

基于分层图限制博弈Shapley值的大湾区跨境数据要素研究

燕文捷

上海理工大学管理学院, 上海

收稿日期: 2025年12月12日; 录用日期: 2026年1月30日; 发布日期: 2026年2月9日

摘要

数字经济迅速发展使数据成为重要的生产要素。粤港澳大湾区在“一国两制三法域”之下存在合规摩擦, 传统的合作博弈难以刻画通道约束。因此本文提出了一种基于图限制博弈的分层蒙特卡洛Shapley值(SGR-MC-Shapley)。在嵌套CES生产函数中引入数据要素后, 利用Myerson值将城市间合规通道表示为合作图边并内化其成本权重, 构造图限制特征函数, 并进一步设计分层抽样和截断策略来降低大规模估值方差。以2024年深圳、香港、澳门等地数据为基础开展情景模拟。研究表明, 深圳体量虽仍居核心, 但合规壁垒下降时香港作为跨境“联系人”的边际贡献提升最显著。研究为跨境数据流动机制和生态补偿制度提供量化的依据。

关键词

CES生产函数, 图限制博弈, 粤港澳大湾区, Shapley值, 跨境数据流动

A Study on Cross-Border Data Elements in the GBA Based on the Shapley Value of Hierarchical Graph-Constrained Games

Wenjie Yan

Business School, University of Shanghai for Science and Technology, Shanghai

Received: December 12, 2025; accepted: January 30, 2026; published: February 9, 2026

Abstract

The rapid growth of the digital economy has caused data to become a factor of production. Within the Guangdong-Hong Kong-Macao Greater Bay Area, due to institutional differences that arise under the

framework of “one country, two systems, three legal jurisdictions”, there are compliance frictions that are unable to fully be accounted for by traditional cooperative games models that do not restrict the formation of coalitions. So this study put forth a graph-restricted Monte Carlo Shapley value with stratification (SGR-MC-Shapley). After introducing the data factor into the nested CES production function, Myerson value on inter-city compliance channels is represented with a weighted edge on a cooperation graph with endogenized regulatory cost to obtain a graph restricted characteristic function. In order to solve for the variance to scale from a large scale valuation, in order to further do a stratified sampling and truncation. Scenarios are set up using Shenzhen, Hong Kong, Macao in 2024 as the data source. Result shows although Shenzhen remains the main source because of its economic scale, due to lower compliance barrier, the biggest increase in marginal contribution from Hong Kong takes place, which shows its role in cross-border connecting. This gives a quantitative basis for designing cross-border data flow mechanism and digital ecosystem compensation programmes.

Keywords

CES Production Function, Graph-Constrained Games, Guangdong-Hong Kong-Macao Greater Bay Area, Shapley Value, Cross-Border Data Flows

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 绪论

在数字化浪潮之下，数据已从信息载体跃升为驱动增长的关键生产要素[1]。粤港澳大湾区作为国家战略高地，汇聚深圳的科技创新、香港的国际金融与澳门的文旅服务，其要素流动效率直接关系区域协同与数字经济一体化水平[2][3]。然而，数据要素具有非竞争性、可低成本复制与强外部性等特点，更重要的是大湾区处于“一国两制三法域”的制度结构之下，内地《个人信息保护法》对数据出境设置了安全评估与标准合同等要求，香港、澳门则分别适用不同的数据治理体系[4][5]，因此形成隐性的制度壁垒与合规摩擦，使得跨境数据合作不是任意主体可无摩擦的结盟，而更接近于受合规通道约束的网络化合作的问题。

在此现实约束下，跨区域数据协同增量怎样被科学衡量并公平分配，面临三重挑战：第一，数据同资本、劳动之间存在替代和互补并存的非线性关系，需要更具解释力的生产函数刻画；第二，制度性通道的限制使联盟的价值取决于节点的连通性，传统 Shapley 值的完全图假设无法处理该类拓扑约束；第三，随着参与主体的增加，精确计算 Shapley 值的复杂度呈指数爆炸，在图限制条件下叠加连通分量判定，计算负担进一步加重。

为回应上述问题，本文构造了分层图限制蒙特卡洛 Shapley 框架。将数据要素引入嵌套 CES 生产函数，并纳入合规成本参数，以图限制博弈描述跨法域合规通道对联盟价值生成的约束；结合分层抽样与截断蒙特卡洛来设计高效的近似估值算法；最后利用大湾区核心城市的宏观数据进行不同的政策情景模拟，量化制度成本变化对价值分配结构的影响，从而给跨境数据流动机制与收益分配制度设计提供可操作的量化依据。

2. 理论基础

2.1. 数据要素与 CES 生产函数

Cobb-Douglas 函数作为经济学中描述投入与产出关系的核心工具[6]，因其形式简单被广泛使用，但

在数字经济时代其隐含的单位替代弹性($\sigma=1$)假设显得过于严苛。Arrow 等(1961)提出的常替代弹性(CES)生产函数允许要素间的替代弹性为任意非负常数,能更灵活地刻画现实场景[7]。近年来,学界开始将数据作为独立要素纳入 CES 框架。Jones 和 Tonetti (2020)认为数据的非竞争性使其具有显著的规模报酬递增性[8]。而在大湾区内部,不同节点的生产函数参数可能存在显著差异。本文将采用嵌套 CES 结构,以区分资本-劳动合成要素与数据要素之间的层级替代关系。

2.2. 图限制博弈与 Shapley 值

Shapley 值(1953)是解决合作博弈中利益分配问题的经典方法[9],根据参与者的平均边际贡献进行分配,然而经典的 Shapley 值假设联盟结构是全连通的。但在现实经济网络中,合作往往受限于通信架构或制度条件。Myerson (1977)引入了图限制博弈(Graph-Restricted Games)的概念,定义了 Myerson 值[10]。在 Myerson 的框架下,如果两个参与者之间没有直接或间接的路径相连,他们就无法在同一个联盟中通过合作创造额外价值。考虑大湾区跨境数据需通过特定通道传输的现实场景,此时合作博弈的特征函数 $v(S)$ 被修正为 $v^G(S)$,即联盟 S 在图 G 中各个连通分量上的价值之和。

2.3. 近似计算算法

对于大规模数据集或网络,精确计算 Shapley 值属于 #P-complete 问题,其中蒙特卡洛采样是最主流的近似方法,用随机抽取排列来估计边际贡献的期望。Maleki 等(2013)引入分层采样[11],将排列空间按联盟大小进行分层,并在层内进行独立采样,进一步提高收敛速度并降低方差,可以有效地降低估计误差。另外在大规模分类任务中,截断蒙特卡洛方法会忽略掉那些边际贡献接近于零的样本,进一步提升了计算效率。本文将结合这些算法思想,针对图限制博弈的特殊性来对传统计算模型进行重构。

3. 模型构建

3.1. 模型基本假设

假设一个由 $N=1,2,\dots,n$ 个城市(或行业节点)构成的经济系统。每个节点 i 拥有初始禀赋,大湾区的制度环境可以被建模成一个赋权无向图 $G=(N,E)$ 。详细禀赋设定见表 1。

Table 1. Initial endowment setting

表 1. 初始禀赋设定

禀赋	定义
传统资本 (K_i)	基础设施、厂房、设备等存量资产
劳动投入 (L_i)	就业人口数量,可进一步细分为普通劳动和高技能劳动
数据要素 (D_i)	可数字化的信息资源存量,经标准化处理后的规模指标
节点 N	代表大湾区城市(如深圳、香港、澳门、广州等)
边 E	代表城市间可行的数据流通通道
权重 w_{ij}	代表数据在 i 和 j 之间流动的“摩擦成本”,包括合规评估费用、技术改造成本及法律风险溢价等。

3.2. 嵌套 CES 生产函数设定

为了捕捉数字经济的特征,我们构建一个两层嵌套的生产函数。

1) 内层:传统要素合成

首先，资本 K_i 和劳动 L_i 结合生成传统经济产出潜能 Z_i 。假设二者为互补关系(替代弹性 $\sigma_1 < 1$):

$$Z_i = A_{1,i} [\alpha_i K_i^{\rho_1} + (1 - \alpha_i) L_i^{\rho_1}]^{\frac{1}{\rho_1}}$$

其中 $\rho_1 = \frac{\sigma_1 - 1}{\sigma_1}$ ， α_i 为资本的分配参数。

2) 外层：数字赋能产出

数据要素 D 与传统合成要素 Z 结合，生成最终产出 Y_i 。此处引入网络效应参数 λ 。当多个城市组成联盟 S 时，其共享的数据总量并非简单的相加，而是具有规模效应：

$$D_{pool}(S) = \left(\sum_{j \in S} D_j \right)^\lambda$$

其中 $\lambda \geq 1$ 体现了梅特卡夫定律类型的网络价值递增。则最终产出函数为：

$$Y_i(S) = A_{2,i} [\beta_i \cdot (D_i(S))^{\rho_2} + (1 - \beta_i) \cdot Z_i^{\rho_2}]^{\frac{1}{\rho_2}}$$

其中 $\sigma_2 = \frac{1}{1 - \rho_2}$ 为数据与传统要素的替代弹性； $\theta_i(S)$ 为数据渗透率，表示联盟 S 中的数据有多少能被节点 i 实际吸收利用。

3.3. 基于图限制的特征函数

在 Myerson 框架下，联盟 S 的价值由它在图 G 中的连通性来决定。设 $S|G$ 为联盟 S 在图 G 上的诱导子图。该子图可以由若干个不相交的连通分量 C_1, C_2, \dots, C_k 组成，即 $S = \bigcup_{r=1}^k C_r$ 。则对于任意连通分量 C_r ，其产生的总经济盈余 $V(C_r)$ 为各成员产出之和减去维持连接的合规成本：

$$V(C_r) = \sum_{i \in C_r} Y_i(C_r) - \sum_{(i,j) \in E(C_r)} Cost(w_{ij})$$

其中 $Cost(w_{ij})$ 是关于边权重的函数。若 i, j 属于同一法域(如深圳 - 广州)， $w_{ij} \rightarrow 0$ ；若跨法域(如深圳 - 香港)， w_{ij} 显著为正。

整个联盟 S 的特征函数定义为各连通分量价值之和：

$$v^G(S) = \sum_{r=1}^k V(C_r)$$

可以得出在 Myerson 值中，如果两个节点在联盟中无法连通，那么它们就无法产生协同效应，其贡献退化为各自独立或局部连通时的产出。

3.4. 制度摩擦系数

3.4.1. 摩擦系数构造

在大湾区“一国两制三法域”的特殊情境下，图 $G = (N, E)$ 中的边权重 w_{ij} 并非简单的物理距离，而是代表了数据跨境流动的综合制度摩擦系数。我们将其解构为“政策距离”与“技术合规成本”的非线性函数。

定义节点 i 与节点 j 之间的摩擦系数 w_{ij} 如下：

$$w_{ij} = \underbrace{\delta(R_i, R_j) \cdot (1 - \eta_{policy}) \cdot C_{legal}}_{\text{制度性合规成本}} + \underbrace{C_{tech}(D_{flow})}_{\text{技术性传输成本}}$$

其中：

R_i, R_j : 表示节点所属的法域, 取值集合为内地、香港、澳门; $\delta(R_i, R_j)$: 法域指示函数。

当同为内地城市, $R_i = R_j$, $\delta = 0$, 表示不存在跨境法律障碍; 若 $R_i \neq R_j$, 如深港之间, $\delta = 1$, 则激活跨境合规成本。

C_{legal} : 基准合规成本。代表在当前法律框架下, 出境安全评估、标准合同备案所需的法律咨询、审计及时间成本的货币化估值。

$\eta_{policy} \in [0,1)$: 政策优惠因子; $\eta = 0$: 代表基准现状; $\eta \rightarrow 1$: 代表“数字自贸区”或“白名单”机制下的极简审批。《大湾区标准合同》的实施可视为 $\eta \approx 0.5$ 。

$C_{tech}(D_{flow})$: 与数据流量 D_{flow} 相关的技术成本, 通常设定为线性函数 $k \cdot D_{ij}$ 。

基于此摩擦系数, 联盟 S 的净价值函数修正为:

$$v^G(S) = \sum_{r=1}^k \left(\sum_{i \in C_r} Y_i(C_r) - \sum_{(i,j) \in MST(C_r)} w_{ij} \right)$$

注: 此处采用了最小生成树(Minimum Spanning Tree, MST)假设, 在连通分量 C_r 内部, 数据流动会自动选择总摩擦成本最低的路径进行传输。

3.4.2. 非对称性摩擦的处理

在现实中, 数据“流入”与“流出”的合规要求可能不同, 即 $w_{i \rightarrow j} \neq w_{j \rightarrow i}$ 。为在不改变本文图限制博弈无向合作图主体框架的情况下吸收该特征, 本文拟采用保守合成规则, 将无向边权定义为两方向摩擦的上确界:

$$\tilde{w}_{ij} = \max \{ w_{i \rightarrow j}, w_{j \rightarrow i} \}$$

其经济含义为: 只要任一方向受到较强合规约束, 双边协同就会被显著抑制。为避免低估跨境约束, 该处理在政策评估中更为审慎。

3.5. 风险惩罚项

仅考虑产出增益 - 合规成本的联盟价值可能高估大规模跨境协作的净收益。为体现统筹发展与安全的政策约束, 本文在连通分量价值函数中引入风险惩罚项, 将数据安全事件的预期损失从总收益中扣除。设联盟 S 在图 G 上诱导子图的连通分量为 $\{C_r\}$, 则将分量价值从原先形式扩展为:

$$V(C_r) = \sum_{i \in C_r} Y_i(C_r) - \sum_{(i,j) \in MST(C_r)} w_{ij} - \mathbb{E}[Loss(C_r)]$$

其中 $\mathbb{E}[Loss(C_r)]$ 为风险损失的期望值。我们将其拆为发生概率 \times 损失规模:

$$\mathbb{E}[Loss(C_r)] = p(C_r) \cdot L(C_r)$$

考虑到联盟规模扩大与跨境连接增多会增加攻击面与治理复杂度, 本文采用可校准的指数型风险函数:

$$p(C_r) = 1 - \exp \{ -\gamma_1 |C_r| - \gamma_2 m_r^{cross} \}, \quad \gamma_1, \gamma_2 \geq 0$$

其中 $|C_r|$ 为分量节点数, m_r^{cross} 为分量内部跨法域连接的数量, 对应值取 $MST(C_r)$ 上的跨境边数, γ_1 刻画规模效应带来的风险上升, γ_2 刻画跨境复杂度带来的额外风险。

风险事件的损失规模可与联盟数据规模或经济产出挂钩, 本文给出对应设定, 在实证中作为稳健性对照:

$$L(C_r) = \ell_D \cdot D_{C_r}^{pool}$$

其中 ℓ_D 为单位数据池风险损失系数。

4. 分层图限制蒙特卡洛 Shapley 框架

鉴于大湾区城市节点之间的复杂连接关系及嵌套 CES 函数的非线性，直接计算 Shapley 值在计算上是不可行的。在此提出一种改进的估值算法。

4.1. 算法逻辑框架

Shapley 值的本质就是计算参与者 i 在所有可能的排列中平均边际贡献。而对于图限制博弈，关键是判断当 i 加入现有联盟 S 时，是否改变了 S 的连通性，例如，连接了两个原本独立的连通分量，又或是仅仅是依附于一个现有的分量。

SGR-MC-Shapley 算法流程，流程图见图 1。

1) 分层：将所有可能的联盟按 Size 分为 N 个层级。第 k 层包含所有大小为 k 的联盟。根据 Shapley 公式，不同大小的联盟权重不同且相关研究表明，数据要素的边际贡献在联盟构建之初和末期往往波动最大，采用分层抽样可以有效地降低这些区域的方差。

2) 蒙特卡洛采样：在每一层 k 中，随机抽取 M_k 个排列或子集。对于每个样本，计算节点 i 的边际贡献：

$$\Delta_i(S) = v^G(S \cup \{i\}) - v^G(S)$$

计算 v^G 时，使用连通分量搜索算法 BFS，以确定加入 i 后图结构的变化。

3) 截断优化：当联盟规模达到一定程度，或者当新加入节点的合规成本远超其带来的数据增量价值时，边际贡献 $\Delta_i(S)$ 会迅速趋近于零。设阈值 ϵ ，连续多次采样 $\Delta_i < \epsilon$ 时，则停止该路径的深入搜索。

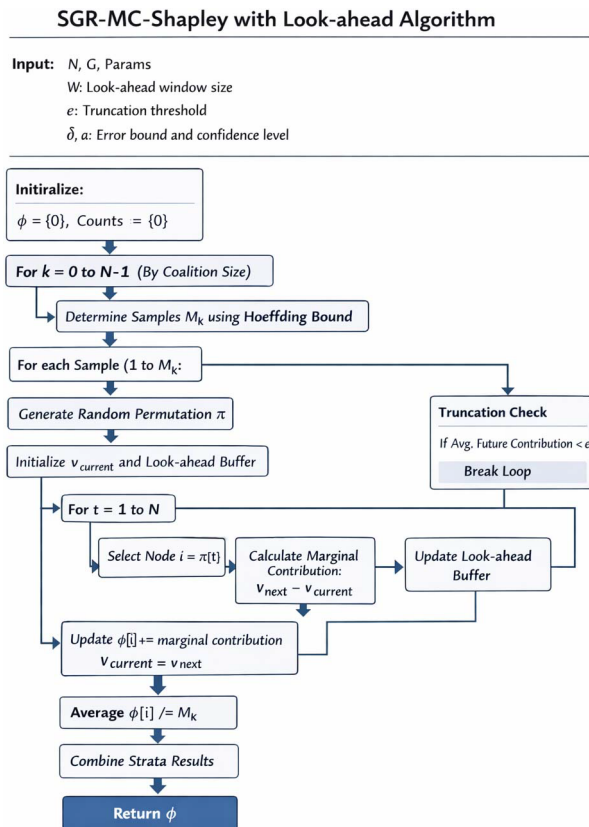


Figure 1. Flowchart of the SGR-MC-Shapley algorithm

图 1. SGR-MC-Shapley 算法流程图

4.2. 基于前瞻性窗口的截断蒙特卡洛算法

传统的截断 Shapley 算法一般在发现某一步的边际贡献 $\Delta_i(S)$ 低于阈值 ϵ 的时候就停止计算。然而在本研究嵌套 CES 生产函数中，存在显著的网络效应参数 $\lambda > 1$ 。这意味数据价值可能呈现“S型”曲线——在联盟形成的平台期，数据量的增加带来的边际贡献较小，直到达到某一临界规模之后，才会出现价值跃升。采用简单的即时截断，很容易在平台期过早停止，从而系统性地低估了后加入节点，尤其是长尾节点在“相变点”附近的巨大贡献。因此，我们设计了自适应前瞻性截断机制。

4.2.1. 前瞻性截断准则

定义 W 为前瞻窗口长度。算法并非在 $\Delta_i(S_t) < \epsilon$ 时立即停止，而是计算未来 W 步的移动平均边际贡献：

$$\bar{\Delta}_{t,W} = \frac{1}{W} \sum_{\tau=0}^{W-1} \Delta_i(S_{t+\tau})$$

仅当 $\bar{\Delta}_{t,W} < \epsilon$ 且当前的联盟规模 $|S_t| > N_{\min}$ 时，才会触发截断。该机制使算法可以暂时的跨越低贡献区，去探测后续因跨法域节点加入并连通子图后可能带来的价值突增。

在图限制博弈中，如果想要连接到包含香港的金融数据池，一个孤立节点或小联盟往往需要经过若干“桥梁节点”才能到达高价值的大型连通分量。因此， W 的理论下界应与图的局部连通特征相关，具体取值应为网络中的平均最短路径长度或关键桥接路径的跳数。

在大湾区城市群网络中，核心节点(深、港、广、澳)之间的连接一般呈现“三角形”或“菱形”结构。例如从惠州连接到澳门，可能需要经过“惠州 → 深圳 → 香港 → 澳门”或“惠州 → 深圳 → 珠海 → 澳门”。这一路径通常包含 2 到 3 个跳跃。如果 $W < 2$ ，算法在加入惠州后，可能未连接到澳门就停止计算，从而错失了整个链条打通后的价值，如图 2 所示。因此，考虑到大湾区城市群网络的平均路径长度特征约为 2.1 以及嵌套 CES 生产函数中网络效应的滞后性，本文将前瞻性窗口设定为 $W = 3$ 。

即算法遇到边际贡献低于阈值 ϵ 的情况时，会继续模拟 3 个随机节点的加入。如果在这三步内，平均贡献回升至 ϵ 以上，则继续采样；否则，认为该路径已进入边际收益递减的尾部，执行截断。

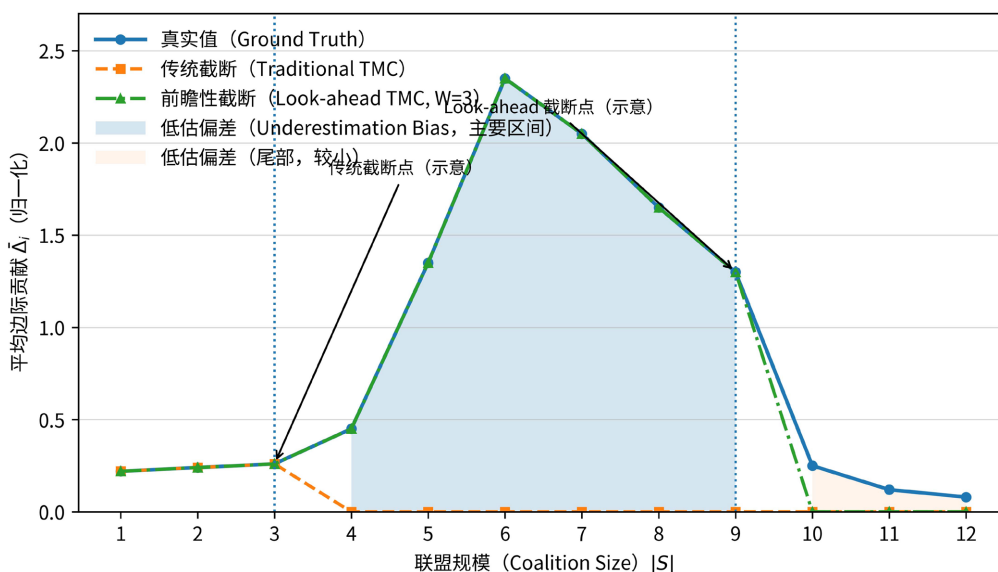


Figure 2. Look-ahead versus traditional truncation mechanisms
图 2. Look-ahead 与传统截断机制对比

4.2.2. 基于霍夫丁界限的采样终止条件

为了平衡计算效率与估值精度，本文不再对于每一层 k 的蒙特卡洛采样次数 M_k 设定固定值，而是使用霍夫丁不等式来动态地控制。

假设要以 $1-\alpha$ 的置信度，使 Shapley 值的估计误差控制小于 δ 。对于取值范围在 $[a,b]$ 之间的边际贡献，所需的最小样本量 M_k 应满足：

$$P\left(|\hat{\phi}_i - \phi_i| \geq \delta\right) \leq 2 \exp\left(-\frac{2M_k \delta^2}{(b-a)^2}\right) \leq \alpha$$

由此推导出的动态终止条件为：

$$M_k \geq \frac{(b-a)^2}{2\delta^2} \ln\left(\frac{2}{\alpha}\right)$$

实时监控样本方差，当采样次数满足上述不等式时，该层级的计算自动收敛并停止，从而确保算法在主要贡献者和次要贡献者的估值上具有同等的统计显著性，避免由于采样不足造成的随机排序误差。

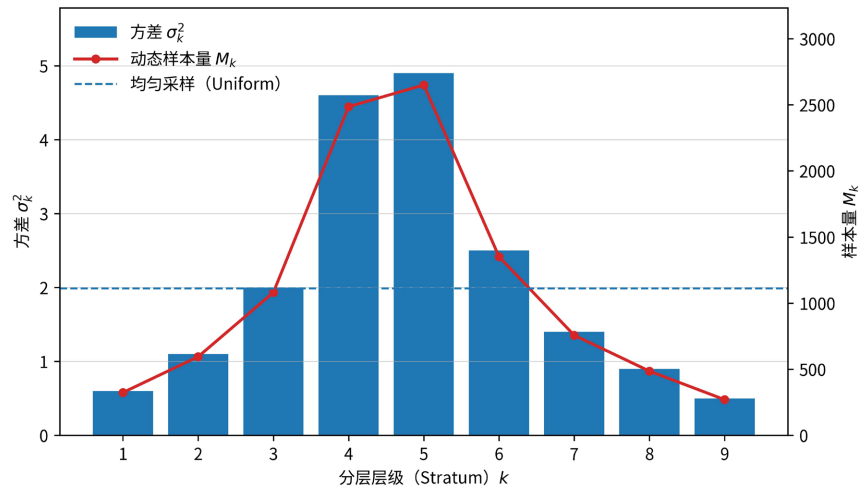


Figure 3. Dynamic sample size distribution under the Hoeffding bound
图 3. 基于霍夫丁界限的样本量动态分布图

图 3 中给出了基于霍夫丁界限的动态采样策略，算法依据分层估值的方差 σ_k^2 自适应分配样本量 M_k 。可以看到，在中间层级(如 $k = 4, 5$)由于联盟组合与图连通结构方差达到峰值，算法应相应提高该层样本量；而在两端层级方差较小，样本量显著低于均匀采样基线。与均匀采样相比，动态策略将约一半左右的计算资源集中于关键层级，从而在给定置信参数下实现更高的总体计算效率。

5. 实证分析

本章节利用 2024 年的大湾区经济数据，应用上述模型对大湾区核心城市的数字经济价值进行实证测算。

5.1. 数据来源与参数校准

5.1.1. 基础经济数据

我们选取大湾区最具代表性的四个节点——深圳、香港、澳门、广州，并以其他城市为背景节点，

构建一个简化但典型的网络结构，标准化设定见表 2。

深圳：根据深圳市统计局及有关报告，2024 年深圳市 GDP 为 3.68 万亿元人民币，同比增长 5.8%。数字经济核心产业增加值占 GDP 比重极高，战略性新兴产业占比约 41.9%。我们将深圳设定为“高数据禀赋、高技术资本”节点。

广州：2024 年广州实现地区生产总值 3.1 万亿元，其第三产业占比超 70%，对经济增长的贡献率达到 91.4%，呈现出商贸 + 先进制造的双轮驱动特征，同时积累有庞大的供应链与消费交易数据。我们将广州设定为“高综合资本、强主要劳动供给、泛流通数据”节点。

香港：2024 年 GDP 约为 2.9 万亿元人民币，服务业占比超 93%。虽然传统制造业空心化，但其在金融科技、跨境贸易数据方面拥有高质量存量。我们将香港设定为“高金融资本、高质量垂直数据”节点。

澳门：2024 年 GDP 约 4033 亿澳门元(约 3660 亿元人民币)。澳门产业结构高度单一，但在“1 + 4”适度多元发展策略下，高新技术产业正在起步。设定为“低数据总量、特定场景(旅游)数据”节点。

Table 2. Standardized factor endowment settings of core nodes in the GBA

表 2. 大湾区核心节点要素禀赋设定标准化处理

城市	传统资本	劳动投入	数据存量	产业特征
深圳(基准)	100	100	100	硬件、互联网、AI
香港	90	45	60	金融、贸易、专业服务
广州	95	92	85	商贸、制造、政务
澳门	15	5	10	旅游、博彩、葡语系服务

5.1.2. 生产函数参数校准

本研究采用嵌套 CES 生产函数，涉及关键参数包括：内层替代弹性 σ_1 (资本 - 劳动)、外层替代弹性 σ_2 (传统要素 - 数据)、分配参数 α, β 以及网络效应参数 λ 。

1) 内层替代弹性 σ_1

宏观经济学主流文献(Antràs, 2004; Klump *et al.*, 2007)发现[12] [13]，在发达经济体中，总资本与总劳动的替代弹性显著小于 1，通常在 0.4~0.7 之间，表现为互补关系。即单纯增加资本，如果没有相应的劳动力配合，产出增长有限。考虑到深圳、香港的高技术产业特性，资本与高技能劳动的互补性更强。

设定 $\sigma_1 = 0.6$ ， $\rho_1 = \frac{0.6-1}{0.6} \approx -0.67$ ，体现了较强的互补刚性。

2) 外层替代弹性 σ_2

数据要素作为新型生产力，其主要价值在于对传统决策流程的优化和替代。因此，数据与传统合成要素之间应该呈现替代关系， $\sigma_2 > 1$ 。参考 Jones & Tonetti (2020)关于数据非竞争性的模型设定[8]，以及相关学者关于 AI 对劳动替代率的测算(Acemoglu & Restrepo, 2018) [14]，数字技术对低效传统要素的替代弹性一般设定在 1.2 至 1.5 之间。

设定 $\sigma_2 = 1.3$ ，则 $\rho_2 = \frac{1.3-1}{1.3} \approx 0.23$ ，可以保证引入数据要素能显著提升全要素生产率，且不存在边际报酬过快速递减的问题。

3) 网络效应参数 λ

梅特卡夫定律认为网络价值与节点平方成正比($\lambda = 2$)，但这通常适用于通信网络。对于数据挖掘而言，数据量翻倍带来的模型精度提升通常服从对数律或幂律。根据 Hestness *et al.* (2017)在深度学习领域的“缩放定律”，模型性能与数据量的幂律指数通常在 0.07 到 0.5 之间，但这仅指单一模型精度[15]。从宏观经济

角度来看，数据存在局部规模报酬递增，而多场景复用会放大这一效应。保守设定 $\lambda = 1.1$ ，即数据总量增加 10%，能带来约 11% 的有效信息增量。这既体现了规模优势，又避免了 $\lambda \geq 2$ 导致的模型数值爆炸。

4) 城市异质性参数 β_i

利用各城市数字经济核心产业增加值占 GDP 比重作为代理变量进行反推城市异质性参数。深圳的数字经济占比高(约 42%)，设定 $\beta_{SZ} = 0.4$ ；香港以服务业为主，金融科技依赖度高，设定 $\beta_{HK} = 0.35$ ；澳门产业结构较传统(博彩服务)，数字化转型初期，设定 $\beta_{MO} = 0.15$ ；广州及其他城市设定为平均水平 $\beta_{Others} = 0.25$ 。

5.2. 模拟情景设计

为了分析制度安排对价值分配的影响，我们设计了三种情景：

1) 基准情景：维持当前的合规成本现状， $w_{cross} = 0.30$ 。

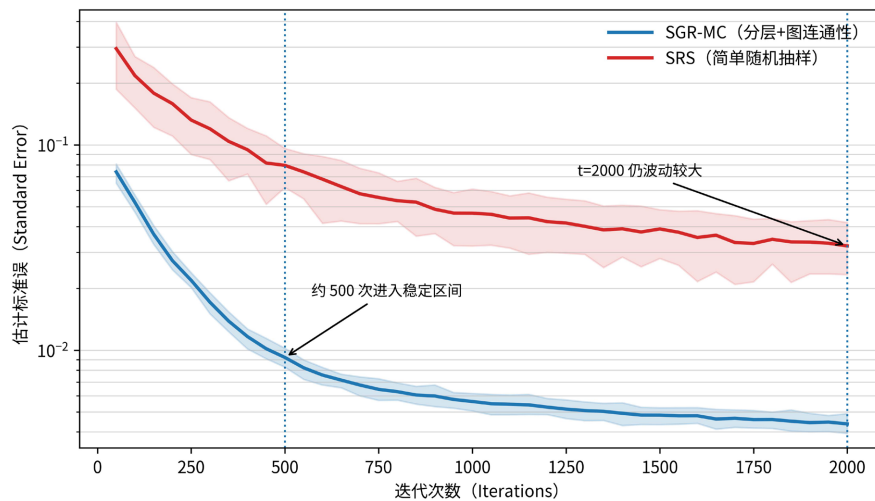
2) 大湾区标准合同情景：假设《大湾区标准合同》全面实施，大幅简化 PIPL 下的出境评估流程，跨境合规成本约降低 50%， $w_{cross} = 0.15$ 。

3) 数字自贸区情景：假设在特定区域(如河套、横琴)实现“白名单”制度，数据自由流动，成本降至技术底线， $w_{cross} = 0.05$ 。

5.3. 实证结果与分析

5.3.1. 算法性能分析

为证明 SGR-MC-Shapley 算法在运筹学上具有优势，本文对比了 SGR-MC-Shapley 算法与传统简单随机抽样 SRS 的收敛速度，见图 4。



注：阴影带表示多次独立重复实验下的标准误分布区间。

Figure 4. Comparison of algorithm convergence

图 4. 算法收敛性对比

在本研究的图限制博弈框架中，节点的边际贡献取决于其加入后是否实质性改变联盟的可连通结构。对大湾区跨境数据协作而言，城市间连边存在显著异质性，部分内地城市间连通紧密，而跨境边会受到制度摩擦系数 w_{ij} 的强约束，导致在许多排列联盟状态下，新增节点并不会带来有效连通性提升，边际贡献接近 0。而 SRS 在多数为零和少数极大值的混合分布下产生高方差，造成 SE 下降缓慢并长期波动。

SGR-MC-Shapley 算法的核心改进在于通过分层抽样与图连通性判定，以及相应的前瞻截断机制。根据联盟规模与结构复杂度分层抽样，在靠近 $N/2$ 处投入更多样本；连通性检查则减少对估值无信息的无效排列，使得单位样本的信息增量更大，两者同时作用使得 SGR-MC 的方差显著降低，收敛速度加快，波动变窄。

5.3.2. 不同摩擦系数下的网络连通性与 Shapley 值

如图 5 所示，颜色映射表示香港节点的 Shapley 值占比，叠加等高线用于刻画非线性梯度与交互效应。图中标注点(0.30, 0.30)表示“现状”附近参数区间。当合规成本降低时， w_{HK} 从 0.30 向左下降，香港价值占比上升速度显著加快，结果呈现明显非线性；同时，等高线呈弯曲而非严格竖直或水平，表明存在港澳制度摩擦的交互影响，单方面降低 w_{HK} 的同时，也会通过网络连通性改善带来对 w_{MO} 方向的间接敏感性。

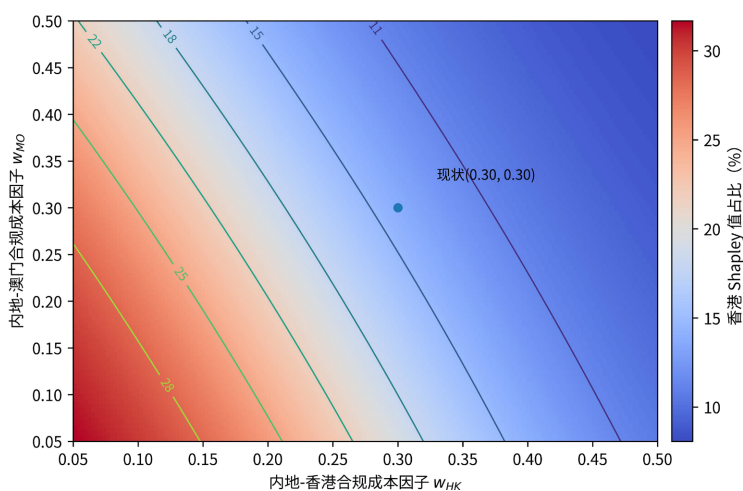


Figure 5. Heatmap of sensitivity to the institutional friction coefficient
图 5. 制度摩擦系数的敏感度热力图(以香港节点为例)

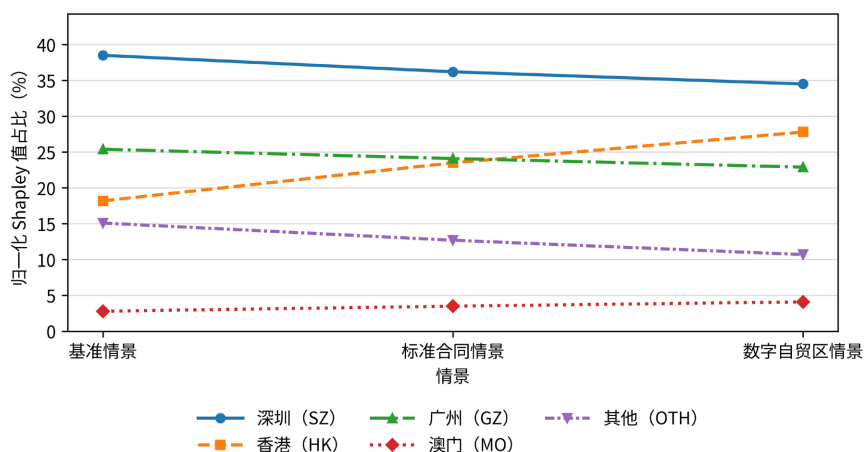
5.3.3. Shapley 值分配结果

利用 SGR-MC-Shapley 算法进行 10,000 次迭代模拟，得到的归一化 Shapley 值，即各城市对大湾区数字经济总盈余的贡献占比如表 3 所示。

Table 3. Scenario-based changes in Shapley Value proportions

表 3. 不同情景下的 shapley 值占比变化

城市	基准情景	标准合同情景	数字自贸区情景	变化趋势分析
深圳	38.5%	36.2%	34.5%	占比微降，总量上升。因其自身体量大，对外部依赖相对较小。
香港	18.2%	23.5%	27.8%	显著上升。作为连接器，壁垒降低使其能更频繁地加入内地产业链联盟，释放其金融数据价值。
广州	25.4%	24.1%	22.9%	同深圳类似，作为内地枢纽，原有的“内循环”优势在“双循环”开启后相对被拉平。
澳门	2.8%	3.5%	4.1%	小幅上升。主要受益于与珠海(横琴)的深度绑定。
其他	15.1%	12.7%	10.7%	占比下降但绝对值随总盈余上升而上升



注：归一化指各城市 Shapley 值除以同一情景下联盟总盈余后得到的百分比；OTH 为合并项，便于在简化网络中保持总和一致。

Figure 6. Variation in cities' normalized Shapley contribution shares across compliance cost scenarios
图 6. 不同合规成本情景下各城市归一化 Shapley 贡献占比变化

观察图 6 可知，随着制度摩擦系数的降低，大湾区数据价值分配格局发生了结构性变化。最显著的特征是香港 Shapley 值占比的攀升，从基准情景的 18.2% 跃升至自贸区情景的 27.8%。说明香港处于网络割点或桥接位置的关键节点，其价值对连接成本具有极高的敏感度。相比之下，深圳和广州虽然其相对份额呈现下降趋势，但绝对收益随着总盈余同步增长。

研究表明，建立数据自贸区本质上是一个帕累托改进过程——它在不损害甚至增加了内地城市绝对收益的前提下，极大释放了港澳被制度所压抑的潜在价值，从而推动区域发展更加均衡。

6. 结论与建议

6.1. 结论

研究表明，制度摩擦是制约大湾区数据协同效率的核心因素。跨法域合规成本会大大削弱联盟连通性，降低整体协同产出，从而在量化层面揭示了现行数据出境监管对区域数字一体化的约束效应。进一步分析可知，节点在网络中的结构位置决定了其数据价值。以香港为例，尽管其数据规模有限，但由于处于内地与国际数据流动的桥接位置，在合规成本下降的情景下，其边际贡献提升最为明显，说明结构性枢纽城市对制度变动高度敏感。情景模拟结果表明，降低跨境数据合规成本可以提升联盟整体福利，且不会削弱深圳、广州等内地核心城市的绝对收益，体现出数据跨境便利化的帕累托改进特征。本文提出的 SGR-MC-Shapley 估值框架在图限制博弈环境下既保证了估值精度与计算效率，同时给区域层面数据要素价值核算与分配提供了一种可行的技术路径。

6.2. 建议

本文认为，大湾区跨境数据治理的重点不是简单的放松监管，而应该通过制度设计降低有效的合规摩擦，并构建与网络结构相匹配的收益分配机制。具体而言，应在现有制度的基础上进一步明确香港的跨境数据枢纽地位，在低风险领域扩大数据跨境试点，以放大其对区域数据协同的正向外溢效应；同时，引入基于 Shapley 贡献度的数据收益分配和补偿机制，通过制度化安排增强各参与方的数据共享激励；对澳门等数据规模较小但具有场景或区位优势节点，应通过差异化制度设计提高其在局部子网络中的不可替代性；从长期看，有必要把数据跨境治理纳入区域协同发展的整体制度框架，由静态合规审查转向

动态风险分级管理，为大湾区数字一体化提供稳定、可预期的制度环境。

参考文献

- [1] 刘涛雄, 张亚迪, 戎珂, 等. 数据要素成为中国经济增长新动能的机制探析[J]. 经济研究, 2024, 59(10): 19-36.
- [2] 申明浩, 赵雨. 粤港澳大湾区跨境数据流动的的制度创新与全球治理启示[J]. 新经济, 2025(7): 48-56.
- [3] 陶珏. 粤港澳大湾区创新金融数据要素市场化配置研究[J]. 江南论坛, 2025(1): 25-29.
- [4] 华劼. 移动互联网时代个人信息隐私保护——对《香港个人资料(私隐)条例》的解读及启示[J]. 重庆邮电大学学报(社会科学版), 2017, 29(5): 40-47.
- [5] 杨翱宇. 个人信息保护的特别机制研究——以澳门《个人资料保护法》为考察样本[J]. 图书馆, 2018(3): 68-74.
- [6] 董晓花, 王欣, 陈利. 柯布-道格拉斯生产函数理论研究综述[J]. 生产力研究, 2008(3): 148-150.
- [7] Arrow, K.J., Chenery, H.B., Minhas, B.S. and Solow, R.M. (1961) Capital-Labor Substitution and Economic Efficiency. *The Review of Economics and Statistics*, **43**, 225-250. <https://doi.org/10.2307/1927286>
- [8] Jones, C.I. and Tonetti, C. (2020) Nonrivalry and the Economics of Data. *American Economic Review*, **110**, 2819-2858. <https://doi.org/10.1257/aer.20191330>
- [9] Shapley, L.S. (1953) A Value for n-Person Games. *Annals of Mathematical Studies*, **28**, 307-317.
- [10] Myerson, R.B. (1977) Graphs and Cooperation in Games. *Mathematics of Operations Research*, **2**, 225-229. <https://doi.org/10.1287/moor.2.3.225>
- [11] Maleki, S., Tran-Thanh, L., Hines, G., et al. (2013) Bounding the Estimation Error of Sampling-Based Shapley Value Approximation.
- [12] Antràs, P. (2004) Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution. *The B.E. Journal of Macroeconomics*, **4**, 1-36.
- [13] Klump, R., McAdam, P. and Willman, A. (2007) Factor Substitution and Factor-Augmenting Technical Progress in the United States: A Normalized Supply-Side System Approach. *Review of Economics and Statistics*, **89**, 183-192. <https://doi.org/10.1162/rest.89.1.183>
- [14] Acemoglu, D. and Restrepo, P. (2018) The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, **108**, 1488-1542. <https://doi.org/10.1257/aer.20160696>
- [15] Hestness, J., Narang, S., Ardalan, N., et al. (2017) Deep Learning Scaling Is Predictable, Empirically.