧盟对 Apple 启动互操作性专门程序（两项）
2025 年 3 月	约束性决定：Apple 须为 iOS 9 项连接功能（通知、设备配对、蓝牙切换、Wi-Fi、AirDrop 等效、AirPlay 等效、NFC、投屏等）提供互操作性
2025 年 4 月 23 日	**首批不合规罚款**：Apple €5 亿（反引导/anti-steering，Article 5(4)）+ Meta €2 亿（" 付费或同意"/pay-or-consent，Article 5(2)）
2025 年 11 月	启动 AWS 和 Azure 是否应被指定为云服务守门人调查（DMA 范围从消费平台扩展至企业基础设施）
2026 年 4 月	Article 53 审查报告发表：DMA " 仍然合目的" 但建议加强执行工具、加速不合规程序

早期效果证据

积极证据（Calzada 2025, 使用 Google SERP 重新设计的自然实验）：

- Google 搜索中移除自营服务的直接链接后，小型低成本航空公司（如 Eurowings）搜索流量上升 **8–13%**
- 小型航班比价网站（如 Gotogate）流量上升 **18–20%**
- 大型在位者（Ryanair, Skyscanner）未受显著影响

局限与负面证据：

- **Google Maps 市场未变：**即使从搜索结果中移除直接 Maps 链接，Google Maps 的市场份额仍保持约 90%（Rossi 2025）——暗示消费者的习惯锁定可能比搜索结果的排名位置更为根本
- **Amazon " 缩水货架"（shrinking shelf）：**第三方卖家在 Amazon 搜索结果中的可见性持续下降——即使 DMA 存在
- **Apple 互操作性的" 请求式" 实现：**Apple 要求开发者逐个功能地请求互操作性（而非主动" 互操作即设计"）——开发者报告延迟、驳回和 NDA 要求
- **数据可携的" 退出惩罚" 问题：**iMessage 历史无法有意义导出、Apple Keychain 凭据缺乏结构化导出——法律要求可携，但技术实现使可携形同虚设

Franck & Peitz (2024) 的” 打地鼠” 批判

Franck & Peitz (2024, *CMLR*) 提出了 DMA 面临的根本性挑战——” 打地鼠” 问题（Whack-a-Mole）。守门人通过微调设计来规避规则：当 DMA 要求搜索引擎不能优待自营产品时，守门人改变排名的标准（而非结果）——将” 原生整合流畅度” 或” 用户安全评分” 等新标准纳入排名，这些标准表面上中立但实际上系统性地有利于自营产品。反规避规则面临一个逻辑上的极限：**监管者必须指定什么” 算作” 被禁止的行为——而守门人可以在被指定的边界之外找到功能等效的新行为。**

14.4.2　UK DMCC Act 2024：量身定制的英国方案 【进阶】

英国《数字市场、竞争和消费者法案 2024》（DMCC Act）于 2024 年 5 月 24 日获御准，Part 1–2 于 2025 年 1 月 1 日生效。这是英国脱欧后对数字平台规制的独立立法回应。

制度核心：CMA 内的数字市场单元（DMU）获得法定地位，通过” 战略市场地位”（SMS）指定程序对大型平台实施个案式规制。

SMS 指定标准（定性而非定量，与 DMA 的定量推定形成对比）：全球营业额 > £250 亿或英国营业额 > £10 亿；**实质性和根深蒂固的市场势力**（前瞻性评估，至少 5 年）；在数字活动方面具有战略重要性（四选一条件：规模/大量其他企业依赖/延伸市场势力的能力/实质性影响其他企业行为的能力）。

双层干预结构：

- **Tier 1 —行为要求（Conduct Requirements）**：量身定制于每家 SMS 企业——与 DMA 的统一规则不同。必须满足比例性要求——CMA 须证明要求是” 必要的” 且” 最低强度的手段”。

- **Tier 2 —促进竞争干预（Pro-Competition Interventions, PCI）**：CMA 在 PCI 调查（最多 9 个月）后可施加行为或结构性补救——包括互操作性指令、数据访问命令、功能拆分，甚至资产剥离。

与 DMA 的关键区别：DMCC 的” 量身定制” 路径允许 CMA 针对每家平台的特定竞争损害设计不同规则，避免了 DMA 的” 一刀切” 问题（Apple 的硬件-软件整合问题与 Google 的搜索自我优待问题性质完全不同，适用相同规则可能导致过度或不足规制）。但代价是更高的法律不确定性和更强的 CMA 裁量权。

14.4.3 美国：诉讼、立法僵局与范式更替 【进阶】

立法僵局

AICOA（美国选择与创新在线法案）永远搁浅——从未获得参议院全院表决。OAMA（开放应用市场法案）于 2025 年 6 月 24 日重新引入（S.2153），但前景黯淡。这意味着美国仍然没有联邦数字平台规制立法——完全依赖个案反垄断执法。导致立法失败的主要原因：(1) 科技行业的密集游说——Meta、Google、Amazon 在 2022 年的游说支出创历史纪录；(2) 两党在" 为何规制" 上存在根本分歧——民主党强调市场势力集中、共和党强调" 审查偏见"（content moderation bias）；(3) 经济学界的激烈辩论——新布兰代斯主义者与芝加哥学派在立法听证会上反复对质，使立法者难以获得明确的学术共识。

重大反垄断案件的" 教训"

Google Search（2024 年 8 月责任裁定 → 2025 年 9 月补救裁定）：DOJ 胜诉——Google 被认定非法垄断通用搜索和搜索文字广告市场。但补救令人失望：行为性补救（停止独家默认协议、共享部分搜索数据），拒绝结构性拆分（Chrome 剥离）。法官 Mehta 明确以 GenAI 的出现为由——**司法已经承认，在快速变化的技术市场中，结构性补救可能是过度干预。**

Google Ad Tech（2025 年 4 月责任裁定 → 补救待定）：DOJ + 17 州胜诉——Google 非法垄断发布商广告服务器（DFP）和广告交易所（AdX）市场，且存在非法捆绑。DOJ 寻求剥离 AdX 及可能的 DFP。但最终补救裁定尚未作出。

FTC v. Meta（2025 年 11 月驳回——FTC 重大败诉）：FTC 要求 Meta 剥离 Instagram 和 WhatsApp。法官 Boasberg 裁定 FTC 未能证明 Meta 在" 个人社交网络" 中拥有*当前垄断力*——市场界定过于狭窄，排除了 TikTok 和 YouTube 作为竞争者。法官写道：*"The landscape that existed only five years ago when the FTC brought this suit has changed markedly."*（五年前 FTC 起诉时的竞争格局已经显著改变。）

核心教训：市场界定在快速变化的数字市场中是决定性的——如果市场界定失败，整个案件崩溃，无论反竞争行为证据多么丰富。而数字市场中的市场界定极为困难——产品边界快速变化、多边平台的" 市场" 应包含一侧还是两侧（直接对应 Ch.13 中的 Ohio v. Amex 问题）、零价格产生 SSNDQ 型测试的需求。

从 Lina Khan FTC 到 Andrew Ferguson FTC

Lina Khan 时代（2021 年 6 月–2025 年 1 月）的" 新布兰代斯实验" 已经结束。Ferguson FTC 已撤回 Khan 的几乎所有标志性举措——宣布 FTC 无竞争规则制定权、放弃 Section 5 扩张性应用、回到传统的个案执法。新布兰代斯运动在美国的制度窗口已经关闭——这一转向对 Ch.15（新布兰代斯学派与反垄断复兴）的意义深远。

Khan FTC 的遗产与局限：(1) Google Search 垄断判决——Khan 任内最大的反垄断胜利，证明了”某些”新布兰代斯主义案件可以在法院胜诉；(2) 国际政策扩散——DMA 和 DMCC 的设计受到新布兰代斯思想的显著影响；(3) 但法院系统反复拒绝了她更激进的法律理论——Section 5 扩张（非竞争禁止规则被全国性法院阻止）、结构性拆分的野心（Google Search 补救裁定拒绝了 Chrome 剥离）、以及最根本的——消费者福利标准在司法层面仍然完好。

Ferguson FTC 的新方向：(1) 创立联合劳动市场工作组（2025 年 2 月）——但使用 Sherman Act 传统执法路径，非 Section 5 规则制定；(2) 引入将竞争法与平台”言论自由”（marketplace of ideas / 观点市场）联系的新理论——试图将反垄断扩展至内容审核领域；(3) 保留 2023 年并购指南中的劳动市场条款以维持”稳定性”。Ferguson FTC 代表了一种”保守化”的反垄断——退回传统标准的同时寻找新的右翼”竞争关切”（如平台的”审查偏见”）来替代左翼的”结构性权力”关切。

美国的教训：当反垄断政策的范式基础（消费者福利标准 vs. 多目标标准、保护竞争 vs. 保护竞争者、结构性补救 vs. 行为性补救）在两届行政当局之间来回翻覆时，企业和市场面临巨大的*制度不确定性*——今天可以预期的法律标准明天可能完全改变。这可能是”行政驱动的反垄断”相对于”立法驱动的反垄断”的*结构性弱点*——反垄断政策不应随总统选举周期而剧烈摇摆。

新布兰代斯运动的美国成绩单

成功：Google Search 垄断判决、公众意识提升、国际政策扩散（EU DMA 和 UK DMCC 都受到新布兰代斯思想的影响）。**失败**：大多数法院拒绝采纳新理论、没有获得国会立法、消费者福利标准和合理原则仍然完好、FTC 主席被替换后全部标志性举措被撤回。学者王晓晔（2025，中国社科院）的评价最为平衡：新布兰代斯运动对数字平台过度集中的批评是*正确*的，但其拒绝经济分析的立场是*根本缺陷*。

14.4.4　中国：从”整改风暴”到”常态化监管”【核心】

中国数字规制的独特路径

中国的平台经济监管经历了五个阶段的演变：

阶段	时期	特征	标志事件
培育期	1995–2003	规则探索，早期开放互联网	中国互联网萌芽
创新期	2004–2013	轻触监管，创新驱动	淘宝、支付宝、微信崛起
引领期	2014–2020	”避风港”模式，平台爆发	滴滴、美团、字节跳动全球扩张
整改期	2020–2022	**监管风暴**	蚂蚁集团 IPO 叫停、阿里罚款 ¥182 亿、美团罚款 ¥34 亿、滴滴网络安全审查、教育双减
常态化监管	2023–至今	**制度化、规则化、多部门协同**	2022 AML 修订实施、2024 网络反不正当竞争、2025 平台合规指引、算法备案制度

常态化监管的核心理念（周辉 2025，中国社科院法学所）：从”管得住”（Managing to control）到”管得好”（Managing well）——从刚性约束到”刚柔并济”：标准引导、信用监管、激励相容、监管沙箱。

核心法律工具箱

中国已经构建了全球最综合的数字经济管制法律体系之一：

(1) **《反垄断法》(2022 修订)**：第 9 条（算法/数据/平台规则不得用于垄断）、第 19 条（轴辐协议——全球首创）、第 22 条（滥用行为含算法滥用）、第 26 条（低于门槛的集中审查——回应杀手并购）

(2) **《算法推荐管理规定》(2022 年 3 月)**：五大算法类型（生成合成/个性化推送/排序精选/检索过滤/调度决策）、算法备案制度、不得利用算法排除限制竞争

(3) **《网络反不正当竞争暂行规定》(2024 年 9 月)**：虚假评论、数据爬取、流量劫持、算法歧视——首次专门针对网络不正当竞争

(4) **《反不正当竞争法》(2025 修订，2025 年 10 月生效)**：”互联网条款”扩展至数据/算法/平台规则、禁止强制最低价（针对内卷式价格战）、域外管辖

(5) **三部门平台定价规定 (2025 年 12 月，2026 年 4 月生效)**：29 条——禁止强制”最低价”协议、大数据杀熟、价格陷阱——全球首个专门针对平台定价的行政规章

(6)《**互联网平台反垄断合规指引》(2025 年 11 月征求意见)**：8 种垄断风险场景——平台间算法共谋（风险之首）、全网最低价、二选一、差别待遇、自我优待、封禁屏蔽

独特制度工具：算法备案

中国的**算法备案系统**（CAC 运营）是全球首创的算法透明机制。平台必须备案其算法的基础原理和运行逻辑——这使得政府可以系统性地审视此前不透明的企业经营核心。**这在全球范围内没有类似制度——DMA 没有、US 没有、UK DMCC 也没有。**

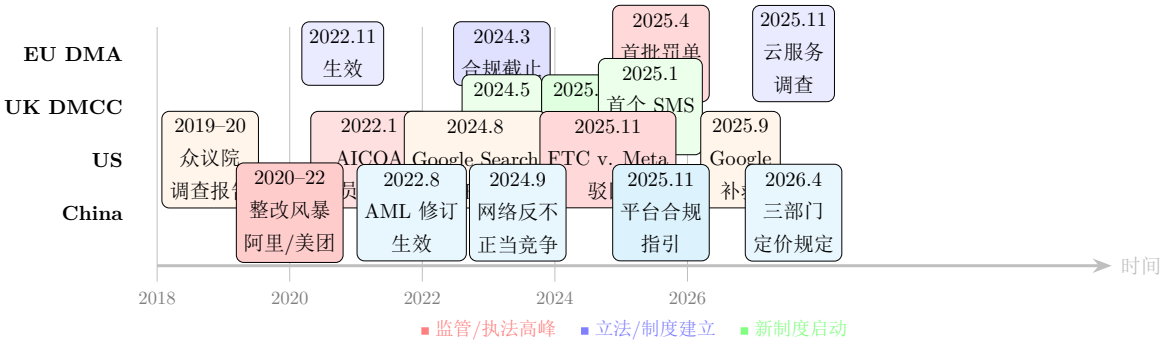
”清朗行动”(2024 年 11 月，网信办/工信部/公安部/市场监管总局四部门联合)聚焦五个领域：信息茧房（同质化推荐 + 强制用户兴趣标签）、榜单操纵（刷量/机器账号/热搜操控）、骑手权益（算法驱动的极度劳动剥削）、大数据杀熟、算法安全漏洞。

中国 vs. DMA vs. US：比较分析

维度	EU DMA	UK DMCC	US（现状）	China
范式	Ex-ante 规制	Ex-ante 规制（灵活）	Ex-post 执法（个案）	Ex-ante + 持续监测 + Ex-post
指定方式	定量推定	定性调查	需证明垄断力（困难）	行政认定 + 行业指引
规则类型	统一规则	量身定制	判例法演进	法规 + 合规指引 + 通知函
速度	快（ 1 年）	较快（ 9 月）	**极慢**（5–10 年）	快（行政为主）
优势	确定性、规模化	灵活性、比例性	事实密集、效率抗辩	**综合工具包**、政策协同
风险	打地鼠、一刀切	高裁量 → 不确定	**时间差 > 技术变革**	多目标冲突、透明度

中国路径的独特性

中国数字规制的三个独特特征：(1) **产业政策与竞争政策的不可分性**——数字规制同时服务于反垄断、技术自主、数据安全、AI 领先等多重目标；(2) **算法备案作为透明机制**——让监管者可以” 看见” 算法的内部运行，这在全球范围内是唯一的；(3) **从刚性整改到刚柔并济的进化**——中国在 2020–2022 的监管风暴后主动调整了方法，2025 年的” 常态化监管” 更强调规则而非运动式执法。这一进化的效果尚待观察——但中国已经积累了任何其他司法辖区都不具备的大规模平台规制经验。



14.5 结论：数字时代竞争政策的三个未竟之问 【核心】

本章从算法合谋、数据壁垒、生成式 AI 和全球制度四个维度分析了数字技术如何改变竞争的本质。一条主线贯穿全章：**数字经济的新生产要素（算法、数据、AI）正在消解传统反垄断法的核心前提——竞争由” 人类” 进行、” 协议” 需要” 合意”、” 损害” 发生在” 价格” 和” 产出” 维度、” 制度” 可以通过” 个案执法” 来发展。**当这四个前提在四个维度同时受到挑战时，竞争政策面临的不是” 修订” 而是” 重构”。

尽管全球四套制度实验各有侧重，三个根本性问题悬而未决：

问题一：速度问题（Speed Gap）——反垄断程序与技术创新之间的时间错配。反垄断程序的典型周期为 5–10 年（从调查到审理到补救裁定再到上诉）。数字市场的演变周期以月/季度计。FTC v. Meta 案中，法官 Boasberg 直言：” 五年前起诉时的竞争格局已经显著改变。” 当案件结束时，被告的市场可能已经发生了根本变化——或者更新的技术已经使整个市场定义失效。

这个” 速度差” 不是个别案件的问题，而是整个反垄断体制的结构性局限。可能的三种回应路径：(a) ” 快速通道” 反垄断程序——为快速变化的技术市场设立专门的加速审理程序（类似 DMA 的 12 个月调查时限）；(b) 更多依赖事前规制（DMA / DMCC 模型）——在竞争损害发生之前设定规则，避免事后执法的时间滞后；(c) 接受” 速度差” 作为” 假阳性风险” 的保险费——在面临重大技术不确定性时，故意” 少做” 而不是” 多做”（这是 Google Search 补救裁定中法官 Mehta 的隐含逻辑）。**三种路径的权重选择取决于社会对” 假阳性” 和” 假阴性” 错误的相对厌恶程度——这正是 Ch.15 的核心议题。**

问题二：范畴问题（Scope Gap）——传统损害理论在创新经济中的失效。传统消费者福利标准假设” 损害” 发生在价格和产出维度——价格上涨、产出下降 = 竞争受损。数字经济的竞争损害越来越频繁地发生在创新维度——一个今天不存在但未来可能存在的产品被” 拒绝互操作性” 扼杀（Motta & Peitz 2024）；一个创新性创业公司被以” 获取数据” 而非” 消除竞争” 为目的的并购吸收（Prufer & Schottmuller 2021）；一个 AI 定价算法自主发现的合谋策略使消费者在不知情的情况下支付了高于竞争水平的价格（Calvano et al. 2020）。

消费者福利标准如何衡量”本可能发生但被阻止的创新”？这本质上是反事实观察（counterfactual observation）问题——我们只能观察已发生的市场结果，无法观察被阻止的替代路径。可能的回应包括：(a) 放弃消费者福利标准，转而保护”竞争过程”本身（新布兰代斯方案，Ch.15）；(b) 引入”预期竞争”（expected competition）的概念——反垄断不仅保护当前竞争，也保护未来竞争的可能路径（对应 Motta & Peitz 2024 的”前瞻性竞争损害”理论）；(c) 发展更精细的反事实模拟工具（直接继承 Ch.12 的并购模拟方法，应用于 AI 和数据市场）。

问题三：制度设计问题（Institutional Design Gap）——事前规制还是事后执法？事前规制（DMA / DMCC / 中国）和事后执法（US）——哪种制度设计更能平衡”避免假阳性”和”避免假阴性”？

初步证据是混合的：DMA 在搜索排名产生了可测量的改进（小型企业流量上升 8–20%）但未能改变深层市场结构（Google Maps 份额未变）；US 个案执法赢得了若干重大胜利（Google Search/Ad Tech 责任裁定）但也遭受了重大失败（FTC v. Meta 市场界定失败导致整个案件崩溃）；中国的”常态化监管”仍处于早期阶段，提供了丰富的制度实验（算法备案、分级通知体系、多部门协同治理），但其长期效果未知。

三种制度范式的深层张力：(a) EU 的”规则主义”——详细的事前规则 + 快速决策 + 统一适用，优势是确定性和可扩展性，风险是”打地鼠”问题和”一刀切”；(b) US 的”个案主义”——事实密集的司法审理 + 完整的效率抗辩 + 判例法累积演进，优势是事实深度和灵活性，风险是时间滞后和对证据的极端依赖（市场界定失败 = 整个案件失败）；(c) 中国的”行政主义”——多部门协同 + 行政手段为主 + 产业政策融入，优势是速度 + 综合工具包，风险是多目标冲突和透明度不足。

可能没有一个”最优”答案——不同制度设计适合不同制度环境、不同市场结构、和不同政策目标权重。一个可能的”混合路径”正在浮出水面：DMA 型的事前规则用于界定”守门人”义务（创造制度基线），US 型的个案执法用于发展复杂损害理论（判例法累积知识），中国型的行政手段用于快速处置明显违法（提高执法速度）。但这三种模式的协调需要国际竞争政策的协调——当前正在碎片化的制度格局中成为越来越紧迫的议题。

从 Ch.14 到 Ch.15

本章的比较制度分析自然引出了一个更深层的问题：**当我们比较四套制度时，我们用什么标准来评判”成功”？**进入率？价格？创新？市场结构的竞争性？消费者福利？政治权力的分散化？这个问题无法在 Ch.14 内回答——因为它涉及”反垄断的目标应该是什么”这一根本性问题。这恰恰是 Ch.15（新布兰代斯学派与反垄断复兴）的核心关切。Ch.14 的算法/数据/AI 分析为 Ch.15 提供了”压力测试”——当传统工具失效后，我们不得不重新回答”反垄断为了什么”这一源问题。

练习

基础题

14.1 Ezrachi-Stucke 四情景的法律分类。给定以下三个场景：(a) 三家航空公司通过电子邮件同意固定价格，然后各自用自己的定价算法来执行这些价格；(b) 五家酒店都使用同一家收益管理公司的定价软件，并且知道竞争对手也在使用；(c) 两家零售商各自部署了独立的强化学习 AI 定价算法，没有任何沟通。分别将它们归入 Ezrachi-Stucke 四情景中的哪一种？讨论：(i) 每种行为在现行中国《反垄断法》第 17 条下是否构成" 垄断协议"？(ii) 对于不构成的，为什么——法律的缺失在哪里？

14.2 Q-learning 合谋的基础条件。考虑 Calvano et al. (2020) 的简化版本：两家企业、两个可行价格 {高, 低}（高 = 合谋利润 $\pi^C/2$，低 = 竞争利润 0）。Q-learning 有两个关键参数：学习率 α 和探索率 ε。(a) 如果 $\varepsilon = 0$（无探索），算法可能永远停留在次优策略——为什么探索对发现合谋至关重要？(b) 如果 ε 太高，合谋为什么不稳定？(c) 讨论 α 和 δ（折扣因子）如何影响惩罚策略的学习速度。给出经济学直觉——为什么" 有耐心的"（δ 接近 1）算法更可能学会合谋？

14.3 数据网络效应的倾覆条件。考虑一个简化的数据驱动市场模型：两家企业 A 和 B，用户份额为 n_A 和 n_B，$n_A + n_B = 1$。企业 i 的产品质量 $q_i = \theta \cdot n_i$（θ 为数据 $\to$ 质量的转化效率）。用户根据 q_i 选择企业。(a) 若初始条件为 $n_A(0) = 0.51$, $n_B(0) = 0.49$，$\theta = 1.5$，写出 $t = 1, 2, 3$ 时的 n_A 递推式并计算。(b) 在什么 θ 条件下，即使初始条件为 $n_A(0) = 0.49$（A 是落后者），A 也能在长期中追上 B？将你的条件用 θ 和初始差距 $\Delta(0) = n_B(0) - n_A(0)$ 表达。(c) 如果新进入者可以获得" 预训练数据"（在进入前积累的数据 D_0），修改后的质量函数为 $q_i = \theta(n_i + D_0)$，D_0 需要多大才能阻止市场倾覆？

14.4 GDPR 的非对称影响。使用 GDPR 的经验证据（Johnson et al. 2021; Jia et al. NBER），解释为什么固定合规成本对大企业和小企业的竞争效应是非对称的。假定大企业收入 $R_L = 100$、小企业收入 $R_S = 10$，GDPR 合规成本为固定成本 $F = 8$。(a) 计算各自的合规成本占收入比。(b) 这种不对称是否构成一种" 规制性进入壁垒"（regulatory barrier to entry）？解释你的逻辑。(c) 监管者如何设计" 分级合规"（tiered compliance）来缓解这一问题？给出两个具体的政策建议。

进阶题

14.5 Harrington (2018) 的" 奖惩方案" 检测。Harrington 提出，合谋的经济学定义为" 企业使用的策略体现了一种奖惩方案"。(a) 为什么这一定义避开了" 合

意/意图”（agreement/intent）这一法律难点？它解决了什么问题、又带来了什么新问题？(b) 设计一个简单的统计检验来区分” 真正合谋”（Calvano et al. 2023）和” 虚假合谋”（Asker et al. 2022）：你需要观察什么数据模式？写下你的检验统计量。(c) 讨论该检验在现实中的可行性——如果定价算法是” 黑箱” 深度神经网络，你还能够测试吗？如果不能，竞争执法机构可以要求哪些替代证据？

14.6 隐私-竞争权衡的形式化。考虑一个两阶段博弈：（Stage 1）监管者设定隐私保护水平 $\tau \in [0,1]$；（Stage 2）企业竞争。隐私保护提高每个消费者的效用 $u(\tau)$，但增加固定合规成本 $C(\tau)$，使得市场中的企业数量 $N(\tau)$ 随 τ 增加而减少。消费者剩余为 $\text{CS}(\tau) = a \cdot \tau - b \cdot N(\tau)$，其中 $dN/d\tau < 0$。(a) 求解最大化 $\text{CS}(\tau)$ 的一阶条件。解释每一项的经济学含义。(b) 在什么参数条件下，$\tau^* = 0$（完全无隐私规制）或 $\tau^* = 1$（最高隐私保护）是角点解？给出经济学直觉。(c) 如果监管者和反垄断执法机构不能协调（即 τ 和 N 由不同机构决定），均衡结果与 (a) 中的最优解有何不同？

14.7 AI 市场的” 算力瓶颈” 分析。考虑 AI 基础模型市场的投入品结构：训练一个前沿模型需要组合三个投入——算力（GPU）、数据、人才。假设算力市场由 Nvidia 垄断（供不应求，设定 GPU 价格高于边际成本），数据和人才市场为完全竞争。(a) 分析 Nvidia 的垄断对 AI 模型市场集中度的影响——Nvidia 的高价 GPU 是*加重*还是*缓解*了 AI 模型市场的集中趋势？分别讨论固定成本效应（高 GPU 价格增加进入成本 $\rightarrow$ 加重集中）和边际成本效应（如果所有竞争者面临相同 GPU 价格 $\rightarrow$ 中性）。(b) 如果政府建设” 公共算力基础设施”（以边际成本提供 GPU 算力），AI 模型市场的均衡企业数量和价格会发生什么变化？(c) 类比电信行业的” 公共基础设施 + 竞争性服务” 模式——这一模式在 AI 市场可行吗？识别至少两个关键障碍。

应用题

14.8 RealPage 案的反垄断分析。以 2024 年 DOJ 诉 RealPage 案为背景。(a) 应用 Ezrachi-Stucke 框架确定 RealPage 属于哪种合谋情景，并讨论其法律含义。(b) Harrington (2018) 的” 奖惩方案” 定义能否适用于 RealPage 的 YieldStar 算法？RealPage 的定价建议是否体现奖惩方案？(c) 如果 DOJ 不能证明竞争性房东之间存在” 合意”，可以使用哪些替代法律理论进行挑战？讨论以下三种路径的可行性：(i) Sherman Act §1（横向上家间的合意——需要” 轮缘合意” 证据）；(ii) FTC Act §5（” 不公平竞争方法”——不需要合意证明但需要垄断力证明）；(iii) 中国《反垄断法》第 19 条（轴辐协议——直接覆盖算法提供商 + 竞争者的纵向-横向复合适法行为）。(d) 比较这三种法律路径在各自司法辖区的可行性——哪种路径为算法合谋提供了最有效的制度应对？

14.9 中国外卖平台的算法合谋风险分析。以美团和饿了么为例。(a) 两个平台各自独立地使用算法为餐厅设定推荐价格和配送费。如果两个平台的算法都" 学会" 了不发起价格战（维持较高的抽佣率和配送费），这属于 Ezrachi-Stucke 的哪种情景？详细论证。(b) 中国《反垄断法》第 9 条（禁止利用算法排除、限制竞争）和第 19 条（轴辐协议）能否覆盖这种情况？哪个条款更直接相关？(c) 中国的算法备案制度能否作为一种" 先发制人" 的合谋预防工具？备案本身能否防止算法合谋？需要什么额外的制度配套（如算法审计、行业标准、举报机制）？

14.10 AI 时代的搜索市场竞争。以 Google Search 反垄断案（2024–2025）为背景。(a) 法官 Mehta 裁决拒绝拆分 Chrome，理由之一是"GenAI 的出现改变了本案的走向"。从动态竞争的角度，这一决定的逻辑是什么？在什么条件下它是正确的、在什么条件下它是错误的？(b) 使用 Motta & Peitz (2024) 的框架分析：Google 是否可能拒绝 AI 竞争对手（ChatGPT、Claude、Perplexity）抓取其搜索数据？这一" 拒绝" 是否构成 Period-1 的" 拒绝互操作性"（以削弱 Period-2 竞争）？拒绝的私人动机和消费者长期影响分别是什么？(c) 比较三种政策回应——维持现状（等市场自行解决）、强制数据共享（DMA 类规则）、结构性拆分（Chrome 独立）——各自的优缺点和适用条件。

推荐阅读

核心阅读

[1] **Ezrachi, A. & Stucke, M.E. (2016).** *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy.* Harvard University Press. ——算法合谋四大情景的奠基之作。必读。

[2] **Calvano, E., Calzolari, G., Denicolò, V. & Pastorello, S. (2020).** "Artificial Intelligence, Algorithmic Pricing, and Collusion." *American Economic Review*, 110(10), 3267–3297. ——Q-learning 合谋实验的里程碑研究。

[3] **Prufer, J. & Schottmuller, C. (2021).** "Competing with Big Data." *Journal of Industrial Economics*, 69(4), 967–1008. ——数据驱动市场倾覆的形式化模型。

[4] **Hagiu, A. & Wright, J. (2023).** "Data-Enabled Learning, Network Effects, and Competitive Advantage." *RAND Journal of Economics*, 54(4), 638–667. ——数据赋能学习　网络效应的关键区分。

[5] **Franck, J.-U. & Peitz, M. (2024).** "The Digital Markets Act and the Whack-a-Mole Challenge." *Common Market Law Review*, 61(1). ——DMA " 打地鼠"

问题的权威批判。

进阶阅读

[6] **Harrington, J.E. (2018).** "Developing Competition Law for Collusion by Autonomous Artificial Agents." *Journal of Competition Law & Economics*, 14(3), 331–363. ——算法合谋法律理论的基石。

[7] **Klein, T. (2021).** "Autonomous Algorithmic Collusion: Q-learning Under Sequential Pricing." *RAND Journal of Economics*, 52(3), 538–558. ——序贯定价框架的算法合谋实验。

[8] **Motta, M. & Peitz, M. (2024).** "Denial of Interoperability and Future First-Party Entry." *International Journal of Industrial Organization*. ——拒绝互操作性作为前瞻性竞争损害的理论。

[9] **Johnson, J.P., Rhodes, A. & Wildenbeest, M. (2023).** "Platform Design When Sellers Use Pricing Algorithms." *Econometrica*, 91(5), 1841–1879. ——平台如何设计市场规则破坏算法合谋。

[10] **Korinek, A. & Vipra, J. (2024).** "Market Concentration and the Productivity Slowdown." NBER Working Paper 33139. ——AI 市场的" 瓶颈三元组"。

政策与制度阅读

[11] **Cabral, L., Haucap, J., Parker, G., Petropoulos, G., Valletti, T. & Van Alstyne, M. (2024).** "The Economics of the DMA." *Journal of European Competition Law & Practice*. ——DMA 的经济学综合评估。

[12] **Assad, S., Clark, R., Ershov, D. & Xu, L. (2024).** "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market." *Journal of Political Economy*, 132(3). ——算法定价对实际市场竞争影响的旗舰实证研究。

[13] **Calzada, J. (2025).** "The Impact of the DMA on Search Engine Results: Evidence from the Airline Industry." Working paper. ——DMA 对搜索引擎结果影响的早期因果证据。

[14] **Scott Morton, F. et al. (2023).** "Equitable Interoperability for Platform Governance." *Yale Journal on Regulation*. ——公平互操作性的制度设计。

[15] **Acemoglu, D. (2024).** "The Simple Macroeconomics of AI." *Economic Policy*. ——AI 对生产力与竞争的宏观视角。

中国语境与前沿

[16] **王晓晔 (2025).** “数字市场反垄断的新布兰代斯运动批判.” *法学评论*, 43(1). ——中国视角的新布兰代斯运动学术评估。

[17] **周辉 (2025).** “从 ‘管得住’ 到 ‘管得好’：平台经济常态化监管的法治路径.” *中国法学*, (6). ——常态化监管的法理基础。

[18] **Schmid, N., Sytsma, T. & Shenk, R. (2024).** “Is AI a Natural Monopoly?” RAND Corporation Working Paper. ——AI 基础模型市场的自然垄断评估。

[19] **Goldfarb, A. & Tucker, C. (2011).** “Privacy Regulation and Online Advertising.” *Management Science*, 57(1), 57–71. ——隐私规制降低广告效果的早期经典证据。

第十五章　新布兰代斯学派与反垄断复兴

章首语

1978 年，Robert Bork 出版了《反垄断悖论》（*The Antitrust Paradox*），论证反垄断法的唯一正当目标是” 消费者福利最大化”——且仅应以价格和产出衡量。此后四十年，这一命题从学术异端变为司法正统，再变为” 默认设置”——如此根深蒂固，以至于大多数反垄断律师和法官已不再将其视为” 芝加哥学派的立场”，而仅将其视为” 反垄断法的本意”。

2017 年，一个 27 岁的耶鲁法学院学生发表了一篇 93 页的法律评论笔记，标题直接回击 Bork——*Amazon's Antitrust Paradox*。Lina Khan 的核心论证简单而锋利：亚马逊几乎完美地满足了消费者福利标准（低价、丰富选择），同时积累了对整个在线商务基础设施的结构性控制力。如果反垄断法只看价格，它就会对亚马逊式的垄断——**结构性权力不通过高价格行使，而通过控制他人必须依赖的基础设施来行使**——完全失明。

这篇学生笔记引爆了一场全球性的反垄断思想运动。2021 年，Khan 以 32 岁的年龄成为联邦贸易委员会（FTC）史上最年轻的主席。从法学院笔记到国家执法议程——这一轨迹在美国法律史上独一无二。2025 年，Khan 离任；2026 年，新布兰代斯运动的制度遗产正在被激烈辩论。

本章是全书的终点。第 ~1~ 章问了” 市场势力从何而来？”——第 ~2~ 章至 第 ~14~ 章提供了回答这个问题的理论、方法和制度工具。但所有这些工具服务于一个更根本的判断：**我们使用反垄断来追求什么？**芝加哥学派说：消费者福利（价格和产出）。新布兰代斯学派说：竞争过程、经济民主、分散权力。欧盟说：公平竞争与单一市场整合。中国说：社会主义市场经济健康发展——资本必须在党-国定义的边界内运作。

这四种回答不是技术性的——它们不能从一阶条件或计量识别中推导出来。它们是关于” 好社会” 的政治经济学判断。本章通过系统比较这四种回答，将读者从 IO 的技术细节带回其政治经济学根基——因为**反垄断是什么，最终取决于我们认为经济生活应该如何组织**。

15.1　芝加哥学派的”反垄断遗产”

15.1.1　从 Brandeis 到 Bork：反垄断法的范式转换

要理解新布兰代斯运动在反对什么，必须先理解它所反对的对象。芝加哥学派的反垄断思想——特别是消费者福利标准（Consumer Welfare Standard, CWS）和错误成本框架（Error Cost Framework）——从 1970 年代末到 2010 年代主导了美国反垄断法的思维与实践，其影响之深远，以至于大多数当代反垄断律师已不再视其为” 一个学派的立场”，而仅将其视为” 反垄断法的本意”。

布兰代斯时代的反垄断（1914–1970s）。在芝加哥革命之前，美国反垄断思维由哈佛学派的结构-行为-绩效范式（SCP, 第 ~3~ 章–第 ~4~ 章）主导。这一时期的核心关切不是” 消费者支付的价格是否最低”，而是” 经济权力是否过度集中”。Louis Brandeis 大法官（1856–1941）的名言——” 大之诅咒”（The Curse of Bigness）——概括了这一立场：过度的企业规模本身就是一种社会和政治问题，不仅因为大企业可能提高价格，更因为集中的经济权力必然转化为集中的政治权力，威胁民主自治。这一时期的标志性反垄断行动包括 Standard Oil 拆分（1911）、AT&T 拆分（1984）和 *Brown Shoe* 案（1962）中最高法院对合并的严格审查——即使合并后市场份额仅为 5–6%。

芝加哥革命的智识起源。这场革命的智识核心是 Aaron Director（1901–2004），芝加哥大学法学院教授，被 George Stigler 誉为” 海德园的苏格拉底”。Director 本人几乎没有发表任何论文——Stigler 调侃道：”Aaron 的大部分文章都以他同事的名字发表了”——但他的影响力通过与 Edward Levi 合开的反垄断课程（被公认为” 法律与经济学” 运动的发源地）以及培养的一代学生（包括 Robert Bork、Richard Posner、Frank Easterbrook）扩散到了整个法律和经济学界。

芝加哥学派的思想经历了一次戏剧性的自我否定：1930 年代的 Henry Simons（Director 的老师）主张政府积极干预以打破集中经济权力，但 Director 和他的学生们在经验研究中逐渐发现，许多被传统反垄断法谴责的商业实践——转售价格维持（RPM）、搭售（tying）、排他性交易（exclusive dealing）——实际上可能具有效率理据。这一” 皈依过程” 成为芝加哥学派思想力量的最有力广告。

芝加哥学派的影响力通过四条渠道扩散：(1) Levi-Director 反垄断课程培养了几代法官和律师；(2) *Journal of Law & Economics*（1958 年创刊）提供了挑战干预主义假设的学术平台；(3) 里根总统任命的芝加哥学派法官（Posner 进入第七巡回法院、Easterbrook 进入第七巡回法院、Douglas Ginsburg 进入 D.C. 巡回法院）从审判席上应用这些理念；(4) 朝圣山学社（Mont Pelerin Society）将芝加哥学者与全球自由市场知识分子网络连接起来。

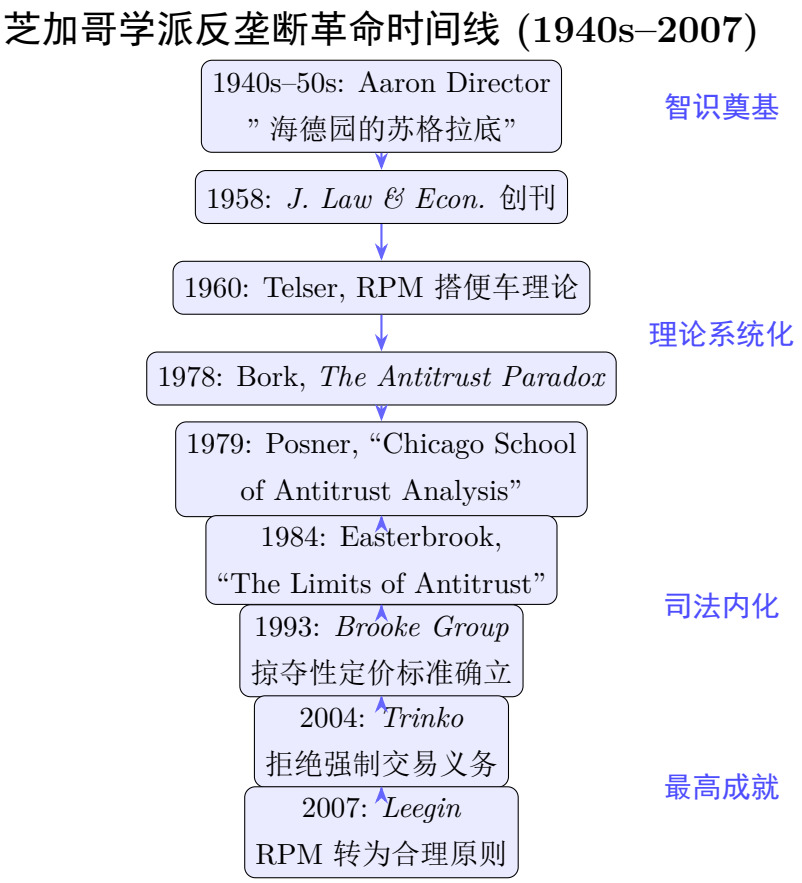


图 15.1: 芝加哥学派反垄断革命的关键节点：从智识奠基（1940s–1958）到理论系统化（1960–1984）到司法内化（1993–2004）到最高成就（2007 *Leegin*）

15.1.2 消费者福利标准（Consumer Welfare Standard, CWS）【核心】

消费者福利标准是芝加哥学派最核心的制度遗产。Bork 在《反垄断悖论》中提出：”**反垄断法的唯一正当目标是消费者福利的最大化。**”（The only legitimate goal of antitrust is the maximization of consumer welfare.）这一命题看似简洁，实则包含了一个深层的术语模糊性——Bork 所说的” 消费者福利”（consumer welfare）在当代经济学中更准确的对应概念是” 总福利”（total welfare）：他关注的是整体效率和产出最大化，而不仅仅是消费者剩余的分配。这一模糊性本身成为后来学术争论的焦点之一。

CWS 排除了什么？ 在 Bork 的框架下，反垄断法明确不应关注以下曾被其前身视为合法目标的考量：

(1) **小企业保护（Small Business Protection）**。传统反垄断法中” 保护小商人免受大企业挤压” 的目标被视为无效率——如果大企业通过效率获得更低成本，让小企业退出市场正是竞争应该产生的结果。

(2) **收入分配（Income Distribution）**。反垄断不是再分配工具——税收和转移支付才是。

(3) **政治权力分散（Political Power Dispersion）**。经济集中是否威胁民主不应由反垄断法处理——这是竞选财务改革和反腐败法律的任务。

(4) **工人福利（Worker Welfare）**。劳动力市场问题由劳动法处理，非反垄断法。

(5) **”竞争过程”本身**。保护”竞争过程”而不看结果（价格和产出），等于保护竞争者而非竞争。

CWS 的三个核心论证。拥护者提出了三个互相补充的论证来支持将反垄断目标限定为消费者福利：

论证一：可测量性与可操作性（*Measurability and Administrability*）。消费者福利——通过价格效应、产出效应和质量效应来衡量——是反垄断唯一可以客观且一致地裁决的标准。多目标反垄断（multi-goal antitrust）将使法院和执法机构面临不可调和的权衡：当保护小企业意味着消费者支付更高价格时，法官应如何裁决？多目标标准使反垄断成为”为所欲为的许可证”。

论证二：立法史（*Legislative History*）。Bork 和 Posner 主张，国会制定《谢尔曼法》（1890）时的唯一意图是促进消费者福利。他们认为立法史中关于保护小企业和分散政治权力的表述仅是修辞（rhetoric），不应作为司法标准。这一论证在 §15.3.1 中将受到详细挑战。

论证三：制度能力（*Institutional Competence*）。反垄断法院和执法机构——由法官和经济学家组成——不具备处理收入分配、政治权力分散等宏大社会问题的制度能力。这些问题应由立法机关（legislature）而非法院（courts）来处理。将反垄断拉伸到超出其制度能力范围，将导致任意执法和法律不确定性的危险组合。

维度	哈佛 SCP 范式 (1950s–1970s)	芝加哥学派 (1980s–1990s)
因果逻辑	Structure → Conduct → Performance（前向因果）	Performance → Conduct → Structure（反向因果：效率决定结构）
对集中的看法	推定有害；集中 ≈ 合谋	反映效率；大不等于坏
反垄断目标	多重目标（政治、社会、消费者）	仅消费者福利/效率
政府角色	积极干预；结构性救济（拆分）	最小化；市场自我纠正
分析方法	结构推定（本身违法优先）	效果导向的合理原则
进入壁垒	广泛：规模经济、资本要求均构成壁垒	窄：只有政府制造的壁垒才是真正的壁垒
纵向限制	推定反竞争	推定有效率（搭便车解决方案）
主导时期	1950s–1970s（Warren Court 时代）	1980s–2000s（Reagan 至 Roberts Court 早期）

表 15.1: 芝加哥学派 vs. 哈佛 SCP 范式：两种反垄断哲学的系统比较

15.1.3 错误成本框架（Error Cost Framework）【核心】

Frank Easterbrook 于 1984 年在《德克萨斯法律评论》发表的《反垄断的限度》（"The Limits of Antitrust," 63 *Texas Law Review* 1–40）引入了反垄断法中最具影响力的分析框架之一——错误成本框架。该框架的核心非对称性假设是：**在反垄断语境下，Type I 错误（假阳性——将促进竞争的行为定罪）比 Type II 错误（假阴性——放过了反竞争的行为）成本更高。**

Easterbrook 的推理链条由三个命题构成：

命题一：Type I 错误是持久的且具有传染性。一个错误的定罪判决成为先例（stare decisis），使同一行为在整个经济体中变成违法。被禁止的效率行为的收益永久丧失。任何其他使用该行为的企业——无论其对消费者的益处如何——都面临制裁风险。"If the court errs by condemning a beneficial practice, the benefits may be lost for good. Any other firm that uses the condemned practice faces sanctions in the name of stare decisis, no matter the benefits."

命题二：Type II 错误是自我纠正的。卡特尔是囚徒困境，倾向于自行崩溃。垄断利润吸引新进入者，竞争会侵蚀市场力量。市场力量随时间递减——"the welfare loss decreases over time"。"If the court errs by permitting a deleterious practice, though, the welfare loss decreases over time."

命题三：法院结构上易犯 Type I 错误。被告难以解释通过试错演化出来的商业实践的促进竞争理据；法官和陪审团难以理解复杂的经济理论和经验证据；企业尝试

数十种实践（大多数失败），使得区分良性试验和反竞争行为变得困难。

Easterbrook 框架的制度含义是强烈的**司法谦抑（judicial restraint）**：(1) 从本身违法（per se rules）转向合理原则（rule of reason）；(2) 法院应使用” 过滤器”（filters）在全面审理前驳回弱势案件；(3) 举证责任应由原告承担且设定在较高水平。Wright & Mungan (2021) 将 Easterbrook 的核心假设形式化为”Easterbrook 定理”：如果 Type I 确实成本更高，最优证明标准应**高于**优势证据标准（preponderance of the evidence）。

最高法院的采纳。错误成本框架已被美国最高法院深度嵌入反垄断法理，在四个关键案件中发挥了决定性作用：*Weyerhaeuser* (2007, 垄断化案件中的审慎标准)、*Leegin* (2007, 推翻 RPM 本身违法规则)、*Trinko* (2004, 拒绝强制交易义务) 和 *LinkLine* (2009, 拒绝价格挤压主张)。

教义领域	芝加哥学派的论证	关键论文/理论	最高法院采纳
掠夺性定价	经济上非理性；回收不可能	McGee (1958); Easterbrook (1981)	*Brooke Group* (1993)：两阶段测试
RPM（转售价格维持）	解决搭便车问题；促进品牌间竞争	Telser (1960)	*Leegin* (2007)：推翻 96 年本身违法先例
搭售（Tying）	不能杠杆化垄断（单一垄断利润定理）	Posner (1979)	*Jefferson Parish* (1984)：收紧搭售标准
纵向一体化	由效率驱动；消除双重边际化	Posner (1976)	*GTE Sylvania* (1977)：非价格纵向限制转为合理原则
合并效率抗辩	真实效率应获” 完全信用”	Bork (1978); Williamson (1968)	1982/1984 合并指南纳入效率考量
市场定义	是分析工具而非分析目的	Stigler (1982)	2010《横向合并指南》弱化市场定义的门槛角色

表 15.2: 芝加哥学派的关键教义及其在最高法院的采纳

15.1.4　芝加哥学派的盲点与批评

芝加哥学派的霸权地位从 1990 年代开始受到来自内部和外部的系统挑战。

Post-Chicago 学派的博弈论反驳。后芝加哥学派（以 Herbert Hovenkamp 为桥梁人物，其 1985 年《芝加哥之后的反垄断政策》超过 70 页）使用博弈论和信息经济学来证明，在更复杂的假设下，反竞争结果比芝加哥学派承认的更为普遍。三个核心理论贡献：(1) **提高对手成本**（Raising Rivals' Costs, Krattenmaker & Salop 1986）——企业可以通过操纵投入市场而非仅是价格来排除竞争；(2) **纵向排他**（Ordover, Saloner & Salop 1990）——纵向并购在一定条件下可以产生排他效应，芝加哥的" 单一垄断利润" 定理并非普遍成立；(3) **策略性进入威慑**——在位者可以通过过度投资产能、捆绑等策略可信地威慑进入，进入并不总是如芝加哥学派假设的那样" 自由"。

Shapiro (2018) 的内部批评。Carl Shapiro 在《民粹时代的反垄断》（*IJIO*, 2018）中代表了主流经济学内部的改革立场：他辩护消费者福利标准作为最理性框架，承认过去 20–30 年许多行业集中度适度上升、企业利润占 GDP 份额大幅上升，主张在现有框架内" 适度加强" 合并执法，并呼吁对占主导地位的企业（包括科技企业）进行密切反垄断审查。Shapiro 的立场可概括为" 一个辩护、三个行动建议"——保留 CWS，但更积极使用已有的结构推定。

历史准确性批评。Bork 的立法史论证——" 国会制定《谢尔曼法》时的唯一意图是促进消费者福利"——受到法律史家的系统挑战。Thomas Hazlett (1992) 对国会辩论的重新审查表明，支持《谢尔曼法》的政治联盟" 追求多重社会目标，并未忠实追求经济效率"。Robert Lande 指出 Bork 的消费者福利解读是对立法史的" 发明"。Barak Orbach 论证 Bork 的" 消费者福利" 概念本身存在内部矛盾——他实际主张的是总福利标准，而非消费者剩余保护。这些历史批评将在 §15.3.1 中进一步展开。

芝加哥学派的持久遗产

即使在 2020 年代的" 反垄断复兴" 中，芝加哥学派的四项遗产仍然是任何严肃反垄断讨论的起点：(1) 经济推理的必要性——反垄断不能回归纯粹的法律形式主义；(2) 对错误成本的关注——任何执法改革必须考虑其误判成本，即便非对称性假设被调整；(3) " 保护竞争，非保护竞争者"——这一区分即使在新布兰代斯框架内也不会被完全放弃；(4) 效果导向分析的框架——即便新布兰代斯想恢复某些结构推定，也必须论证这些推定更好地服务于效果评估。

15.2　新布兰代斯运动的兴起

15.2.1　布兰代斯的精神源头：为什么" 大" 本身就是问题？

Louis D. Brandeis（1856–1941），美国最高法院大法官（1916–1939 年在任），被誉为" 人民的律师"（People's Attorney）。他的反垄断思想——被芝加哥革命埋葬后又在新布兰代斯运动中被重新发掘——是理解当前辩论不可或缺的智识来源。

Brandeis 的核心信条可以概括为三项相互关联的命题。第一，" **大之诅咒**"（**The**

Curse of Bigness）：过度的企业规模本身就是一种社会和政治问题，不仅仅是经济问题。他写道：" 竞争与大规模生产和分配根本不相容。"（Competition is in no sense consistent with large-scale production and distribution.）集中的经济权力必然转化为集中的政治权力，威胁民主自治——这一洞见直接连接了反垄断法与民主理论。

第二，**垄断不等于效率**。Brandeis 反对" 托拉斯是规模经济所必需" 的论点——企业可能" 大到无法有效率"（too big to be efficient），这预示了后来的有限理性（bounded rationality）和 X-无效率（X-inefficiency）概念。他的论证不是反资本主义的——他信仰竞争性资本主义，但认为垄断是人为消灭竞争的产物，而非市场竞争的自然结果。

第三，**功能分离（Separation of Functions）**。金融家不应同时控制银行和工业企业——" 阳光是最好的消毒剂"（Sunlight is the best disinfectant）和透明度的强调——这一原则后来被称为结构性分离的先驱，并直接影响了 1933 年《格拉斯-斯蒂格尔法》（Glass-Steagall Act）将商业银行与投资银行分离的制度设计。

从 Brandeis 到 Bork 再到 Khan：一个历史的钟摆。Brandeis 哲学曾主导美国反垄断思想超过半个世纪（1914–1970s），后来被称为反垄断的" 哈佛学派"。Bork (1978) 用消费者福利标准取代了这一传统。四十年后，Khan (2017) 的文章标题——*Amazon's Antitrust Paradox*——是对 Bork 的直接回击，将 Brandeis 的关切重新引入 21 世纪话语。这不是一个简单的循环——每一次回归都带着新的分析工具和新的经验证据。

15.2.2　催化时刻：Khan (2017) 与 Wu (2018)

Khan (2017)：《亚马逊的反垄断悖论》。2017 年 1 月，时年 27 岁的耶鲁法学院三年级学生 Lina M. Khan 在《耶鲁法律杂志》发表了约 93 页的学生笔记——*Amazon's Antitrust Paradox*（126 *Yale Law Journal* 710–805），成为法律评论历史上阅读量最高的文章之一。

Khan 的核心论证是一个精心构造的双重悖论。**第一层悖论（回击 Bork）：**正是 Bork 将反垄断缩减为消费者福利，才使亚马逊这样的平台垄断免受审查——" 仿佛贝索斯先绘制了反垄断法的地图，然后规划了巧妙绕行的路线"（It is as if Bezos charted the company's growth by first drawing a map of antitrust laws, and then devising routes to smoothly bypass them.）。**第二层悖论（亚马逊本身）：**亚马逊以消费者福利标准的" 满分学生" 面貌出现——无休止地降低价格、丰富选择——同时通过四个反竞争维度积累了对在线商务基础设施的结构性控制力：

(1) **低于成本的定价（Below-cost Pricing）**。以电子书为例——亚马逊愿意为增长放弃利润，辅以耐心资本市场的补贴，使得掠夺性定价高度理性，与芝加哥学派的假设完全相反。

(2) **收购竞争对手（Acquisition of Competitors）**。在价格战重创 Quidsi/Diapers.com 后将其收购——利用了芝加哥学派关于市场自我纠正的错误假设。

(3) **跨领域杠杆化支配地位（Leveraging Dominance Across Sectors）**。如亚马逊物流/配送——在一个市场中的权力被用来征服相邻市场。

(4) **利用第三方卖家数据（Data Exploitation）**。使用来自 Marketplace 第三方卖家的销售数据识别热门产品，然后推出竞争的亚马逊自有品牌——平台同时担任市场运营者和参与者的经典利益冲突。

Khan 提出了两种改革方案：(1) 恢复传统结构主义反垄断——认真对待掠夺性定价和垂直整合，考虑市场结构而非仅是价格/产出；(2) 公共承运人/公用事业监管——将主导平台视为关键基础设施，施加非歧视义务。

Wu (2018)：《大之诅咒》。Tim Wu（哥伦比亚大学法学院教授，" 网络中立性" 概念的提出者）同年出版了 154 页的宣言式著作——*The Curse of Bigness: Antitrust in the New Gilded Age*（Columbia Global Reports）。Wu 的核心历史论证是：我们经历过这一切。镀金时代（Rockefeller, Morgan, Carnegie）的极端企业集中和财富不平等催生了进步时代的反垄断回应（Sherman Act, Clayton Act, FTC Act）。但芝加哥学派在 1970–80 年代系统地拆除了这些保护，导致了" 新镀金时代"。Wu 警告：容忍极端企业集中和不平等在历史上催生了民粹主义、民族主义甚至法西斯主义——" 如果我们不通过法律来分散经济权力，人们最终会通过更危险的方式来寻求对经济权力的控制。"

15.2.3　新布兰代斯学派的六大核心纲领

新布兰代斯运动由多位学者、律师和活动家组成，并非铁板一块——从激进结构主义者到温和改革派。但其核心共识可以归纳为六大纲领：

纲领 1： **拒绝消费者福利作为反垄断的唯一标准。**CWS——以短期价格效应衡量——是不充分的、在经验上已失败的。它不能涵盖：(a) 创新损害（扼杀式收购——在位者收购初创企业以消灭未来竞争，详见 第 ~14~ 章）；(b) 隐私损害（平台以数据作为" 免费" 服务的价格，但 CWS 只看到零价格）；(c) 劳动力市场损害（买方垄断势力压低工资——Azar, Marinescu & Steinbaum 2018）；(d) 质量退化（"enshittification"——平台随着竞争减少而恶化）；(e) 政治权力——集中的经济权力对民主治理的腐蚀效应。替代方案：一个更广泛的" 竞争标准"（competition standard），保护竞争过程（competitive process）和市场结构（market structure），而不仅仅是消费者价格。

纲领 2： **扩展反垄断受保护的利益。**反垄断应回到其最初的广泛使命——保护竞争过程、小企业和创业者的竞争机会、创新（防止现有企业通过收购和复制

扼杀新进入者）、政治民主（防止经济权力转化为政治权力）、工人（买方垄断势力、非竞争条款）、地方社区（防止兼并对地方经济的侵蚀）。

纲领 3： **错误成本反转：在集中市场中，Type II 错误（假阴性/执法不足）的成本远高于 Type I 错误（假阳性/过度执法）。**这是对 Easterbrook (1984) 核心非对称性假设的直接反转。论证：(a) 集中市场**不会**自我纠正——网络效应、数据飞轮和高转换成本使市场势力具有黏性（Google 搜索、Facebook 社交网络、Amazon 电商的主导地位已持续 15–20 年）；(b) 四十年执法不足的累积效应已经成为系统性的经济结构转型——集中度上升、创业率下降、加价率上升、工资停滞——这些不是自我纠正的现象；(c) 假阴性的不可逆性：错误批准的合并无法" 解蛋"（unscrambling the egg, Kwoka & Valletti 2020）。

纲领 4： **结构性救济（拆分、业务线分离）优先于行为性救济（行为规制）。**行为性救济在平台市场的记录是灾难性的——Microsoft (2001) 互操作性披露未能恢复浏览器竞争，EU Google Android (2018) 选择屏幕对市场份额几乎无影响，FTC Google (2011) 搜索偏见同意令被广泛认为无效，FTC Facebook (2019) $50 亿美元和解未改变任何结构。EU 委员会 2025 年发布的 20 年救济措施回顾研究证实：结构性救济" 始终实施且普遍有效"，而" 三分之二纯粹行为性救济部分或完全无效"。

纲领 5： **大型平台应被视为" 关键设施"（Essential Facilities）或" 公共承运人"（Common Carriers），承担非歧视接入义务。**主导数字平台——Amazon Marketplace、Google Search、Apple App Store、AWS 云基础设施——类似于 19 世纪的铁路和 20 世纪的电话网络。它们已经成为其他企业必须依赖的基础设施（"bottleneck"），平台所有者同时作为市场参与者的双重角色产生了无法抗拒的利益冲突。这一框架植根于 1912 年的关键设施原则（*United States v. Terminal Railroad Association*）。

纲领 6： **举证责任转移：一旦原告证明被告拥有市场势力，被告应证明其行为是促进竞争的。**现行框架要求原告（通常是政府）证明被告行为具有反竞争效果——这是一项沉重且昂贵的举证责任。新布兰代斯学派主张：一旦证明市场势力，应将举证责任转移给被告——由被告证明其行为更可能是促进竞争而非反竞争。这一主张的经济学逻辑是信息不对称：被告知道自己的成本、效率和策略动机，原告必须从可观察的市场结果中推断——而市场结果是模糊的。

理论坦诚：替代标准的构建仍在进行中。新布兰代斯阵营内部承认的一个理论弱点是：" 否定的共识" 已经形成（CWS 不行），但" 肯定的共识"（替代方案是什么）尚在构建阶段。Steinbaum & Stucke (2020) 在《芝加哥大学法律评论》发表的" 有效

竞争标准”（The Effective Competition Standard）是最系统的替代方案尝试——它主张将反垄断的目标恢复为” 保护和促进整个经济体系中的有效竞争，包括供应链和劳动力市场”。但该标准在可操作性上仍受到质疑：如何界定” 有效竞争”？法院如何在实际案件中以可预测的方式适用？

图 15.2: 新布兰代斯运动的关键事件时间线：从 2017 年 Khan 的学生笔记到 2026 年 Amazon 审判

维度	芝加哥学派	新布兰代斯学派
反垄断的目标	消费者福利（价格/产出/效率）——单一目标	竞争 + 民主 + 多元主义 + 劳动者 + 小企业 + 地方社区——多重目标
对集中的看法	通常是效率的结果；市场会自我纠正	推定具有危险性；集中市场不会自我纠正
分析方法	效果导向、一案一议的经济分析	明确规则（bright-line rules）、结构推定
执法姿态	节制/谦抑（偏好执法不足）	能动主义（执法不足是更大的风险）
对错误的偏好	避免 Type I 错误（假阳性/过度执法）	避免 Type II 错误（假阴性/执法不足），尤其在集中市场
救济偏好	行为性救济（行为规制）	结构性救济（拆分/剥离）优先
经济推理	微观经济学价格理论	结构主义 + 制度经济学 + 政治经济学
对垂直整合	推定有效率（消除双重边际化）	推定怀疑（尤其在平台语境中的利益冲突）
举证责任	原告证明反竞争效果	被告证明促进竞争效果（一旦市场势力被证明）
范围	狭窄（经济效率）	广泛（经济 + 政治 + 社会 + 劳动）

表 15.3: 芝加哥学派 vs. 新布兰代斯学派：十个核心维度的系统对比

15.2.4 从象牙塔到制度权力：2021–2025 的制度实验

”三头政治”（The Troika）。2021 年，拜登总统将新布兰代斯运动的三位领军人物同时安插入美国反垄断的三大关键岗位：Lina Khan 于 2021 年 6 月 15 日被确认为 FTC 主席（32 岁，史上最年轻），Tim Wu 于 2021 年 3 月 5 日被任命为白宫国家经济委员会技术与竞争政策特别助理（无需参议院确认），Jonathan Kanter 于 2021 年 11 月 16 日被确认为 DOJ 反垄断局助理检察长（68–29 票）。三人均为自我标识的新布兰代斯主义者，据报均由参议员 Elizabeth Warren 力荐。

EO 14036 (2021.7.9)。拜登签署《促进美国经济中的竞争》行政令——一项”全政府”（whole-of-government）努力，涵盖 72 项具体举措跨越十余个联邦机构，明确包含新布兰代斯语言：谴责” 过度兼并”，呼吁重新审视已完成的兼并，将劳动力市场纳入竞争政策。该行政令由 Wu 和 Bharat Ramamurti 起草，被广泛描述为” 如果

布兰代斯主义的' 大之诅咒' 是一个政策平台，那么这就是它本身。"2025 年 8 月被特朗普行政令废除。

重大案件记分卡。2020–2026 年间，美国政府对主要科技平台发起了自 1998 年 Microsoft 案以来最激进的反垄断诉讼浪潮：

案件	机构	提起	指控	结果 (截至 2026.5)
DOJ v. Google (Search)	DOJ	2020.10	非法维持通用搜索垄断 (Sherman Act §2)	[胜诉] 2024.8 胜诉；[部分] 2025 补救裁定拒绝拆分 Chrome，施加行为性救济
DOJ v. Google (Ad Tech)	DOJ	2023.1	非法垄断数字广告技术栈	[胜诉] 2025 裁定违法；[部分] 补救裁定令政府失望；上诉中
FTC v. Meta (Instagram/WhatsApp)	FTC	2020.12 首次；2021.8 修改	收购 Instagram 和 WhatsApp 非法维持个人社交网络垄断	[败诉] 2025.11.18 败诉：法院裁定 FTC 未能证明 Meta 在当下拥有垄断权力
FTC v. Amazon	FTC + 17 州	2023.9.26	非法维持" 在线超市" 和" 在线市场服务" 垄断	[审理中] 审判定于 2026.10
DOJ v. Apple	DOJ + 16 州	2024.3	非法维持" 性能智能手机" 市场垄断	[审理中] 驳回动议被拒，案件进行中
DOJ/州 v. Live Nation	DOJ + 30 州	2024.5	非法维持演唱会推广和票务垄断	[审理中] 进行中
FTC v. Kroger/Albertsons	FTC	2024	超市合并	[胜诉] FTC 在法院胜诉，合并被阻止
FTC v. Tapestry/Capri	FTC	2024	奢侈品合并	[胜诉] FTC 在法院胜诉
FTC v. PepsiCo	FTC	2024	价格歧视（Robinson-Patman Act）	[败诉] 2025 被新任 FTC 主席 Ferguson 撤销

表 15.4: 2020–2026 年美国重大反垄断案件结果：新布兰代斯时代的司法检验

FTC v. Meta 败诉的深入分析。2025 年 11 月 18 日，James E. Boasberg 法官裁定 FTC 未能证明 Meta 在当下拥有垄断权力——这是新布兰代斯运动最惨痛的司法失败。五个关键教训：

(1) **市场定义决定案件命运**。FTC 将相关市场定义为" 个人社交网络"（Personal Social Networking）——仅包括 Facebook、Instagram、Snapchat 和一个平台 MeWe。Boasberg 裁定 TikTok 和 YouTube 显然属于同一市场。到 2025 年，Facebook 用户仅 17% 的时间花在" 熟人内容" 上，Instagram 仅为 7%——绝大多数时间在观看 AI 推荐的短视频。(2) **静态理论在动态市场中不堪一击**。法官以赫拉克利特的名言开篇——" 人不能两次踏入同一条河"——强调市场已完全转型。(3) **前沿专家证据可能适得其反**。法院对 FTC 专家经济学家 NYU Scott Hemphill 表达了对其" 身兼举报人与专家" 双重角色的怀疑。(4) **FTC 内部逻辑矛盾被尖锐揭露**。FTC 一方面主张 Meta 在" 熟人内容" 上投资不足（证明行使垄断权），另一方面主张 Meta 维持了对该功能的垄断——被法官称为" 毫无道理"。(5) **" 购买或埋葬" 理论受严重打击**。经过监管审查通过的旧收购，在十年后市场已完全改变的情况下，极难被追溯推翻。

维度	评估	说明
学术议程	[成功]	消费者福利标准不再不可挑战；结构主义重新进入可敬的学术话语
人事任命	[历史性突破]	Khan、Kanter、Wu 三人同时掌控三大关键反垄断岗位
立法影响	[有限]	国会未通过重大反垄断立法（Klobuchar-Grassley 等技术反垄断法案未成法）
司法胜诉	[混合]	Google Search/Ad Tech 胜诉但补救有限；Meta 完全败诉；合并案有胜有负
制度变革	[部分逆转]	2023 合并指南保留，但执行力度减弱；非竞争条款禁令被撤销；HSR 规则被推翻
结构性救济	[未实现]	没有一家大型科技公司被拆分
行政令	[已废除]	EO 14036 被特朗普废除
持久影响	[评估中]	私人反垄断诉讼继续扩散；部分转移到州层面（加州 Cartwright Act 修改）；" 奥弗顿窗口" 被移动

表 15.5: 新布兰代斯运动的制度遗产评估（截至 2026 年 5 月）

15.3　核心争论点

反垄断改革不仅是一个政策问题——它是一个深刻的智识争论，涉及经济学、法律、历史和政治理论的交叉。本节系统梳理芝加哥学派和新布兰代斯学派之间的五个核心争论。每个争论都同时呈现双方的最强版本——因为理解" 为什么理性的、聪明的人会不同意" 比知道" 谁是对的" 更具智识价值。

15.3.1　争论一：反垄断的目标是什么？

这是所有其他争论的根源。如果反垄断的目标不能达成共识，那么关于错误成本、救济方式和举证责任的争论就无法获得原则性的解决。

芝加哥/CWS 立场：单一目标是制度的必要条件。支持者（Bork, Easterbrook, Ginsburg, Wright, Melamed, Carlton）的核心论证是制度性的而非实质性的：即便人们认为小企业保护、收入再分配和政治权力分散是值得追求的社会目标，反垄断法也**不是**追求这些目标的合适工具。原因有三：(1) 多目标标准意味着不可裁决的权衡——当保护小企业要求消费者支付更高价格时，法官没有原则性的方式来裁决；(2) 多目标标准使反垄断成为行政和司法自由裁量权的" 罗夏测试"——不同党派将各自偏好的社会目标投射其上（右派针对" 觉醒资本"，左派针对大型科技公司）；(3) 没有更好的替代方案——新布兰代斯提议的" 竞争过程保护" 或" 有效竞争标准" 一旦被操作化，往往退化到消费者福利或变得无法操作。

Shapiro (2019) 的警告精准地概括了这一立场：要求 FTC 或 DOJ 基于政治权力考虑来评估合并将" 为腐败打开方便之门"（would invite corruption），并破坏反垄断执法的法治基础。

新布兰代斯/多目标立场：CWS 从来不是谢尔曼法的原始意图。批评者（Khan, Wu, Steinbaum & Stucke, Vaheesan）提出两个层面的反驳。**经验层面：**CWS 在经验上已经失败——它系统性地允许了没有立即可见价格效应的反竞争行为（扼杀式收购、数据提取、隐私退化、劳动市场集中）。**历史层面：**Bork 的立法史解读是选择性的历史修正。Senator John Sherman 本人的原话清楚地表明了多重关切：" 如果我们不能容忍一个国王作为政治权力，我们也不应容忍一个国王来控制生活必需品的生产、运输和销售"（21 *Cong. Rec.* 2456）。Hazlett (1992) 对国会辩论的经济计量重新审查和 Grandy (1993) 对" 消费者福利假说" 的再检验都支持了这一多重目标的解读。

Justice William O. Douglas 在 *United States v. Columbia Steel Co.* (1948) 中的反对意见表达了反集中立场的最有力司法表述：" 谢尔曼法的哲学是——集中的工业权力不应存在。工业权力应该被分散到许多人手中，使人民的命运不依赖于少数几个自我任命的人的偏见、政治偏好和情绪稳定。这就是谢尔曼法的哲学和命令。它建立在一种敌视私人手中集中的权力如此之大以至于只有人民的政府才应拥有它的理论之上。"

中间立场。Baker (2019) 代表了一个微妙的中间位置：保留 CWS 作为操作标准

但将其解释得更广泛——包括创新、质量、隐私和劳动市场效应——并显著加强执法力度。Hovenkamp 指出 CWS 本身就是再分配性的——它偏向消费者而非生产者——因此不是中性的效率标准。两者的共同点是：问题不在于 CWS 本身，而在于它被诠释和执行得过于狭窄和胆怯。

15.3.2 争论二：谁的错误更昂贵？

芝加哥立场（Type I > Type II）已在 §15.1.3 中详细说明。其核心非对称性假设——假阳性通过先例扩散至全经济且不可逆，假阴性被市场自我纠正——在传统制造业和服务业市场中具有一定的经验基础。

Baker (2015) 的系统拆解。Jonathan Baker 的"Taking the Error Out of 'Error Cost' Analysis"（80 *Antitrust Law Journal* 1, 2015）是对芝加哥错误成本框架最全面的学术回应。Baker 论证芝加哥学派的错误成本分析依赖于五个系统性地错误的经验假设：

芝加哥假设	为何此假设关键	经验现实
1. 市场通过进入自我纠正	如果市场自我纠正，Type II 错误是暂时的	进入壁垒（网络效应、数据优势、转换成本）使市场势力持续数十年。Google、Facebook、Amazon 的主导地位已持续 15–20+ 年
2. 寡头竞争；卡特尔不稳定	如果卡特尔自行崩溃，不需要执法	许多卡特尔长期存在（赖氨酸、维他命、LCD 面板）。算法合谋（第 ~14~ 章）比人类卡特尔更稳定
3. 垄断者创新最有效	如果垄断促进创新，不应拆分	竞争通常最大化创新（Arrow 1962 替代效应）。垄断者通过扼杀式收购消灭未来竞争
4. 单一垄断利润定理成立	如果搭配销售不能杠杆化垄断，纵向限制是有效的	仅在狭窄、非现实条件下成立。平台捆绑/搭售可将垄断延伸至相邻市场
5. 竞争市场中常见的实践不会损害竞争	如果常见则可能有效率	相同的策略（排他性交易、捆绑）在被具有市场势力的企业使用时可以产生排除效应

表 15.6: Baker (2015)：芝加哥学派错误成本框架依赖的五个系统性错误的经验假设

Baker 的底线结论：反垄断执法的收益至少是成本的 50 倍。芝加哥错误成本框架系统性地高估了 Type I 错误成本并低估了 Type II 错误成本，产生了四十年的系统性执法不足，其累积效应现在可见于美国经济的集中度上升和创业活力下降。

新布兰代斯对错误成本的论证更进一步：(1) 数字市场不会自我修正——网络效应 + 数据飞轮 + 高转换成本 → 市场势力黏性强；(2) 假阴性的不可逆性：错误批准的合并无法” 解蛋”（Kwoka & Valletti 2020）；(3) 四十年执法不足的累积效应是永久性的——集中度上升、加价率上升、劳动份额下降、创业率下降——这不是自我纠正的现象。

核心实证难题。整个错误成本争论面临一个根本性的实证空白：没有公认的方法来测量两种错误的相对成本。我们观察不到反事实——不知道哪些被阻止的合并会产生效率，也不知道哪些被批准的合并会产生市场势力。这使得该争论在可预见的未来将仍然主要是理论性和意识形态性的。

15.3.3　争论三：结构性救济还是行为性救济？

结构性救济和行为性救济的区别、各自的利弊已在前文讨论（纲领 4）。本小节聚焦于最关键的**经验证据**和**混合方案**。

EU 委员会 2025 年救济措施研究（20 年回顾）提供了迄今为止最全面的经验证据。该研究审查了 EU 反垄断执法 20 年的救济措施，发现：(1) 结构性救济始终实施且普遍有效；(2) 三分之二纯粹行为性救济部分或完全无效；(3) 带有结构元素的行为性救济效果混合。研究建议废除 EU 法律中使结构性救济从属于行为性救济的层级制度——目前结构性剥离仅是” 最后手段”。

四种混合方案。日益形成的共识——即使在偏好结构性救济的执法者中——承认纯结构性方案存在实际限制。(1) **结构性分离 + 行为性保障**——如 Schinkel & van Oosten (2025) 的”Smart Cuts” 方案，将推荐系统与商业服务分离但不拆分整个企业。(2) **行为性规则 + 结构性救济备用**——UK DMCC 法案的” 升级阶梯”（escalation ladder），对具有” 战略市场地位”（SMS）的企业施加行为性义务，但保留结构性分离权力作为后备。(3) **德国” 第四支柱”**——GWB 第 11 修正案拟议 §32f，授权联邦卡特尔局在行业调查后施加结构性救济，**即使未发现具体违法行为**。(4) **DMA 式事前规则**——通过事前指定义务（互操作性、数据可携带性、禁止自我优待）短路径绕过救济争论，使其成为默认规则而非逐案谈判的救济。

维度	结构性救济	行为性救济
机制	拆分企业或资产，消除利益冲突的根源	施加行为规则（非歧视、互操作性、数据访问）
持续性	永久且自我执行；一旦剥离完成无需持续监控	需要持续监控和执行；监管者需要永久人员配置
可规避性	无可规避（结构已改变）	高度可规避——”打地鼠”问题（Franck & Peitz 2024）；当一种实践被禁止时，稍作修改的替代品出现
效率保全	可能损失整合效率（消除双重边际化、降低交易成本）	保留整合效率
精确性	可能过于宽泛——伤害无辜的业务线	更有针对性——可针对特定行为
证据记录 (EU 2025)	始终实施且普遍有效	三分之二纯粹行为性救济部分或完全无效
可行性	政治和司法阻力大；”划线”问题——在哪割？	政治和司法阻力较小；更容易获得法院批准
适用场景	存在内生利益冲突的平台（同时是市场运营者和参与者）	当整合效率显著且行为可被清晰指定和监控时

表 15.7: 结构性救济 vs. 行为性救济：八个维度的系统比较

15.3.4 争论四：谁应该承担举证责任？

现行框架：传统三步举证责任转移。在合理原则（Rule of Reason）下：(1) 原告（或政府）必须首先证明反竞争效果——通常通过定义相关市场和显示受挑战行为造成了实际或可能的竞争损害；(2) 如果原告满足了表面证据负担，举证责任转移给被告——由被告提出促进竞争的正当理由或效率抗辩；(3) 举证责任再次转移给原告——原告可以证明声称的效率可以通过更少反竞争的方式合理实现（”更少限制的替代方案”测试）。

Hovenkamp 的批判：举证责任实际上从未转移。Hovenkamp 有力地论证了现行合理原则”过于偏袒被告”，因为最高法院”将所有证明要求都装进了案件的初始部分”。”结果是，举证责任从未转移到被告。这使得合理原则案件非常非常难以证

明。原告很少赢得这些案件。” 这实际上是使合理原则” 几乎无用”，驱使原告寻求本身违法处理——这产生其自身的扭曲，因为本身违法可能谴责在合理原则下会存活的行为。

改革提案：Klobuchar 法案 (2021)。《竞争和反垄断法执法改革法案》提议在以下情况下将举证责任转移给被告：(1) 市场份额超过 50% 的公司——被告必须证明交易不会损害竞争；(2) 价值超过 $50 亿的” 超级合并”——被告证明无损害；(3) 年收入或市值超过 $1000 亿的公司进行超过 $5000 万的交易——被告证明无损害。经济学逻辑是信息不对称：被告知道自己的成本、效率和策略动机，而原告必须从可观察的市场结果中推断。标准的证据经济学理论（Hay & Spier 1997）表明举证责任应落在具有更优信息获取能力的当事方。

Easterbrook 反论。一篇 2022 年的《国际法律与经济学评论》论文形式化了此权衡：如果 Type I 错误（假阳性）确实比 Type II 错误（假阴性）成本更高（Easterbrook 的前提），则最优证明标准应**高于**优势证据标准——框架应保持偏向被告——即使被告具有更好的信息。这再次表明，整个举证责任争论的解决取决于错误成本争论的解决。

15.3.5　争论五：经济学在反垄断中应扮演什么角色？

技术官僚观（芝加哥）。反垄断应是一项专家事业——经济学提供客观工具来评估竞争效果。从本身违法规则到合理原则的转变 = 从法律形式主义到经济现实。法院应适用严格的经济分析，而非关于” 大” 的” 民粹主义” 直觉。

民主观（新布兰代斯）。反垄断根本上是关于如何组织经济生活的**政治问题**——它不能被还原为经济技术。消费者福利标准假装是中立的技术标准，但嵌入了价值判断（只有消费者重要，效率压倒一切，结构无关紧要）。国会——而非法院或经济学家——应做出根本性的价值选择。反垄断是一种民主制度，而非仅是效率最大化工具。

” 实证空白” 问题。整个争论的一个突出特征是缺乏决定性经验证据：我们没有公认的方法来测量 Type I vs. Type II 错误的相对成本、确定集中度上升是由效率还是进入壁垒驱动的、理解反垄断执法的威慑效应（多少反竞争行为被执法威胁所阻止 versus 多少促进竞争行为被假阳性恐惧所寒蝉化）。这些实证空白意味着争论在可预见的未来将保持主要是理论性和意识形态性的——双方都能（并且确实）引用符合各自立场的证据，但双方都不能声称具有决定性的经验支持。

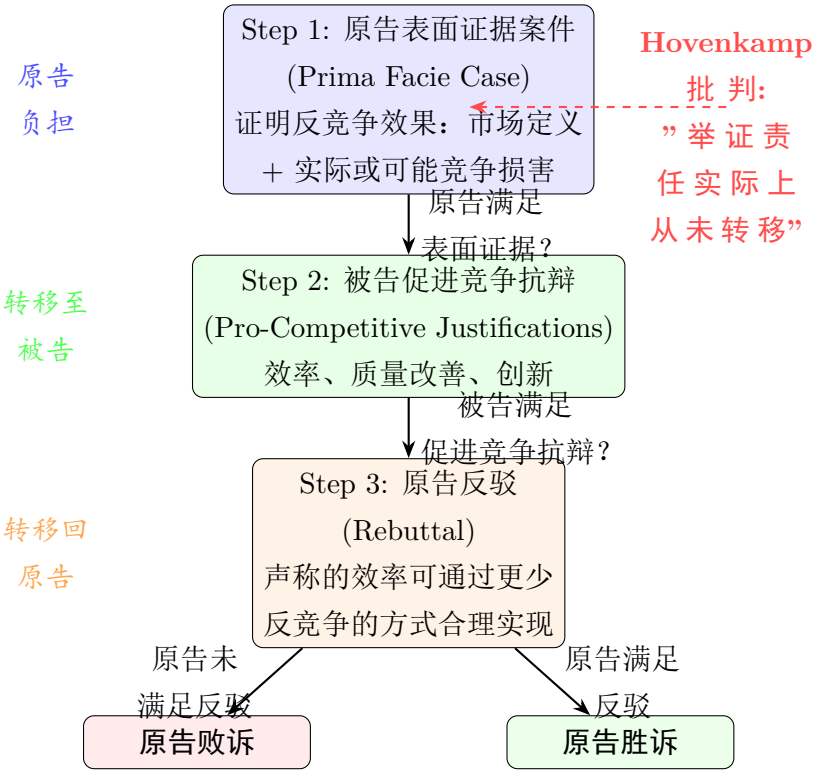


图 15.3: Rule of Reason 举证责任转移的三步框架与 Hovenkamp 批判

15.4　制度现实与学术回应

15.4.1　法院作为制度壁垒

新布兰代斯运动从学术批评转化为制度执行的过程中，遇到的最强大的制度壁垒不是政治反对——而是**法院**。美国反垄断法由法院塑造（普通法传统），而联邦法官——终身任职、不可选举、政治光谱多样——对经济学范式转换的吸收极其缓慢。

芝加哥学派从学术共识到司法内化用了 20–30 年（Posner 1979 论文 → *Trinko* 2004）。相比之下，新布兰代斯运动从 Khan 2017 到今天仅 9 年。法院在几个关键案件中展现了对新布兰代斯理论的深层抵制：

- **FTC v. Meta (2025)**：法院拒绝将”个人社交网络”窄化为排除 TikTok 和 YouTube 的市场，并明确批评以过去收购时（而非现在）的市场条件来评估垄断的理论。
- **Google 搜索补救裁定 (2025)**：虽然裁定 Google 违法，但法院拒绝了 DOJ 要求强制出售 Chrome 浏览器的最激进补救要求，引用了 AI（ChatGPT 等）对搜索市场的”颠覆性冲击”作为理由。
- **非竞争条款禁令**：FTC 的全国性非竞争条款禁令在 2024 年被全国禁制令阻止，反映出法院对行政机构通过规则制定权（rulemaking）进行重大经济政策变革

的抵触。

核心张力：新布兰代斯运动试图通过行政机构（FTC, DOJ）推动变革，但美国反垄断法的最终仲裁者是**法院**——而法院仍然深深嵌入芝加哥学派的消费者福利框架中。DMA 式的事前规则路径之所以在欧洲可行，而美国立法停滞（Klobuchar-Grassley 等技术反垄断法案未能成法），部分原因就在于这种制度结构的差异。

15.4.2　学术的四类立场

当前反垄断学术争论并非简单的二元对立。至少存在四个可辨识的立场：

流派	关键人物	核心主张	政策建议
芝加哥最低限度主义者	Easterbrook, Bork 的精神继承人 (GAI/ICLE): Wright, Ginsburg	反垄断只在最窄范围内介入。CWS（狭义 = 短期价格效应）是唯一可操作标准。Type I 错误绝对更严重。	维持现状；限制合并审查；本身违法仅适用于核心卡特尔
主流改革派	Shapiro, Hovenkamp, Baker, Scott Morton	保留 CWS 作为操作标准但强化执法。承认非价格维度。Type I 和 Type II 都严重，相对成本取决于市场情境。	更积极的合并执法；关注劳动买方垄断；更严格的效率分析；使用已有结构推定
新布兰代斯结构主义者	Khan, Wu, Steinbaum & Stucke, Vaheesan, Teachout	CWS 应被替代。保护竞争过程而非仅是消费者福利。Type II 被系统性低估。结构性救济应为默认。	替代 CWS；结构性分离平台；举证责任转移给被告；恢复结构推定；拆分主导企业
现状微调派	Ginsburg, Wright, Melamed, Carlton	框架基本正确。问题不在标准而在执行——更好的执法工具和资源即可。	增加 FTC/DOJ 预算和人员；改善经济分析质量；不赞成根本性法律改革

表 15.8: 反垄断学术争论的四类立场：关键人物、核心主张与政策建议

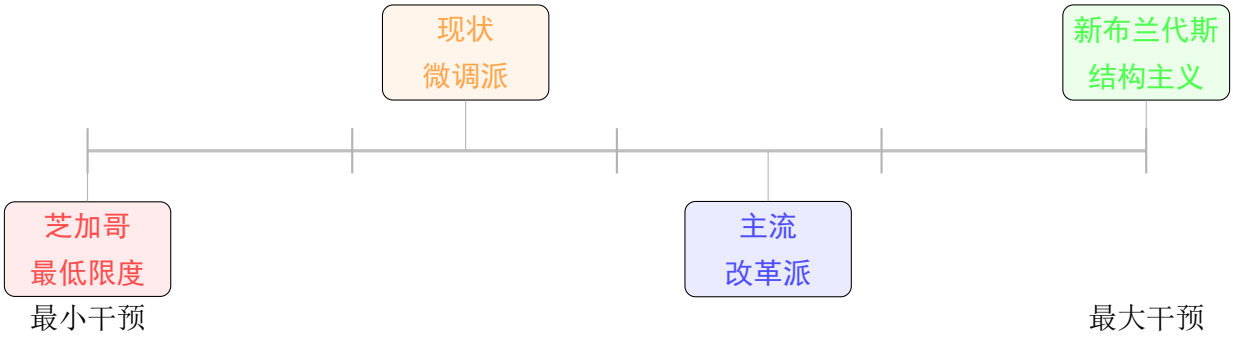


图 15.4: 反垄断学术立场光谱：从” 最小干预” 的芝加哥最低限度主义到” 最大干预” 的新布兰代斯结构主义

15.4.3 2025–2026 年的制度剧变

2025 年 1 月，特朗普任命的 Andrew Ferguson 接替 Khan 出任 FTC 主席。Khan 于 2025 年 1 月 31 日完全辞去 FTC 职务。Ferguson 宣布保留 2023 年合并指南（出人意料），但表示将以” 务实” 方式执行。FTC 撤销了针对 PepsiCo 的 Robinson-Patman 案（FTC 以 3–2 票撤销，两党分歧），但保留了对 Southern Glazer's 的平行诉讼。2025 年 8 月，EO 14036 被正式废除。2026 年 2 月，德克萨斯联邦法院撤销了 2024 年 HSR 申报规则，FTC 选择不上诉。

” 大解锁”（The Great Unlocking）——2026 年 Q1 国内合并价值同比增长约 40%。然而，私人反垄断诉讼继续扩散，新布兰代斯议程部分转移到州层面（加州推动修改 Cartwright Act 以对齐 EU 式竞争法），FTC 诉 Amazon 的审判（2026 年 10 月）仍是新布兰代斯议程的最终检验。

15.4.4 国际比较：美国是异类

一个经常被忽视的关键发现：**多目标反垄断是国际常态，而非异类**。EU 竞争法同时追求不受扭曲的竞争（ordoliberal 遗产）、市场整合（一个独特的欧洲目标）和消费者福利。德国竞争法根植于 Freiburg 学派的秩序自由主义——” 绩效竞争”（*Leistungswettbewerb*）概念本身就是将竞争作为权力分散工具，而非仅是效率工具。中国《反垄断法》第 1 条列举了七重目标（§15.5 详述）。英国、日本、韩国——没有一个将反垄断法缩减为单一目标。

在这一比较视角下，美国坚持反垄断应有单一目标（消费者福利）是**国际异常**——不是规则。这意味着新布兰代斯学派主张回归广泛目标并非如在国内争议中看起来那样激进——它只是主张美国重新加入国际主流。

15.5　中国的位置：第三条道路？

本章以上讨论的芝加哥-新布兰代斯争论——消费者福利还是多重目标？效率还是结构？行为性救济还是结构性救济？——是**西方市场经济内部的争论**。中国提供了一个根本不同的参照系：一种由政治经济学定义而非由竞争法理论定义的竞争政策路径。本节不仅是对中国反垄断实践的介绍——它是全书的**智识收尾**，将读者从 IO 的技术细节带回一个更根本的问题：**市场势力的管制最终是关于国家、资本和市场之间关系的政治选择**。

15.5.1　”社会主义市场经济”的反垄断法

中国《反垄断法》（Anti-Monopoly Law, AML）于 2008 年颁布，2022 年首次修订，其第 1 条以立法设计的七重目标开篇：

> ”为了**预防和制止垄断行为，保护市场公平竞争，鼓励创新，提高经济运行效率，维护消费者利益和社会公共利益，促进社会主义市场经济健康发展**，制定本法。”

这七重目标不是按层级排列的，且包含内在张力。”经济效率”可能与”社会公共利益”冲突；”保护公平竞争”在政府偏爱特定行业或企业时可能与”促进社会主义市场经济健康发展”矛盾。最关键的术语——”社会主义市场经济”——是**有意模糊**的，允许党和国家根据当前政治信念和优先发展事项来解释。在实践中，”社会主义”主导”市场经济”：国家保留关于哪些市场如何运作的最终权威。竞争被工具性地重视——作为效率和创新的手段——而非作为目标本身。

”被偶然的布兰代斯主义”？表面上有亲和性：中国和新布兰代斯学派都拒绝单一目标的芝加哥模式，都明确追求消费者福利之外的多重目标，都对集中的经济权力持怀疑态度（至少在某些情境下），都将竞争法视为更广泛社会政策的工具。但差异是根本性的：(1) 权威来源——新布兰代斯来自民主问责和反垄断作为共和自治的工具，中国来自党-国领导和 AML 作为社会主义发展的工具；(2) 靶标——新布兰代斯瞄准”集中的私人经济权力”本身，中国瞄准”资本在党-国定义的边界之外运作”；(3) 解药——新布兰代斯提出结构性分离和复兴反垄断执法（由独立机构执行），中国提出**资本服从党-国**，通过校准执法、监管架构和直接行政干预来实现。

维度	芝加哥 CWS（美国传统）	中国 AML 七重目标
目标数量	一个（消费者福利/总福利）	七个（法律明文），”社会主义市场经济”作为总体框架
多重目标 vs. 单一目标	单一目标——唯一可操作的框架	多重目标——有意不按层级排列；”社会公共利益”可作为灵活调整的压倒性理由
最终参照点	效率/消费者剩余	党领导下的社会主义市场经济健康发展
经济分析的角色	核心：效果导向，严谨	模糊：工具性使用但可被”公共利益”推翻
对”大”的处理	大就是有效率，除非证明反之	大可以接受——如果它服务于国家目标；如果它挑战国家权威则有问题
新布兰代斯亲和性	极低（单一目标，仅效率）	表面亲和（多重目标，怀疑不受约束的平台权力）

表 15.9: 中国 AML 七重目标 vs. 芝加哥 CWS 单一目标：不是”被偶然的布兰代斯主义”

15.5.2 ”防止资本无序扩张”——中国独有的反垄断话语

2020 年 12 月，中央经济工作会议将”强化反垄断和防止资本无序扩张”列为 2021 年八大重点任务之一。这一措辞——”防止资本无序扩张”——迅速成为统一反垄断执法、金融监管、数据治理和社会政策的总概念（master concept）。

”无序扩张”被刻画为包含五个维度：(1) **市场垄断**——”二选一”排他性交易、掠夺性定价、算法价格歧视（大数据杀熟）、搭售/捆绑；(2) **监管套利**——平台企业跨界进入未经许可的金融服务，在行业监管之外运营；(3) **创新方向偏差**——资本流向流量变现、社区团购补贴和寻租，而非”卡脖子”技术的基础研发；(4) **系统性风险**

——通过平台借贷进行影子银行、房地产市场高杠杆、可能的连锁失败；(5) **社会危害**——不平等加剧、挤压小商贩和传统企业、隐私侵犯。

这一措辞——" **资本**" 的无序扩张——而非" 市场势力" 或" 垄断"——具有政治策略性。它使国家能够：(1) 针对特定企业和行业，而不宣布所有私营企业可疑；(2) 在竞争法标准难以满足时（市场定义、损害证明、效率抗辩）绕过竞争法的技术要求，通过金融监管、数据安全审查或直接行政命令进行干预；(3) 向国内外受众传达中国并不反市场或反私营本身，而是对资本施加必要纪律——一种共鸣于马克思主义政治经济学同时又可被务实商业受众接受的框架；(4) 将中国的方法与西方反垄断区分开来——后者被框定为技术官僚经济规制而非对资本的政治控制。

" 防止资本无序扩张" vs. 新布兰代斯——系统对比：

维度	美国新布兰代斯	中国" 防止资本无序扩张"
核心问题	集中的私人经济权力固有地威胁民主治理	在党-国定义边界外扩张的资本是" 无序" 且有害的
诊断	市场结构导致政治不平等	平台行为威胁国家监管能力和社会稳定
处方	拆分垄断；复兴结构反垄断；分离商业与平台	让资本服从国家目标；校准执法；将投资重新导向战略行业
谁决定？	独立机构（FTC, DOJ）+ 法院	党领导 + 跨部门行政协调
好坏的区别标准	所有过度的私人集中都可疑	服务国家目标的大（硬科技、EV、半导体）是好的；挑战国家权威或提取租金的大是坏的

表 15.10: " 防止资本无序扩张" vs. 新布兰代斯诊断：表面亲和 vs. 根本分歧

15.5.3　2020–2022 整改风暴与标志性案件

时间线。这一密集执行期的时间线揭示了政治授权——而非预先存在的执法计划——是执法的约束性约束：2020 年 10 月 Ma 的 Bund Summit 演讲（批评监管的" 当铺思维"）→ 2020 年 11 月 Ant Group 创纪录的 $370 亿 IPO 被暂停 → 2020 年 12 月中央经济工作会议语言 → 2021 年 4 月阿里巴巴 182.28 亿元罚款 → 2021 年 7 月 Didi 从应用商店下架（数据安全审查）和" 双减" 教育政策 → 2021 年 10 月美团

34.42 亿元罚款。

统一逻辑：五项行动——跨越反垄断（阿里巴巴、美团）、金融监管（Ant Group）、数据主权（Didi）、社会/产业政策（教育" 双减"）——不是随意或报复性的。统一框架是" 防止资本无序扩张"：北京得出结论，不受约束的消费互联网平台产生了系统性金融风险、加剧了社会不平等、创造了数据安全脆弱性，并将资本从战略上至关重要的" 硬科技" 转移。

案件	时间	法律依据	核心行为与救济	主要逻辑
阿里巴巴	2021.4	AML 第 17/47 条	"二选一" 排他性交易；罚款 182.28 亿元 (4% × 2019 国内销售额)；三年合规监控	竞争法
美团	2021.10	AML 第 17/47 条	"二选一" 排他性交易；罚款 34.42 亿元 (3%)；退还 12.89 亿元独家合作保证金	竞争法 + 再分配
Ant Group	2020.11	金融监管授权	IPO 暂停；强制重组为金融控股公司；全面纳入 PBOC 监管	金融稳定
Didi	2021.7	《网络安全审查办法》	应用商店下架；暂停新用户注册；网络安全审查	数据主权/国家安全
教育" 双减"	2021.7	国务院政策	禁止核心学校科目的营利性辅导；禁止外资；要求非营利重组	社会/产业政策
腾讯音乐	2021	AML 第 48 条	放弃独家音乐版权；罚款（未申报收购）	竞争法 + 产业政策
虎牙/斗鱼	2021	AML 第 28 条	首个数字领域合并禁止（合并后 > 70% 游戏直播市场份额）	竞争法 + 产业政策

表 15.11: 2020–2022 整改风暴关键案例：竞争法还是产业政策？

关键洞见：即使是" 真正的竞争法案件"（阿里巴巴、美团）也是由政治授权来执行的。中国反垄断的约束性约束从来不是法律工具的充分性——2008 年 AML 一直具有广泛的适用范围和实质性处罚。约束性约束是**使用这些工具对付国内冠军企业的政治授权**（2020 年 12 月获得）。一旦获得授权，执法迅速而有效。这与美国（约束性约束通常是教义性的——在消费者福利标准下证明损害——或司法性的——说服持怀疑态度的法院）和 EU（约束性约束通常是程序性的——正当程序、抗辩权、司法审查）根本不同。

15.5.4　从” 强监管” 到” 常态化监管”

中国平台监管经历了四个阶段：(1) **探索期**（2008–2016）——数字市场无明确反垄断执法；” 避风港”；(2) **包容审慎监管**（2017–2020）——轻触式，鼓励创新；VIE 合并被容忍未申报；(3) **强监管**（2021）——运动式；” 反垄断和防止资本无序扩张”；(4) **常态化监管**（2022–今）——系统化、基于规则、可预测；” 规范 AND 发展” 并重。

常态化监管的关键要素包括：可预测规则取代临时性镇压（” 杜绝选择性执法、运动式执法”）、全链条监管（事前合并审查/公平竞争审查 + 事中市场监控/算法审计 + 事后处罚/救济）、梯次化监管工具（” 三书一函” 制度——提醒函 → 监管约谈 → 行政指导 → 正式调查和罚款）、跨部门协调（SAMR + CAC + PBOC + MIIT + NDRC + MPS）。2024 年 7 月，二十届三中全会进一步确立了” 促进平台经济创新发展，健全平台经济常态化监管制度” 的方针。

” 第三条道路” 还是策略性校准？。支持” 第三条道路” 论点的理由是：中国 AML 自 2008 年以来保持了一贯的多目标框架，以” 社会主义市场经济” 作为总体原则。2020–2022 年镇压——虽然表面上突然——与 AML 的法定设计一致：法律始终提供了公共利益压倒和状态校准的空间。中国的竞争政策（黄勇倡导的” 竞争政策基础地位” 辩论和 Jiang Feitao 的” 功能协调” 综合）与产业政策关系的学术辩论是一项严肃、持续的参与。反对” 第三条道路” 论点的理由是：执法强度急剧波动——从对平台的近乎零执法（2008–2020）到创纪录罚款（2021）再到近乎零执法（2023 年 SAMR 未公开报告数字领域垄断案）——与连贯的竞争法逻辑不一致。反垄断执法已被用于远远超出竞争法的目标：金融稳定（Ant）、数据主权（Didi）、社会政策（教育）、收入再分配（共同富裕）。该体制缺乏真正法治反垄断特征的制度独立性——SAMR 服从国务院，后者服从党。

综合判断：最具说服力的解释是，中国具有连贯的**政治经济学逻辑**（私人资本服从于党-国发展目标），通过一个**制度灵活且政治校准**的反垄断机构来体现。逻辑是连贯的，执法是机会主义的。这既不同于美国（执法受法院和教义一致性约束），也不同于 EU（执法受条约义务和委员会独立性约束），但它是由政治经济学定义的” 第三条道路”，而非由独特的竞争法理论定义的。根本问题不是” 消费者福利还是民主结构？”——而是” **这家企业的市场势力是服务于还是挑战了国家权威和发展目标？**”

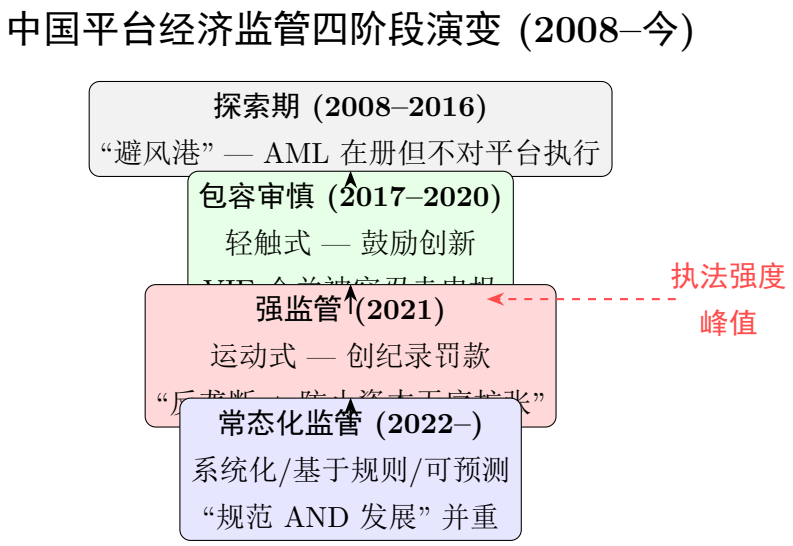


图 15.5: 中国平台经济监管四阶段演变：从” 避风港” 到” 常态化监管”

15.5.5 中国、欧盟、美国：三种反垄断哲学

Anu Bradford 在《数字帝国》（*Digital Empires*, Oxford 2023）中提供了一种有助于理解分歧的类型学：美国 = **市场驱动**——效率、消费者福利；法院对干预持怀疑态度；高证据负担。EU = **权利驱动**——公平、市场整合、数据保护；机构驱动；事前守门人监管（DMA）。中国 = **国家驱动**——产业政策、国家安全、国家对资本的控制；政党校准的执法；反垄断作为众多工具之一。

维度	美国（芝加哥/后芝加哥）	EU（秩序自由/权利导向）	中国（国家校准）
反垄断目标	CWS（消费者福利）作为主导操作原则	CWS + 市场整合 + 公平 + 数据保护；”数字主权”（新兴）	七项法定目标；”社会主义市场经济”作为总体框架
损害标准	效果导向：需要实际或可能的消费者损害证明	效果导向但门槛更低；潜在竞争损害足够；公平和非价格维度相关	形式上效果导向但”公共利益”允许推翻；对国家监管目标的损害可替代市场损害
制度结构	法院（法官造法）；独立机构（FTC, DOJ）受司法审查	欧盟委员会（行政）；EU 法院司法审查；日益多机构（DMA 执行者）	SAMR（行政）+ 跨部门协调（CAC, MIIT, PBOC, NDRC, MPS）；党领导提供政治方向
执法方式	事后、逐案、对抗性诉讼；高证据负担；”错误成本”框架倾向于假阴性	事后（第 101/102 条 TFEU）+ 事前（DMA 守门人义务）；机构调查和罚款	事后（AML）+ 事前（低于阈值合并审查、公平竞争审查）；行政执法规导；梯次化工具
与产业政策的关系	反垄断传统上反对产业政策；CHIPS 法案和 IRA 代表新产业政策转向产生张力	竞争政策历史上主导；产业政策重新出现（绿色协议、半导体补贴）；DMA/DSA 作为”通过规制的数字产业政策”	产业政策历史上主导；”竞争政策基础地位”是理想目标；反垄断被校准以服务产业政策目标
平台处理	历史上宽容（自 Microsoft 2001 以来无重大平台拆分）；新布兰代斯 FTC 信号转向但面临司法怀疑主义	高额罚款（Google ~€82.5 亿）；DMA 施加事前守门人义务；数据主导地位作为竞争关切	最初宽容（培育国内冠军）；2021 镇压；现”常态化监管”——平台必须在国家定义边界内运营
最终权威	法院 + 民主过程；反垄断作为法治	EU 条约 + 委员会 + 司法审查；反垄断作为超国家市场治理	党领导；反垄断作为社会主义市场经济治理的工具
关键知识影响	芝加哥学派 → 后芝加哥[476] → 新布兰代斯（有争议）	秩序自由主义 → ”更加经济的方法” → 权利导向的数字	马克思主义资本政治经济学 + EU 竞争法移植 + 中国

表 15.12: 中国 vs. 欧盟 vs. 美国：三种反垄断哲学的系统比较

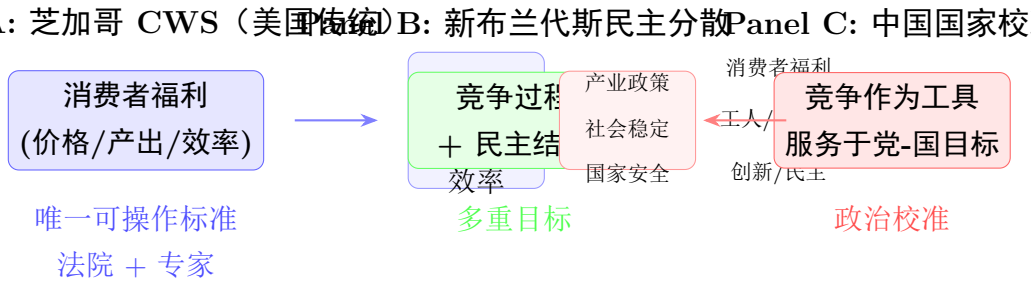


图 15.6: 三种反垄断哲学概念图：芝加哥 CWS（效率/价格）vs. 新布兰代斯（民主分散/多重目标）vs. 中国（国家校准/工具性竞争）

15.5.6 全书收尾：回到 Ch.1 的核心问题

第 ~1~ 章提出了全书的核心问题：”**市场势力从何而来？**” 第 ~2~ 章至 第 ~14~ 章提供了回答这个问题的理论、方法和制度工具——从 Cournot 的一阶条件到 BLP 的结构估计，从 SCP 的集中度回归到并购模拟的 UPP/GUPPI 公式，从双边市场的修正 Lerner 公式到算法合谋的 Q-learning 实验。

但所有这些工具服务于一个更根本的判断——**我们使用这些工具来追求什么？**芝加哥学派给出的回答是：消费者福利——以价格和产出衡量。这一标准具有智识上的优雅性（一个单一的可最大化目标）、制度上的可操作性（法院可以裁决价格效应）和政治上的表面中立性（” 我们只是在执行经济学”）。

本章——以及新布兰代斯运动——揭示了这一回答的深层不完整性。消费者福利标准不能告诉你如何评估扼杀式收购（消费者价格从未上升，因为竞争者在产生任何竞争之前就被消灭了）。它不能告诉你如何评估隐私退化（” 免费” 服务的价格是零——$5\% \times 0 = 0$）。它不能告诉你如何评估买方垄断势力对工资的压制（工人不是消费者福利标准意义上的” 消费者”）。最重要的是，它不能回答一个先于经济学的问题：**我们为什么重视竞争？**——是因为它产生低价格，还是因为它分散经济权力、保护机会的开放性和维持民主自治的条件？

IO 三百年的钟摆。从古典工业结构分析到博弈论策略行为，从 SCP 实证到结构估计，从单边市场到双边平台，从人类定价到算法合谋——IO 经济学的核心张力一直是**结构**与**行为**之间、**效率**与**权力**之间的张力。芝加哥学派将钟摆推向行为/效率的极端——只看行为的实际效果，只看价格和产出。新布兰代斯学派试图将钟摆推回结构/权力的方向——重新关注市场结构本身是否威胁了竞争过程和经济民主。数字时代没有改变这一根本张力——它只是将其推到了新的极端：当算法可以在没有人类沟通的情况下达成合谋、当数据飞轮可以创造不可逾越的进入壁垒、当 AI 堆栈的纵向一体化可以同时控制基础设施和应用层时，” 结构” 和” 行为” 之间的界限变得前所未有的模糊。

三个未完成的制度实验。美国新布兰代斯实验（通过行政机构推动变革，面临法院的结构性抵制）、EU DMA/事前监管实验（通过立法指定守门人义务，短路径绕过救济争论）、中国”第三条道路”实验（将竞争法作为国家发展目标的校准工具）——这三个实验各自尚未完成，各自存在内部张力，各自代表着对”反垄断是为了什么？”这一问题的不同回答。

终章命题：反垄断是什么？不是一个可以通过一阶条件求解的技术问题。它是一个关于我们应该如何组织经济生活的、永无止境的民主辩论。芝加哥学派、新布兰代斯学派、欧盟秩序自由主义、中国国家校准——每一种都提供了部分答案。IO 经济学的智识任务不是选择其中一个而抛弃其他，而是理解每一者在什么条件下成立、在什么条件下失败，以及——最重要地——在每一具体的市场和制度情境中，哪种考虑应该占据优先地位。

全书终章：IO 经济学的根本命题

本书始于一个问题——”市场势力从何而来？”——终于一个更深刻的追问——”我们应该如何治理市场势力？”第一个问题有答案：市场势力来自进入壁垒、产品差异化、规模经济、网络效应、信息不对称、策略行为和制度。第二个问题没有终极答案：它是每一个世代必须根据其自身的市场结构、政治价值和制度能力重新回答的问题。IO 经济学不能代替这个判断——但它可以使这个判断更加清晰、更加诚实地面对权衡、更加审慎地设计制度。这是本书 15 章的终极目的。

习题

题 1 Bork vs. Brandeis。Bork 为什么认为保护小企业不应是反垄断的目标？Brandeis 会如何回应？请引用 Bork (1978) 第 2 章和 Brandeis (1914) 的具体论证。（概念理解）

题 2 消费者福利标准的操作化。如果平台对消费者免费（价格 =0），消费者福利标准如何评估竞争损害？请用 Google Search 或 Meta/Facebook 的例子说明。Khan (2017) 对此提出了什么替代方案？（概念理解）

题 3 错误成本假设。Easterbrook (1984) 的非对称性假设（Type I > Type II）在什么市场条件下最可能成立？在什么条件下最可能失败？请结合 第 ~14~ 章中讨论的网络效应和数据壁垒进行分析。（概念理解 + 跨章联立）

题 4 Khan (2017) 的逻辑。选择 Amazon 的四个反竞争维度之一：（a）低于成本定价，（b）收购竞争对手，（c）跨领域杠杆化支配地位，（d）利用第三方卖家数

据。解释为什么以短期价格效应衡量的消费者福利标准无法捕获这种损害。请提供具体案例细节。（分析应用）

题 5 Meta 案的市场定义。FTC v. Meta (2025) 中，Boasberg 法官将 TikTok 和 YouTube 纳入相关市场，裁定 FTC 未能证明 Meta 在当下拥有垄断权力。你认为这一裁决合理吗？这一裁决对未来的平台反垄断——特别是涉及动态市场定义和" 扼杀式收购" 理论的案件——有何含义？（分析应用）

题 6 结构性 vs. 行为性救济。比较 Microsoft (2001) 的行为性救济（互操作性披露）与 AT&T (1984) 的结构性救济（拆分为七家地区性贝尔运营公司和一家长话公司）。哪种更成功地恢复了竞争？为什么？请引用 EU 委员会 (2025) 救济措施回顾研究发现。（分析应用 + 跨章联立）

题 7 中国" 二选一"。阿里巴巴" 二选一" 案（2021, 182.28 亿元罚款）与美国反垄断法下的排他性交易（exclusive dealing）有何异同？中国执法与美国执法在哪些根本维度上不同？请结合 §15.5 的框架和 第 ~12~ 章的反垄断分析工具进行讨论。（制度比较 + 跨章联立）

题 8 常态化监管 vs. 司法中心主义。中国的" 常态化监管" 模式（行政主导、梯次化工具、跨部门协调）与美国反垄断的" 司法中心主义" 模式（法院主导、逐案诉讼、高证据负担）有何根本区别？各有什么利弊？在什么条件下行政模式更优？在什么条件下司法模式更优？（制度比较）

题 9 反垄断的目标。假设你被任命为中国《反垄断法》修订委员会主席。第 1 条目前列举了七项法定目标（包括" 促进社会主义市场经济健康发展"）。你会在修订版中保留、删除或增添哪些目标？请将这些目标排序并说明你的排序原理。你的最终框架更接近芝加哥、新布兰代斯、还是中国" 第三条道路"？（开放性思考）

题 10 全书反思。回顾从 第 ~1~ 章的" 市场势力从何而来？ " 到本章的" 反垄断究竟为了什么？ " 这一完整的 15 章智识旅程。IO 经济学的核心张力是什么？这一张力——结构与行为之间、效率与权力之间——最终可以解决吗？为什么能或为什么不能？（开放性思考）

推荐阅读

芝加哥学派经典

(1) Bork, R.H. (1978). *The Antitrust Paradox: A Policy at War with Itself.* New York: Basic Books. [第 1–4 章：理论基础与纵向限制]

(2) Easterbrook, F.H. (1984). "The Limits of Antitrust." *Texas Law Review*, 63: 1–40. [错误成本框架的奠基性文本]

(3) Posner, R.A. (1979). "The Chicago School of Antitrust Analysis." *University of Pennsylvania Law Review*, 127: 925–948. [芝加哥学派反垄断分析最系统的阐述]

Post-Chicago 与内部批评

(4) Hovenkamp, H. (1985). "Antitrust Policy After Chicago." *Michigan Law Review*, 84: 213–284. [Post-Chicago 的奠基性批评：超过 70 页]

(5) Baker, J.B. (2015). "Taking the Error Out of 'Error Cost' Analysis: What's Wrong with Antitrust's Right." *Antitrust Law Journal*, 80(1): 1–62. [对芝加哥错误成本框架最全面的学术拆解]

(6) Shapiro, C. (2019). "Antitrust in a Time of Populism." *International Journal of Industrial Organization*, 61: 714–748. [主流经济学家在民粹时刻的平衡评估]

新布兰代斯核心文本

(7) Khan, L.M. (2017). "Amazon's Antitrust Paradox." *Yale Law Journal*, 126(3): 710–805. [引爆运动的学生笔记：约 93 页]

(8) Wu, T. (2018). *The Curse of Bigness: Antitrust in the New Gilded Age.* New York: Columbia Global Reports. [154 页宣言式著作]

(9) Khan, L.M. (2019). "The Separation of Platforms and Commerce." *Columbia Law Review*, 119(4): 973–1098. [为新布兰代斯的核心救济方案——结构性分离——提供技术和历史论证]

(10) Steinbaum, M. & Stucke, M.E. (2020). "The Effective Competition Standard: A New Standard for Antitrust." *University of Chicago Law Review*, 87(2): 595–623. [最系统的新布兰代斯替代标准方案——在芝加哥学派的大本营发表，具有象征意义]

综合与中间立场

(11) Baker, J.B. (2019). *The Antitrust Paradigm: Restoring a Competitive Economy.* Cambridge, MA: Harvard University Press. [349 页——被广泛认为是当代反垄断改革辩论中最重要的单卷著作]

(12) Hovenkamp, H. (2019). "Is Antitrust's Consumer Welfare Principle Imperiled?" *Journal of Corporation Law*, 45: 65–100. [消费者福利标准阵营内部最老练的批评和最强大的辩护的统一]

比较与国际视角

(13) Bradford, A. (2023). *Digital Empires: The Global Battle to Regulate Technology.* Oxford: Oxford University Press. [第 7–9 章：三种模式——市场驱动、权利驱动、国家驱动]

(14) European Commission (2025). *Study on the Effectiveness of Antitrust Remedies* (20-year retrospective). [里程碑式实证研究：三分之二行为性救济部分或完全无效]

(15) Schinkel, M.P. & van Oosten, R. (2025). "Breaking up Big Tech: Scissor-Line Suggestions for Smart Cuts." *Journal of Antitrust Enforcement*, 13(2): 237–272. [定向结构性分离：学术上最精细的拆分方案]

中国相关

(16) 王晓晔 (2025). "反垄断新思潮新布兰代斯主义批判." 法学评论 (Law Review). [中国反垄断法第一人对新布兰代斯运动最系统的中文评估]

(17) Collier, A. (2022). *China's Technology War: Why Beijing Took Down Its Tech Giants.* London: Palgrave Macmillan. [2020–2022 整改风暴的英文专著]

历史与立法意图

(18) Hazlett, T.W. (1992). "The Legislative History of the Sherman Act Re-Examined." *Economic Inquiry*, 30(2): 263–276. [对 Bork 立法史论证的经济计量重新审查]

(19) Crane, D.A. (2019). "How Much Brandeis Do the Neo-Brandeisians Want?" *The Antitrust Bulletin*, 64(4): 531–539. [新布兰代斯最老练的批评者提出的三个尖锐挑战——关于一致性、权衡和经验主义]

附录 A　数学附录：核心模型推导与计量方法速查

本章为全书核心模型的数学推导提供集中参考，并汇总博弈论和计量经济学的基础工具。建议在阅读正文过程中随时回来查阅。

A.1　博弈论基础

A.1.1　策略式博弈与 Nash 均衡

一个**策略式博弈**（strategic form game）由三个要素构成：玩家集合 $N = \{1, 2, \ldots, n\}$、每个玩家 i 的策略空间 S_i、以及每个玩家 i 的支付函数 $u_i : S_1 \times \cdots \times S_n \to \mathbb{R}$。

Nash 均衡（Nash Equilibrium, NE）：一个策略组合 $s^* = (s_1^*, \ldots, s_n^*)$ 是 Nash 均衡，如果对每个玩家 i 和每个备选策略 $s_i \in S_i$，

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*) \tag{A.1}$$

其中 s_{-i}^* 表示除 i 外所有玩家的均衡策略。

在连续策略空间中的求解：通常用一阶条件法。假设 $S_i \subseteq \mathbb{R}$ 且 u_i 可微，在内部解处：

$$\frac{\partial u_i(s_i, s_{-i})}{\partial s_i} = 0 \tag{A.2}$$

一阶条件给出玩家 i 的**反应函数**（reaction function）$s_i = R_i(s_{-i})$。Nash 均衡是反应函数的不动点：$s_i^* = R_i(s_{-i}^*)$ 对所有 i 成立。

例题：Cournot 双寡头。$n = 2$，$u_i = (a - b(q_1 + q_2))q_i - cq_i$。一阶条件：

$$\frac{\partial u_1}{\partial q_1} = a - 2bq_1 - bq_2 - c = 0 \tag{A.3}$$

$$\frac{\partial u_2}{\partial q_2} = a - 2bq_2 - bq_1 - c = 0 \tag{A.4}$$

反应函数：

$$R_1(q_2) = \frac{a - c - bq_2}{2b}, \quad R_2(q_1) = \frac{a - c - bq_1}{2b} \tag{A.5}$$

解不动点得 Nash 均衡：$q_1^* = q_2^* = (a - c)/(3b)$。

A.1.2　扩展式博弈与子博弈完美均衡

扩展式博弈（extensive form game）用博弈树（game tree）表示，明确博弈的顺序和信息结构。

子博弈完美 Nash 均衡（Subgame Perfect Nash Equilibrium, SPNE）：一个策略组合是 SPNE，如果它在**每一个**子博弈（subgame）中都构成 Nash 均衡。SPNE 通过逆向归纳法（backward induction）求解：从博弈树的末端开始，在每个决策节点上选择最优行动，逐步向前推导。

例题：Stackelberg 领导模型。企业 1（领导者）先选择 q_1，企业 2（追随者）观察到 q_1 后选择 q_2。

- **第 2 阶段**：给定 q_1，企业 2 求解 $\max_{q_2}(a-b(q_1+q_2))q_2-cq_2$。一阶条件给出反应函数 $q_2=R_2(q_1)=(a-c-bq_1)/(2b)$。

- **第 1 阶段**：企业 1 预见到企业 2 的反应，求解 $\max_{q_1}(a-b(q_1+R_2(q_1)))q_1-cq_1$。代入 R_2 并求导：

$$q_1^*=\frac{a-c}{2b},\quad q_2^*=\frac{a-c}{4b},\quad Q^*=\frac{3(a-c)}{4b} \tag{A.6}$$

与 Cournot 相比，领导者产量更高，追随者产量更低，总产量更高，价格更低——这是**先发优势**（first-mover advantage）的直接体现。

A.1.3　不完全信息博弈与贝叶斯 Nash 均衡

当玩家拥有私人信息时，使用**贝叶斯博弈**（Bayesian game）。每个玩家 i 有一个**类型**（type）$\theta_i\in\Theta_i$，只有自己知道。类型分布 $F(\theta)$ 是共同知识。

贝叶斯 Nash 均衡（Bayesian Nash Equilibrium, BNE）：每个类型的每个玩家在给定其信念（关于其他玩家类型的后验分布）下最大化期望支付。

显示原理（Revelation Principle）：任何贝叶斯 Nash 均衡的结果都可以通过一个**直接显示机制**（direct revelation mechanism）实现——每个玩家如实报告自己的类型，并且如实报告是一个均衡。这个原理是 Laffont-Tirole 激励规制理论（第 ~9~ 章）的核心工具。

A.1.4　重复博弈与 Folk 定理

无限重复博弈：阶段博弈（stage game）无限次重复，玩家使用贴现因子 $\delta\in(0,1)$ 对未来支付进行贴现。

触发策略（trigger strategy）：在每一期合作，直到有人背叛；一旦有人背叛，永远转向阶段博弈的 Nash 均衡（" 永久惩罚"）。

设 π^C 为每期合谋利润，π^D 为单期背叛利润，π^N 为 Nash 均衡利润。合谋的激励相容条件为：

$$\frac{\pi^C}{1-\delta} \geq \pi^D + \frac{\delta \pi^N}{1-\delta} \tag{A.7}$$

整理得：

$$\delta \geq \frac{\pi^D - \pi^C}{\pi^D - \pi^N} \tag{A.8}$$

即贴现因子必须足够大（玩家足够耐心），合谋才能维持。

Folk 定理（Folk Theorem）：在无限重复博弈中，如果玩家足够耐心（$\delta \to 1$），任何满足个体理性约束的可行支付向量都可以作为 SPNE 的结果维持。这意味着**合谋有多种方式可以实现**——无穷多种均衡。

A.2　核心模型完整推导

A.2.1　Cournot 寡头：n 个企业的一般解

设定：n 个企业生产同质产品。逆需求函数 $P(Q) = a - bQ$，$Q = \sum_{i=1}^{n} q_i$。企业 i 的成本函数 $C_i(q_i) = c_i q_i$（常数边际成本）。

推导：企业 i 的利润最大化问题：

$$\max_{q_i} \ \pi_i = \big(a - b(q_i + Q_{-i})\big) q_i - c_i q_i \tag{A.9}$$

其中 $Q_{-i} = \sum_{j \neq i} q_j$。一阶条件：

$$a - 2bq_i - bQ_{-i} - c_i = 0 \tag{A.10}$$

这定义了企业 i 的反应函数。对于对称情况 $c_i = c \ \forall i$，利用对称性 $q_i^* = q^*$：

$$a - 2bq^* - b(n-1)q^* - c = 0 \tag{A.11}$$

$$a - b(n+1)q^* - c = 0 \tag{A.12}$$

$$q^* = \frac{a-c}{b(n+1)} \tag{A.13}$$

市场总产量、价格和 Lerner 指数：

$$Q^* = nq^* = \frac{n(a-c)}{b(n+1)} \tag{A.14}$$

$$P^* = a - bQ^* = \frac{a + nc}{n+1} \tag{A.15}$$

$$L = \frac{P^* - c}{P^*} = \frac{a-c}{a+nc} \tag{A.16}$$

收敛性质：

- $n = 1$：垄断 $Q^* = (a-c)/(2b)$，$P^* = (a+c)/2$

- $n = 2$：双寡头 $Q^* = 2(a-c)/(3b)$
- $n \to \infty$：$P^* \to c$（完全竞争极限）

福利损失（Corchón, 2008）：在对称线性 Cournot 模型中，相对于完全竞争的福利损失百分比为：

$$\text{PWL} = \frac{1}{(n+1)^2} \tag{A.17}$$

福利损失随 n 的平方迅速衰减——这是集中度与效率关系的理论基准。

A.2.2　Bertrand 均衡：差异化产品

同质产品：两个企业以相同边际成本 c 生产同质产品，同时选择价格。唯一的 Nash 均衡是 $p_1^* = p_2^* = c$——Bertrand 悖论。

差异化产品：需求系统为 $q_i = D_i(p_i, p_j) = \alpha - \beta p_i + \gamma p_j$，其中 $\beta > \gamma > 0$（自身价格效应大于交叉价格效应）。边际成本 c。

企业 i 的利润：$\pi_i = (p_i - c)(\alpha - \beta p_i + \gamma p_j)$。一阶条件：

$$\alpha - 2\beta p_i + \gamma p_j + \beta c = 0 \tag{A.18}$$

利用对称性 $p_i = p_j = p^*$：

$$\alpha - 2\beta p^* + \gamma p^* + \beta c = 0 \tag{A.19}$$

$$\alpha - (2\beta - \gamma)p^* + \beta c = 0 \tag{A.20}$$

$$p^* = \frac{\alpha + \beta c}{2\beta - \gamma} \tag{A.21}$$

当 $\gamma \to 0$（产品完全独立），$p^* \to (\alpha + \beta c)/(2\beta)$（各自的垄断价格）。当 $\gamma \to \beta$（产品几乎同质），$p^* \to (\alpha + \beta c)/\beta$——需要检查 β 和 α 的符号来解读。关键是：**差异化程度 γ/β 决定了高于边际成本的加成**。

A.2.3　Hotelling 线性城市：完整推导

设定：消费者均匀分布在 $[0, 1]$ 上。企业 A 在 $x = 0$，企业 B 在 $x = 1$，分别定价 p_A, p_B。交通成本：距离 d 处的消费者支付 td^2（二次成本，确保纯策略均衡存在）。

位于 x 的消费者从 A 购买获得效用 $v - p_A - tx^2$，从 B 购买获得 $v - p_B - t(1-x)^2$。无差异消费者 $\hat{x}$ 满足：

$$v - p_A - t\hat{x}^2 = v - p_B - t(1-\hat{x})^2 \tag{A.22}$$

解出：

$$\hat{x} = \frac{1}{2} + \frac{p_B - p_A}{2t} \tag{A.23}$$

企业 A 面临的需求为 $D_A = \hat{x}$（$[0, \hat{x}]$ 上的消费者）。利润 $\pi_A = (p_A - c)\hat{x}$。一阶条件：

$$\frac{\partial \pi_A}{\partial p_A} = \hat{x} + (p_A - c)\left(-\frac{1}{2t}\right) = 0 \tag{A.24}$$

利用对称性 $p_A = p_B = p^*$：

$$\frac{1}{2} + (p^* - c)\left(-\frac{1}{2t}\right) = 0 \tag{A.25}$$

$$p^* = c + t \tag{A.26}$$

关键结果：均衡价格高于边际成本，加成等于差异化参数 t——产品差异化削弱价格竞争。这一结果的普遍性在差异化产品市场的现代模型（如 BLP，第 ~11~ 章）中持续出现。

A.2.4　Dixit 产能承诺模型

两阶段博弈：

- 第 1 阶段：在位者（企业 1）选择产能 K_1，该选择不可逆
- 第 2 阶段：潜在进入者（企业 2）观察到 K_1，决定是否进入；若进入，双方进行 Cournot 产量竞争

用逆向归纳法求解。在第 2 阶段，若企业 2 进入，则进行 Cournot 双寡头博弈。企业 1 受产能约束：$q_1 \leq K_1$。

关键洞见：企业 1 在第 1 阶段选择 K_1 时，可以通过选择 $K_1 > q_1^{\text{Cournot}}$（过度投资）来承诺在第 2 阶段生产更多，从而降低企业 2 的进入利润。当进入的固定成本 F 足够大时，企业 1 可以完全阻止进入。这是 **Spence-Dixit 进入遏制**的核心机制（第 ~8~ 章 §8.1）。

A.2.5　Laffont-Tirole 激励规制模型

设定：企业成本 $C = \beta - e$，其中 β 是效率类型（低成本 $= \underline{\beta}$，高成本 $= \overline{\beta}$），e 是降低成本的不可观测努力。努力的负效用为 $\psi(e)$（$\psi' > 0, \psi'' > 0$）。规制者向企业支付 t。

企业的参与约束和激励相容约束分别为：

$$t - \psi(\beta - C) \geq 0 \quad \text{(IR)} \tag{A.27}$$

$$t - \psi(\beta - C) \geq \tilde{t} - \psi(\beta - \tilde{C}) \quad \text{(IC)} \tag{A.28}$$

其中 $(\tilde{t}, \tilde{C})$ 是对另一类型的合同。

最优合同菜单：规制者提供 $\{(\underline{t}, \underline{C}), (\overline{t}, \overline{C})\}$。

- 高效率类型（$\underline{\beta}$）：选择固定价格合同（C 固定，企业保留全部成本节约）——”高能激励”
- 低效率类型（$\overline{\beta}$）：选择成本加成合同（C 可变，成本由规制者报销）——” 低能激励”

高效率类型获得正的信息租金：$\underline{U} = \psi(\overline{\beta} - \overline{C}) - \psi(\overline{\beta} - \underline{C}) > 0$。低效率类型获得零租金。这个” 租金-效率” 权衡（rent-efficiency trade-off）是激励规制的核心（第 ~9~ 章）。

A.3　计量方法速查

A.3.1　广义矩方法（GMM）

矩条件：设 $g(z_i, \theta)$ 是 $m \times 1$ 的矩函数向量，满足总体矩条件 $\mathbb{E}[g(z_i, \theta_0)] = 0$。当 $m > k$（维度 θ）时，系统是过度识别的。

GMM 估计量：

$$\hat{\theta}_{\text{GMM}} = \arg\min_{\theta} \left(\frac{1}{n} \sum_{i=1}^{n} g(z_i, \theta) \right)' W \left(\frac{1}{n} \sum_{i=1}^{n} g(z_i, \theta) \right) \tag{A.29}$$

其中 W 是 $m \times m$ 的权重矩阵。

最优权重矩阵：$W = \mathbb{E}[g(z_i, \theta_0) g(z_i, \theta_0)']^{-1}$（Hansen, 1982）。两步 GMM：

1. 用 $W = I$（或任何正定矩阵）获得第一步估计 $\hat{\theta}_1$
2. 用 $\hat{\theta}_1$ 估计最优 W，获得 $\hat{\theta}_2$（有效 GMM）

过度识别检验（Hansen J 检验）：H_0: 矩条件正确设定。

$$J = n \cdot \left(\frac{1}{n} \sum_{i} g(z_i, \hat{\theta}) \right)' \hat{W} \left(\frac{1}{n} \sum_{i} g(z_i, \hat{\theta}) \right) \sim \chi^2_{m-k} \tag{A.30}$$

在 BLP 中的应用（第 ~11~ 章）：矩条件为 $\mathbb{E}[\xi_{jt} \cdot Z_{jt}] = 0$，其中 ξ_{jt} 是不可观测产品特征，Z_{jt} 是工具变量（如 BLP instruments：竞争对手的产品特征）。

A.3.2　非线性最小二乘（NLS）

模型：$y_i = f(x_i, \theta) + \varepsilon_i$，其中 f 是 θ 的非线性函数。

NLS 估计量：

$$\hat{\theta}_{\text{NLS}} = \arg\min_{\theta} \sum_{i=1}^{n} (y_i - f(x_i, \theta))^2 \tag{A.31}$$

通常使用数值优化方法（Gauss-Newton、Levenberg-Marquardt）求解。

在 NEIO 中的应用（第 ~10~ 章）：供给关系 $P = MC + \theta \cdot P'(Q) \cdot Q/n$ 是 θ 和成本参数的非线性函数。Bresnahan (1982) 用 NLS 联合估计需求和供给方程。

A.3.3　收缩映射（Contraction Mapping）

定义：函数 $T : X \to X$（X 是完备度量空间）是压缩映射，如果存在 $0 \leq \beta < 1$ 使得 $d(T(x), T(y)) \leq \beta \cdot d(x, y)$ 对所有 $x, y \in X$ 成立。

Banach 不动点定理：每个压缩映射有唯一不动点 $x^* = T(x^*)$，且对任意初始值 x_0，迭代 $x_{n+1} = T(x_n)$ 收敛到 x^*。

在 BLP 中的应用：Berry (1994) 证明，在 Logit 需求系统中，均值效用 δ_{jt} 的更新函数是压缩映射：

$$\delta_{jt}^{(k+1)} = \delta_{jt}^{(k)} + \ln s_{jt}^{\text{obs}} - \ln s_{jt}(\delta^{(k)}, \theta) \tag{A.32}$$

其中 s_{jt}^{obs} 是观测市场份额，$s_{jt}(\delta, \theta)$ 是模型预测市场份额。这一性质保证了内循环（给定 θ 求 δ）的快速收敛。

A.3.4　Bootstrap 标准误

当渐近标准误公式不可靠（如小样本或复杂非线性模型），使用 Bootstrap 进行推断。

非参数 Bootstrap：

1. 从原始样本（n 个观测值）中有放回地抽取 B 个 Bootstrap 样本，每个大小为 n
2. 在每个 Bootstrap 样本 $b = 1, \ldots, B$ 上计算估计量 $\hat{\theta}_b$
3. Bootstrap 标准误：$\widehat{\text{SE}}_{\text{boot}} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b - \bar{\theta})^2}$，其中 $\bar{\theta} = B^{-1}\sum_b \hat{\theta}_b$

在 IO 中的应用：BLP 估计中，标准误通常通过 Bootstrap 或对 GMM 目标函数的数值 Hessian 来计算。由于收缩映射和模拟积分的嵌套结构，解析标准误不易获得。

A.3.5　数值优化方法概述

表 A.1: BLP 和相关实证 IO 模型中常用的优化算法

算法	类型	IO 中的典型应用
BLP 收缩映射	不动点迭代	给定 θ，求解需求侧的均值效用 δ
MPEC (Su-Judd 2012)	约束优化	将 δ 作为优化变量，同时求解 θ 和 δ
NFP (Nevo 2000)	梯度法	嵌套不动点：外循环用梯度法搜索 θ，内循环用收缩映射求 δ
SMM	模拟矩	当市场均衡无解析解时，用模拟比较矩
Parallel BLP	并行计算	将市场份额的模拟积分分配到多核/GPU

A.4　BLP 收缩映射的压缩性证明

命题（Berry, 1994）：在 Logit 模型中，更新函数

$$T(\delta_{jt}) = \delta_{jt} + \ln s_{jt}^{\text{obs}} - \ln s_{jt}(\delta) \tag{A.33}$$

是定义在 $\mathbb{R}^J$ 上的压缩映射。

证明概要：令 $s(\delta)$ 为由 Logit 选择概率 $s_j = \exp(\delta_j)/\sum_k \exp(\delta_k)$ 给出的市场份额向量。s 关于 δ 的 Jacobian 矩阵为：

$$\frac{\partial s_j}{\partial \delta_k} = \begin{cases} s_j(1 - s_j) & j = k \\ -s_j s_k & j \neq k \end{cases} \tag{A.34}$$

$T(\delta)$ 的 Jacobian 为 $I - D^{-1}\nabla_\delta s$，其中 $D = \text{diag}(s_1, \ldots, s_J)$。可以证明该矩阵的谱半径（最大特征值的绝对值）严格小于 1，因此 T 在 sup-norm 下是压缩的。Banach 不动点定理保证了从任意 δ_0 出发的迭代指数级收敛到唯一 δ^*。

在随机系数 Logit 中的推广：当存在随机系数时，市场份额需要模拟积分。BLP (1995) 证明了在模拟矩条件下，收缩映射性质仍然成立，保证了 BLP 内循环的数值稳定性。

这些工具的更多应用细节和 IO 语境中的讨论，请参阅正文相关章节。

附录 B　中国产业组织的特殊议题

本附录聚焦中国产业组织与竞争政策的独特问题——国有企业的竞争与垄断、反垄断法的演变、平台经济的竞争格局、产业政策与竞争政策的张力，以及中国 IO 研究的前沿。

B.1　国有企业的竞争与垄断

B.1.1　国有企业的双重身份

中国国有企业（State-Owned Enterprises, SOEs）在法律上和实践中具有双重身份：作为商业主体，它们参与市场竞争、追求利润；作为政策工具，它们执行国家战略、提供公共服务。这种双重性使传统 IO 的企业分类（利润最大化的私人企业）不能完全适用。

截至 2024 年，中国约有 96 家中央企业（由国资委监管），加上数千家地方国有企业。央企在石油石化（中石油、中石化、中海油）、电力（国家电网、华能、大唐）、电信（中国移动、中国联通、中国电信）、铁路（国铁集团）、航空（国航、东航、南航）和金融（工、农、中、建、交五大行）等领域占据主导地位。

B.1.2　行政垄断 vs. 市场垄断

中国竞争政策面临的一个独特挑战是区分**行政垄断**（administrative monopoly）和**市场垄断**（market monopoly）：

- **行政垄断**：政府机关或授权的组织滥用行政权力，排除或限制竞争。例如，地方政府通过采购或补贴偏好保护本地企业（地方保护主义）、行业主管部门对特定企业给予排他性许可或准入优惠。
- **市场垄断**：企业通过市场行为——如并购、排他性合同、掠夺性定价——获得或维持市场支配地位。这与标准 IO 分析中的垄断概念一致。

2008 年《反垄断法》首次明确禁止行政垄断（第五章），但执法工具仅限于向有关上级机关” 建议”——没有罚款权。2022 年修订强化了公平竞争审查制度（Fair

Competition Review System），要求制定涉及市场主体经济活动的政策时进行竞争影响评估。

B.1.3　混合所有制改革

2013 年以来的混合所有制改革（mixed-ownership reform）试图通过引入私人资本来改善国有企业的治理和效率。从 IO 视角看，混合所有制改革提出了一系列研究问题：

- 混合所有制企业在定价和产量决策上与纯私有企业有系统性差异吗？
- 私有资本的引入是否降低了行政垄断行为？
- 混合所有制的治理结构如何影响企业的创新投入和动态效率？

现有研究（如陈林和王志强，2016；Hsieh & Song, 2015）表明，部分私有化确实提高了生产效率，但对市场势力行使的影响更为复杂，取决于行业特征和监管环境。

B.2　中国反垄断法的演变与执法

B.2.1　2008-2020：从无到有的制度建设

中国《反垄断法》（Anti-Monopoly Law, AML）于 2008 年 8 月 1 日正式施行，结束了中国没有综合竞争法的历史。2008-2020 年间，三大反垄断执法机构（商务部、发改委、工商总局）分头执法——商务部负责经营者集中审查，发改委负责价格相关垄断行为，工商总局负责非价格垄断行为。2018 年国务院机构改革将三者合并为国家市场监督管理总局（SAMR），实现了统一执法。

B.2.2　2022 年修订：数字经济的回应

2022 年 6 月修订（2022 年 8 月 1 日生效）是《反垄断法》施行 14 年来的首次重大修订。关键变化包括：

1. **数字经济条款**：新增第 9 条——”经营者不得利用数据和算法、技术、资本优势以及平台规则等从事本法禁止的垄断行为。”首次在法律层面明确数据和算法可以构成垄断行为的工具。

2. **处罚力度大幅提升**：垄断协议和滥用市场支配地位的罚款上限从年销售额的 10% 提高——对于情节特别严重的，罚款可达上一年度销售额的 50% 以上，并可能处以销售额 5 倍的惩罚性罚款。

3. **公平竞争审查入法**：将公平竞争审查制度写入法律（第 5 条），从行政规范升级为法定义务。

4. ”安全港”规则：对于纵向垄断协议，市场份额低于规定标准的企业可以适用”安全港”豁免（第 18 条）。

5. **个人责任**：首次引入对垄断行为直接负责的个人处罚（法定代表人和主要负责人，最高 100 万元罚款）。

B.2.3 标志性执法案例

表 B.1: 中国反垄断标志性执法案例

案件	年份	罚款	核心问题
高通垄断案	2015	60.88 亿元	滥用标准必要专利（SEP）市场支配地位——收取不公平高价许可费、捆绑销售。中国反垄断史上第一个涉及知识产权滥用的重大案件。
阿里巴巴”二选一”	2021	182.28 亿元	滥用市场支配地位——对平台内商家实施”二选一”排他性要求。罚款 = 2019 年中国境内销售额的 4%。是中国反垄断史上最高罚款。
美团”二选一”	2021	34.42 亿元	滥用市场支配地位——对餐饮外卖商家实施”二选一”。要求退还独家合作保证金 12.89 亿元。
知网垄断案	2022	8760 万元	滥用市场支配地位——对学术数据库服务收取不公平高价。罚款 = 2021 年中国境内销售额的 5%。

B.2.4 中国反垄断法的哲学基础

与芝加哥的”消费者福利标准”和新布兰代斯的”多目标标准”不同，中国的竞争政策哲学体现出特有的混合特征。中国《反垄断法》第 1 条将立法目标定为”保护公平竞争、提高经济运行效率、维护消费者利益和社会公共利益、促进社会主义市场经济健康发展”——涵盖公平、效率、消费者利益、公共利益和发展五个维度。这种”多目标”框架——原则上是新布兰代斯所赞同的——在中国被嵌入一个非常不同的制度语境和政治经济体系中（见 第 ~15~ 章 §15.5 的讨论）。

B.3　中国平台经济的竞争格局

B.3.1　市场结构演变

中国平台经济的市场结构在过去十年经历了快速演变。从早期的 BAT 三足鼎立（百度、阿里、腾讯），到 ATM（阿里、腾讯、美团），再到字节跳动（抖音/TikTok）和拼多多的崛起，竞争格局一直在高速动态变化。

- **电商**：阿里巴巴（淘宝、天猫）vs. 京东 vs. 拼多多 vs. 抖音电商——四强格局，拼多多凭借社交裂变和低价策略在 2018–2024 年间从阿里手中抢走了巨大市场份额
- **社交/通讯**：微信（腾讯）占据即时通讯领域的压倒性份额，但面临抖音私信和办公通讯（钉钉、飞书）在部分场景的竞争
- **本地生活**：美团在外卖领域保持高份额（约 65–70%），但抖音本地生活（到店团购）从 2021 年起迅速增长，成为美团在到店业务上的重要竞争者
- **短视频**：抖音 vs. 快手——双寡头格局
- **移动支付**：支付宝 vs. 微信支付——双寡头格局

B.3.2　”二选一”——中国版本的排他性行为

”二选一”（either-or choice）是中国平台经济中出现频率最高的反垄断议题。大型平台要求商家不得同时在其他平台上运营，或以其他平台为条件限制优惠资源分配。从 IO 理论看，”二选一”类似于排他性交易（exclusive dealing），但其竞争效应评估取决于具体条件：在平台市场，排他性可能（a）阻止竞争对手达到临界规模（反竞争）或（b）保护平台的投资激励（效率抗辩）。

2021 年阿里和美团的”二选一”处罚标志着中国反垄断执法对平台排他性行为的系统性打击。此后，”二选一”行为大幅减少。

B.3.3　平台用工：IO 与劳动经济学的交叉

中国平台经济还提出了 IO 与劳动经济学交叉的新问题。外卖骑手、网约车司机和快递员等平台劳动者是否属于”独立承包商”还是”雇员”？平台的算法管理（定价算法、调度算法、绩效评分算法）对劳动者的收入和劳动条件有什么影响？这些问题涉及买方垄断（monopsony）——平台作为劳动力的单一或主要买家可能压低收入——这一概念在 第 ~2~ 章 §2.6（Robinson 的买方垄断分析）和现代劳动力市场集中度研究（如 Azar et al., 2020）中有详细讨论。

B.4 产业政策与竞争政策的张力

中国的竞争政策在一个产业政策（industrial policy）始终占据重要地位的经济体中运行。这两个政策框架之间存在根本性的张力：

- **产业政策的逻辑**：市场存在协调失灵——政府需要通过补贴、保护、准入管制和选择性支持来培育战略产业、推动技术追赶、维护经济安全
- **竞争政策的逻辑**：市场势力损害效率和消费者——政府需要通过禁止垄断行为、审查并购、拆除壁垒来维护竞争过程

B.4.1 典型案例

新能源汽车：中国通过补贴、政府采购和准入管制培育了全球最大的新能源汽车市场（比亚迪在 2024 年成为全球最大的电动车制造商）。然而，电池和整车制造领域的高度集中也带来了竞争担忧。

半导体：面对美国的技术出口管制，中国加大了对半导体产业的国家支持（国家大基金）。这在逻辑上可能导致更加集中的市场结构和更高的进入壁垒——两者都在传统 IO 分析中被视为竞争风险。

AI 大模型：2023–2025 年中国出现了 100 多个大语言模型（LLM）——百度文心一言、阿里通义千问、字节云雀等——形成竞争性市场结构。产业政策在这一领域是”促竞争”（没有指定唯一”冠军”企业）还是会在未来转向集中化支持，仍待观察。

B.4.2 ”全国统一大市场”

2022 年 4 月，中共中央、国务院发布《关于加快建设全国统一大市场的意见》，旨在消除地方保护主义、统一市场规则、促进要素自由流动。从 IO 角度看，建设统一大市场的本质是通过降低行政性进入壁垒来解决地方市场的碎片化——这与 SCP 范式（第 ~3~ 章）和可竞争市场理论（第 ~6~ 章）的核心关切高度吻合。如果成功，全国统一大市场将降低企业的合规成本、扩大有效市场规模，并增强跨区域竞争。

B.5 中国 IO 研究前沿学者与机构

中国的产业组织经济学研究在过去二十年取得了显著进展。以下列出具有代表性的研究机构和学者（按拼音排序，非详尽）：

主要研究机构

- **中国社会科学院**（CASS）：数量经济与技术经济研究所、工业经济研究所

- **北京大学**：光华管理学院（周黎安、刘俏）、国家发展研究院（姚洋、黄益平）、经济学院
- **清华大学**：经济管理学院（钱颖一、白重恩）、五道口金融学院
- **复旦大学**：经济学院（陈钊、陆铭）
- **中国人民大学**：经济学院（刘守英、杨瑞龙）、汉青经济与金融高级研究院
- **上海交通大学**：安泰经济与管理学院（陆铭）、上海高级金融学院
- **香港中文大学**：经济系（宋铮、龚启圣）

代表性研究方向

- **结构估计与生产率**：将 Olley-Pakes、De Loecker 等方法应用于中国工业企业数据，估计加成率和资源配置效率（Hsieh & Klenow, 2009; 尹恒、李世刚等）
- **反垄断经济学的中国路径**：探讨中国反垄断执法的经济学基础——如何在中国语境中平衡消费者福利、效率和产业发展（黄勇、王晓晔等）
- **平台经济规制**：中国平台经济的竞争分析、数据产权与竞争秩序（荆文君、韩伟等）
- **国有企业改革与竞争**：混合所有制改革对市场竞争和企业行为的影响

推荐阅读（中国语境）

- 吴汉洪 (2012). 《产业组织理论》. 中国人民大学出版社. ——中国学者撰写的 IO 研究生教材，将国际 IO 框架与中国案例结合。
- Hsieh, C.-T. & Klenow, P. (2009). "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4), 1403–1448. ——用微观数据度量中国制造业的资源配置扭曲，是结构估计方法在中国数据上的经典应用。
- 荆文君等 (2025). "Platform Cross-Border Operation, Market Power Changes and Regional Economy High-Quality Development." 《重庆大学学报》. ——平台企业跨境经营对区域经济的影响。
- 韩伟 (2022). "2022 年《反垄断法》对平台经济的回应与展望." 《信息通信技术与政策》. ——中国 2022 年反垄断法修订的详细分析。

- Huang, Y. & Wang, X. (2019). "Antitrust in China: A Historical and Comparative Perspective." In *Research Handbook on Asian Competition Law*. ——中国反垄断的历史考察与国际比较。

B.6 中国数据资源与实证 IO 研究机会

对于有志于用中国数据从事 IO 实证研究的读者，以下数据资源可能是富有成果的起点：

- **中国工业企业数据库**（CIE Database）：1998–2013 年国有及规模以上非国有工业企业的面板数据。是生产率估计（OP/LP 方法，第 ~10~ 章）、加成率测算（De Loecker 方法）、市场集中度（CR4, HHI）和进入/退出动态的标准数据库。
- **中国海关贸易数据库**（Chinese Customs Trade Database）：2000–2016 年进出口交易级数据。可与工业企业数据库匹配，用于分析贸易自由化对加成率和市场竞争的影响。
- **国家统计局普查数据**：2004、2008、2013 和 2018 年经济普查微观数据。
- **上市公司数据**：CSMAR 和 Wind 数据库，包含详细的财务报表和治理信息。
- **平台数据**：阿里巴巴、京东、美团、抖音等平台的交易数据（部分可通过研究合作获取）。
- **反垄断公示**：国家市场监督管理总局（SAMR）反垄断局定期公布的经营者集中审查决定、行政处罚决定书——其中的市场界定和经济分析提供了了解中国竞争执法的原始材料。

回到正文：→ 第 ~1~ 章 什么是产业组织学