arXiv:2601.01871v1 [math.ST] 5 Jan 2026

关于 非同步观测点过程的滞后估计程

Takaaki Shiotani* Takaki Hayashi† Yuta Koike*

2026年1月6日

摘要

本文提出了一个用于分析点过程之间领先—滞后关系的新理论框架，特别关注其在高频金融数据中的应用。具体而言，我们关注两组订单到达时间戳序列之间的领先—滞后关系。Dobrev 和 Schaumburg 的开创性研究基于时间戳的交叉计数，提出了无需模型假设的跨市场交易活动度量。尽管已知该方法能得出可靠结果，但其原始表述本质上依赖于离散时间观测，这一局限正是本研究要解决的问题。具体来说，我们将估计两个点过程之间领先—滞后关系的问题，表述为估计双变量平稳点过程的交叉对相关函数（CPCF）的形状；这一量在神经科学和空间统计学文献中已有深入研究。在此框架下，主导领先—滞后时间被定义为 CPCF 最尖锐峰值的位置。按照这种解释，Dobrev 和 Schaumburg 的跨市场活动度量中的峰值位置，可视为上述意义下领先—滞后时间的一个估计量。此外，我们提出了一种基于核密度估计的替代性领先-滞后时间估计量，并证明它具有理想的理论性质，且数值表现更优。高频金融数据的实证证据表明，我们提出的方法行之有效。

关键词：带宽选择；交叉相关直方图；交叉配对相关函数；高频数据；非同步性；领先—滞后效应。

1 引言

关于两个金融时间序列之间领先—滞后关系的实证研究，长期以来一直是金融学领域活跃的研究方向。识别这种关系对于理解价格发现至关重要，也可能为从业者提供获取超额利润的机会。在现代金融市场中，这类关系只能在极短的时间范围内持续，甚至短至一毫秒或更短。因此，较低频率或粗粒度聚合的数据必然无法发现这些关系的精细结构。这促使研究者使用逐笔数据，即记录所有交易、并按照交易随机到达且非同步发生的原始高频数据。尤其是在利用此类数据估计领先—滞后关系时，处理非同步性是一个核心问题。

*东京大学数理科学研究科，日本东京目黑区驹场3-8-1，邮编153-8914
†庆应义塾大学经营管理研究科，日本横滨日吉4-1-1，邮编223-8526

现有大多数研究都使用价格序列考察高频领先—滞后动态。具有代表性的方法包括基于估计交叉协方差函数的方法、 [26, 43, 45], 小波分析[38, 39], 局部谱估计 [50], 基于Hawkes过程的多资产模型 [5, 22] 以及 [16]提出的多资产滞后调整模型。在这些方法中，Hoffmann 等人 [43] 引入了一种简单的交叉协方差估计量，可以直接根据非同步观测收益计算，并提出通过寻找其最大化点来估计主导领先—滞后时间。尽管由于其直观的解释，该方法能够得出合理的实证结论 [2, 7, 10, 25, 45, 59], 但所得的领先—滞后时间估计往往不稳定且不可靠 [7, 37, 45], 这可能是因为高频价格序列受到市场微观结构噪声的影响。作为一种替代方法，Dobrev & Schaumburg [27]提出了基于订单到达交叉计数的两个资产之间领先—滞后关系的无模型测度。实践表明，其领先—滞后时间估计量能够产生高度稳定且可靠的估计结果；参见 [27, 37]。

然而，Dobrev–Schaumburg方法本质上是描述性的，并且该方法实际估计的基础量究竟是什么，并不显而易见。[1] 事实上，正如我们在第2节中所展示的，在某些情况下，他们的方法在实践中表现不佳，尤其是在数据包含相对较少观测值时。此外，实现他们的方法需要将观测期划分为等间隔分桶，而桶大小的选择会对结果产生重大影响。然而，由于该方法是“无模型”的，因此无法从统计角度解释这种敏感性为何会出现。

为解决这些问题，我们从点过程的角度重新表述Dobrev–Schaumburg方法。这一视角揭示出，他们的测量本质上是在估计由订单到达生成的二元点过程的交叉成对相关函数 （CPCF）的形状特征；定义见第3节。因此，Dobrev–Schaumburg的领先—滞后时间估计量可以解释为CPCF最尖峰位置的估计量。这一解释还澄清了，他们的方法之所以表现出不稳定性，是因为当桶大小相对于根据底层数据性质所允许的范围而言过小时，便会出现这种情况。同时，由于他们的估计量只能取桶大小的整数倍，使用较大的桶大小会导致估计过于粗糙。为克服这些局限，我们提出了一种领先—滞后时间的非参数、基于核的估计量，以及一种数据驱动的带宽选择程序。我们从理论和实证两方面表明，即使在Dobrev–Schaumburg方法失效的情形下，这一新估计量仍能产生稳定且准确的结果。

本文其余部分安排如下。第2节详细介绍 Dobrev–Schaumburg方法。第3节引入点过程框架，以阐明 Dobrev–Schaumburg方法的理论含义。第4节在该框架下提出一种领先—滞后时间的替代估计量，并推导其理论性质。第5节通过一项全面的蒙特卡洛研究展示其优异的数值表现。第6节通过一项实证应用，利用真实数据说明所提估计量的有效性。第7节总结

[1]尽管 [28] 讨论了两个时间戳序列相互独立时其测量值的一些渐近性质，但他们并未明确说明其背后的目标估计量。

了我们的主要贡献，并探讨未来研究方向。附录包含数学证明和更多实现细节。

记号 有限集 S 的基数记为 $|S|$。Leb 表示 Lebesgue测度。拓扑空间 S 的 Borelσ-代数记为 $\mathcal{B}(S)$。对于定义在集合 S上的实值函数 f，我们设 $\|f\|_\infty = \sup_{x\in S}|f(x)|$。此外，我们用 $\arg\max_{x\in S} f(x)$ 表示 f 在 S 上的最大化点集合。对于随机变量 X 和 $p \geq 1$，我们设 $\|X\|_p := (\mathrm{E}[|X|^p])^{1/p}$。底层概率空间记为 $(\Omega, \mathcal{F}, \mathrm{P})$。按照惯例，我们将 $1/0 = \infty$ 解释为。

2 Dobrev–Schaumburg方法

假设我们有两种金融资产的逐笔数据，其时间戳分别为 $0 \leq t_1^1 < \cdots < t_{n_1}^1 \leq T$和 $0 \leq t_1^2 < \cdots < t_{n_2}^2 \leq T$。全文中，我们假设 $T \geq 1$ 为整数，并且在处理真实数据时，时间戳以秒为单位；因此，我们可以将 T 视为一个较大的整数。Dobrev & Schaumburg [27] 提出了以下方法，用于测量两种资产之间的领先—滞后关系。

首先，将观测区间 $[0, T]$ 划分为等间隔的时间桶 $I_k^h := [kh, (k+1)h)$, $k = 0, 1, \ldots, T/h - 1$，其中 $h > 0$ 的选取满足 $h^{-1} \in \mathbb{N}$。我们将资产2的事件在资产1的事件之后滞后 $\ell \in \mathbb{Z}$ 发生的情形，解释为存在时间戳 t_i^1、t_j^2 以及一个桶索引 $k \in \{|\ell|, |\ell|+1, \ldots, T/h - 1 - |\ell|\}$，使得 $t_i^1 \in I_k^h$ 且 $t_j^2 \in I_{k+\ell}^h$。通过计算这类桶索引的数量，我们可以得到资产1对资产2、滞后为 ℓ时领先—滞后效应的度量。基于这一思路，Dobrev & Schaumburg [27] 引入了偏移量为 ℓ 的原始跨市场活动，其定义为

$$\mathcal{X}_h^{\mathrm{raw}}(\ell) := \sum_{k=|\ell|}^{T/h-1-|\ell|} 1_{\{\exists t_i^1 \in I_k^h,\, \exists t_j^2 \in I_{k+\ell}^h\}}.$$

为调整自由度，他们还将相对跨市场活动 定义为

$$\mathcal{X}_h^{\mathrm{rel}}(\ell) := \frac{\mathcal{X}_h^{\mathrm{raw}}(\ell)}{\min\left\{\sum_{k=|\ell|}^{T/h-1-|\ell|} 1_{\{\exists t_i^1 \in I_k^h\}},\ \sum_{k=|\ell|}^{T/h-1-|\ell|} 1_{\{\exists t_{i+\ell}^2 \in I_k^h\}}\right\}}.$$

计算 $\mathcal{X}_h^{\mathrm{raw}}(\ell)$ 和 $\mathcal{X}_{\mathrm{rel}h}(\ell)$ 需要选择一个 h 的取值。在 [27] 中，h 被设为 1 毫秒，这一取值是通过考虑数据的时间粒度，以临时方式选定的。在这里，1 毫秒是其标准普尔500现货市场数据集中可用的最小时间单位。

根据定义，$\mathcal{X}_h^{\mathrm{rel}}(\ell)$ 的值越大，表示领先—滞后效应越强，即资产2在延迟 ℓh 后跟随资产1。受此启发，Dobrev & Schaumburg [27] 提出通过在映射 $\ell \mapsto \mathcal{X}_h^{\mathrm{rel}}(\ell)$ 中定位峰值，来识别占主导地位的领先—滞后时间。具体而言，给定搜索网格 $\mathcal{G}_h$，领先—滞后时间估计为

$$\hat{\theta}_h^{DS} = \hat{\ell} h, \quad \text{where } \hat{\ell} \in \arg\max_{\ell \in \mathcal{G}_h} \mathcal{X}_h^{\mathrm{rel}}(\ell).$$

我们将$\hat{\theta}_h^{DS}$称为*DS*估计量。下文中，我们假设真实的领先—滞后时间位于区间$(-r, r)$内，其中 r 是某个已知的正常数，并定义 $\mathcal{G}_h := \{\ell \in \mathbb{Z} : |\ell h| \leq r\}$。

如引言中所述，多项实证研究报告称，DS估计量能够对领先—滞后时间给出稳定且易于解释的估计值。然而，在某些情况下，该估计量的表现并不理想，尤其是数据集中的观测值相对较少时。图1展示了这一问题：对于2015年8月12日的MNST股票，图中给出了根据美国两家证券交易所 NASDAQ 和 BATS 的最佳报价更新计算出的相对跨市场活动度量。在此例中，跨市场活动度量的数值剧烈波动，因此难以判断峰值的可靠性。然而，由于 Dobrev–Schaumburg 方法具有“无模型”的特性，现有方法并未对这种不稳定性的成因提供统计学解释。本研究的目标之一是为该方法奠定理论基础，弥补这一空白。

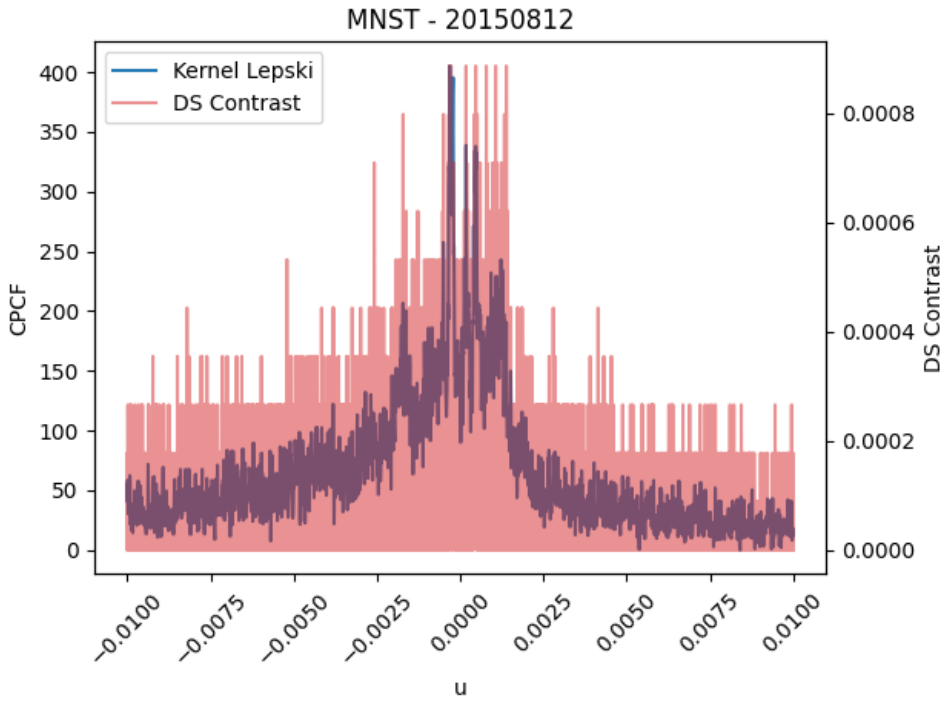

图 1：对比函数：DS 与我们的方法。MNST、NASDAQ 与 BATS（报价，2015 年 8 月 12 日）。跨市场活动度量 $\mathcal{X}_h^{\mathrm{rel}}(\ell)$ 由红线表示，而采用 Lepski 选择带宽的核密度估计量 $g_{\hat{h}}(u)$（见第 4 节）由蓝线表示。横轴单位为秒。

3 所提出的框架

为阐明 Dobrev–Schaumburg 方法的统计学含义，我们将观测到的时间戳 $(t_i^1)_{i=1}^{n_1}$ 和 $(t_j^2)_{j=1}^{n_2}$ 视为实数轴上一个二元点过程的实现。具体而言，对于每个 $a = 1,\ 2$，我们将时间戳 (t_i^a)视为在区间$[0, T]$上观测点过程 N_a 于 $\mathbb{R}$ 的结果。也就是说，对于博雷尔集 $A \subset [0, T]$，有 $N_a(A) = |\{i : t_i^a \in A\}|$。下文主要沿用 [23, 24] 中对点过程的数学表述，对于未作解释的概念和符号，请参阅这些专著（总结见 [65, 第2节。2] ）。按照这一表述，我们可以

将 $\mathcal{X}_h^{\mathrm{rel}}(\ell)$ 改写为

$$\mathcal{X}_h^{\mathrm{rel}}(\ell) = \frac{\sum_{k=|\ell|}^{T/h-1-|\ell|} 1_{\{N_1(I_k^h)>0,\, N_2(I_{k+\ell}^h)>0\}}}{\min\left\{\sum_{k=|\ell|}^{T/h-1-|\ell|} 1_{\{N_1(I_k^h)>0\}},\, \sum_{k=|\ell|}^{T/h-1-|\ell|} 1_{\{N_2(I_{k+\ell}^h)>0\}}\right\}}.$$

利用这个表达式，我们可以将 $\mathcal{X}_{\mathrm{rel}h}(\ell)$ 与 N 的成对交相关函数（*CPCF*）联系起来。

为了正式陈述理论结果，我们引入若干假设和记号。我们假设$N=(N_1,N_2)$ 是 $\mathbb{R}$ 上的简单平稳二元点过程，其强度为 $\lambda_1,\lambda_2\in(0,\infty)$。请注意，我们使用的"强度"一词与 [23] 中的含义相同（参见同上第 47 页）。根据 [23, 命题 3.3.IV]，，我们有 $\lambda_a = \mathrm{E}[N_a((0,1])]$，其中 $a=1,2$。我们还假设存在一个局部可积函数$\lambda_{12}\colon \mathbb{R}\to[0,\infty]$，使得

$$\mathrm{E}[N_1(A_1)N_2(A_2)] = \int_{A_1\times A_2} \lambda_{12}(y-x)dxdy \tag{3.1}$$

对于任意有界的 $A_1,A_2\in\mathcal{B}(\mathbb{R})$。我们将 λ_{12}称为 N 的交叉强度函数[2]。 N 的 CPCF 定义为函数 $g\colon \mathbb{R}\to[0,\infty]$，其定义如下：

$$g(u) = \frac{\lambda_{12}(u)}{\lambda_1\lambda_2} \qquad (u\in\mathbb{R}).$$

交叉强度函数可以通过以下方式与 N_1 和 N_2 的无穷小增量之间的交叉协方差函数联系起来（参见（A.11））：

$$\nu_{12}(u) := \lim_{h\downarrow 0} \frac{\mathrm{Cov}(N_1(0,h], N_2(u,u+h])}{h^2} = \lambda_{12}(u) - \lambda_1\lambda_2 \quad \text{a.e. } u.$$

函数 ν_{12} 称为 N的协方差密度。从这个意义上说， $g(u)$度量了 $N_1(\cdot)$与 $N_2(\cdot+u)$之间的交叉协变。

我们还需要点过程的 α-混合性（或强混合）概念。回顾一下，两个子-σ-代数 $\mathcal{G}$ 和 $\mathcal{H}$ （属于 $\mathcal{F}$）的 α-混合系数定义为

$$\alpha(\mathcal{G},\mathcal{H}) := \sup\{|\,\mathrm{P}(C\cap D) - \mathrm{P}(C)\,\mathrm{P}(D)| : C\in\mathcal{G}, D\in\mathcal{H}\}.$$

对于 $E\in\mathcal{B}(\mathbb{R})$，我们用 $N\cap E = (N_i\cap E)_{i=1}^2$表示 N在 E上的限制，即当 $i=1,2$ 且 $A\in\mathcal{B}(\mathbb{R})$ 时，$(N_i\cap E)(A) = N_i(A\cap E)$。此外，对于某个 $y\in E\}$， $E\oplus r := \{x\in\mathbb{R}\colon |y-x|<r$表示 E的 r-扩张。此外，给定定义在 $\mathbb{R}$上的二元点过程 $M=(M_1,M_2)$， $\sigma(M)$表示由$\bigcup_{i=1}^2\{M_i(A)\colon A\in\mathcal{B}(\mathbb{R})\}$生成的 σ-代数。作为 N的 α-混合性系数，我们采用以下定义：

$$\begin{aligned}\alpha^N_{c_1,c_2}(m;r) = \sup\Big\{\alpha(\sigma(N\cap E_1),\sigma(N\cap E_2)) : E_1 = \bigcup_{j\in J_1} I_j\oplus r,\ E_2 = \bigcup_{j\in J_2} I_j\oplus r,\\ |J_1|\le c_1,\ |J_2|\le c_2,\ d(J_1,J_2)\ge m,\ J_1,J_2\subset\mathbb{Z}\Big\}, \quad m,c_1,c_2,r\ge 0,\end{aligned} \tag{3.2}$$

[2]在神经科学中，"交叉强度函数"通常指函数 $\lambda_{12}(u)/\lambda_1$ 或 $\lambda_{12}(u)/\lambda_2$（例如参见 [15]）。我们遵循空间统计学中的术语用法 [41, 42]。按照点过程理论的术语， λ_{12} 是 N 的约化交叉矩测度的密度（参见 [23, 第 8.3]节）。

其中 $I_j := I_j^1 = (j, j+1]$ 和 $d(J_1, J_2) := \inf\{|j_1 - j_2| : j_1 \in J_1, j_2 \in J_2\}$。这个定义是 [65], 中所用定义的一个小变体；他们使用 $(j - \frac{1}{2}, j + \frac{1}{2}]$ 而非 I_j。这种差异无关紧要，我们采用当前定义只是为了（稍微）简化技术论证。

注 3.1 （将混合系数视为连续时间过程）。 由于 N 可以视为一个以 $\mathbb{R}$为索引的随机过程，我们也可以据此定义 α-混合系数 [14]:

$$\alpha^N_{\text{proc}}(\tau) := \sup_{t\in\mathbb{R}} \alpha\left(\sigma(N \cap (-\infty, t)),\, \sigma(N \cap (t+\tau, \infty))\right), \qquad \tau \geq 0.$$

我们的定义比这个版本弱，因为我们有

$$\alpha^N_{c_1,c_2}(m; r) \leq \alpha^N_{\text{proc}}(m - 2r) \tag{3.3}$$

对所有 $c_1, c_2, r \geq 0$和 $m \geq 2r$均成立。对于点过程的情形，采用较弱的版本很重要，因为有时难以对 (3.3) 的左侧关于 $c_1, c_2 \geq 0$一致地进行界定；例如，参见 [57,命题 2.8] 和 [65, 引理 10.13]。

我们对 N施加以下正则性条件。

[A1] (i) 对每个 $p \geq 1$，存在一个常数 $B_p > 0$，使得 $\max_{i=1,2} \lambda_i^{-1} \|N_i((0,1])\|_p \leq B_p$。 (ii) 对任意 $p, q \geq 1$，存在一个常数 $B_{p,q} > 0$，使得 $\alpha^N_{p,p}(m;\, r_1) \leq B_{p,q} m^{-q}$对所有 $m \in \mathbb{N}$均成立，其中 $r_1 := r + 1$。

这一假设在相关文献中相当合理，并且许多标准点过程模型都满足该假设，例如 Hawkes过程和 Neyman-Scott过程，只要它们的核函数满足某些正则性条件；详见第 5.1 节。

在 [A1] 及其他技术假设下，我们得到 Dobrev–Schaumburg 的跨市场活动度量的如下渐近表示:

命题 3.1。 假设 [A1]。另外，假设 g 有界，且

$$\begin{aligned} \max_{\ell\in\mathcal{G}_h} \mathrm{E}[N_1(I_0^h)\{N_1(I_0^h) - 1\}N_2(I_\ell^h)] &= o(h^{1+\varpi}), \\ \max_{\ell\in\mathcal{G}_h} \mathrm{E}[N_1(I_0^h)N_2(I_\ell^h)\{N_2(I_\ell^h) - 1\}] &= o(h^{1+\varpi}) \end{aligned} \tag{3.4}$$

当 $h \to 0$ 时，对于 $\varpi = 1$成立。此外，假设对于某个 $0 < \gamma < 1$，当 $T \to \infty$ 时有 $h = h_T \asymp T^{-\gamma}$。则

$$\max_{\ell\in\mathcal{G}_h} \left| \frac{\mathcal{X}_h^{\text{rel}}(\ell)}{h} - (\lambda_1 \vee \lambda_2) \int_{\mathbb{R}} \frac{1}{h} K^{\text{tri}}\left(\frac{u - \ell h}{h}\right) g(u)du \right| \to^p 0,$$

其中 $K^{\text{tri}}(x) = (1 - |x|)1_{[-1,1]}(x)$。

注3.2 （关于条件（3.4））。 （3.4）左端的量可以与 N 的(2,1)阶和(1,2)阶阶乘矩测度联系起来（定义见 [65] 第2.3节）。特别地，如果这些测度相对于Lebesgue测度具有有界密度，则 $\varpi = 1$ 满足（3.4）。由于每个阶乘矩测度都可以通过 [65,表示为阶乘累积量测度之和

式（2.5）]（另见 [12, 式（3.21）]），因此，对于Hawkes过程，其密度可通过 [48, 式（39）]计算；对于Neyman-Scott过程，则可通过 [65, 式（5.2）], 计算。因此，原则上可以在对其核函数作适当假设的条件下验证（3.4）。我们不再进一步讨论这一点，因为对于下一节提出的新估计量的理论发展而言，该条件并非必要。

在命题 3.1 的假设下，若 g 连续，则有 $\max_{\ell\in\mathcal{G}_h}|\mathcal{X}_h^{\mathrm{rel}}(\ell)/h-(\lambda_1\vee\lambda_2)g(\ell h)|\to^p 0$。因此，跨市场活动度量可解释为 CPCF 的一个估计量，差一个乘法常数。此外，若 g 在 $(-r,\ r)$ 上有唯一的最大化点 θ^*，则上述结果表明，当 $T\to\infty$ 时，$\hat{}\theta_h^{DS}$ 是 θ^* 的相合估计量。为了进一步理解 $\hat{}\theta_h^{DS}$ 的理论性质，我们研究其收敛速率。为此，我们对 g 提出以下假设：

[A2] 存在常数 $\theta^*\in(-r,\ r)$、$\alpha\in(0,\ 1)$、$\cup(1,\ \infty)$、$b>1$ 和 $\delta>0$，使得以下条件成立：

(i) 若 $\alpha>1$，则 $\sup_{u\in\mathbb{R}}g(u)\le b$，并且

$$\min\left\{\sup_{0<u-\theta^*<\delta}\frac{g(\theta^*)-g(u)}{|u-\theta^*|^{\alpha-1}},\ \sup_{0<\theta^*-u<\delta}\frac{g(\theta^*)-g(u)}{|u-\theta^*|^{\alpha-1}}\right\}\le b$$

and

$$\inf_{0<|u-\theta^*|<\delta}\frac{g(\theta^*)-g(u)}{|u-\theta^*|^{\alpha-1}}\ge\frac{1}{b}$$

and

$$\sup_{|u-\theta^*|\ge\delta}g(u)\le g(\theta^*)-\frac{1}{b}.$$

（ii）如果 $\alpha<1$,

$$\max\left\{\inf_{0<u-\theta^*<\delta}\frac{g(u)}{|u-\theta^*|^{\alpha-1}},\ \inf_{-\delta<u-\theta^*<0}\frac{g(u)}{|u-\theta^*|^{\alpha-1}}\right\}\ge\frac{1}{b}.$$

此外，存在一个常数 $\alpha<\alpha_0\le 1$ 和一个可测函数 $g_0:\ \mathbb{R}\to[0,\infty]$，使得$\|g_0\|_{L^{1/(1-\alpha_0)}([-r_1,r_1])}\le b$以及

$$g(u)\le b\left(g_0(u)+|u-\theta^*|^{\alpha-1}\right)\quad\text{for all } u\in[-r_1,r_1].$$

在 [A2]（i）下，g 有界，且 θ^* 是 g的唯一最大化点。相反，在 [A2]（ii）下，g 在 θ^* 处发散，因此无界。我们通过辅助函数 g_0允许 g 具有除 θ^* 之外的极点；然而，条件$\|g_0\|_{L^{1/(1-\alpha_0)}([-r_1,r_1])}<\infty$ 确保 θ^* 仍然是g中“最尖锐”的峰值位置。我们将 θ^* 称为领先—滞后（时间）参数。允许 g 无界，是基于若干经验观测结果：（i）Dobrev–Schaumburg跨市场活动度量通常呈现极其尖锐的峰值（例如参见 [27, 图 7]）；（ii）Shiotani & Yoshida [65] 发现，当CPCF无界时，其半参数模型对日本股票市场数据的拟合效果更好；（iii）Rambaldi 等 [60] 报告称，不同外汇汇率之间的强度突发事件呈现领先—滞后关系。

备注 3.3（与众数估计的关系）。 将 CPCF 替换为密度后，我们对超前-滞后参数估计问题的表述便自然地与概率密度函数的众数估计联系起来。 [A2] 正是受这一观察启发。事实上，Wegman [71]将单峰分布的众数

估计问题分为三类：密度有界的情形标记为 I型，密度无界的情形标记为 II型，没有密度的情形标记为 III型。在这一分类中， [A2](i) 对应 I型，而 [A2](ii) 对应 II型。为分析收敛速率，我们假设 CPCF 在 θ^* 的邻域内呈幂函数 $|u-\theta^*|^{\alpha-1}$ 的形式。当$\alpha > 1$时，这一假设与 [3] 中研究的条件类似（参见同书第 1.1 节）。当 $\alpha < 1$时，它对应于 [9] 中条件 [**H-III**] 的类似情形。注意，情形 $\alpha = 1$ 被排除，仅仅是因为此时幂函数 $|u-\theta^*|^{\alpha-1}$ 变为常数。

在 [A2], 下，我们设

$$\beta_\alpha := \alpha \vee (2\alpha - 1) = \begin{cases} \alpha & \text{if } \alpha < 1, \\ 2\alpha - 1 & \text{if } \alpha > 1. \end{cases}$$

定理 3.1。 假设 [A1], [A2] 且 $h = h_T \asymp T^{-\gamma}$ 如 $T \to \infty$ 对于某个 $0 < \gamma < 1/\beta_\alpha$成立。此外，假设当 $\varpi = \alpha$时满足 (3.4) 。那么，当 $T \to \infty$时，$\hat{}\theta_h^{DS} = \theta^* + O_p(h)$ 。

由于按构造，$\hat{}\theta_h^{DS}$是 h的整数倍，定理 3.1 意味着其收敛速率恰好为 $O(h)$的阶。特别地，随着 h 变小，$\hat{}\theta_h^{DS}$的精度会提高。然而，根据 $\mathcal{X}_h^{\mathrm{rel}}(\ell)$的定义，选择小于数据中最小时间单位的 h 毫无意义。从这个意义上说，Dobrev & Schaumburg [27] 将 h 设为最小时间分辨率是合理的。

然而，定理 3.1 也对 h的阶施加了约束，要求 h 趋于 0 的速度比 T^{-1/β_α}更慢。此外，这一约束并非证明中的技术性产物，而是至关重要的，因为我们证明了，在定理 3.1 的假设下， T^{-1/β_α} 给出了任何关于 θ^*的估计量收敛速率的极小极大下界，至少在 $\alpha < 2$ 时如此（见注 4.2）。因此，当 $h \ll T^{-1/\beta_\alpha}$时，定理 3.1 的结论在理论上不可能成立。

注意，对于每个 $a = 1, 2$，在 [A1]下， $n_a/T \to \lambda_a > 0$ as $T \to \infty$ 几乎必然成立。因此，为使$\hat{}\theta_h^{DS}$表现良好，必须根据 n_1， n_2 （样本量）以及 α （CPCF峰值的尖锐度）来选择 h 。具体而言，当 n_1 和 n_2 较小和/或 α 较大时，应增大 h 。这为图1所示示例中Dobrev–Schaumburg方法表现不佳提供了一种解释。也就是说，在该数据集中，考虑到观测值数量相对较少（n_1， n_2）（28048，11287）， $h = 1$ μs可能设得过小。

这些观测结果表明， h 的选择对DS估计量的实现起着至关重要的作用。然而，DS估计量应被理解为对包含领先—滞后参数的一个区间进行估计，而不是对参数本身进行估计，因此很难通过数据驱动的方式选择 h 。基于这些原因，我们将在下一节提出一种基于核密度估计的新型领先—滞后时间估计量，并说明它能够克服这一问题。

4 新估计量

命题 3.1 表明， $\mathcal{X}_h^{\mathrm{rel}}(\ell)$ 渐近等价于基于三角核、针对 g 的核密度估计量的离散化版本。这自然促使我们直接考虑针对 g 的核密度估计量。

形式上，设 $K\colon \mathbb{R}\to\mathbb{R}$ 为核函数， $h=h_T>0$ 为带宽参数，且当 $h\to 0$ 时 $T\to\infty$。接下来考虑如下统计量:

$$\hat{g}_h(u)=\frac{T}{n_1n_2}\int_{(0,T]^2}K_h(y-x-u)N_1(dx)N_2(dy),\qquad u\in\mathbb{R},$$

其中，对于 $t\in\mathbb{R}$， $K_h(t)=h^{-1}K(t/h)$。我们通过取 $\hat{g}_h$ 的最大化点来估计 θ^*。也就是说，我们定义一个满足以下条件的 $\hat{}$ 随机变量θ_h

$$\hat{\theta}_h\in\operatorname*{arg\,max}_{u\in[-r,r]}\hat{g}_h(u).$$

计算过程 $\hat{}\ \theta_h$ 及其计算复杂度的具体计算方法见附录B .

当使用均匀核作为核函数时，$\hat{g}_h$ 本质上等价于神经科学中所谓的互相关直方图，长期以来一直被用于研究神经元尖峰之间的关系（例如参见 [15]）。这项研究促使统计学家开始探究 $\hat{}$ g_h [13, 14, 20, 32] 的理论性质。然而，据我们所知，尚无先前研究涉及 g 无界的情形，也没有研究 $\hat{g}_h$ 的最大化点的渐近行为。

我们对核函数作出如下假设。

[K] K 非负，在 0 处连续，具有有界变差，且支撑于 $[-1,1]$，满足 $K(0)>0$， $\int_{-\infty}^{\infty}K(t)dt=1$，并且 $\operatorname{arg\,max}^{u\in[-r,r]}\hat{g}_h(u)\neq\emptyset$ a.s.

例如，假设 [K] 对均匀核 $K=\frac{1}{2}1_{[-1,1]}$ 和三角核 $K=K^{\mathrm{tri}}$成立。

定理 4.1. 假设 [A1], [A2] 和 [K]。令 $\eta>0$ 为一个常数。那么，存在常数 $A>1$ 和$0<h_0<1$，它们仅取决于 α、 α_0、 δ、 b 和 K，使得

$$\mathrm{P}(|\hat{\theta}_h-\theta^*|>Ah)\le\frac{C}{\sqrt{Th^{\beta_\alpha+\varepsilon}}}\tag{4.1}$$

对所有 $h\le h_0\wedge T^{-\eta}$和 $\varepsilon>0$，其中 $C>0$是一个仅取决于 r、 α、 δ、 b、$(B_p)^{p\ge1}$、$(B_{p,q})_{p,q\ge1}$、 $\hat{}\varepsilon$、 η和 $\|K\|_\infty$的常数。特别地， $\theta_h=\theta^*+O_p(h)$ 当 $T\to\infty$ 时成立，若某个 $0<\gamma<1/\beta_\alpha$满足 $h=h_T\asymp T^{-\gamma}$。

根据定理 4.1，$\hat{\theta}_h$ 以与 $\hat{\theta}_h^{DS}$相同的收敛速率估计 θ^*。此外，与 DS 估计量不同，我们不需要假设（3.4）。特别是，通过适当选择带宽，我们的估计量几乎达到极小极大最优收敛速率 T^{-1/β_α}（见定理 4.3）。下一小节将讨论以数据驱动方式选择带宽的方法。

注释 4.1（与核模态估计的关系）。根据备注 3.3，我们的估计量与通过最大化核密度估计量得到的模态估计量密切相关。按照 [17], ，我们将这类估计量称为核众数估计量。

对于具有密度 f 和唯一总体模态 θ^*的独立同分布数据，Parzen [55] 证明了当 f 属于 C^2 类且 $f''(\theta^*)<0$时，核众数估计量具有渐近正态性。该设定是 [A2] 且 $\alpha=3$的一种特例。值得注意的是，通过恰当的带宽选择，核众数估计量几乎可以达到收敛速率 $n^{-1/5}$，其中 n 为样本量。由于 $\beta_3=5$，我们的估计量具有一种

类似的性质。Has'minskii [36] 证明，在此设定下，率参数 $n^{-1/5}$ 是极小极大最优的；但在对密度和核施加额外光滑性假设后，该速率可以进一步提高；例如参见 [31, 49, 70]。

在不对密度作光滑性假设的情况下，Abraham 等 [1] 以及 Herrmann 和 Ziegler [40]在类似于 [A2](i) 的条件下，得出了核众数估计量及其变体的收敛速率。在这种情形下，Arias-Castro 等 [3]确定了极小极大最优率，并提出了一种自适应估计方法。此情形下的收敛速率为 $n^{-1/(2\alpha-1)}$，这同样与我们的结果类似。据我们所知，除 [9], 的研究外，对于类似于 [A2](ii) 的假设下的众数估计，尚无其他结果。Bercu 等 [9] 研究了一种直方图型估计量，它可以视为离散化的核众数估计量，并证明通过适当选择箱宽，其收敛速率可以任意接近 $n^{-1/\alpha}$。这一性质也与我们的估计量类似。

最后，据我们所知，对于非光滑密度，核众数估计量的渐近分布仍然未知。我们猜想，当 $\alpha < 3/2$ 时，它可能是非高斯的；这与存在密度奇异性时的位置参数估计情形相呼应（参见 [46, 第六章]）。

4.1 利用 *Lepski* 方法进行带宽选择

在独立同分布观测和核密度估计的经典设定中，带宽选择通常基于最小化核估计量的平均积分平方误差（MISE）；例如，参见 [69, 第 1]章。对于点过程的矩密度函数进行核估计时，类似的 MISE 型准则也很常见 [34, 47]。

然而，就我们的目的而言，诸如 MISE 这样的全局损失准则并不完全令人满意，因为我们关注的对象不是函数 g 本身，而是其峰值的位置 θ^*。直观地说，全局准则旨在拟合整条曲线，包括 g 仅为中等大小或较小的区域；而 θ^* 的估计误差几乎可能完全取决于峰值小邻域内 g 的局部行为，在条件 [A2](ii)下，该行为甚至可能是奇异的。非参数众数回归领域的文献也已认识到一个密切相关的问题，该领域采用独立同分布观测；例如，参见 Chen [18] 的第 4.2 节。受这一领域近期研究的启发 [19, 72], ，我们在附录 C 中探讨基于交叉验证的损失最小化方法。

这里，我们改为采用一种基于 Lepski 型方法的自适应估计策略，即对具有不同带宽的估计量进行成对比较。我们的方法尤其受到 Klemel¨a[49], 的启发，他提出了一种基于独立同分布数据进行众数估计的 Lepski 型方法。关于 Lepski 方法的教科书论述，请参阅 [33,第 8 节2]。

固定常数 $a > 1$、$\gamma_{\max} > 0$ 和 $j_{\min} \in \mathbb{N}$。我们考虑以下集合，将其中的元素作为带宽候选值：

$$\mathcal{H}_T := \{a^{-j} : j_{\min} \le j \le \lceil \log_a(T^{\gamma_{\max}}) \rceil, j \in \mathbb{Z}\}.$$

对于每个 $h \in \mathcal{H}_T$，令

$$\mathcal{M}_h := \operatorname*{arg\,max}_{u \in [-r,r]} \hat{g}_h(u).$$

我们定义

$$\hat{h} := \min\left\{h \in \mathcal{H}_T : \bar{d}(\mathcal{M}_h, \mathcal{M}_{h'}) \le A_T h' \text{ for all } h' \in \mathcal{H}_T \text{ with } h' \ge h\right\},$$

其中 A_T 是一个正的常数，且

$$\bar{d}(\mathcal{M}_h, \mathcal{M}_{h'}) := \sup\{|x-y| : x \in \mathcal{M}_h,\, y \in \mathcal{M}_{h'}\}.$$

定理 4.2。 假设 [A1], [A2] 且 [K]。此外，假设 $1/\beta_\alpha \le \gamma_{\max}$。进一步，假设 $A_T \to \infty$ 且 $\hat{} A_T = o(T^c)$ 对于任意 $c > 0$，当 $T \to \infty$ 时成立。那么，对于任意 $0 < \gamma < 1/\beta_\alpha$， $\theta_{\hat{h}} = \theta^* + O_p(T^{-\gamma})$ 成立。

定理 4.2 表明，估计量 $\hat{} \theta_{\hat{h}}$ 几乎达到了最优收敛速率 T^{-1/β_α}，且无需知道 α的精确值。我们在第 5.3 节通过数值结果验证了这一结论，并评估了估计量对调节参数 A_T选择的稳健性。我们将看到，将A_T设为 log log T的一个常数倍，效果相当不错。

4.2 收敛速率的极小极大下界

在本小节中，我们证明，当 [A1] 和 [A2] 成立时， T^{-1/β_α} 给出了 θ*任意估计量收敛速率的极小极大下界。为此，我们考虑 N 的一类子模型，它们满足[A1] 和 [A2], ，具体定义如下。给定定义在 $\mathbb{R}$上的概率密度函数 g，我们考虑 $(\Omega\ \mathcal{F})$ 上的一个概率测度 P_g，它具有以下性质：

(i) $N_1 = \sum_{i=1}^{\infty} \delta_{t_i}$ 在 $\mathbb{R}$ 上是一个在 P_g下具有单位强度的泊松过程。(ii) N_2 具有 $N_2 = \sum_{i=1}^{\infty} \delta_{t_i+\gamma_i}$的形式，其中 $(\gamma_i)_{i=1}^{\infty}$ 是一个与 N_1 独立的独立同分布随机变量序列，并且在 P_g下， γ_1 的分布具有密度 g。

在 P_g下， N 是定义在 $\mathbb{R}$ 上的二元泊松过程，其含义与 [23, 例 6.3(e)], 中的相同；在该例的记号中，我们有 $Q_1 = Q_2 = 0$ 和 $Q_3(dxdy) = g(y-x)\,dxdy$。特别地，在 P_g下， N 的 CPCF 由 g 给出。

我们将 $\mathcal{G}(\theta^*, \alpha, \delta, b)$ 记为概率密度函数类 g： $\mathbb{R} \to [0, \infty]$，其支撑在 $[-1, 1]$上，并满足 [A2]的条件。注意，如果 $g \in \mathcal{G}(\theta^*, \alpha, \delta, b)$，则 N满足 [A1] 和 [A2]，其中常数族为 $(B_p)_{p\ge 1}$。事实上， [A2] 显然成立，而 [A1](i) 则由 N_1 和 N_2 都是定义在 $\mathbb{R}$ 上且具有单位强度的泊松过程这一事实推出。最后，由于 g 的支撑在 $[-1, 1]$上， $\alpha_{c_1}^N$、$_{c_2}(m;\ r_1) = 0$对 $m \ge m_0$成立，其中 m_0 仅取决于 r。因此， [A1](ii) 也得到满足。基于上述考虑，以下定理表明， T^{-1/β_α} 给出了在 [A1] 和 [A2]下针对 θ^* 的任何估计量的收敛速率的极小极大下界：

定理 4.3。 对于任意 $\alpha \in (0,\ 1) \cup (1,\ \infty)$，存在一个常数 $b > 0$ 使得

$$\liminf_{T\to\infty} \inf_{\hat{\theta}_T} \sup_{|\theta| \le 2\rho_T} \sup_{g \in \mathcal{G}(\theta, \alpha, 1/2, b)} \mathrm{P}_g\left(|\hat{\theta}_T - \theta| \ge \rho_T\right) > 0,$$

其中 $\rho_T := T^{-1/\beta_\alpha}$，下确界取遍所有基于 $N \cap [0, T]$ 的估计量，即所有 $\sigma(N \cap [0, T])$-可测随机变量。

注 4.2. 由于当 $p \wedge q > 1$时， N的阶数为（p， q）的累积量测度为零，因此只要 $\varpi = 1$，条件（3.4）就成立；当 g 有界时，该条件对 $\varpi < 2$ 也成立。因此，当 $\alpha < 2$时，在定理 3.1 的假设下， T^{-1/β_α} 也是 θ^* 的任何估计量收敛速率的极小极大下界。

5 模拟研究

带有伽马核的双变量 Hawkes 过程和奈曼–斯科特过程用于对高频金融数据中两个时间戳序列之间的关系建模；例如，分别参见 [58] 和[65], 。在我们的实验中，我们采用这些模型的“滞后”变体作为数据生成过程。利用模拟数据，我们通过数值实验考察 DS估计量和我们提出的估计量的准确性、收敛速率以及对调节参数的敏感性。本文所有数值实验中的核方法均采用三角核 K_{tri}。

5.1 模型

5.1.1 具有伽马核的滞后双变量霍克斯过程

首先，我们介绍一个双变量霍克斯过程，其条件强度函数为

$$\lambda_i(t) = \mu_i + \sum_{t_{k,1}<t} \phi_{i1}(t - t_{k,1}) + \sum_{t_{k,2}<t} \phi_{i2}(t - t_{k,2}), \qquad i = 1, 2,$$

其中， $\{t_{k,i}\}$ 表示第 i个分量中的第 k个事件时间。我们将核函数参数化为$\phi_{ij}(t) = \alpha_{ij} h_{ij}(t)$，其中 $t > 0$， $\alpha_{ij} > 0$ 为分枝比， $h_{ij}(t)$是定义在 $(0, \infty)$ 上的概率密度函数。

在本研究中，我们采用伽马核。具体而言，我们假设 h_{ij} 服从伽马密度$\Gamma(D_{ij}, \beta_{ij})$:

$$h_{ij}(t) = \frac{\beta_{ij}^{D_{ij}}}{\Gamma(D_{ij})} t^{D_{ij}-1} e^{-\beta_{ij} t}, \qquad t > 0,$$

其中， $D_{ij} > 0$ 是形状参数， $\beta_{ij} > 0$ 是率参数。当 $D_{ij} = 1$时，该设定退化为经典的指数核，因此指数霍克斯模型是其一个特例。我们用 $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$ 表示基准强度向量，并分别用 $\boldsymbol{\alpha} = (\alpha_{ij})_{1\le i,\ j\le 2}$、 $\boldsymbol{\beta} = (\beta_{ij})_{1\le i,\ j\le 2}$和$\boldsymbol{D} = (D_{ij})_{1\le i,\ j\le 2}$ 表示分枝比、率参数和形状参数矩阵。随后，我们将所有核参数和基准参数汇总为

$$\eta = (\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}) \in (0, \infty)^2 \times (0, \infty)^{2\times 2} \times (0, \infty)^{2\times 2} \times (0, \infty)^{2\times 2}.$$

我们假设 $\boldsymbol{\alpha}$ 的谱半径小于 1，以确保平稳性。遵循 Bacry 等人 [4], 的做法，强度为

$$\Lambda = (\lambda_1(\eta), \lambda_2(\eta))^\top = (I_2 - \boldsymbol{\alpha})^{-1} \boldsymbol{\mu}$$

（ [4]中的式（3））， CPCF 为

$$g(u; \eta, \theta) = 1 + \frac{\nu_{12}(u; \eta)}{\lambda_1(\eta)\, \lambda_2(\eta)}, \qquad u \in \mathbb{R},$$

其中， ν_{12} 是双变量霍克斯过程（ [4]中的式（8））的无穷小协方差矩阵 ν 的（1，2）分量:

$$\nu_{12}(u) = \Big(\Psi(u)\Sigma + \Sigma\Psi^\top(-u) + \widetilde{\Psi} * \Sigma\Psi^\top(u) \Big)_{1,2} \qquad u \in \mathbb{R}, \tag{5.1}$$

其中，$\Sigma = \mathrm{diag}\{\lambda_1, \lambda_2\}$, $\Psi(u) = (\Psi_{ij}(u))_{1\le i,\ j\le 2}$ 定义为

$$\Psi(u) = \sum_{m=1}^{\infty} \Phi^{(*m)}(u), \qquad \Phi^{(*m)} = \underbrace{\Phi * \cdots * \Phi}_{m \text{ times}},$$

核矩阵为 $\Phi(u) = \big(\phi_{ij}(u)\big)_{1\le i,\ j\le 2}$，其中 $*$ 表示矩阵卷积，而 $\widetilde{\Psi}(u) = \Psi(-u)$ $u \in \mathbb{R}$。

此外，令 $\theta \in \mathbb{R}$ 表示领先—滞后参数。给定双变量霍克斯过程$N = (N_1^0,\ N_2^0)$的一个实现，我们将平移后的过程 $(N_1,\ N_2) = (N_1^0,\ N_2^0(\cdot - \theta))$ 称为具有伽马核的滞后双变量霍克斯过程（*LBHPG*）。其分布记为

$$\mathrm{LBHPG}(\eta, \theta).$$

对于平移后的过程 LBHPG($\eta,\ \theta$)，其强度仍为 Λ，而交叉协方差密度仅发生平移：

$$\nu_{12}(u - \theta; \eta), \qquad u \in \mathbb{R}.$$

此外，LBHPG($\eta,\ \theta$)的交叉对相关函数（CPCF）为

$$g(u; \eta, \theta) = 1 + \frac{\nu_{12}(u - \theta; \eta)}{\lambda_1(\eta)\,\lambda_2(\eta)}, \qquad u \in \mathbb{R}. \tag{5.2}$$

在模拟研究中，我们仅考虑速率参数相同的情形，即对于所有 $i,\ j \in \{1, 2\}$，均有 $\beta_{ij} \equiv \beta$。在这种情况下，LBHPG($\eta,\ \theta$) 满足 [A2](ii)，其中 $\alpha = \min\{D_{12}, D_{21}\}$，如果 $\min\{D_{12}, D_{21}\} < 1$（发散的伽马核），并满足 [A2](i)，其中 $\alpha = 2$，如果 $D_{11} = D_{21} = D_{12} = D_{22} = 1$（指数核）。前者可由伽马密度的再生性以及双边伽马密度在原点处的局部行为推出，见 [51, 定理 6.1]。详情见附录 A.7。后者则由具有指数核的双变量霍克斯过程的 CPCF 显式公式推出，该公式可由 [6, 例 3]得到。由于 [52]中的定理 1，矩条件 [A1](i) 得以保证。此外，由于伽马核呈几何衰减，根据 [11], 中的定理 3.1，我们还可得到强混合速率 [A1](ii) 的一个上界。因此，在这些特殊情形下，LBHPG($\eta,\ \theta$) 满足定理 4.1 中的假设，且所用的（平滑）核满足假设 [K]。

5.1.2 具有伽马核的滞后二元 *Neyman-Scott* 过程

首先，我们按照 Shiotani & Yoshida [65] 的第 5.1 节，回顾在 $\mathbb{R}$ 上构造双变量 Neyman-Scott 过程的方法。设 $\mathcal{C}$ 是定义在 $\mathbb{R}$ 上、强度为 $\lambda > 0$ 的齐次泊松（母）过程。对于 $i = 1$、2 以及每个 $c \in \mathcal{C}$，令 $M_i(c)$ 表示来自母点 c 的子代中属于分量 i 的数量。假设 $\{M_i(c)\}_{c\in\mathcal{C}}$ 是取值于 $\mathbb{Z}_{\ge 0}$ 的随机变量 M_i 的独立同分布副本，且其均值 $\sigma_i = \mathbb{E}[M_i] \in (0, \infty)$ $i = 1$ 有限，2。为简单起见，假设 M_i 服从泊松分布 Poi(σ_i) $i = 1, 2$。给定 $M_i(c)$ 后，子代的时间偏移量 $\{d_i(c, m)\}_{m=1}^{M_i(c)}$ 独立同分布，密度为 f_i；它们在 i、c、m 之间相互独立，并且独立于 $\mathcal{C}$ 和 $\{M_i(c)\}$。接着，令

$$N_i^0 = \sum_{c\in\mathcal{C}} \sum_{m=1}^{M_i(c)} \delta_{c + d_i(c,m)}, \qquad i = 1, 2.$$

我们称 $N^0 = (N_1^0,\ N_2^0)$ 为双变量 Neyman-Scott 过程。

N^0 是平稳的，强度为

$$\lambda_i(\xi) = \lambda\,\sigma_i, \qquad i = 1, 2.$$

此外，$(N_1^0,\ N_2^0)$的交叉强度为

$$\lambda_{12}(u;\xi) = \lambda^2\sigma_1\sigma_2 + \lambda\sigma_1\sigma_2 \int_{\mathbb{R}} f_1(s;\tau_1) f_2(u+s;\tau_2)\, ds, \qquad u \in \mathbb{R},$$

参见式（5.2）及 [65]中的后续计算。因此，双变量 Neyman-Scott 过程的交叉成对相关函数（CPCF）为

$$g(u;\xi) = \frac{\lambda_{12}(u;\xi)}{\lambda_1(\xi)\lambda_2(\xi)} = 1 + \frac{1}{\lambda}\int_{\mathbb{R}} f_1(s;\tau_1) f_2(u+s;\tau_2)\, ds, \qquad u \in \mathbb{R},$$

其中 ξ 汇集了上文引入的所有参数。

在这项模拟研究中，我们采用 Shiotani 和 Yoshida [65] 的 NBNSP-G 模型所用的伽马扩散核。具体而言，对于 $i = 1$，2，我们假设

$$f_i(u;\tau_i) = \frac{l_i^{\alpha_i}}{\Gamma(\alpha_i)} u^{\alpha_i - 1} e^{-l_i u} 1_{(0,\infty)}(u), \qquad u \in \mathbb{R},$$

其中 $\tau_i = (\alpha_i,\ l_i)$，$\alpha_i > 0$ 是伽马分布的形状参数，$l_i > 0$ 是其率参数。然后，我们将这些参数汇集为

$$\xi = (\lambda, \sigma_1, \sigma_2, \alpha_1, \alpha_2, l_1, l_2) \in (0,\infty)^7.$$

令 $\theta \in \mathbb{R}$ 表示超前–滞后参数。给定上述伽马核双变量 Neyman-Scott 过程的一个实现 $N^0 = (N_1^0,\ N_2^0)$，我们通过平移第二个分量来定义滞后过程：

$$(N_1, N_2) = \big(N_1^0,\ N_2^0(\,\cdot\, - \theta)\big).$$

我们将 $N = (N_1, N_2)$ 称为伽马核滞后双变量内曼－斯科特过程（*LBNSPG*），并将其分布记为

$$\mathrm{LBNSPG}(\xi, \theta).$$

LBNSPG(ξ，θ) 的强度仍由 $\lambda_1(\xi)$ 和 $\lambda_2(\xi)$ 给出，而交叉强度仅作平移：

$$\lambda_{12}(u - \theta;\xi), \qquad u \in \mathbb{R}.$$

因此，LBNSPG(ξ，θ) 的 CPCF 为

$$g(u;\xi,\theta) = \frac{\lambda_{12}(u-\theta;\xi)}{\lambda_1(\xi)\lambda_2(\xi)} = 1 + \frac{1}{\lambda} p(u-\theta;\tau_1,\tau_2),$$
$$p(u;\tau_1,\tau_2) = \int_{\mathbb{R}} f_1(s;\tau_1) f_2(u+s;\tau_2)\, ds, \qquad u \in \mathbb{R}.$$

Name	族	α	β_α	收敛速率 T^{-1/β_α}
对称 Hawkes Gamma	LBHPG	0.4	0.4	$T^{-5/2}$
非对称 Hawkes Gamma	LBHPG	0.4	0.4	$T^{-5/2}$
Hawkes 指数	LBHPG	2	3	$T^{-1/2}$
NS Gamma 1	LBNSPG	0.8	0.8	$T^{-5/4}$
NS Gamma 2	LBNSPG	1.6	2.2	$T^{-5/11}$
ns gamma 3	LBNSPG	3	5	$T^{-1/5}$

表 1：各模型的 α （CPCF g 在 [A2]中的“尖锐度”参数；值越小，函数越尖锐）、 $\beta_\alpha = \alpha \vee$ （$2\alpha - 1$），以及定理 4.3 中收敛速率 T^{-1/β_α} 的极小极大下界。

函数 $p(\cdot;\, \alpha_1,\ \alpha_2,\ l_1,\ l_2)$ 是双边伽马分布的密度。K¨uchler 和 Tappe [51]详细分析了双边伽马密度的形状，包括单峰性以及零点附近的局部行为（参见其中的定理 6.1）。结合他们的结果与上述表示可知，当参数对称，即 $\alpha_1 = \alpha_2$ and$l_1 = l_2$ 时， $g(\cdot; \xi,\ \theta)$ 严格单峰，且峰值位于 θ。在模拟实验中，我们将注意力限制在此类对称情形下。在这一限制下，如果 $\alpha_1 + \alpha_2 < 1$，则 LBNSPG($\xi,\ \theta$) 满足 [A2](ii)；如果 $\alpha_1 + \alpha_2 > 1$，则满足 [A2](i)，两种情况下均有 $\alpha = \min\{\alpha_1 + \alpha_2, 3\}$。

注意，LBNSPG($\xi,\ 0$) 是第 6.1 节 [65] 中引入的 NBNSP-G 模型的特例（将噪声过程设为零，并令 M_i的分布为泊松分布）。因此，LBNSPG($\xi,\ 0$) 满足 [65], 中的条件 [NS]，据此该处的引理 10.13 给出了 LBNSPG($\xi,\ \theta$) 的 α 混合系数与扩散核的尾概率之间的界：对所有 $c_1 \geq 0$ 和 $m \geq 2r + 2$,

$$\alpha^N_{c_1,\infty}(m; r) := \sup_{c_2 \geq 0} \alpha^N_{c_1,c_2}(m; r) \leq 8m\lambda\,(m + 1 + 2r) \sum_{i=1}^{2} \sigma_i \int_{|z| \geq m/2-2r} f_i(z; \tau_i)\, dz.$$

由于伽马核具有几何衰减的尾部，这意味着 $\alpha_{c_1,\ \infty}(m; r_1)$ 的衰减速度快于 m 的任意次幂，因此 LBNSPG($\xi,\ \theta$) 满足假设 [A1](ii)。此外，由于服从泊松分布的 M_i 具有任意阶矩，引理 10.14 [65] 推得$N_i((0, 1])$ 具有任意阶有限矩，因此满足假设 [A1](i)。因此，若 $\alpha_1 = \alpha_2$、 $l_1 = l_2$ 和 $\alpha_1 + \alpha_2 \neq 1$，则 LBNSPG($\xi,\ \theta$) 满足假设 [A2] 和 [A1]，因而适用于定理 4.1。

5.1.3 模型及模拟研究中的参数设定

我们考虑两类二元平稳点过程模型，每类模型均包含三组参数值。这些模型分别标记为对称 Hawkes Gamma、非对称 Hawkes Gamma、Hawkes 指数、NS Gamma 1、NS Gamma 2 和 ns gamma 3。

所有 LBHPG 模型都具有

$$\boldsymbol{\mu} = (0.2, 0.2)^\top, \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} 10 & 10 \\ 10 & 10 \end{pmatrix}.$$

我们使用三个特定的 形状矩阵 $\boldsymbol{D}$ 的设定，以涵盖不同程度的正则性和不对称性：

$$\texttt{hawkes_gamma_sym}: \quad \boldsymbol{D} = \begin{pmatrix} 0.4 & 0.4 \\ 0.4 & 0.4 \end{pmatrix};$$

$$\texttt{hawkes_gamma_asym}: \quad \boldsymbol{D} = \begin{pmatrix} 0.4 & 0.4 \\ 0.8 & 0.4 \end{pmatrix};$$

$$\texttt{hawkes_exp}: \quad \boldsymbol{D} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

对于 LBNSPG 模型，我们固定

$$\lambda = 0.1, \qquad \sigma_1 = \sigma_2 = 4$$

并采用对称的扩散形状参数和速率参数。我们选择三种设定，以改变 CPCF 在峰值附近的平滑程度：

$$\begin{aligned}
\texttt{ns_gamma_1}: &\quad (\alpha_1, \alpha_2) = (0.4, 0.4), \quad (l_1, l_2) = (10, 10), \\
\texttt{ns_gamma_2}: &\quad (\alpha_1, \alpha_2) = (0.8, 0.8), \quad (l_1, l_2) = (10, 10), \\
\texttt{ns_gamma_3}: &\quad (\alpha_1, \alpha_2) = (2.0, 2.0), \quad (l_1, l_2) = (100, 100).
\end{aligned}$$

5.2 准确性

在本实验中，我们使用均方根误差（RMSE）比较六种情景下领先—滞后时间估计量的准确性。

对于每个观测时间间隔长度 $T \in \{1000, 2000, 4000, 8000\}$ 和估计量组合，我们生成 5000 条样本路径的蒙特卡洛重复。在每次重复中，真实的领先—滞后时间均抽取为 $\hat{}\theta^* \sim \mathcal{U}(-0.1,\ 0。1)$。对于给定的估计量 θ，我们报告

$$\mathrm{RMSE} = \left(\frac{1}{5000} \sum_{b=1}^{5000} \big(\hat{\theta}^{(b)} - \theta^{*(b)} \big)^2 \right)^{1/2},$$

其中，$\hat{}\theta^{(b)}$ 和 $\theta^{*(b)}$ 分别表示第 b次重复实验中的估计领先—滞后时间和真实领先—滞后时间。在各次重复实验之间随机化 θ^*，可以总结在一系列领先—滞后时间取值上的平均性能，而不是单一固定θ^*下的性能。领先—滞后时间参数应位于（-1，1）内，即我们设定 $r = 1$。对于 Lepski 方法和 DS 估计量，带宽网格均为 $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$。对于 Lepski 方法，我们设定 $A_T = \log\log T$。当对比函数在（$-r$，r）上存在多个最大化点时，我们选择其中的最小值作为确定性的平局判定规则。

图 2 展示了两种系统性规律。首先，DS估计量的性能对分桶宽度 h 敏感。对于每个 T，均方根误差随 h 变化的曲线通常呈 U 形：分桶过小会因联合激活稀少而导致方差较高，而分桶过大则会引入离散化偏差，因为领先—滞后时间被限制在较粗的网格上。此外，使均方根误差最小的 h 值

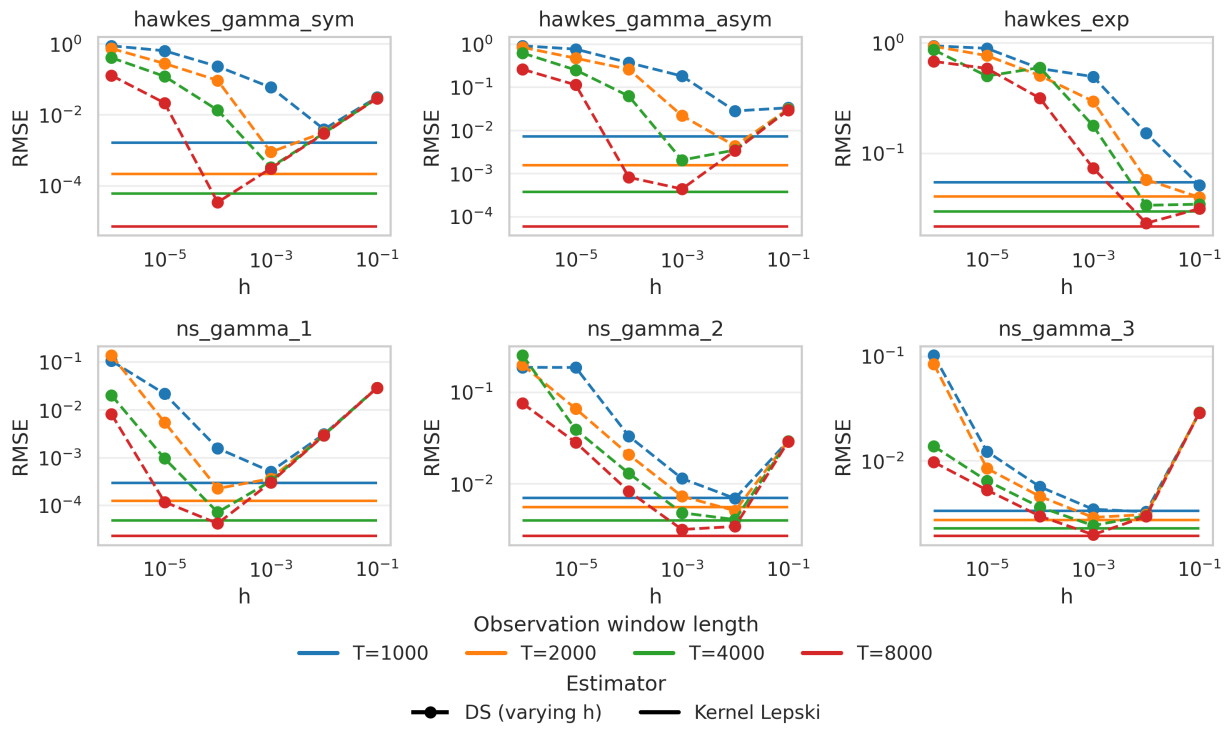

图 2：六种情景下领先—滞后时间估计量的均方根误差。每个面板均采用双对数坐标轴，横轴为带宽 h，纵轴为均方根误差。带标记的虚线描绘了 DS估计量在带宽网格上的表现，实心水平线则表示使用 Lepski 选定的带宽的核估计量。颜色表示观测窗口长度 $T \in \{1000, 2000, 4000, 8000\}$，并在各面板中保持一致。

取决于底层模型和观测窗口长度 T。这证实，在实践中，DS估计量需要对 h 进行模型特定调参。

其次，对于几乎所有 T 与数据生成过程的组合，采用 Lepski 带宽选择的核估计量都比 DS 估计量具有更低的均方根误差。在图 2 中，每个面板里对应 Lepski 估计量的实心水平线都接近（且通常低于）网格上 DS 估计量的最佳均方根误差。重要的是，这一改进无需手动调节带宽：一旦确定网格 $\mathcal{H}_T$ 和缓慢发散的常数 $A_T = \log\log T$，该方法便会根据数据自动调整平滑程度。这体现了所提出方法相较于 DS 方法的主要实际优势。

5.3 Lepski 方法的收敛速率及其对超参数的依赖

在本实验中，我们研究定理 4.2 所示的 Lepski 估计量的收敛速率，并考察估计量随调节参数 A_T 变化时的表现。

与上一个实验一样，对于每个观测窗口长度 $T \in \{1000, 2000, 4000, 8000\}$ 和每个估计量，我们生成样本路径的 5000 次蒙特卡洛重复。在每次重复中，真实的领先—滞后时间从 $\theta^* \sim \mathcal{U}(-0.1, 0.1)$ 中抽取。领先—滞后时间参数应位于 $(-1, 1)$ 内，即我们设定

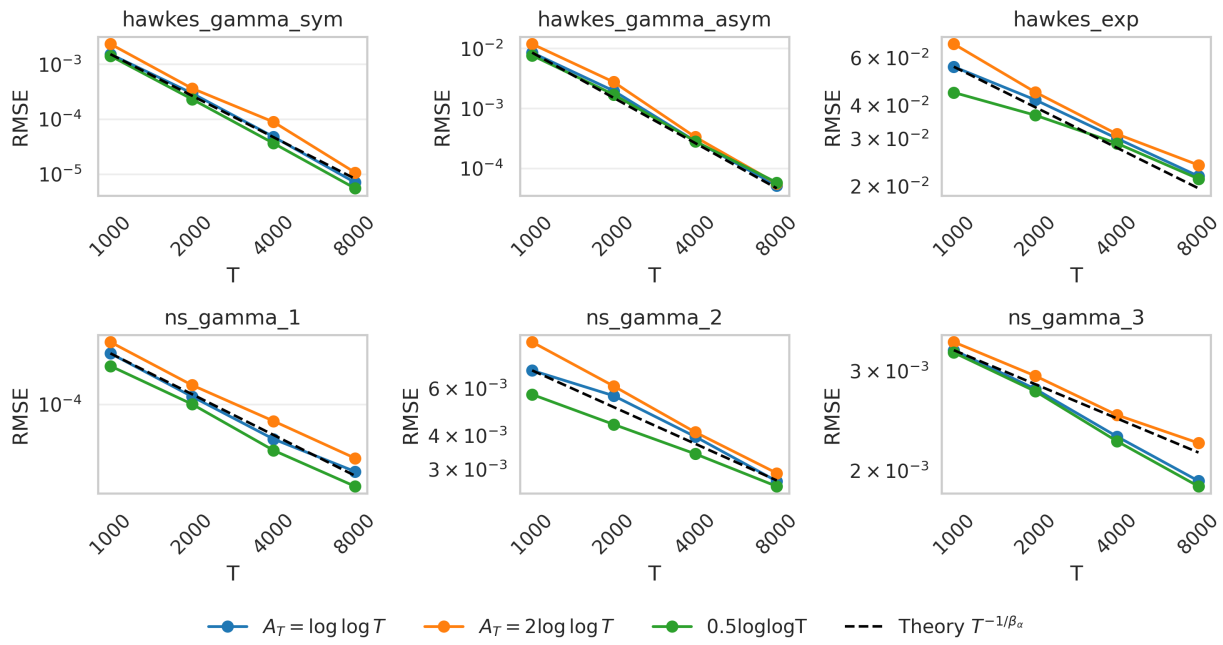

图 3：所有模拟设定和 A_T 种调度下，RMSE 随观测窗口 T 的变化（双对数坐标）。虚线表示理论 T^{-1/β_α} 斜率。

$r = 1$。带宽网格为 $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$。

图 3 考察了 Lepski 估计量的收敛速率。在每种情形下，图中都以双对数坐标绘出了 RMSE 随 T 的变化，并给出了根据定理 4.3 中的极小极大下界推导出的理论斜率 T^{-1/β_α}。在全部六个模型中，Lepski 估计量的经验曲线几乎与参考线平行。因此，所提出的估计量可能达到理论给出的最优收敛速率。

该图还比较了调节参数 A_T 的三种调度方式：0.5 log log T、log log T 和 2 log log T。在全部六个模型中，三条 RMSE 曲线彼此接近，且斜率相似，这表明在此范围内，收敛速率对 A_T 的选择具有稳健性。不同调度方式之间的差异主要体现为整体水平的偏移，因此，log log T 前面的具体常数对于达到极小极大速率并不关键。

6 实证说明

Dobrev & Schaumburg [27] 所描述的方法有一项突出特点：DS估计量能够有效捕捉地理位置相隔较远的两个市场之间最快的信息传输速度，而这一速度通常接近光速。具体而言，他们分析了美国10年期国债和标普500指数的现货市场与期货市场之间的领先—滞后关系。DS估计量稳定地检测到跨市场活动度量在5毫秒处出现尖峰，这与期货交易所（位于伊利诺伊州奥罗拉）和现货市场平台（位于新泽西州锡考克斯）之间的光传播时间相符。在本节中，我们将探究这一发现是否同样适用于地理位置更近的市场，因为检测此类关系需要亚毫秒级估计。

这一结论在地理位置更近的市场中是否依然成立；要检测此类关系，需要亚毫秒级估计。

更确切地说，我们考察同一只股票在两个不同交易所的报价之间的领先—滞后关系：NASDAQ（位于新泽西州卡特里特）和 BATS（位于新泽西州锡考克斯）。根据 [67, 表 2], , NASDAQ 与 BATS 交易所之间的光传播时间约为 0.1 毫秒。我们从 Daily TAQ 数据集中获取 2015年8月各交易所最佳报价更新的时间戳。该数据集提供综合交易协会（CTA）和非上市交易特权（UTP）的所有参与者向综合行情带报告的每一条报价。根据[8], , 该数据集于 2015年8月6日全面启用微秒时间戳。因此，我们将研究样本期限定为从 2015年8月6日开始的 18 个交易日。我们的分析涵盖 2015 年 NASDAQ-100指数的成分股，共 108 只股票。我们仅关注 9:45 至 15:45 之间发生的交易，剔除交易日开盘和收盘前后的 15 分钟，以避免开盘和收盘时常见的非平稳性。

在DailyTAQ数据中，每条报价记录包含两个时间戳：Time和参与者时间戳，分别对应证券信息处理商（SIPs）和交易所撮合引擎发布的时间戳。按照 [8], 的说法，我们将前者称为*SIP*时间戳，后者称为参与者时间戳。有关制度背景，请参见 [8, 第 2] 节和 [35, 第 4] 节。在本研究中，我们使用参与者时间戳，因为初步分析表明，SIP时间戳受到 [8]术语中的报告延迟严重污染。具体而言，即使多个市场事件同时发生并共享同一个参与者时间戳，它们通常也会获得不同的SIP时间戳，从而导致市场事件之间出现人为错位 [56, 62, 68]。此外，报告延迟会因各种延迟来源而动态波动，且其分布具有厚尾特征（参见 [44, 56] 和[8, 附录]），因此很难将其影响与真实的领先—滞后关系区分开来。基于这些原因，我们采用参与者时间戳。开发一种能够对这类时间戳污染保持稳健性的领先—滞后估计量，将是未来研究中一个有趣的方向。

ˆ 对于每个交易日和每只股票，我们分别计算核估计量 θ_h 和 DS估计量 $\hat{\theta}_h^{DS}$。对于核估计量，我们采用三角核，并通过第 4.1 节提出的 Lepski型方法选择带宽，其中 $\mathcal{H}_T = \{1\,\mu\text{s}\ \ 10\,\mu\text{s}\ \ 100\,\mu\text{s}\ \ 1000\,\mu\text{s}\}$， $A_T = \log\log\ T$，以及 $T = 21600$（6 个交易时段中的秒数）。对于 DS估计量，我们考虑两种桶大小： $h = 1\ \mu\text{s}$ 和 $h = 100\ \mu\text{s}$。前者对应于 [27], 中建议的最小时间单位，而后者的设定依据是两个交易所之间约 $100\ \mu\text{s}$ 的物理传输时间（参见 [67,表 2]）。我们将领先—滞后参数的搜索范围设为 $r = 10\ \text{ms}$。

ˆ 图4展示了三种估计量的领先—滞后时间估计小提琴图。 $\theta_{\hat{h}}$的估计值聚集在几个数值附近，最明显的是 95 μs和 130 μs附近。这些数值具有明确的物理解释：根据 [67, 表 2], , NASDAQ与BATS之间的传输时间通过混合激光链路约为95 μs，通过光纤电缆约为 128 μs。这还表明，在我们的样本中，NASDAQ通常领先于BATS更新最佳报价；鉴于所分析的所有股票均在NASDAQ上市，且NASDAQ往往具有更高的市场参与度，这一结果符合直觉。

对于带有 $h = 1\ \mu$s的DS估计量，我们观察到类似的聚集模式，不过这些估计值似乎更多地聚集在 130 μs附近，而不是 95 μs附近。然而，它们还表现出一些负的“离群值”，这是核估计量中不存在的问题。使用更大的桶大小 $h = 100\ \mu$s可以消除这类离群值，但DS估计量固有的粗粒度离散化会显著扭曲所估计的领先—滞后参数。

为突出我们的估计量与采用 $h = 1\ \mu$s的DS估计量之间的差异，图5展示了两组估计值的散点图，并按 $\sqrt{n_1 n_2}$（样本量的几何平均数）着色。在许多情况下，这两种估计量得到的数值相近，但当 n_1 和 n_2中的一个或两个变小时，它们的估计值便会出现差异。这一观察与我们的理论发现一致：样本量减小时，应增大DS估计量中的桶大小（参见定理 3.1之后的讨论）。

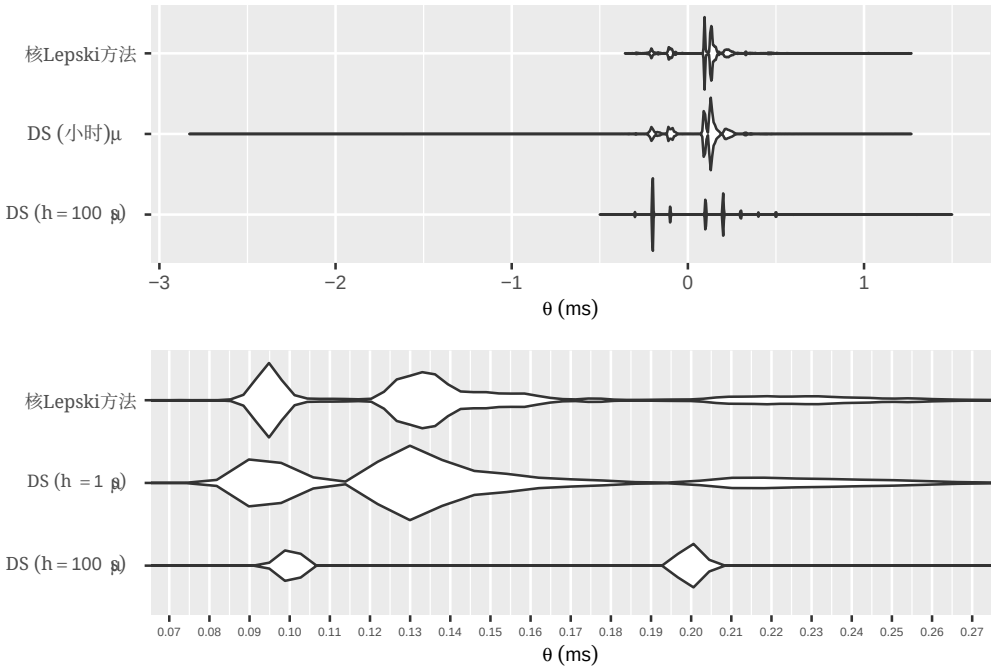

图4：2015年8月6日至31日期间，针对NASDAQ-100指数的所有成分股计算的NASDAQ和BATS交易所报价之间每日领先—滞后时间估计的小提琴图。上图显示完整图形，下图放大展示了0.07毫秒至0.27毫秒的区间。小提琴图的平滑带宽采用Sheather & Jones [64]方法选择，该方法通过R函数bw.SJ()实现。横轴以毫秒为单位。正估计值表示NASDAQ领先于BATS，反之亦然。

7 总结

本文从点过程视角为基于时间戳的领先—滞后分析奠定了理论基础。在这一框架下，Dobrev & Schaumburg [27]用于分析领先—滞后关系的方法可解释为对由两个时间戳序列生成的二元点过程的交叉对相关函数（CPCF）的估计量。因此，主导领先—滞后时间自然定义为该 CPCF 最尖锐峰值的位置。我们提出通过最大化 CPCF 的核密度估计量来估计这一

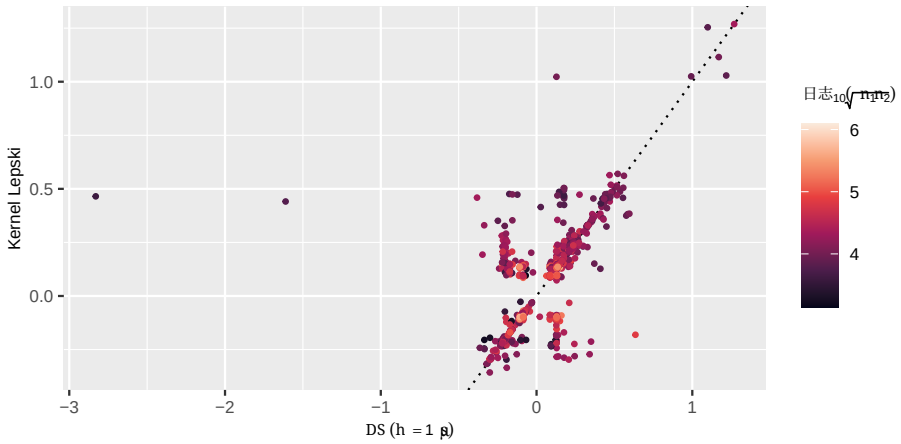

图5：$\hat{}\,\theta_h^{DS}$与 $h = 1\ \mu$s的估计值相对于$\hat{}\,\theta_{\hat{h}}$的散点图，按样本量的几何平均数以不同颜色编码。数值以毫秒表示。虚线为45度线。

领先—滞后时间。理论上，只要适当选择核估计量的带宽，我们的估计量就几乎达到了领先—滞后时间估计的最优收敛速率。为此，我们引入了一种 Lepski型带宽选择方法。在实践中，我们的方法解决了 Dobrev–Schaumburg估计量因其离散性质而产生的若干不足。我们通过全面的模拟研究和一项说明性实证分析，展示了该估计量的优越性能。

最后，我们讨论若干未来研究方向，以拓展本框架的适用范围。首先，将所提出的方法推广至非平稳情形具有实际意义。由于金融市场活动通常随时间变化，本研究所采用的平稳性假设可能会限制某些实证应用。一种有前景的途径是采用转移不变交叉对相关函数的概念（例如，参见 [63]），该概念允许强度随时间变化，同时保留仅取决于时间滞后的稳定领先—滞后结构。针对这类非齐次情形开发类似方法，将有助于更准确地估计领先—滞后时间。其次，在实际应用中，时间戳序列可能受到观测噪声污染。在这种情况下，需要一种对噪声具有稳健性的领先—滞后参数估计量。这个问题与反卷积密切相关； [21] 在点过程强度函数估计的背景下研究了这一问题，而 [54] 则在独立同分布数据众数估计的背景下对其进行了研究。第三，高频时间戳的经验 CPCF 估计通常会呈现多个峰值。这些峰值自然产生，是因为两个市场之间的信息传输具有双向性（如 [27] 所述），也因为不同市场参与者的信息传输速度存在差异。本文仅关注最高的峰值，但识别所有“显著”峰值也值得研究。众数估计的经典文献中研究过类似问题；参见 [17, 第 3] 节及其中的参考文献。第四，研究领先—滞后参数如何受到订单类型、规模以及价差和深度等市场状况的影响，也很有价值。通过将这些属性视为标记，我们可以分析标记点过程之间的领先—滞后关系。

附录

A 证明

在全文讨论中，我们采用以下记号和约定。对于两个实数 x, y，若存在某个仅依赖于 $r, \alpha, \alpha_0, \delta, b, (B_p)_{p\geq 1}, (B_{p,q})_{p,q\geq 1}, \varepsilon$和 $\|K\|_\infty$ 的常数 $C>0$，使得 $x \leq Cy$，则记作 $x \lesssim y$。对于定义在 $[-r_1, r_1]$ 上的实值函数 f，我们将$\|f\|_{L^p} = \|f\|_{L^p([-r_1,r_1])}$简记为每个$p \in [1,\infty]$。当 [A2](i) 成立时，我们令 $g_0 \equiv 1$。

A.1 预备知识

首先，根据交叉强度函数（3.1）的定义，我们很容易推得

$$\mathrm{E}\left[\int_{D\times\mathbb{R}} \varphi(y-x) N_1(dx) N_2(dy)\right] = \mathrm{Leb}(D)\lambda_1\lambda_2 \int_{\mathbb{R}} \varphi(u) g(u) du$$

对任意博雷尔函数 $\varphi:\ \mathbb{R}\to[0,\infty]$ 和 $D \in \mathcal{B}(\mathbb{R})$。下文将这一恒等式称为坎贝尔公式。

接下来，我们证明两个在全文讨论中起基础作用的辅助估计。第一个是关于相依随机变量之和的矩不等式，它以 α-混合系数（3.2）表示，并可通过截断论证从 [30, 定理 2]直接推出。

引理 A.1。设$(X_j)_{j=0}^{T-1}$是随机变量序列，使得对所有 j，X_j都是 $\sigma(N\cap(I_j\oplus r_1))$-可测的。另设$\max_{0\leq j\leq T-1}\mathrm{E}[X_j^q] < \infty$，其中 $q \geq 2$为偶整数。则存在一个仅依赖于 q 的常数 C_q，使得对于任意 M、$\tau > 0$,

$$\left\|\sum_{j=0}^{T-1}(X_j - \mathrm{E}[X_j])\right\|_q \leq C_q \left\{ \left(T\tau M \max_{0\leq j\leq T-1} \mathrm{E}[|X_j|] + TM^2 \sum_{m=\tau}^{\infty} \alpha_{q,q}^N(m; r_1) \right)^{1/2} + T^{1/q} M \left(\sum_{m=0}^{\infty} (m+1)^{q-2} \alpha_{q,q}^N(m; r_1) \right)^{1/q} \right\} + 2T \max_{0\leq j\leq T-1} \left\| X_j 1_{\{|X_j|>M\}} \right\|_q .$$

证明。对于 $j=0$，令 $Y_j := X_j 1_{\{|X_j|\leq M\}}${style id='6'>，1，.....，$T-1$。根据三角不等式，

$$\begin{aligned} \left\|\sum_{j=0}^{T-1}(X_j - \mathrm{E}[X_j]) - \sum_{j=0}^{T-1}(Y_j - \mathrm{E}[Y_j])\right\|_q &\leq \sum_{j=0}^{T-1} \left\| X_j 1_{\{|X_j|>M\}} - \mathrm{E}[X_j 1_{\{|X_j|>M\}}] \right\|_q \\ &\leq 2T \max_{0\leq j\leq T-1} \left\| X_j 1_{\{|X_j|>M\}} \right\|_q . \end{aligned} \tag{A.1}$$

同时，对于满足 $j_{k+1} - j_k = m$的任意整数 $0 \leq j_1 \leq \cdots \leq j_p \leq T-1$，其中 $0 \leq k < p$和$m \geq 0$满足上述条件，我们有

$$|\operatorname{Cov}[Y_{j_1}\cdots Y_{j_k}, Y_{j_{k+1}}\cdots Y_{j_p}]| \leq \min\left\{ 2M^{p-1} \max_{0\leq j\leq T-1} \mathrm{E}[|X_j|],\ 4M^p \alpha_{p,p}^N(m; r_1) \right\},$$

其中，第二个上界由强混合条件下的协方差不等式推出（例如参见 [29, Section 1.2] 中的引理 3）。因此，我们可以将 [30, Theorem 2] 应用于 Y_j，其中 $M = M, \gamma = C = 1$ 以及

$$\theta_m = \min\left\{ M \max_{0\leq j\leq T-1} \mathrm{E}[|X_j|],\ 4M^2 \alpha_{q,q}^N(m; r_1) \right\}$$

在他们的记号中。因此，存在一个仅依赖于 q 的常数 C_q，使得

$$\begin{aligned}\mathrm{E}\left[\left|\sum_{j=0}^{T-1}(Y_j-\mathrm{E}[Y_j])\right|^q\right] &\le C_q\left\{\left(T\sum_{m=0}^{T-1}\theta_m\right)^{q/2}+M^{q-2}T\sum_{m=0}^{T-1}(m+1)^{q-2}\theta_m\right\}\\ &\le C_q\left\{\left(2T\tau M\max_{0\le j\le T-1}\mathrm{E}[|X_j|]+4TM^2\sum_{m=\tau}^{\infty}\alpha_{q,q}^N(m;r_1)\right)^{q/2}\right.\\ &\qquad\left.+4TM^q\sum_{m=0}^{\infty}(m+1)^{q-2}\alpha_{q,q}^N(m;r_1)\right\}.\end{aligned}$$

将此结果与 (A.1) 结合，即得所需结论。 □

第二个是对核平滑 CPCF 的估计。

引理 A.2. 假设 K 有界且支撑于 $[-1,1]$。对于 $h\in(0,1]$，，定义函数$f_h:\mathbb{R}\to[0,\infty)$ 如下

$$f_h(u)=\int_{\mathbb{R}}K_h(v-u)g(v)dv=\int_{\mathbb{R}}K(t)g(u+ht)dt,\qquad u\in\mathbb{R}. \tag{A.2}$$

再假设存在常数 $\theta^*\in[-r,r]$, $\tilde\alpha\in(0,1]$, $b>1$ 以及函数 $g_0\in L^{\frac{1}{1-\tilde\alpha}}([-r_1,r_1])$，使得 ——

$$g(u)\le b\left\{g_0(u)+|u-\theta^*|^{\tilde\alpha-1}\right\}\quad \textit{for all } u\in[-r_1,r_1]. \tag{A.3}$$

那么我们有

$$f_h(u)\le 2^{\tilde\alpha}\|K\|_\infty b\left(\|g_0\|_{L^{1/(1-\tilde\alpha)}}+\frac{2}{\tilde\alpha}\right)h^{\tilde\alpha-1}\quad \textit{for all } u\in[-r,r].$$

证明。固定 $u\in[-r,r]$。(A.3) 给出

$$f_h(u)\le b\left(\int_{\mathbb{R}}K_h(v-u)g_0(v)dv+\int_{\mathbb{R}}K(t)|u-\theta^*+ht|^{\tilde\alpha-1}dt\right)=:b(I+II).$$

由于 K_h的支撑位于 $[-h,h]\subset[-1,1]$上，我们可以将 I改写为

$$I=\int_{-r_1}^{r_1}K_h(v-u)g_0(v)dv.$$

因此，由杨氏卷积不等式可得 $I\le\|K_h\|_{L^{1/\tilde\alpha}}\|g_0\|_{L^{1/(1-\tilde\alpha)}}$。由于

$$\|K_h\|_{L^{1/\tilde\alpha}}\le\left(2h\|K_h\|_\infty^{1/\tilde\alpha}\right)^{\tilde\alpha}\le 2^{\tilde\alpha}h^{\tilde\alpha-1}\|K\|_\infty,$$

我们得到

$$I\le 2^{\tilde\alpha}h^{\tilde\alpha-1}\|K\|_\infty\|g_0\|_{L^{1/(1-\tilde\alpha)}}. \tag{A.4}$$

同时，对于任意 $c\in\mathbb{R}$，直接计算可得

$$\int_{-1}^{1}|c+ht|^{\tilde\alpha-1}dt=\begin{cases}\dfrac{\left||c+h|^{\tilde\alpha}-|c-h|^{\tilde\alpha}\right|}{\tilde\alpha h} & \text{if } |c|>h,\\ \dfrac{|c+h|^{\tilde\alpha}+|c-h|^{\tilde\alpha}}{\tilde\alpha h} & \text{if } |c|\le h.\end{cases}$$

注意到对于任意 $x,\ y \geq 0$，我们得到

$$\sup_{c\in\mathbb{R}} \int_{-1}^{1} |c+ht|^{\tilde{\alpha}-1} dt \leq \frac{2^{1+\tilde{\alpha}}}{\tilde{\alpha}} h^{\tilde{\alpha}-1}. \tag{A.5}$$

因此，$II \leq \|K\|_\infty \frac{2^{1+\tilde{\alpha}}}{\tilde{\alpha}} h^{\tilde{\alpha}-1}$ 由于 K 的支撑在 $[-1,1]$ 上。结合 (A.4) 即得所需结果。

□

A.2 命题 *3.1* 的证明

在以下两个引理中，我们分别处理 $\mathcal{X}_h^{\mathrm{rel}}(\ell)$ 的分子和分母。

引理 A.3. 假设 [A1]。设存在常数 $\theta^* \in [-r,r]$, $\tilde{\alpha} \in (0,1]$, $b>1$ 以及一个函数$g_0 \in L^{1/(1-\tilde{\alpha})}([-r_1,r_1])$，使得 (A.3) 成立。还假设，当 $T\to\infty$ 时，$h = h_T \asymp T^{-\gamma}$，其中 $\gamma>0$。那么，对于任意 $\varepsilon>0$,

$$\mathrm{E}\left[\max_{\ell\in\mathcal{G}_h}\left|\mathcal{X}_h^{\mathrm{raw}}(\ell) - \frac{T}{h}\,\mathrm{E}[1_{\{N_1(I_0^h)>0, N_2(I_\ell^h)>0\}}]\right|\right] = O\left(1+\sqrt{Th^{\tilde{\alpha}-\varepsilon}}\right). \tag{A.6}$$

证明。令

$$\mathcal{X}_h^0(\ell) := \sum_{k=0}^{T/h-1} 1_{\{N_1(I_k^h)>0,\ N_2(I_{k+\ell}^h)>0\}}.$$

由于

$$\begin{aligned}|\mathcal{X}_h^{\mathrm{raw}}(\ell) - \mathcal{X}_h^0(\ell)| &\leq \sum_{k=0}^{|\ell|-1} 1_{\{N_1(I_k^h)>0\}} + \sum_{k=T/h-|\ell|}^{T/h-1} 1_{\{N_1(I_k^h)>0\}} \\ &\leq N_1((0,|\ell|h]) + N_1((T-|\ell|h,T]),\end{aligned}$$

我们有

$$\mathrm{E}\left[\max_{\ell\in\mathcal{G}_h}|\mathcal{X}_h^{\mathrm{raw}}(\ell) - \mathcal{X}_h^0(\ell)|\right] \leq \mathrm{E}[N_1((0,r])] + \mathrm{E}[N_1((T-r,T])] = 2\lambda_1 r = O(1). \tag{A.7}$$

此外，根据平稳性，$\mathrm{E}[\mathcal{X}_h^0(\ell)] = (T/h)\,\mathrm{E}[1_{\{N_1(I_0^h)>0,\ N_2(I_\ell^h)>0\}}]$。因此，只要证明

$$\mathrm{E}\left[\max_{\ell\in\mathcal{G}_h}\left|\mathcal{X}_h^0(\ell) - \mathrm{E}[\mathcal{X}_h^0(\ell)]\right|\right] = O\left(\sqrt{Th^{\tilde{\alpha}-\varepsilon}}\right).$$

对于任意 $p>1$，由詹森不等式可得

$$\begin{aligned}\mathrm{E}\left[\max_{\ell\in\mathcal{G}_h}\left|\mathcal{X}_h^0(\ell) - \mathrm{E}[\mathcal{X}_h^0(\ell)]\right|\right] &\leq \left\|\max_{\ell\in\mathcal{G}_h}\left|\mathcal{X}_h^0(\ell) - \mathrm{E}[\mathcal{X}_h^0(\ell)]\right|\right\|_p \\ &\leq |\mathcal{G}_h|^{1/p} \max_{\ell\in\mathcal{G}_h}\left\|\mathcal{X}_h^0(\ell) - \mathrm{E}[\mathcal{X}_h^0(\ell)]\right\|_p.\end{aligned}$$

设 p 是一个偶整数，使得 $p \geq \frac{4}{\varepsilon(1-\tilde{\alpha}\gamma)}$ 那么我们有 $|\mathcal{G}_h|^{1/p} = O(h^{-\varepsilon/4})$。因此，只需证明

$$\max_{\ell\in\mathcal{G}_h}\left\|\mathcal{X}_h^0(\ell) - \mathrm{E}[\mathcal{X}_h^0(\ell)]\right\|_p = O\left(\sqrt{Th^{\tilde{\alpha}-\varepsilon/2}}\right). \tag{A.8}$$

对每个 $\ell \in \mathcal{G}_h$，令

$$Y_j(\ell) := \sum_{k=j/h}^{(j+1)/h-1} 1_{\{N_1(I_k^h)>0, N_2(I_{k+\ell}^h)>0\}},$$

从而 $\mathcal{X}_h^0(\ell) = \sum_{j=0}^{T-1} Y_j(\ell)$。注意，$Y_j(\ell)$ 是 $\sigma(N \cap (I_j \oplus r))$-可测的。此外，由于

$$Y_j(\ell) \le \sum_{k=j/h}^{(j+1)/h-1} N_1(I_k^h) = N_1(I_j),$$

[A1](i) 得出

$$\sup_j \|Y_j(\ell)\|_q \le \|N_1(I_j)\|_q \lesssim \lambda_1 \tag{A.9}$$

对任意 $q \ge 1$。因此，将引理 A.1 应用于 $(Y_j(\ell))_{j=0}^{T-1}$，其中 $M = h^{-\varepsilon/4}$ 和 $\tau = \lfloor h^{-\varepsilon/4} \rfloor$，可得

$$\begin{aligned}\left\|\mathcal{X}_h^0(\ell) - \mathrm{E}[\mathcal{X}_h^0(\ell)]\right\|_p \le C_p \Bigg\{ &\left(Th^{-\varepsilon/2}\,\mathrm{E}[Y_0(\ell)] + Th^{-\varepsilon/2} \sum_{m=\lfloor h^{-\varepsilon/4} \rfloor}^{\infty} \alpha_{p,p}^N(m; r_1) \right)^{1/2} \\ &+ T^{1/p} h^{-\varepsilon/4} \left(\sum_{m=0}^{\infty} (m+1)^{p-2} \alpha_{p,p}^N(m; r_1) \right)^{1/p} \Bigg\} + 2T \left\| Y_0(\ell) 1_{\{Y_0(\ell) > h^{-\varepsilon/4}\}} \right\|_p,\end{aligned}$$

其中，C_p 是一个仅依赖于 p 和 λ_1 的常数。根据 [A1](ii),

$$\sum_{m=\lfloor h^{-\varepsilon/4} \rfloor}^{\infty} \alpha_{p,p}^N(m; r_1) = O(h^{\tilde{\alpha}}) \quad \text{and} \quad \sum_{m=0}^{\infty} (m+1)^{p-2} \alpha_{p,p}^N(m; r_1) = O(1).$$

此外，根据 (A.9)，我们有 $\|Y_0(\ell) 1_{\{Y_0(\ell) > h^{-\varepsilon/4}\}}\|_p \le h^{p\varepsilon/4} (\mathrm{E}[Y_0(\ell)^{p^2+p}])^{1/p} = O(T^{-1})$。因此，只要证明

$$\max_{\ell} \mathrm{E}[Y_0(\ell)] = O\left(h^{\tilde{\alpha}}\right). \tag{A.10}$$

注意到

$$\mathrm{E}[Y_0(\ell)] = h^{-1}\,\mathrm{E}[1_{\{N_1(I_0^h)>0, N_2(I_\ell^h)>0\}}] \le h^{-1}\,\mathrm{E}[N_1(I_0^h) N_2(I_\ell^h)]$$

and

$$\begin{aligned}\mathrm{E}[N_1(I_0^h) N_2(I_\ell^h)] &= \lambda_1 \lambda_2 \int_{\mathbb{R}^2} 1_{I_0^h}(x) 1_{I_\ell^h}(x+u) g(u) dx du \\ &= \lambda_1 \lambda_2 h^2 \int_{\mathbb{R}} K_h^{\mathrm{tri}}(u - \ell h) g(u) du. \end{aligned} \tag{A.11}$$

因此，(A.10) 由引理 A.2 得出。 □

引理 A.4. 假设 [A1]。那么，

$$\begin{aligned}\max_{\ell \in \mathcal{G}_h} \left| \frac{1}{T} \sum_{k=|\ell|}^{T/h-1-|\ell|} 1_{\{N_1(I_k^h)>0\}} - \frac{\mathrm{P}(N_1(I_0^h) > 0)}{h} \right| &= O_p\left(\frac{1}{\sqrt{T}}\right), \\ \max_{\ell \in \mathcal{G}_h} \left| \frac{1}{T} \sum_{k=|\ell|}^{T/h-1-|\ell|} 1_{\{N_2(I_{k+\ell}^h)>0\}} - \frac{\mathrm{P}(N_2(I_0^h) > 0)}{h} \right| &= O_p\left(\frac{1}{\sqrt{T}}\right)\end{aligned} \tag{A.12}$$

and

$$\max_{\ell\in\mathcal{G}_h}\left|\frac{1}{T}\min\left\{\sum_{k=|\ell|}^{T/h-1-|\ell|}1_{\{N_1(I_k^h)>0\}},\sum_{k=|\ell|}^{T/h-1-|\ell|}1_{\{N_2(I_{k+\ell}^h)>0\}}\right\}-\lambda_1\wedge\lambda_2\right|\to^p 0 \tag{A.13}$$

如 $T\to\infty$。

证明。由 (A.12) 和强度的定义（参见 [23] 中的式 (3.3.4)），可得 (A.13)，因此只需证明 (A.12)。我们只证明 (A.12) 的第一个等式，因为第二个等式可以用几乎相同的论证得到。

首先，按照 (A.7) 的证明中的相同论证可得

$$\mathrm{E}\left[\max_{\ell\in\mathcal{G}_h}\left|\sum_{k=|\ell|}^{T/h-1-|\ell|}1_{\{N_1(I_k^h)>0\}}-\sum_{k=0}^{T/h-1}1_{\{N_1(I_k^h)>0\}}\right|\right]=O(1).$$

因此，只需证明

$$\frac{1}{T}\sum_{k=0}^{T/h-1}1_{\{N_1(I_k^h)>0\}}=\frac{\mathrm{P}(N_1(I_0^h)>0)}{h}+O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A.14}$$

我们将左端改写为

$$\frac{1}{T}\sum_{k=0}^{T/h-1}1_{\{N_1(I_k^h)>0\}}=\frac{1}{T}\sum_{j=0}^{T-1}X_j,$$

其中 $X_j:=\sum_{k=j/h}^{(j+1)/h-1}1_{\{N_1(I_k^h)>0\}}$。注意，$X_j$ 是 $\sigma(N\cap I_j)$-可测的。因此，

$$\mathrm{Var}\left[\sum_{j=0}^{T-1}X_j\right]\le\sum_{j,m=0}^{T-1}|\operatorname{Cov}[X_j,X_{j+m}]|\le 8\sum_{j,m=0}^{T-1}\|X_j\|_4\|X_m\|_4\sqrt{\alpha_{1,1}(m;0)}\lesssim T\|X_0\|_4^2,$$

其中第二个不等式由 [29, 第 1 节中的定理 32] 得出，第三个不等式由 [A1](ii) 得出。由于 $|X_0|\le N_1(I_0)$，我们得到

$$\sum_{j=0}^{T-1}X_j=\sum_{j=0}^{T-1}\mathrm{E}[X_j]+O_p(\sqrt{T})=\frac{T}{h}\mathrm{E}[1_{\{N_1(I_0^h)>0\}}]+O_p(\sqrt{T}).$$

由于 $\mathrm{E}[1_{\{N_1(I_0^h)>0\}}]=\mathrm{P}(N_1(I_0^h)>0)$，我们得到 (A.14)。 □

命题 *3.1* 的证明。由于 g 有界，(A.3) 对 $b=\sup_{x\in\mathbb{R}}|g(x)|$, $\tilde{\alpha}=1$ 和 $g_0\equiv 0$ 成立（θ^* 任意）。此外，根据假设，当 $T\to\infty$ 时，$Th^{\tilde{\alpha}+\varepsilon}\to\infty$，其中某个 $\varepsilon>0$。因此，由引理 A.3 可得

$$\max_{\ell\in\mathcal{G}_h}\left|\frac{\mathcal{X}_h^{\mathrm{raw}}(\ell)}{Th}-\frac{\mathrm{E}[1_{\{N_1(I_0^h)>0,N_2(I_\ell^h)>0\}}]}{h^2}\right|\to^p 0. \tag{A.15}$$

特别地，这给出 $\max_{\ell\in\mathcal{G}_h}\mathcal{X}_{\mathrm{rawh}}(\ell)/Th=O_p(1)$。将其与 (A.13) 结合可得

$$\max_{\ell\in\mathcal{G}_h}\left|\frac{\mathcal{X}_h^{\mathrm{rel}}(\ell)}{h}-\frac{1}{\lambda_1\wedge\lambda_2}\frac{\mathrm{E}[1_{\{N_1(I_0^h)>0,N_2(I_\ell^h)>0\}}]}{h^2}\right|\to^p 0.$$

由于 $\lambda_1\lambda_2/(\lambda_1 \wedge \lambda_2) = \lambda_1 \vee \lambda_2$，只要证明

$$\max_{\ell \in \mathcal{G}_h} \left| \frac{\mathrm{E}[1_{\{N_1(I_0^h)>0, N_2(I_\ell^h)>0\}}]}{h^2} - \lambda_1\lambda_2 \int_{\mathbb{R}} K_h^{\mathrm{tri}}(u-\ell h) g(u) du \right| = o(1).$$

注意到

$$\begin{aligned}
0 &\le \mathrm{E}[N_1(I_0^h)N_2(I_\ell^h)] - \mathrm{E}[1_{\{N_1(I_0^h)>0, N_2(I_\ell^h)>0\}}] \\
&= \mathrm{E}[(N_1(I_0^h)-1)_+ \cdot N_2(I_\ell^h)] + \mathrm{E}[1_{\{N_1(I_0^h)>0\}}(N_2(I_\ell^h)-1)_+] \\
&\le \mathrm{E}[N_1(I_0^h)(N_1(I_0^h)-1)_+ \cdot N_2(I_\ell^h)] + \mathrm{E}[N_1(I_0^h)N_2(I_\ell^h)(N_2(I_\ell^h)-1)_+] \\
&= \mathrm{E}[N_1(I_0^h)(N_1(I_0^h)-1)N_2(I_\ell^h)] + \mathrm{E}[N_1(I_0^h)N_2(I_\ell^h)(N_2(I_\ell^h)-1)].
\end{aligned}$$

由于我们假设 $\varpi = 1$ 满足 (3.4)，因此得到

$$\max_{\ell \in \mathcal{G}_h} \left| \mathrm{E}[N_1(I_0^h)N_2(I_\ell^h)] - \mathrm{E}[1_{\{N_1(I_0^h)>0, N_2(I_\ell^h)>0\}}] \right| = o(h^2).$$

将其与 (A.11) 结合，即可得到所需结果。 □

A.3 定理 *3.1* 的证明

以下引理总结了 [A2] 所蕴含的可识别性条件。

引理 A.5. 假设 [K]，并定义函数 f_h，其形式如(A.2)所示。

(a) 假设 [A2](i)。那么，对于某个 $\sigma \in \{-1, 1\}$，存在常数 $A > 1$ 和 $0 < h_0 < 1$，它们仅取决于 α、b、δ，使得

$$f_h(\theta^* + \sigma v) - \sup_{u \in [-r,r]: |u-\theta^*| > Ah} f_h(u) \ge h^{\alpha-1} \quad \textit{for all } h < h_0 \textit{ and } v \in [h, 2h]. \tag{A.16}$$

(b) 假设 [A2](ii)。那么，对于某个 $\sigma \in \{-1, 1\}$，存在常数 $A > 1$、$c > 0$ 和 $0 < h_0 < 1$，它们仅取决于 α、α_0、b、δ、K，使得

$$f_h(\theta^* + \sigma v) - \sup_{u \in [-r,r]: |u-\theta^*| > Ah} f_h(u) \ge ch^{\alpha-1} \quad \textit{for all } h < h_0 \textit{ and } v \in [0, h]. \tag{A.17}$$

证明。（a）根据假设，我们有

$$\sup_{0 < \sigma(u-\theta^*) < \delta} \frac{g(\theta^*) - g(u)}{|u-\theta^*|^{\alpha-1}} \le b \tag{A.18}$$

对于某个 $\sigma \in \{-1, 1\}$。此外，我们有

$$g(u) \le g(\theta^*) - b^{-1} \min\left\{1, |u-\theta^*|^{\alpha-1}\right\} \tag{A.19}$$

对所有 $u \in \mathbb{R}$ 均成立。现在，对于任意 $A > 1$，根据 [K] 和（A.19），我们有

$$\sup_{u \in [-r,r]: |u-\theta^*| > Ah} f_h(u) \le g(\theta^*) - b^{-1} \sup_{u \in [-r,r]: |u-\theta^*| > Ah} \int_{-1}^{1} K(t) \min\left\{1, |u-\theta^*+ht|^{\alpha-1}\right\} dt$$

$$\leq g(\theta^*) - b^{-1}\min\left\{1, (A-1)^{\alpha-1}h^{\alpha-1}\right\}.$$

同时，根据（A.18），对于所有 $h < \delta/3$ 和 $v \in [h, 2h]$，我们有

$$f_h(\theta^* + \sigma v) \geq g(\theta^*) - b\int_{-1}^{1} K(t)|\sigma v + ht|^{\alpha-1}dt \geq g(\theta^*) - b(3h)^{\alpha-1}.$$

结合这些估计可得

$$f_h(\theta^* + \sigma v) - \sup_{u\in[-r,r]:|u-\theta^*|>Ah} f_h(u) \geq b^{-1}\min\left\{1, (A-1)^{\alpha-1}h^{\alpha-1}\right\} - b(3h)^{\alpha-1}.$$

因此，如果 $A \geq 1 + (b + 3^{\alpha-1}b^2)^{1/(\alpha-1)}$ 且 $h < \min\{\delta/3, \{b(3^{\alpha-1}b+1)\}^{-1/(\alpha-1)}\}$，则有 (A.16)。

(b) 根据假设，有

$$\inf_{0<\sigma(u-\theta^*)<\delta} \frac{g(u)}{|u-\theta^*|^{\alpha-1}} \geq \frac{1}{b} \tag{A.20}$$

对某个 $\sigma \in \{-1, 1\}$成立。此外，根据 [K]，对于某个 $0 < \delta_0 < 1$，我们有 $\inf_{t\in[-\delta_0,\delta_0]} K(t) \geq K(0)/2 > 0$。结合 (A.20)，对于所有 $h < \delta/2$ 和 $v \in [0, h]$，有

$$f_h(\theta^* + \sigma v) \geq \frac{K(0)}{2b}\int_0^{\delta_0} |v + ht|^{\alpha-1}dt \geq \frac{K(0)}{2b}h^{\alpha-1}\int_0^{\delta_0}(1+t)^{\alpha-1}dt \geq \frac{K(0)\delta_0}{4b}h^{\alpha-1}.$$

同时，将 (A.4) 应用于 $\tilde{\alpha} = \alpha_0$ 可得

$$\int_{\mathbb{R}} K_h(v-u)g_0(v)dv \leq 2^{\alpha_0}h^{\alpha_0-1}\|K\|_\infty b.$$

因此，对于任意 $A > 1$，根据 [A2] 和 [K]，有

$$\begin{aligned}\sup_{u\in[-r,r]:|u-\theta^*|>Ah} f_h(u) &\leq b\|K\|_\infty\left(2^{\alpha_0}h^{\alpha_0-1}b + \sup_{u\in[-r,r]:|u-\theta^*|>Ah}\int_{-1}^{1}|u-\theta^*+ht|^{\alpha-1}dt\right)\\ &\leq 2b\|K\|_\infty\left(bh^{\alpha_0-1} + \frac{h^{\alpha-1}}{(A-1)^{1-\alpha}}\right).\end{aligned}$$

因此，若 A 足够大，使得

$$\frac{2\|K\|_\infty b}{(A-1)^{1-\alpha}} \leq \frac{K(0)\delta_0}{8b},$$

我们有

$$f_h(\theta^* + \sigma v) - \sup_{u\in[-r,r]:|u-\theta^*|>Ah} f_h(u) \geq \frac{K(0)\delta_0}{8b}h^{\alpha-1} - 2b^2\|K\|_\infty h^{\alpha_0-1}.$$

因此，若满足以下条件，则 (A.17) 成立

$$h_0 \leq \min\left\{\frac{\delta}{2}, \left(\frac{K(0)\delta_0}{32\|K\|_\infty b^3}\right)^{1/(\alpha_0-\alpha)}\right\} \quad \text{and} \quad c \leq \frac{K(0)\delta_0}{16b}.$$

证明至此完成。 □

定理的证明　推论 *3.1.* 注意，在 $\tilde{\alpha} = \alpha \wedge 1$ [A2] 下，(A.3) 成立。因此，由引理 A.3 可得

$$h^{1-\alpha} \max_{\ell \in \mathcal{G}_h} \left| \frac{\mathcal{X}_h^{\mathrm{raw}}(\ell)}{Th} - \frac{E[1_{\{N_1(I_0^h)>0, N_2(I_\ell^h)>0\}}]}{h^2} \right| = O\left(\frac{h^\alpha}{T} + \frac{1}{\sqrt{Th^{\beta_\alpha + \varepsilon}}} \right) \quad \text{for any } \varepsilon > 0,$$

其中 我们使用了恒等式 $2\alpha - \alpha \wedge 1 = \beta_\alpha$。此外，根据命题 3.1 的证明以及 (3.4),

$$\max_{\ell \in \mathcal{G}_h} \frac{\left| \mathrm{E}[N_1(I_0^h) N_2(I_\ell^h)] - \mathrm{E}[1_{\{N_1(I_0^h)>0, N_2(I_\ell^h)>0\}}] \right|}{h^2} = o(h^{\alpha-1}).$$

现在，在 (A.2) 中定义函数 f_h，其中 $K = K^{\mathrm{tri}}$。回顾 $h \asymp T^{-\gamma}$，其中 $0 < \gamma < 1/\beta_\alpha$。因此，结合上述两个方程与 (A.11)，我们得到

$$h^{1-\alpha} \max_{\ell \in \mathcal{G}_h} \left| \frac{\mathcal{X}_h^{\mathrm{raw}}(\ell)}{Th} - \lambda_1 \lambda_2 f_h(\ell h) \right| \to^p 0.$$

将其与引理 A.2 结合，特别可得 $\max_{\ell \in \mathcal{G}_h} \mathcal{X}_{\mathrm{raw}h}(\ell)/Th = O_p(h^{\tilde{\alpha}-1})$。因此，由引理 A.4,

$$h^{1-\alpha} \max_{\ell \in \mathcal{G}_h} \left| \frac{\mathrm{P}(N_1(I_0^h) > 0) \wedge \mathrm{P}(N_2(I_0^h) > 0)}{h^2} \mathcal{X}_h^{\mathrm{rel}}(\ell) - \frac{\mathcal{X}_h^{\mathrm{raw}}(\ell)}{Th} \right| \to^p 0.$$

因此，我们有

$$h^{1-\alpha} \bar{\Delta}_T \to^p 0 \qquad \text{as} \quad T \to \infty, \tag{A.21}$$

其中

$$\bar{\Delta}_T := \max_{\ell \in \mathcal{G}_h} \left| \frac{\mathrm{P}(N_1(I_0^h) > 0) \wedge \mathrm{P}(N_2(I_0^h) > 0)}{\lambda_1 \lambda_2 h^2} \mathcal{X}_h^{\mathrm{rel}}(\ell) - f_h(\ell h) \right|.$$

我们转向证明的主体部分。考虑情形 $\alpha > 1$。那么，根据引理 A.5，对于某个$\sigma \in \{-1, 1\}$，存在常数 $A > 1$和 $0 < h_0 < 1$，它们仅取决于 α, b, δ，使得 (A.16) 成立。我们可以找到一个整数 ℓ^*，使得 $\ell^* h = \theta^* + \sigma v$对于某个 $v \in [h, 2h]$成立。注意，当 h充分小时， $\ell^* \in \mathcal{G}_h$成立。于是，由于 $\hat{\theta}_h^{DS}$是 $\mathcal{G}_h \ni \ell \mapsto \mathcal{X}_h^{\mathrm{rel}}(\ell) \in [0, \infty)$ 的最大化点，我们有

$$\begin{aligned}
\mathrm{P}\left(|\hat{\theta}_h^{DS} - \theta^*| > (A+4)h \right) &\le \mathrm{P}\left(|\hat{\theta}_h^{DS} - \ell^* h| > (A+2)h \right) \\
&\le \mathrm{P}\left(\mathcal{X}_h^{\mathrm{rel}}(\ell^*) \le \max_{\ell \in \mathcal{G}_h : |\ell - \ell^*| > A+2} \mathcal{X}_h^{\mathrm{rel}}(\ell) \right) \\
&\le \mathrm{P}\left(f_h(\ell^* h) \le \max_{\ell \in \mathcal{G}_h : |\ell - \ell^*| > A+2} f_h(\ell h) + 2\bar{\Delta}_T \right) \\
&\le \mathrm{P}\left(f_h(\ell^* h) \le \max_{\ell \in \mathcal{G}_h : |\ell h - \theta^*| > Ah} f_h(\ell h) + 2\bar{\Delta}_T \right).
\end{aligned}$$

因此，(A.16) 给出 $\mathrm{P}(|\hat{\theta}_h^{DS} - \theta^*| > (A+4)h) \le \mathrm{P}\left(2\bar{\Delta}_T \ge h^{\alpha-1}\right)$。于是，由 (A.21) 可得所需结果。

接下来，考虑 $\alpha < 1$的情形。然后，根据引理 A.5，对于某个 $\sigma \in \{-1, 1\}$，存在常数 $A > 1, c > 0$和 $0 < h_0 < 1$，它们仅取决于 $\alpha, \alpha_0, b, \delta$，使得 (A.17) 成立。我们可以找到一个整数 ℓ^*，使得 $\ell^* h = \theta^* + \sigma v$ 对某个 $v \in [0, h]$成立。然后，类似于上面的论证可知，当 $T \to \infty$时，$\mathrm{P}(|\hat{\theta}_h^{DS} - \theta^*| > (A+2)h) \to 0$。 □

A.4 定理 *4.1* 的证明

令 $\tilde{\alpha} := \alpha \wedge 1$。注意到 $\tilde{\alpha} \leq \beta_\alpha$。由于 (4.1) 的左端总是被 1 所界，我们不妨设 $Th^{\beta_\alpha+\varepsilon} \geq 1$。

考虑以下统计量:

$$\tilde{g}_h(u) := \frac{n_1}{T\lambda_1}\frac{n_2}{T\lambda_2}\hat{g}_h(u) = \frac{1}{T\lambda_1\lambda_2}\int_{(0,T]^2} K_h(y-x-u)N_1(dx)N_2(dy), \quad u \in \mathbb{R}.$$

注意到，θ_h 也是 $u \in [-r,r]$ 上 $\tilde{g}_h(u)$ 的一个最大化点。此外，我们可以将其改写为 $\tilde{g}_h(u) = \sum_{j=0}^{T-1} X_j^0(u)$，其中

$$X_j^0(u) = \frac{1}{T\lambda_1\lambda_2}\int_{I_j\times(0,T]} K_h(y-x-u)N_1(dx)N_2(dy).$$

我们将 $X_j^0(u)$ 的边界校正版本定义为

$$X_j(u) = \frac{1}{T\lambda_1\lambda_2}\int_{I_j\times\mathbb{R}} K_h(y-x-u)N_1(dx)N_2(dy).$$

不难看出，$(X_j(u))_{j\in\mathbb{Z}}$ 是平稳的。我们首先证明，在我们的论证中，用 $X_j(u)$ 替换 $X_j^0(u)$ 并无影响。

Lemma A.6. 假设 [A1]。还假设 K 有界且支撑于 $[-1,1]$。那么，

$$\mathrm{E}\left[\sup_{u\in[-r,r]}\left|\tilde{g}_h(u) - \sum_{j=0}^{T-1} X_j(u)\right|\right] \lesssim \frac{1}{Th}.$$

证明。由于 K_h 的支撑为 $[-h,h] \subset [-1,1]$，对于任意 $u \in [-r,r]$，$X_j^0(u) = X_j(u)$ 当 $r_1 < j < T-1-r_1$ 时成立。因此，

$$\left|\tilde{g}_h(u) - \sum_{j=0}^{T-1} X_j(u)\right| \leq \frac{\|K\|_\infty}{Th\lambda_1\lambda_2} \sum_{0\leq j\leq r_1 \text{ or } T-1-r_1\leq j\leq T-1} N_1(I_j)N_2(I_j \oplus r_1).$$

因此，根据 [A1](i)，所需结果即得。 □

接下来，令

$$\Delta_T(u) := \sum_{j=0}^{T-1}\{X_j(u) - \mathrm{E}[X_j(u)]\}, \qquad u \in [-r,r].$$

我们的下一个目标是建立 $\sup_{u\in[-r,r]}|\Delta_T(u)|$ 足够快的收敛性。我们首先建立逐点矩界。

引理 A.7。假设 [A1] 且 [A2]。进一步假设 K 有界且支撑在 $[-1,1]$ 上。若偶整数 $p > 2$ 满足 $Th^{p\varepsilon/4} \leq 1$，则

$$h^{1-\tilde{\alpha}} \sup_{u\in[-r,r]} \|\Delta_T(u)\|_p \lesssim \frac{C_p}{\sqrt{Th^{\tilde{\alpha}+\varepsilon/2}}},$$

其中 C_p 是一个仅取决于 p的常数。

证明。固定 $u \in [-r, r]$。由于 K_h的支撑集包含于 $[-h, h] \subset [-1, 1]$，我们可以将 $X_j(u)$ 重写为

$$X_j(u) = \frac{1}{T\lambda_1\lambda_2} \int_{I_j \times (I_j \oplus r_1)} K_h(y - x - u) N_1(dx) N_2(dy). \tag{A.22}$$

因此，对于每个 j， $X_j(u)$ 都是 $\sigma(N \cap (I_j \oplus r_1))$-可测的。此外，由于 $\|K_h\|_\infty \le h^{-1}\|K\|_\infty$，我们对任意 $q > 1$都有

$$\max_{0 \le j \le T-1} \|X_j(u)\|_q = \|X_0(u)\|_q \lesssim \frac{1}{Th\lambda_1\lambda_2} \|N_1(I_0) N_2(I_0 \oplus r_1)\|_q \lesssim \frac{1}{Th}. \tag{A.23}$$

因此，将引理 A.1 应用于 $(X_j(u))_{j=0}^{T-1}$，取 $M = h^{-\varepsilon/4}/(Th)$ 和 $\tau = \lfloor h^{-\varepsilon/4} \rfloor$，得到

$$\begin{aligned} \|\Delta_T(u)\|_p \le C_p \Bigg\{ & \left(\frac{h^{-\varepsilon/2}}{h} \mathrm{E}[|X_0(u)|] + \frac{h^{-\varepsilon/2}}{Th^2} \sum_{m=\lfloor h^{-\varepsilon/4} \rfloor}^{\infty} \alpha_{p,p}^N(m; r_1) \right)^{1/2} \\ & + \frac{T^{1/p} h^{-\varepsilon/4}}{Th} \left(\sum_{m=0}^{\infty} (m+1)^{p-2} \alpha_{p,p}^N(m; r_1) \right)^{1/p} \Bigg\} + 2T \left\| X_0(u) 1_{\{X_0(u) > h^{-\varepsilon/4}/(Th)\}} \right\|_p, \end{aligned}$$

其中 C_p 是仅依赖于 p的常数。根据 [A1](ii),

$$\sum_{m=\lfloor h^{-\varepsilon/4} \rfloor}^{\infty} \alpha_{p,p}^N(m; r_1) \lesssim h^{\tilde{\alpha}} \quad \text{and} \quad \sum_{m=0}^{\infty} (m+1)^{p-2} \alpha_{p,p}^N(m; r_1) \lesssim 1.$$

此外，由于 (A.3) 在 [A2]，下成立，坎贝尔公式和引理 A.2 给出

$$\mathrm{E}[X_0(u)] = \frac{1}{T} \int_{-1}^{1} K(t) g(u + ht) dt \lesssim \frac{1}{T} \int_{-1}^{1} \left(1 + |u - \theta^* + ht|^{\tilde{\alpha}-1}\right) dt \lesssim \frac{1}{Th^{1-\tilde{\alpha}}}. \tag{A.24}$$

此外，由 (A.23)，

$$\left\| X_0(u) 1_{\{X_0(u) > h^{-\varepsilon/4}/(Th)\}} \right\|_p \le \left(\frac{Th}{h^{-\varepsilon/4}} \right)^p (\mathrm{E}[X_0(u)^{p^2+p}])^{1/p} \lesssim \frac{h^{p\varepsilon/4}}{Th}.$$

由于 $Th^{p\varepsilon/4} \le 1$，我们得到

$$h^{1-\tilde{\alpha}} \|\Delta_T(u)\|_p \lesssim C_p \left(\frac{1}{\sqrt{Th^{\tilde{\alpha}+\varepsilon/2}}} + \frac{1}{Th^{\tilde{\alpha}+\varepsilon/2}} \right).$$

由于我们假设 $Th^{\tilde{\alpha}+\varepsilon/2} \ge Th^{\beta_\alpha+\varepsilon} \ge 1$，因此得到了所需的结果。 □

为了将引理 A.7 提升为对 $\sup_{u \in [-r,r]} |\Delta_T(u)|$ 的矩界，我们需要以下技术引理。

引理 A.8. 设 F 是定义在 $\mathbb{R}$上的有界单调不减函数。对于任意 h、 $\rho > 0$，存在有限点 $-r = u_0 \le u_1 \le \cdots \le u_N = r$ 和通用常数 C、 $d \ge 1$，使得 $N \le C\rho^{-d}$ 且

$$\sup_{u \in [-r,r]} \min_{0 \le a \le N-1} \int_{-r_1}^{r_1} \{F_h(v - u_a) - F_h(v - u_{a+1})\} g(v) dv \le 2\rho \frac{\|F\|_\infty}{h} \int_{-r_1}^{r_1} g(u) du. \tag{A.25}$$

证明。不失一般性，我们可以假设 $\int_{-r_1}^{r_1} g(u)du > 0$，否则所述结论显然成立。

根据 [33, 命题 3.6.12]的证明，$\mathcal{G} := \{F_h(\cdot - u) : u \in [-r, r]\}$ 是一个 VC 子图函数类。此外，$\mathcal{G}$ 存在包络函数 $\|F\|_\infty/h$。因此，根据 [33, 定理 3.3.9], ，存在有限多个点 $-r = s_0 < s_1 < \cdots < s_L = r$ 以及通用常数 C，$d \geq 1$，使得 $L \leq C\rho^{-d}$ 并且

$$\sup_{u\in[-r,r]} \min_{0\leq a\leq L} \int_{\mathbb{R}} |F_h(v-u) - F_h(v-s_a)|\, Q(dv) \leq \rho \frac{\|F\|_\infty}{h}, \tag{A.26}$$

其中，Q 是定义在$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$上的概率测度，定义为

$$Q(A) = \frac{\int_{A\cap[-r_1,r_1]} g(u)du}{\int_{-r_1}^{r_1} g(u)du}, \qquad A \in \mathcal{B}(\mathbb{R}).$$

接下来，定义函数 $\psi : \mathbb{R} \to \mathbb{R}$ 如下:

$$\psi(u) = \int_{-r_1}^{r_1} F_h(v-u)g(v)dv = \int_{\mathbb{R}} F(t)g(u+ht)1_{[-r_1,r_1]}(u+ht)dt, \qquad u \in \mathbb{R}.$$

于是，(A.26)给出

$$\sup_{u\in[-r,r]} \min_{0\leq a\leq L} |\psi(u) - \psi(s_a)| \leq \rho \frac{\|F\|_\infty}{h} \int_{-r_1}^{r_1} g(u)du.$$

此外，由于 g 非负且 F 单调不减，ψ 单调不增。而且，由于 $g1_{[-r_1,r_1]} \in L^1(\mathbb{R})$且 F 有界，ψ 连续（例如，参见 [53, 引理 1.8.1]）。因此，对于每个 $a = 0, \ldots\ldots, L-1$，都存在一点 $t_a \in [s_a, s_{a+1}]$，使得 $\psi(t_a) = \{\psi(s_a) + \psi(s_{a+1})\}/2$，这由介值定理可得。注意，$\psi(s_a) - \psi(t_a) = \psi(t_a) - \psi(s_{a+1}) = \{\psi(s_a) - \psi(s_{a+1})\}/2$。此外，由于$\psi$ 单调不增，

$$\min_{0\leq a'\leq L} |\psi(t_a) - \psi(s_{a'})| = \{\psi(s_a) - \psi(t_a)\} \wedge \{\psi(t_a) - \psi(s_{a+1})\} = \frac{\psi(s_a) - \psi(s_{a+1})}{2}.$$

因此，令 $u_{2a} = s_a$ 和 $u_{2a+1} = t_a$ 分别取 $a = 0$、……、$L-1$ 和$u_{2L} = s_L$，即可得到所需的点。

□

结合前两个引理，我们可以得到 $\Delta_T(u)$ 的如下矩界:

引理 A.9. 假设 [A1] 且 [A2]。还假设 K 有有界变差且支撑于 $[-1, 1]$。若对于某个 $\eta > 0$，有 $h \leq \min\{T^{-\eta}, 1/2\}$，则

$$\mathrm{E}\left[h^{1-\tilde{\alpha}} \sup_{u\in[-r,r]} |\Delta_T(u)|\right] \lesssim \frac{C_\eta}{\sqrt{Th^{\tilde{\alpha}+\varepsilon}}},$$

其中 C_η 仅取决于 η。

证明。由于 K 具有有界变差且支撑于 $[-1, 1]$，因此存在两个定义在 $\mathbb{R}$ 上的非递减函数F_1，F_2，使得 $K = (F_1 - F_2)1_{[-1,1]}$且 $|F_1| \vee |F_2| \leq \|K\|_\infty$。因此，不失一般性，我们可以假设 K 的形式为 $K = F1_{[-1,1]}$，其中 F是定义在 $\mathbb{R}$上的非递减函数。在证明的其余部分，我们分两步进行。

步骤 1。 对于 $\rho = h/T^{1/\tilde{\alpha}}$，令 $-r = u_0 \le u_1 \le \cdots \le u_N = r$ 和 $d \ge 1$ 如引理 A.8 所述。如有必要，将等距点 $-r + k\rho (k = 1, \ldots\ldots, \lfloor 2r/\rho \rfloor)$ 插入序列 $(u_a)_{a=0}^N$ 中，则可假设 $\max_{0\le a\le N-1}(u_{a+1} - u_a) \le \rho$，同时 (A.25) 仍成立。注意，这一操作至多使点的数量增加 $\lfloor 2r/\rho \rfloor$，因此我们有 $N \lesssim \rho^{-d}$。

对每个 $u \in [-r, r]$，令

$$\widetilde{X}_j(u) := \frac{1}{T\lambda_1\lambda_2} \int_{I_j \times \mathbb{R}} 1_{[-h,h]\oplus\rho}(y - x - u) F_h(y - x - u) N_1(dx) N_2(dy), \quad j = 0, 1, \ldots, T-1$$

and

$$\widetilde{\Delta}_T(u) := \sum_{j=0}^{T-1} \left\{ \widetilde{X}_j(u) - \mathrm{E}[\widetilde{X}_j(u)] \right\}.$$

在步骤 2 中，我们将证明

$$\mathrm{E}\left[h^{1-\tilde{\alpha}} \sup_{u\in[-r,r]} |\Delta_T(u)| \right] \lesssim \mathrm{E}\left[h^{1-\tilde{\alpha}} \max_{0\le a\le N} |\widetilde{\Delta}_T(u_a)| \right] + \frac{1}{\sqrt{Th^{\tilde{\alpha}}}}. \tag{A.27}$$

由此估计，我们可以如下证明引理中的结论。由于 $N \lesssim \rho^{-d} \le h^{-d\{1+1/(\tilde{\alpha}\eta)\}}$，对于一个仅依赖于 α 和 η 的充分大的偶整数 p，有 $N^{1/p} \lesssim h^{-\varepsilon/4}$。注意，詹森不等式给出

$$\mathrm{E}\left[h^{1-\tilde{\alpha}} \max_{0\le a\le N} \left|\widetilde{\Delta}_T(u_a)\right| \right] \le h^{1-\tilde{\alpha}} \left(\mathrm{E}\left[\max_{0\le a\le N} \left|\widetilde{\Delta}_T(u_a)\right|^p \right] \right)^{1/p} \le (N+1)^{1/p} h^{1-\tilde{\alpha}} \max_{0\le a\le N} \left\| \widetilde{\Delta}_T(u_a) \right\|_p.$$

同时，定义函数 $K\colon \mathbb{R} \to \mathbb{R}$，令 $K(u) = 2F(2u) 1_{[-1,1]\oplus T^{-1/\tilde{\alpha}}}(2u)$，其中 $u \in \mathbb{R}$。注意，K 的支撑包含于 $[-1, 1]$，且其有界于 $2\|K\|_\infty$。此外，我们可以将 $X_j(u)$ 改写为

$$\widetilde{X}_j(u) = \frac{1}{T\lambda_1\lambda_2} \int_{I_j \times \mathbb{R}} \tilde{K}_{2h}(y - x - u) N_1(dx) N_2(dy).$$

因此，我们可以对 $\Delta_T(u)$ 应用引理 A.7，从而得到

$$\mathrm{E}\left[h^{1-\tilde{\alpha}} \max_{0\le a\le N} \left|\widetilde{\Delta}_T(u_a)\right| \right] \lesssim \frac{(N+1)^{1/p}}{\sqrt{Th^{\tilde{\alpha}+\varepsilon/2}}} \lesssim \frac{1}{\sqrt{Th^{\tilde{\alpha}+\varepsilon}}}.$$

将其与 (A.27) 结合，即得引理的结论。

步骤 2. 剩下需要证明 (A.27)。固定 $u \in [-r, r]$。我们可以找到一个指标 $0 \le a < N$ 使得 $u_a \le u \le u_{a+1}$。由于 $u - u_a \le u_{a+1} - u_a \le \rho$ 且 F 是非递减的，因此有

$$X_j(u) = \frac{1}{T\lambda_1\lambda_2} \int_{I_j \times \mathbb{R}} 1_{[-h,h]}(y - x - u) F_h(y - x - u) N_1(dx) N_2(dy) \le \widetilde{X}_j(u_a)$$

对所有 j，坎贝尔公式给出

$$\mathrm{E}[\widetilde{X}_j(u_a)] = \frac{1}{T} \int_{\mathbb{R}} 1_{[-h,h]\oplus\rho}(v - u_a) K_h(v - u_a) g(v) dv$$

$$= \frac{1}{T}\int_{-r_1}^{r_1} 1_{[-h,h]\oplus\rho}(v-u_a)K_h(v-u_a)g(v)dv,$$

其中第二个等式由 $u_a \in [-r,r]$ 和 $h+\rho \le 2h \le 1$ 得出。注意，$1_{[-h,h]\oplus\rho}(v-u_a) = 1_{J_{h,u}}(v-u)$，其中 $J_{h,u} := ([-h,h]\oplus\rho) + (u_a - u)$。注意，$J_{h,u} \supset [-h,h]$，因为 $u - u_a \le \rho$。因此，

$$\left|\mathrm{E}[\widetilde{X}_j(u_a)] - \frac{1}{T}\int_{-r_1}^{r_1} 1_{[-h,h]}(v-u)K_h(v-u_a)g(v)dv\right| \le \frac{\|K\|_\infty}{Th}\int_{-r_1}^{r_1} 1_{J_{h,u}\setminus[-h,h]}(v-u)g(v)dv.$$

注意到，$\|g_0\|_{L^{1/(1-\tilde{\alpha}/2)}} \lesssim 1$ 根据詹森不等式（若 $\alpha < 1$）以及 $\int_{-r_1}^{r_1} |u-\theta^*|^{\frac{\alpha-1}{1-\tilde{\alpha}/2}} du \lesssim 1$ 因为 $\frac{\alpha-1}{1-\tilde{\alpha}/2} > -1$，我们由 [A2] 得到 $\|g\|_{L^{1/(1-\tilde{\alpha}/2)}} \lesssim 1$。接着，由于 $\mathrm{Leb}(J_{h,u}\setminus[-h,h]) \lesssim \rho$，杨氏不等式给出

$$\left|\mathrm{E}[\widetilde{X}_j(u_a)] - \frac{1}{T}\int_{-r_1}^{r_1} 1_{[-h,h]}(v-u)K_h(v-u_a)g(v)dv\right| \lesssim \frac{\rho^{\tilde{\alpha}/2}}{Th}.$$

与此同时，再次由坎贝尔公式，

$$\mathrm{E}[X_j(u)] = \frac{1}{T}\int_{\mathbb{R}} 1_{[-h,h]}(v-u)K_h(v-u)g(v)dv.$$

因此，(A.25) 给出

$$\left|\mathrm{E}[X_j(u)] - \frac{1}{T}\int_{-r_1}^{r_1} 1_{[-h,h]}(v-u)K_h(v-u_a)g(v)dv\right| \lesssim \frac{\rho}{Th}.$$

由此可得，

$$X_j(u) - \mathrm{E}[X_j(u)] \le \widetilde{X}_j(u_a) - \mathrm{E}[\widetilde{X}_j(u_a)] + C_0\frac{\rho^{\tilde{\alpha}/2}}{Th},$$

其中，$C_0 > 0$ 是一个仅依赖于 r，α，α_0，δ，b，$(B_p)_{p\ge1}$，$(B_{p,q})_{p,q\ge1}$，ε 和 $\|K\|_\infty$ 的常数。类似地，我们还有

$$X_j(u) - \mathrm{E}[X_j(u)] \ge \widetilde{X}_j(u_{a+1}) - \mathrm{E}[\widetilde{X}_j(u_{a+1})] - C_0\frac{\rho^{\tilde{\alpha}/2}}{Th}.$$

因此，我们得出结论

$$\sup_{u\in[-r,r]} |\Delta_T(u)| \le \max_{0\le a\le N} |\widetilde{\Delta}_T(u_a)| + C_0\frac{\rho^{\tilde{\alpha}/2}}{h}.$$

因此，(A.27) 由 ρ 的定义得出。 □

定理 *4.1* 的证明。由 (A.2) 定义函数 f_h，并令 $\Delta_T := \sup_{u\in[-r,r]} |\tilde{g}_h(u) - f_h(u)|$。由于对所有 j，$\mathrm{E}[X_j(u)] = f_h(u)$，由引理 A.6 可得

$$\mathrm{E}\left[\bar{\Delta}_T\right] \lesssim \mathrm{E}\left[\sup_{u\in[-r,r]} |\Delta_T(u)|\right] + \frac{1}{Th}. \tag{A.28}$$

现在，考虑 $\alpha > 1$ 的情形。由引理 A.5 可知，对于某个 $\sigma \in \{-1,1\}$，存在仅依赖于 α、b、δ 的常数 $\hat{A} > 1$ 和 $0 < h_0 < 1$，使得 (A.16) 成立。由于 θ_h 是 $[-r,r] \ni u \mapsto \tilde{g}_h(u) \in [0,\infty)$ 的最大化点，我们有

$$\mathrm{P}\left(|\hat{\theta}_h - \theta^*| > Ah\right) \le \mathrm{P}\left(\tilde{g}_h(\theta^* + \sigma h) \le \sup_{u\in[-r,r]:|u-\theta^*|>Ah} \tilde{g}_h(u)\right)$$

$$\leq \mathrm{P}\left(f_h(\theta^* + \sigma h) \leq \sup_{u \in [-r,r]: |u-\theta^*| > Ah} f_h(u) + 2\bar{\Delta}_T \right).$$

因此，由 (A.16) 可得 $\mathrm{P}\left(\hat{\theta}_h - \theta^*| > Ah\right) \leq \mathrm{P}\left(2\bar{\Delta}_T \geq h^{\alpha-1}\right)$。于是，由马尔可夫不等式、(A.28) 和引理 A.9 可得所需结果。

接下来，考虑 $\alpha < 1$ 的情形。根据引理 A.5，对于某个 $\sigma \in \{-1, 1\}$，存在常数 $A > 1$, $c > 0$ 和 $0 < h_0 < 1$，它们仅取决于 $\alpha, \alpha_0, b, \delta, K$，使得 (A.17) 成立。由于 θ_h 是 $[-r, r] \ni u \mapsto \tilde{g}_h(u) \in [0, \infty)$ 的最大化点，

$$\begin{aligned}
\mathrm{P}\left(|\hat{\theta}_h - \theta^*| > Ah\right) &\leq \mathrm{P}\left(\tilde{g}_h(\theta^*) \leq \sup_{u \in [-r,r]: |u-\theta^*| > Ah} \tilde{g}_h(u) \right) \\
&\leq \mathrm{P}\left(f_h(\theta^*) \leq \sup_{u \in [-r,r]: |u-\theta^*| > Ah} f_h(u) + 2\bar{\Delta}_T \right) \\
&\leq \mathrm{P}\left(2\bar{\Delta}_T \geq ch^{\alpha-1}\right),
\end{aligned}$$

其中最后一行由 (A.17) 得出。因此，再由马尔可夫不等式、(A.28) 和引理 A.9，即可得到所需结果。 □

A.5 定理 *4.2* 的证明

下文假设 T 足够大，使得 $j_{\min} \leq \log_a(T^{\gamma}_{\max})$。对于 $0 < \gamma \leq \gamma_{\max}$，我们将$h^*(\gamma)$记为满足 $h \leq T^{-\gamma}$ 的最大元素 $h \in \mathcal{H}_T$。注意，由于 $\gamma \leq \gamma_{\max}$， $h^*(\gamma)$ 是定义良好的。此外，根据构造，当 T 足够大时， $ah^*(\gamma) > T^{-\gamma}$，因此当 $T \to \infty$ 时， $h^*(\gamma) \asymp T^{-\gamma}$。

引理 A.10. 在定理 *4.2* 的假设下，对于任意 $0 < \gamma < 1/\beta_\alpha$，$\mathrm{P}(h > T^{-\gamma}) \to 0$ 当$T \to \infty$。

证明。简记 $h^* = h^*(\gamma)$。注意到

$$\begin{aligned}
\mathrm{P}(\hat{h} > T^{-\gamma}) &\leq \mathrm{P}(\hat{h} > h^*) \leq \mathrm{P}(\bar{d}(\mathcal{M}_{h^*}, \mathcal{M}_{h'}) > A_T h' \text{ for some } h' \in \mathcal{H}_T \text{ with } h' \geq h^*) \\
&\leq \sum_{h' \in \mathcal{H}_T : h' \geq h^*} \mathrm{P}(\bar{d}(\mathcal{M}_{h^*}, \mathcal{M}_{h'}) > A_T h').
\end{aligned}$$

由于 $d(\mathcal{M}_{h^*}, \mathcal{M}_{h'}) \leq d(\mathcal{M}_{h^*}, \{\theta^*\}) + d(\mathcal{M}_{h'}, \{\theta^*\})$，因此我们有

$$\begin{aligned}
\mathrm{P}(\hat{h} > T^{-\gamma}) &\leq \sum_{h' \in \mathcal{H}_T : h' \geq h^*} \left\{ \mathrm{P}(\bar{d}(\mathcal{M}_{h^*}, \{\theta^*\}) > A_T h'/2) + \mathrm{P}(\bar{d}(\mathcal{M}_{h'}, \{\theta^*\}) > A_T h'/2) \right\} \\
&\leq 2|\mathcal{H}_T| \max_{h \in \mathcal{H}_T : h \geq h^*} \mathrm{P}(\bar{d}(\mathcal{M}_h, \{\theta^*\}) > A_T h/2).
\end{aligned}$$

对于每个 $h \in \mathcal{H}_T$，我们都可以找到一个随机变量 $\theta_h \in \mathcal{M}_h$，使得 $|\theta_h - \theta^*| > A_T h/2$ 在事件 $d(\mathcal{M}_h, \{\theta^*\}) > A_T h/2$。因此，

$$\mathrm{P}(\hat{h} > T^{-\gamma}) \leq 2|\mathcal{H}_T| \max_{h \in \mathcal{H}_T : h \geq h^*} \mathrm{P}(|\tilde{\theta}_h - \theta^*| > A_T h/2).$$

因此，对于任意 $\varepsilon > 0$，定理 4.1 给出

$$\mathrm{P}(\hat{h}_T > T^{-\gamma}) = O\left(|\mathcal{H}_T| \max_{h \in \mathcal{H}_T : h \geq h^*} \frac{1}{\sqrt{T h^{\beta_\alpha + \varepsilon}}}\right) = O\left(\frac{|\mathcal{H}_T|}{\sqrt{T (h^*)^{\beta_\alpha + \varepsilon}}}\right).$$

由于 $\gamma < 1/\beta_\alpha$ 且 $|\mathcal{H}_T| = O(\log T)$，取 ε 使得 $\gamma(\beta_\alpha + \varepsilon) < 1$，即可得到所需结果。 □

定理 4.2 的证明。 令 γ' 为满足 $\gamma < \gamma' < 1/\beta_\alpha$ 的常数。于是，根据假设，$A_T T^{-\gamma'} = o(T^{-\gamma})$。因此，只需证明 $\mathrm{P}(|\hat{\theta}_{\hat{h}} - \theta^*| > 2A_T T^{-\gamma'}) \to 0$。此外，由引理 A.10，只需证明

$$\mathrm{P}(|\hat{\theta}_{\hat{h}} - \theta^*| > 2A_T T^{-\gamma'}, \hat{h} \leq T^{-\gamma'}) \to 0.$$

在事件 $\hat{h} \leq T^{-\gamma'}$ 上，我们有 $h^* \leq h^*(\gamma')$，因此

$$|\hat{\theta}_{\hat{h}} - \theta^*| \leq |\hat{\theta}_{\hat{h}} - \hat{\theta}_{h^*(\gamma')}| + |\hat{\theta}_{h^*(\gamma')} - \theta^*| \leq A_T h^*(\gamma') + |\hat{\theta}_{h^*(\gamma')} - \theta^*| \leq A_T T^{-\gamma'} + |\hat{\theta}_{h^*(\gamma')} - \theta^*|,$$

，其中第二个不等式由 h 的定义推出。因此，

$$\mathrm{P}(|\hat{\theta}_{\hat{h}} - \theta^*| > 2A_T T^{-\gamma'}, \hat{h} \leq T^{-\gamma'}) \leq \mathrm{P}(|\hat{\theta}_{h^*(\gamma')} - \theta^*| > A_T T^{-\gamma'}).$$

由于根据定理 4.1，$\mathrm{P}(|\hat{\theta}_{h^*(\gamma')} - \theta^*| > A_T T^{-\gamma'}) \to 0$，我们得到所需结果。 □

A.6 定理 *4.3* 的证明

定理 4.3 的证明依赖于 [69], 中的定理 2.2，该定理需要用到海灵格距离的概念。回顾一下，在共同的可测空间 $(\mathcal{X}, \mathcal{A})$ 上定义的两个概率测度 P 和 Q 之间的海灵格距离定义为

$$H(P, Q) := \sqrt{\int_{\mathcal{X}} \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}}\right)^2 d\nu},$$

其中 ν 是 (σ) 上任意一个 σ- 有限测度，该测度支配 P 和 Q 二者。根据 [66],中的引理 2.9 和 2.10(1)

$$H^2(P, Q) = 2\left(1 - \int_{\mathcal{X}} \sqrt{\frac{dQ}{dP}} dP\right), \tag{A.29}$$

其中 $dQ/dP := dQ^a/dP$，且 Q^a 是 Q 相对于 P 的绝对连续部分。注意，Strasser[66] 将海灵格距离定义为 $H(P, Q)/\sqrt{2}$（用我们的记号表示）。

定理 *4.3* 的证明。 定义 $\tilde{N}$ 上的点过程 $\mathbb{R}^2$ 为 $\tilde{N} := \sum_{i=1}^{\infty} \delta_{(t_i, t_i + \gamma_i)}$。由于 $N_1(\cdot) = \tilde{N}(\cdot \times \mathbb{R})$ 且 $N_2(\cdot) = \tilde{N}(\mathbb{R} \times \cdot)$，我们有 $\sigma(N \cap [0, T]) \subset \sigma(\tilde{N} \cap [0, T]^2)$。此外，令 $D_T := [0, T] \times [-1, T + 1]$，显然有 $\sigma(\tilde{N} \cap [0, T]^2) \subset \sigma(\tilde{N} \cap D_T)$。因此，只需证明存在一个常数 $b > 0$ 使得

$$\liminf_{T \to \infty} \inf_{\tilde{\theta}_T} \sup_{|\theta| \leq 2\rho_T} \sup_{g \in \mathcal{G}(\theta, \alpha, 1/2, b)} \mathrm{P}_g\left(|\tilde{\theta}_T - \theta| \geq \rho_T\right) > 0, \tag{A.30}$$

其中，下确界取遍所有基于 $N \cap D_T$ 的估计量。对于 g 上的每个概率密度 $\mathbb{R}$，记 $P_{T,\, g}$ 为在 P_g 下由 $\tilde{N} \cap D_T$ 在（$\mathcal{N}_{\#D_T}$，$\mathcal{B}$（$\mathcal{N}_{\#D_T}$））上诱导的分布，其中 $\mathcal{N}_{D_T}^{\#}$ 表示

赋予 D_T $w^{\#}$-拓扑的所有计数测度组成的空间；详细信息请参见 [23, 附录 A2.6] 和 [24,定义 9.1.II] 。根据式 (2.9) 和 [69], 中的定理 2.2，只要找到$g_T \in \mathcal{G}$（$2\rho_T$， α， $1/2$， b）和 $g_0 \in \mathcal{G}$（0， α， $1/2$， b），使得

$$\limsup_{T\to\infty} H^2(P_{T,g_T}, P_{T,g_0}) < 2. \tag{A.31}$$

现在来计算 H^2($P_{T,\ g_T}$，$P_{T,\ g_0}$)。注意， $\tilde{N}$可以视为定义在 $\mathbb{R}$ 上的簇过程，其中心过程为 N_1 ，分量过程为 $\{\delta_{(t_i,t_i+\gamma_i)}:\ i\in\mathbb{N}\}$。因此，根据 [23], 中的命题 6.3.III， N在 P$_g$ 下对于 $g\in\{g_0,g_T\}$ 的概率生成泛函（p.g.fl）为

$$\begin{aligned} G_g(\varphi) &= \exp\left(-\int_{\mathbb{R}}\left(1-\int_{\mathbb{R}}\varphi(x,y)g(y-x)dy\right)dx\right)\\ &= \exp\left(\int_{\mathbb{R}^2}(\varphi(x,y)-1)\,g(y-x)dxdy\right)\end{aligned}$$

对于每个可测函数 $\varphi:\ \mathbb{R}^2\to(0,1]$ ，使得 $1-\varphi$ 的支撑有界。由于 g 的支撑包含于 $[-1,1]$,

$$G_g(1-1_{D_T}+\varphi 1_{D_T}) = e^{-T}\exp\left(\int_{[0,T]\times\mathbb{R}}\varphi(x,y)g(y-x)dxdy\right).$$

˜因此，根据 [23], 中的式 (5.5.14)， N 在 P$_g$ 下于 D_T 上的局部 Janossy 密度为

$$j_{n,g}((x_1,y_1),\dots,(x_n,y_n)\mid D_T) = e^{-T}\prod_{i=1}^{n} g(y_i-x_i)1_{[0,T]}(x_i)\quad (n=1,2,\dots).$$

由此可得

$$\frac{dP_{T,g_T}}{dP_{T,g_0}}(\tilde{N}) = \prod_{i:t_i\in[0,T]}\frac{g_T(\gamma_i)}{g_0(\gamma_i)}\quad \mathrm{P}_{g_0}\text{-a.s.}$$

这里，回忆一下，$dP_{T,\ g_T}/dP_{T,\ g_0}:=dP^a_{T,\ g_T}/dP_{T,\ g_0}$，其中 $P^a_{T,\ g_T}$是 P_T的绝对连续部分，相对于 $P_{T,\ g_0}$。因此，由 (A.29) 得

$$H^2(P_{T,g_T},P_{T,g_0}) = 2\left(1-\mathrm{E}_{g_0}\left[\prod_{i:t_i\in[0,T]}\sqrt{\frac{g_T(\gamma_i)}{g_0(\gamma_i)}}\right]\right) =: 2(1-a_T),$$

其中 E$_{g_0}$ 表示在 P$_{g_0}$下的期望。因此，只要证明 $\liminf_{T\to\infty} a_T > 0$.即可得到 (A.31)。回顾一下，在 P$_{g_0}$下，$(\gamma_i)_{i=1}^{\infty}$ 是独立同分布的，其共同密度为 g_0 ，且独立于 N_1。因此，

$$a_T = \mathrm{E}_{g_0}\left[\left(\int\sqrt{g_T(x)g_0(x)}dx\right)^{N_1([0,T])}\right].$$

由于在 P$_{g_0}$下， $N_1([0,T])$服从强度为 T的泊松分布，

$$a_T = \exp\left(T\left(\int\sqrt{g_T(x)g_0(x)}dx-1\right)\right) = \exp\left(-\frac{T}{2}\int\left(\sqrt{g_T(x)}-\sqrt{g_0(x)}\right)^2dx\right).$$

The因此，只要证明存在一个常数 $b>0$ 使得 t

$$\int\left(\sqrt{g_T(x)}-\sqrt{g_0(x)}\right)^2dx = O(T^{-1}) \tag{A.32}$$

对于某个 $g_T\in\mathcal{G}$($2\rho_T$， α， $1/2$， b) 和 $g_0\in\mathcal{G}$(0， α， $1/2$， b)。

情形 1： $0 < \alpha < 1$。对于每个 $\theta \in [0,1]$，定义函数 f_θ： $\mathbb{R} \to [0,\infty)$ 为

$$f_\theta(x) = \alpha|x-\theta|^{\alpha-1} 1_{[-1,0)}(x-\theta) \quad (x \in \mathbb{R}).$$

根据构造，显然有 $f_\theta \in \mathcal{G}(\theta,\ \alpha,\ 1/2,\ b)$，其中常数 $b > 0$ 仅取决于 α。此外，由于 f_0 在 $z = 0$ 的邻域内满足 [46, 第六章] 中的公式 (1.9)，且按照其记号，$\alpha = \alpha - 1$、$p \equiv 1$和 $q \equiv 0$，因此 f_0 在 0 处有一个阶为 $\alpha - 1$ 的奇点，其含义见 [46,第六章]中的定义 1.1。因此，[46, 第六章] 中的定理 1.1给出了 $g_T = f_{2\rho_T}$和 $g_0 = f_0$的 (A.32)。

情形 2： $\alpha > 1$。我们采用证明 [3, 定理 2]时所用构造的一个小变体。对于每个 $\theta \in [0, 1/2)$，定义函数 ψ_θ： $\mathbb{R} \to \mathbb{R}$ 以及 f_θ： $\mathbb{R} \to \mathbb{R}$ 如下

$$\psi_\theta(x) = (|x|^{\alpha-1} - (2\theta)^{\alpha-1}) 1_{(-2\theta,0]}(x) + \left(|x|^{\alpha-1} + (2\theta)^{\alpha-1} - 2^\alpha|x-\theta|^{\alpha-1}\right) 1_{(0,2\theta)}(x), \quad x \in \mathbb{R}$$

and

$$f_\theta(x) = \frac{\alpha}{2(\alpha-1)} \left(1 - |x|^{\alpha-1} + \psi_\theta(x)\right) 1_{[-1,1]}(x), \quad x \in \mathbb{R}.$$

注意到 $f_\theta \geq 0$和

$$\int_{-\infty}^{\infty} f_\theta(x)dx = \frac{\alpha}{2\alpha-1} \left(\int_{-1}^{1} (1 - |x|^{\alpha-1})dx + \int_{-2\theta}^{2\theta} |x|^{\alpha-1}dx - 2^\alpha \int_0^{2\theta} |x-\theta|^{\alpha-1}dx \right) = 1.$$

因此，f_θ 是 $\mathbb{R}$上的概率密度函数。此外，对于所有 $x \in [-1,1]$，我们有

$$2^{2-\alpha}|x-\theta|^{\alpha-1} \leq \frac{2(\alpha-1)}{\alpha} \left(f_\theta(\theta) - f_\theta(x)\right) \leq 2^\alpha|x-\theta|^{\alpha-1}. \tag{A.33}$$

事实上，直接计算可得

$$\frac{2(\alpha-1)}{\alpha} \left(f_\theta(\theta) - f_\theta(x)\right) = \begin{cases} 2(2\theta)^{\alpha-1} & \text{if } -2\theta < x \leq 0, \\ 2^\alpha|x-\theta|^{\alpha-1} & \text{if } 0 < x < 2\theta, \\ |x|^{\alpha-1} + (2\theta)^{\alpha-1} & \text{otherwise.} \end{cases}$$

因此，当 $0 < x < 2\theta$时，(A.33) 显然成立。若 $-2\theta < x \leq 0$，则 $\theta \leq |x-\theta| \leq 3\theta$，所以 2(2$\theta$)$^{\alpha-1} \leq 2^\alpha|x-\theta|^{\alpha-1}$和 2(2$\theta$)$^{\alpha-1} \geq 2(2|x-\theta|/3)^{\alpha-1} \geq 2^{2-\alpha}|x-\theta|^{\alpha-1}$。因此，(A.33) 成立。此外，詹森不等式给出$|x-\theta|^{\alpha-1} \leq 2^{\alpha-2}$(|$x$|$^{\alpha-1} + \theta^{\alpha-1}$) $\leq 2^{\alpha-2}$(|x|$^{\alpha-1} + (2\theta)^{\alpha-1}$)。因此，当$|x| \geq 2\theta$时，(A.33) 的下界成立。此外，若 $x \leq -2\theta$，则 $|x| \geq 2\theta$ 和 $|x-\theta| = \theta - x \geq -x = |x|$，所以

$$|x|^{\alpha-1} + (2\theta)^{\alpha-1} \leq 2|x|^{\alpha-1} \leq 2|x-\theta|^{\alpha-1}.$$

如果 $x \geq 2\theta$，那么 $x - \theta \geq \theta$，因此

$$|x|^{\alpha-1} + (2\theta)^{\alpha-1} \leq 2^{\alpha-2}|x-\theta|^{\alpha-1} + 3 \cdot 2^{\alpha-2}\theta^{\alpha-1} \leq 2^\alpha|x-\theta|^{\alpha-1},$$

其中，第一个不等式由詹森不等式得出。因此，当$|x| \geq 2\theta$时，(A.33)的上界同样成立。于是，对于某个仅依赖于 α的常数 $b > 0$，有 $f_\theta \in \mathcal{G}(\theta,\ \alpha,\ 1/2,\ b)$。

现在，由 [69] 的式(2.27)可得

$$\int \left(\sqrt{f_\theta(x)} - \sqrt{f_0(x)}\right)^2 dx \le \int_{-1}^{1} \left(\frac{f_\theta(x)}{f_0(x)} - 1\right)^2 f_0(x)dx$$
$$= \frac{\alpha}{2(\alpha-1)} \int_{-2\theta}^{2\theta} \frac{\psi_\theta(x)^2}{1-|x|^{\alpha-1}} dx \le \frac{\alpha}{2(\alpha-1)} \frac{3(2\theta)^{2\alpha-1}}{1-(2\theta)^{\alpha-1}}.$$

因此， $g_T = f_{2\rho_T}$和 $g_0 = f_0$满足 (A.32)。 □

A.7 具有伽马核的 *LBHPG* 满足假设 [A2]*(ii)* 的证明

本节证明，具有共同速率参数 $\beta_{ij} \equiv \beta > 0, i, j = 1,2$ 的 LBHPG 在 $\min\{D_{12}, D_{21}\} < 1$ 且平稳性假设（矩阵 $\boldsymbol{\alpha}$ 的谱半径 $\rho(\boldsymbol{\alpha})$ 小于 1）成立时，满足 [A2](ii)，其中 $\alpha = \min\{D_{12}, D_{21}\}$。令

$$h_a(t) := \frac{\beta^a}{\Gamma(a)} t^{a-1} e^{-\beta t} \mathbf{1}_{(0,\infty)}(t), \qquad t \in \mathbb{R}, a > 0,$$

$\boldsymbol{H} = (h_{D_{ij}})_{1\le i,j\le 2}$，$\alpha_* = \min\{D_{12}, D_{21}\} (< 1)$，以及 $D_* = \min\{D_{11}, D_{12}, D_{21}, D_{22}\}$。根据定义，我们有$\Phi = \boldsymbol{\alpha} \odot \boldsymbol{H}$，其中 $\odot$为 Hadamard 积。回顾一下，$\Psi = \sum_{m\ge 1} \Phi^{(*m)}$逐分量收敛于 $L^1(\mathbb{R})$。

根据 (5.1) 和 (5.2)，只需证明以下命题:

命题 A.1. 在上述假设下，存在 $C > 0$ 和 $\delta > 0$ 使得

$$\alpha_{ij} h_{D_{ij}}(u) \le \Psi_{ij}(u), \qquad u > 0,\ (i,j) \in \{(1,2),(2,1)\}, \tag{A.34}$$
$$\max\{\Psi_{12}(u), \Psi_{21}(u)\} \le C u^{\alpha_*-1}, \qquad 0 < u < \delta, \tag{A.35}$$
$$0 \le \int_{\mathbb{R}} \Psi_{i1}(s)\Psi_{i2}(s+u)\, ds \le C|u|^{\alpha_*-1}, \qquad 0 < |u| < \delta, i \in \{1,2\}. \tag{A.36}$$

下面，我们先给出若干引理，再利用它们证明命题 A.1。在继续之前，我们将 Ψ分解为

$$\Psi = \Psi^s + \Psi^b, \tag{A.37}$$

其中

$$\Psi^s := \sum_{m=1}^{M-1} \Phi^{(*m)}, \qquad \Psi^b := \sum_{m=M}^{\infty} \Phi^{(*m)}, \qquad M = \lceil 1/D_* \rceil + 1.$$

引理 A.11. 对于 $m \ge 1$， i， $j \in \{1,2\}$ 和 $t > 0$,

$$(\Phi^{(*m)})_{ij}(t) = \sum_{\substack{(i_0,\dots,i_m)\in\{1,2\}^{m+1},\\ i_0=j,\ i_m=i}} \Big(\prod_{\ell=1}^{m} \alpha_{i_\ell i_{\ell-1}}\Big) h_{\sum_{\ell=1}^m D_{i_\ell i_{\ell-1}}}(t).$$

证明。该结果由伽马分布的可再生性得出。 □

引理 A.12. 对于每个 $a \ge 1$， $\|h_a\|_\infty \le \beta$。

证明。固定 $a \geq 1$。由于伽马密度 h_a 在 $[0,\infty)$ 上是对数凹的，我们有

$$\|h_a\|_\infty \leq \frac{1}{\sigma_a}$$

由 [61, 式 (5.8)], ，其中 σ_a 是 h_a的标准差。由于 h_a 的标准差为 $\sigma_a = \sqrt{a}/\beta$，我们得出

$$\|h_a\|_\infty \leq \frac{\beta}{\sqrt{a}} \leq \beta.$$

□

引理 A.13. Ψ^b 的每个分量在 $\mathbb{R}$上都有界。

证明。由引理 A.11，对于 $m \geq M$， $\Phi^{(*m)}$ 的每个分量都是形状参数至少为 $mD_* > 1$的伽马密度。因此，由引理 A.12 可得

$$\|(\Phi^{(*m)})_{ij}\|_\infty \leq \beta \sum_{\substack{(i_0,\ldots,i_m)\in\{1,2\}^{m+1},\\ i_0=j,\ i_m=i}} \prod_{\ell=1}^{m} \alpha_{i_\ell i_{\ell-1}} = \beta(\boldsymbol{\alpha}^m)_{ij}.$$

由于 $\rho(\boldsymbol{\alpha}) < 1$，级数 $\sum_{m\geq 1} \boldsymbol{\alpha}^m$ 逐项收敛，因此 $\sum_{m\geq M}(\boldsymbol{\alpha}^m)_{ij} < \infty$，从而

$$\|\Psi^b_{ij}\|_\infty \leq \sum_{m=M}^{\infty} \|(\Phi^{(*m)})_{ij}\|_\infty \leq \beta \sum_{m\geq M} (\boldsymbol{\alpha}^m)_{ij} < \infty.$$

□

命题 *A.1* 的证明。 在整个证明中， $C > 0$ 和 $\delta > 0$ 表示通用常数。
证明 (A.34)。 由于 $\Psi = \sum_{m\geq 1} \Phi^{(*m)}$ 且每一项都非负，我们有 $(0,\ \infty)$ 上的 $\Psi_{ij} \geq \Phi_{ij} = \alpha_{ij} h_{D_{ij}}$，由此得到 (A.34)。

证明(A.35)。**固定**$(i,\ j) \in \{(1,\ 2),\ (2,\ 1)\}$。由 (A.37),

$$\Psi_{ij}(u) = \Psi^s_{ij}(u) + \Psi^b_{ij}(u), \qquad u > 0.$$

由引理 A.11 及 Ψ^s 作为有限和的定义， Ψ^s_{ij} 是伽马密度的有限非负线性组合 h_a （具有相同的率参数 β），其形状参数满足 $a \geq D_{ij} \geq \alpha_{*\circ}$ 因此，存在 $C_s > 0$，使得对于所有 $u \in (0,\ 1)$， $\Psi^s_{ij}(u) \leq C_s u^{\alpha_*-1}$。此外，引理 A.13 给出 $\|\Psi^b_{ij}\|_\infty < \infty$。选取 $\delta \in (0,\ 1)$，使得 $u^{\alpha_*-1} \geq 1$ 在 $(0,\ \delta)$ 上成立。于是，对于 $u \in (0,\ \delta)$,

$$\Psi_{ij}(u) \leq C_s u^{\alpha_*-1} + \|\Psi^b_{ij}\|_\infty \leq (C_s + \|\Psi^b_{ij}\|_\infty)\, u^{\alpha_*-1}.$$

对 $(i,\ j) = (1,\ 2),\ (2,\ 1)$ 取最大值，得到 (A.35)。

证明 (A.36)。 固定 $i \in \{1,2\}$ 并设

$$I_i(u) := \int_{\mathbb{R}} \Psi_{i1}(s)\Psi_{i2}(s+u)\, ds, \qquad u \in \mathbb{R}.$$

非负性意味着 $I_i(u) \ge 0$。我们将证明上界。

利用 (A.37) 和 $\int_{\mathbb{R}} f(s)g(s+u)\,ds \le \|f\|_1\|g\|_\infty$（对非负的 f，g），可得

$$\begin{aligned} I_i(u) &= \int_{\mathbb{R}} (\Psi^s_{i1} + \Psi^b_{i1})(s)\,(\Psi^s_{i2} + \Psi^b_{i2})(s+u)\,ds \\ &\le \int_{\mathbb{R}} \Psi^s_{i1}(s)\Psi^s_{i2}(s+u)\,ds + \|\Psi_{i1}\|_1\|\Psi^b_{i2}\|_\infty + \|\Psi^b_{i1}\|_\infty\|\Psi_{i2}\|_1, \end{aligned}$$

其中用到了 $\Psi^s_{ik} \le \Psi_{ik}$ 和 $\Psi^b_{ik} \le \Psi_{ik}$。由于 $\|\Phi_{ij}\|_1 = \alpha_{ij}$ 且 $\|(\Phi^{(*m)})_{ij}\|_1 = (\boldsymbol{\alpha}^m)_{ij}$，假设 $\rho(\boldsymbol{\alpha}) < 1$ 推出 $\|\Psi_{ik}\|_1 = \sum_{m\ge1}(\boldsymbol{\alpha}^m)_{ik} < \infty$。由引理 A.13，$\|\Psi^b_{ik}\|_\infty < \infty$。因此，后两个项是与 u 无关的有限常数；又由于 $\alpha_* < 1$，我们可以缩小 $\delta \in (0,\,1)$，使这些常数可被 $C|u|^{\alpha_*-1}$ 在 $0 < |u| < \delta$ 上吸收（在那里使用 $|u|^{\alpha_*-1} \ge 1$）。因此，只需证明

$$\int_{\mathbb{R}} \Psi^s_{i1}(s)\Psi^s_{i2}(s+u)\,ds \le C|u|^{\alpha_*-1}, \qquad 0 < |u| < \delta. \tag{A.38}$$

对于 $a, b > 0$，定义

$$f^{\mathrm{BG}}_{a,b}(u) := \int_{\mathbb{R}} h_a(s)\,h_b(s+u)\,ds, \qquad u \in \mathbb{R}.$$

那么，$f^{\mathrm{BG}}_{a\,b}$ 是参数为 $(\alpha_+, \lambda_+, \alpha_-, \lambda_-) = (b, \beta, a, \beta)$ 的双边伽马分布 [51] 的概率密度函数。根据 Küchler 与 Tappe [51, 定理 6.1]，当 $u \to 0$ 时，密度满足 $f^{\mathrm{BG}}_{a\,b}$，若 $a + b < 1$，则 $(u) = O(|u|^{a+b-1})$；若 $a + b = 1$，则 $f\mathrm{BG}_{a,b}(u) = O(M(|u|))$，其中 M 在 0 处缓慢变化。若 $a+b > 1$，则 $f^{\mathrm{BG}}_{a\,b}$ 在 0 的某个邻域内有界。因此，对于任意满足 $a+b > \alpha_*$ 的 $a, b > 0$，存在 $C_{a,b} > 0$ 和 $\delta_{a,b} \in (0, 1)$，使得

$$f^{\mathrm{BG}}_{a,b}(u) \le C_{a,b}|u|^{\alpha_*-1}, \qquad 0 < |u| < \delta_{a,b} \tag{A.39}$$

因为根据假设有 $\alpha^* < 1$。接下来，由引理 A.11，每个 Ψ^s_{ik} 都是伽马密度 h_a 的有限非负线性组合。因此，(A.38) 的左侧是 $f^{\mathrm{BG}}_{a\,b,}(u)$ 的有限非负线性组合。若 $i = 1$，则 Ψ^s_{12} 仅涉及形状参数 $b \ge D_{12} \ge \alpha_*$，而 Ψ^s_{11} 中的形状参数 a 严格为正；因此 $a + b > \alpha_*$。若 $i = 2$，则 Ψ^s_{21} 仅涉及形状参数 $a \ge D_{21} \ge \alpha_*$，而 Ψ^s_{22} 中的形状参数 b 严格为正；因此仍有 $a + b > \alpha_*$。所以，(A.39) 适用于线性组合中出现的所有配对 (a, b)；我们可以取共同的 $\delta > 0$ 和 $C > 0$，从而得到 (A.38)。这就证明了 (A.36)。 □

B 实现与计算复杂度

本节讨论核密度估计量 $\hat{}g_h(u)$ 的高效计算，以及搜索其最大化点 θ_h 的策略。在我们的实现中，根据带宽为 h 的 $[0, T]$ 上的观测值计算 θ_h 的预期时间复杂度按以下方式缩放：

$$\begin{cases} O\big(T\log T + T^2 h\big), & \text{under Assumption [A2](i),} \\ O\big(T\log T + T^2 h^{\alpha}\big), & \text{under Assumption [A2](ii) with } 0 < \alpha < 1. \end{cases}$$

这比朴素的 $O(T^2)$ 方法更好；该方法会在每个候选 u 处通过对所有时间戳对求和来计算 $\hat{}g_h(u)$，尤其是在带宽 h 较小时。

B.1 在网格上计算 $\hat{}g_h$ 的算法

直接在网格 $\{u_1, \ldots, u_M\} \subset \mathbb{R}$ 上计算 $\hat{}g_h$ 的计算成本可能很高，其规模大致随网格大小与数据对数量的乘积增长。为降低计算成本，我们采用一种算法，遍历相关的时间戳对，并将其核权重分配到附近的网格点上。当带宽 h 相对于网格范围较小时，这种方法尤其高效。

设 N_1 和 N_2 为底层点过程，在窗口 $[0,T]$上观测得到，并令 $\mathcal{T}_i := \{t_{i,1} < \cdots < t_{i,\ n_i}\} \subset [0,T]$，$n_i := N_i([0,T])$ $i = 1,\ 2$ 表示相应的观测事件时间。令 $\mathcal{U} = \{u_1 < \cdots < u_M\}$ 为我们希望计算估计量的排序网格，并定义 $u_{\text{最小}} := u_1$ 和 $u_{\text{最大}} := u_M$。对于任意 $a < b$，定义滞后窗口 $[a,b]$ 中的相关配对集合及相应的差值集合如下：

$$\mathcal{P}(a,b) := \{(x,y) \in \mathcal{T}_1 \times \mathcal{T}_2 :\ y - x \in [a,b]\}, \qquad N_{\text{pairs}}(a,b) := |\mathcal{P}(a,b)|,$$

$$\mathcal{D}(a,b) := \{y - x :\ (x,y) \in \mathcal{P}(a,b)\} \subset [a,b].$$

算法 1 概述了该过程。我们不固定 u并对所有配对求和，而是遍历每个观测时间 $x \in \mathcal{T}_1$。利用 $\mathcal{T}_2$中的二分查找，我们确定 $y \in \mathcal{T}_2$ 的范围，使差值 $d = y - x$ 落在滞后窗口 $[u_{\min} - h,\ u_{\max} + h]$内；也就是说，这些滞后值可通过带宽为 h 的核影响至少一个网格点。对于每个这样的差值 d，我们在 $\mathcal{U} \cap [d-h, d+h]$ 中找出网格点子集（同样通过二分查找），并累加这些网格点处的核贡献。本节始终假设 K 的支撑位于 $[-1,1]$上。

计算复杂度。只有滞后窗口 $[u_{\min} - h,\ u_{\max} + h]$ 中的配对会对在 $\mathcal{U}$上评估的 $\hat{}g_h$ 产生贡献，此类配对的数量为 $N_{\text{pairs}}(u_{\min} - h,\ u_{\max} + h)$。对于每个 $x \in \mathcal{T}_1$，我们通过在 $\mathcal{T}_2$中进行二分查找来确定 $\mathcal{T}_2 \cap [x + u_{\min} - h,\ x + u_{\max} + h]$ 的索引范围，每次操作的开销为 $O(\log n_2)$（因此总开销为 $O(n_1 \log\ n_2)$）。对于每个相关配对 $(x,y) \in \mathcal{P}(u_{\min} - h,\ u_{\max} + h)$，其中 $d = y - x$，我们在 $\mathcal{U}$上执行二分查找以定位 $\mathcal{U} \cap [d-h, d+h]$，开销为 $O(\log\ M)$，然后更新该子数组中的所有网格点。设 $M_h := \sup_{u \in \mathbb{R}} |\mathcal{U} \cap [u-h, u+h]|$ 表示尺度 h下的最大局部网格占用量。因此，总计算复杂度上界为

$$O\big(n_1 \log n_2 + N_{\text{pairs}}(u_{\min} - h,\ u_{\max} + h)\big(\log M + M_h\big)\big)\,.$$

这比朴素的 $O(M \cdot N_{\text{pairs}}(u_{\min} - h,\ \ u_{\max} + h))$ 方法快得多（当 $M_h \ll M$ 时），而这种情况通常出现在带宽 h较小时。正如定理 4.1 所示，实际应用中较小的带宽更为理想，尤其是当 g 呈现尖峰时，此时所提出的实现可带来显著的计算增益。

B.2 寻找 $\hat{}g_h$ 的最大化点

$\hat{}$ 为了计算估计量 θ_h，我们需要在范围 $[-r, r]$ 内找到 $\hat{}g_h(u)$ 的全局最大化点。目标函数 $\hat{}g_h$ 可能存在许多局部最优点，因此标准数值优化难以收敛到全局最大值。不过，当 K 为分段线性时，$\hat{}g_h(u)$ 也为分段线性。

算法 1 在网格上计算 $\hat{}g_h(u)$ 需要：已排序的事件时间 $\mathcal{T}_1, \mathcal{T}_2$，排序网格 $\mathcal{U} = \{u_1, \ldots, u_M\}$，带宽 h，支撑于 $[-1, 1]$ 的核 K。确保：值 $G = (G_1, \ldots, G_M)$ 与 $(\hat{}g_h(u_1) \ldots, \hat{}g_h(u_M))$ 相对应。初始化 $G \leftarrow (0, \ldots, 0) u_{\min} \leftarrow u_1$, $u_{\max} \leftarrow u_M$，对于 $x \in \mathcal{T}_1$，识别索引 $[j_{\text{start}}, j_{\text{end}}]$，对于 $\mathcal{T}_2 \cap [x + u_{\min} - h,\ x + u_{\max} + h]$，从 $j \leftarrow j_{\text{start}}$ 到 j_{end}，执行$y \leftarrow \mathcal{T}_2[j]$ $d \leftarrow y - x$识别索引 $[k_{\text{start}}, k_{\text{end}}]$，对于 $\mathcal{U} \cap [d - h,\ d + h]$，从 $k \leftarrow k_{\text{start}}$ 到 k_{end}，执行$G_k \leftarrow G_k + \frac{1}{h} K\left(\frac{d - u_k}{h}\right)$结束循环，结束循环，结束循环。将 G 缩放 $\frac{T}{n_1 n_2}$ 倍，即 $G \leftarrow \frac{T}{n_1 n_2} G$返回 G

因此，只需在由差值 $d = y - x$、$x \in \mathcal{T}_1$、$y \in \mathcal{T}_2$ 所确定的折点位置处计算 $\hat{}g_h$，即可找到 $[-r, r]$ 上的最大化点。

令 $\mathcal{D}_{r,h} := \mathcal{D}(-r - h,\ r + h)$ 表示窗口 $[-r - h, r + h]$ 内的滞后差。具体而言，对于三角核 $K^{\text{tri}}(x) = (1 - |x|)\mathbf{1}_{[-1,1]}(x)$，函数仅可能在每个滞后差 $d \in \mathcal{D}_{r,h}$ 对应的 $\hat{}u = d - h,\ u = d,\ u = d + h$ 处改变斜率。因此，要得到最大化点 θ_h，只需在候选集上计算 $\hat{}g_h$：

$$\mathcal{U}_{\text{kink}} := \Big(\mathcal{D}_{r,h} \cup \big(\mathcal{D}_{r,h} + h\big) \cup \big(\mathcal{D}_{r,h} - h\big) \cup \{-r, r\}\Big) \cap [-r, r].$$

在实现中，建议先删除 $\mathcal{U}_{\text{kink}}$ 中的重复项，再进行排序；否则网格中可能包含重复点，从而产生不必要的计算。

计算复杂度。枚举相关配对 $\mathcal{P}(-r - h, r + h)$ 时即可构造集合 $\mathcal{D}_{r,h}$，其时间复杂度为 $O(N_{\text{pairs}}(-r - h,\ r + h))$，另有常数因子开销。从 $\mathcal{D}_{r,h}$ 构造 $\mathcal{U}_{\text{kink}}$ 的时间复杂度为线性于 $|\mathcal{U}_{\text{kink}}|$；我们可以对 $\mathcal{U}_{\text{kink}}$ 排序一次，以应用算法 1。使用算法 1 在 $\mathcal{U}_{\text{kink}}$ 上计算 $\hat{}g_h$ 的形式与上文相同，只需将 M 替换为 $|\mathcal{U}_{\text{kink}}|$，并按照上文的定义计算 M_h，但改在网格 $\mathcal{U}_{\text{kink}}$ 上进行：

$$O\big(n_1 \log n_2 + N_{\text{pairs}}(-r - h,\ r + h)\big(\log |\mathcal{U}_{\text{kink}}| + M_h\big)\big)\,.$$

最后，通过单次遍历所有计算出的值即可得到最大值，代价为 $O(|\mathcal{U}_{\text{kink}}|)$。

对 T和 h的缩放。 对于具有强度 λ_1、 λ_2和 CPCF g 的简单平稳二元点过程 $N=(N_1,N_2)$，当 $i=1$时，2，我们有 $\mathrm{E}[n_i]=\lambda_i T$。此外，根据交叉强度函数的定义，我们有

$$\mathrm{E}[N_{\mathrm{pairs}}(a,b)]=\int_{(0,T]^2}\mathbf{1}\{y-x\in[a,b]\}\,\lambda_{12}(y-x)\,dx\,dy\approx\lambda_1\lambda_2\,T\int_a^b g(u)\,du.$$

若 $T\gg r+h$ 且 $[a,b]\subset[-r-h,r+h]$，则 $\mathrm{E}[N_{\mathrm{pairs}}(a,\ b)]\asymp\lambda_1\lambda_2\,T\int_a^b g(u)\,du$。特别是 ar,

$$\mathrm{E}[N_{\mathrm{pairs}}(-r-h,r+h)]=O(T),\qquad \mathrm{E}[n_1\log n_2]=O(T\log T),$$

其中隐藏常数取决于 g 在 $[-r-h,r+h]$ 上的质量。

接下来，注意到 $|\mathcal{U}_{\mathrm{kink}}|\le 3|\mathcal{D}_{r,\ h}|+2\le 3N_{\mathrm{pairs}}(-r-h,\ r+h)+2$，因此其期望为 $|\mathcal{U}_{\mathrm{kink}}|=O(T)$，而一次性排序成本为 $O(T\log T)$。对于折点网格上的局部更新成本，注意对任意 $u\in[-r,r]$，

$$|\mathcal{U}_{\mathrm{kink}}\cap[u-h,u+h]|\ \le\ 3\,|\mathcal{D}_{r,h}\cap[u-2h,u+2h]|+2\ \le\ 3\,N_{\mathrm{pairs}}(u-2h,u+2h)+2.$$

取期望可得

$$\mathrm{E}[|\mathcal{U}_{\mathrm{kink}}\cap[u-h,u+h]|]=O\left(T\int_{u-2h}^{u+2h}g(v)\,dv\right).$$

对 $u\in[-r,r]$ 取上确界，得到最坏情况界

$$M_h=O\left(T\sup_{u\in[-r,r]}\int_{u-2h}^{u+2h}g(v)\,dv\right).$$

特别是，如果 g 有界（例如，在 [A2](i) 的条件下），那么 $\int_{u-2h}^{u+2h}g(v)\,dv=O(h)$ 关于 u 一致，因此 $M_h=O(Th)$，预期时间上界可简化为 $O(T\log T+T^2h)$。如果像 [A2](ii) 中那样，g 在 θ^* 处无界，那么 $\int_{\theta^*-2h}^{\theta^*+2h}g(v)\,dv=O(h^\alpha)$。因此，我们可能有 $M_h=O(Th^\alpha)$，预期时间上界为 $O(T\log T+T^2h^\alpha)$。另一方面，朴素实现的代价为 $O(N_{\mathrm{pairs}}\times|\mathcal{U}_{\mathrm{kink}}|)\approx O(T^2)$。

C 通过交叉验证进行带宽选择

ˆ 本节将介绍基于交叉验证的估计量 θ_h 的其他带宽选择方法，并通过模拟研究评估其性能。

在众数回归的背景下，Chen等人 [19] 提出，通过最小化与众数回归函数相关的预测集大小来选择带宽；Zhou 与 Huang [72]则提出了一种交叉验证准则（CVM），该准则惩罚响应与估计众数集之间的平方距离，并对众数的数量施加显式惩罚。受这些进展的启发，我们设计了一种交叉验证方案，直接评估由点过程训练部分得到的最大化点集对测试部分中观测到的经验滞后差的预测效果。

固定整数 $K_{\mathrm{cv}}\ge 2$，并将观测窗口 $[0,T]$ 划分为 K_{cv} 个长度相等且互不相交的子区间 I_1、... ...、$I_{K_{\mathrm{cv}}}$。此处 K_{cv} 表示折数。对于第 j折，我们将 I_j 视为测试区间，并

将 $[0,T]\setminus I_j$ 视为相应的训练区间。在训练区间上，我们基于 $N\cap([0,T]\setminus I_j)$计算核估计量 $\hat{g}_h^{(-j)}$ 及其（可能为集合值的）最大化点集

$$M_h^{(-j)} := \operatorname*{arg\,max}_{u\in[-r,r]} \hat{g}_h^{(-j)}(u), \qquad h\in\mathcal{H}_T,$$

其中 $\mathcal{H}_T$ 是有限的带宽网格。

对于集合 $M\subset[-r,r]$ 和 $z\in\mathbb{R}$，记

$$d(z,M) := \inf_{u\in M} |z-u|$$

为从 z 到 M的距离。给定测试区间 I_j，我们将限制在 $[-r,r]$ 内观测到的滞后差集合定义为

$$\Delta_j(r) := \big\{y-x; x\in\mathcal{T}_{1,j},\ y\in\mathcal{T}_{2,j},\ -r\le y-x\le r\big\},$$

其中$\mathcal{T}_{i,j}$ 表示落在 I_j 内且满足 $i\in\{1,2\}$的 N_i 的（有限）事件时间集合。我们将 $\Delta_j(r)$视为多重集，即滞后差按重数计数，并用$n_j := |\Delta_j(r)|$ 表示其元素个数。实际应用中，折叠长度应相对于 r 选得足够大，以便每个 I_j 都包含足够多的滞后位于 $[-r,r]$内的配对。当 $n_j\ge 1$时，我们将 $\Delta_j(r)$的元素（按任意顺序）列为$(d_{j,1},\dots,d_{j,n_j})$，并引入距离

$$\delta_{j,\ell}(M) := d(d_{j,\ell},M), \qquad \ell=1,\dots,n_j,$$

其次序统计量为$\delta_{j,(1)}(M)\le\cdots\le\delta_{j,(n_j)}(M)$。固定截尾参数 $\tau\in(0,1]$和最小计数 $n_{\min}\in\mathbb{N}$，并设

$$k_j := \max\{\lceil\tau n_j\rceil,\ n_{\min}\}, \qquad \varepsilon_j(M) := \begin{cases}\delta_{j,(k_j)}(M), & n_j\ge k_j,\\ +\infty, & n_j<k_j.\end{cases}$$

若 $n_j<k_j$，则我们设定 $L_{\mathrm{nearest}}(M;\,I_j)=+\infty$，不论 M 如何，从而实际上舍弃了那些测试折中滞后差数量过少的带宽。采用这一记法，我们考虑定义在I_j上的以下损失函数，候选极大值点集合有限，为 $M\subset[-r,r]$（因此 $|M|$ 是其基数）：

- **均方误差型损失**:

  $$L_{\mathrm{mse}}(M;I_j) := \frac{|M|^2}{n_j}\sum_{\ell=1}^{n_j}\delta_{j,\ell}(M)^2,$$

  按照约定，若某折满足 $n_j=0$，则不纳入交叉验证平均值。因子 $|M|^2$ 会对产生较多最大化点的带宽施加惩罚，发挥稳定作用，这与 Zhou& Huang [72] 的思路一致。

- **最近范围损失**:

  $$L_{\mathrm{nearest}}(M;I_j) := \mathrm{Leb}\big(\{x\in\mathbb{R}: d(x,M)\le\varepsilon_j(M)\}\big).$$

  集合 $\{x\in\mathbb{R}:\ d(x,\ M)\ \le\varepsilon\}$是 M 的 ε 邻域，而 Leb(·) 是其总长度。根据构造，$\varepsilon_j(M)$ 的取值使得至少有 $k_j=\max\{\lceil\tau n_j\rceil,n_{\min}\}$ 个测试滞后差位于

M 的 $\varepsilon_j(M)$ 距离范围内，因此 $L_{\mathrm{nearest}}(M;\, I_j)$ 是覆盖该截尾比例的最小邻域长度。最小计数 $n_{\min}$ 提高了数值稳定性。此损失函数基于 Chen 等人 [19]提出的预测集方法。

令 $J := \{j \in \{1,\dots,K_{\mathrm{cv}}\} :\ n_j \geq 1\}$ 表示在 $[-r, r]$中至少包含一个滞后差的折的集合。给定损失函数 $L \in \{L_{\mathrm{mse}}, L_{\mathrm{nearest}}\}$的选择，我们将 $h \in \mathcal{H}_T$的 K_{cv}折交叉验证评分定义为

$$\mathrm{CV}(h) := \frac{1}{|J|} \sum_{j \in J} L\big(M_h^{(-j)}; I_j\big).$$

随后，我们选择交叉验证带宽为

$$\hat{h}_{\mathrm{CV}} \in \arg\min_{h \in \mathcal{H}_T} \mathrm{CV}(h)$$

ˆ 如果最小值点不唯一，则选择 $\mathcal{H}_T$中最小的 h。最后得到自适应估计量 $\theta_{\hat{h}_{\mathrm{CV}}}$。

为比较不同损失函数及其相应调节参数的表现，我们在以下共同设计下开展一系列模拟实验。候选带宽为 $h \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$, $r = 1$，观测窗口为 $[0, T]$。观测时域取值为 $T \in \{1000, 2000, 4000, 8000\}$；对于模型、估计量和$T$的每种组合，我们生成 5000 次蒙特卡洛重复。所考察的交叉验证准则包括最近范围损失 L_{nearest}（修剪水平为 $\tau \in \{0.01, 0.025, 0.05\}$）和均方误差型损失 L_{mse}。作为对比，我们还纳入了 Lepski 选择器，其中 $A_T = \log\log T$。在图中，不同τ对应的最近范围交叉验证曲线以红色渐变区分；基于均方误差的交叉验证曲线以蓝色显示；Lepski 估计量的曲线以绿色显示。我们设定 $K_{\mathrm{cv}} = 5$ 和 $n_{\min} = 5$。作为参考，每个面板还以黑色虚线显示理论收敛速率 T^{-1/β_α}。

图 6 比较了第 5.1 节所述六种数据生成模型下不同带宽选择方案的有限样本表现。由此可以看出一些系统性规律。

首先，基于均方误差型损失的交叉验证准则有些不稳定。在六种模型中的五种模型下，随着T增加，均方误差型交叉验证估计量的均方根误差仅缓慢下降，有时甚至完全没有下降，这与理论斜率所预示的情况形成对比。这表明，对于我们的情形，均方误差型损失并不适合用于选择带宽。

最近范围损失 L_{nearest} 的表现更为理想，但其性能取决于截尾参数 τ。对于 CPCF 更尖锐的模型，即 α 小于 1 的模型，较小的 τ 值往往效果更好。相比之下，对于具有 $\alpha > 1$的更平滑模型，较大的 τ 值更具竞争力，甚至更为理想。换言之，τ 的最优选择似乎取决于 α：当 g 具有非常尖锐的峰值时，激进截尾会带来益处；而当 g 在其最大值附近较为平坦时，较温和的截尾就已足够。一个值得注意的例外是非对称 Hawkes 模型（hawkes gamma asym）。在这种情况下，最近范围交叉验证估计量随着 T的增加而改善得更慢、也更不规则，且这一现象出现在所考察的各个 τ值中。一个可能的影响因素是，L_{nearest} 基于估计的最大化点集的对称邻域构造，而在这种非对称设定下，有限样本中的组内滞后差可能存在偏斜；在此类情况下，对称邻域对于选择 h的参考价值可能较低。

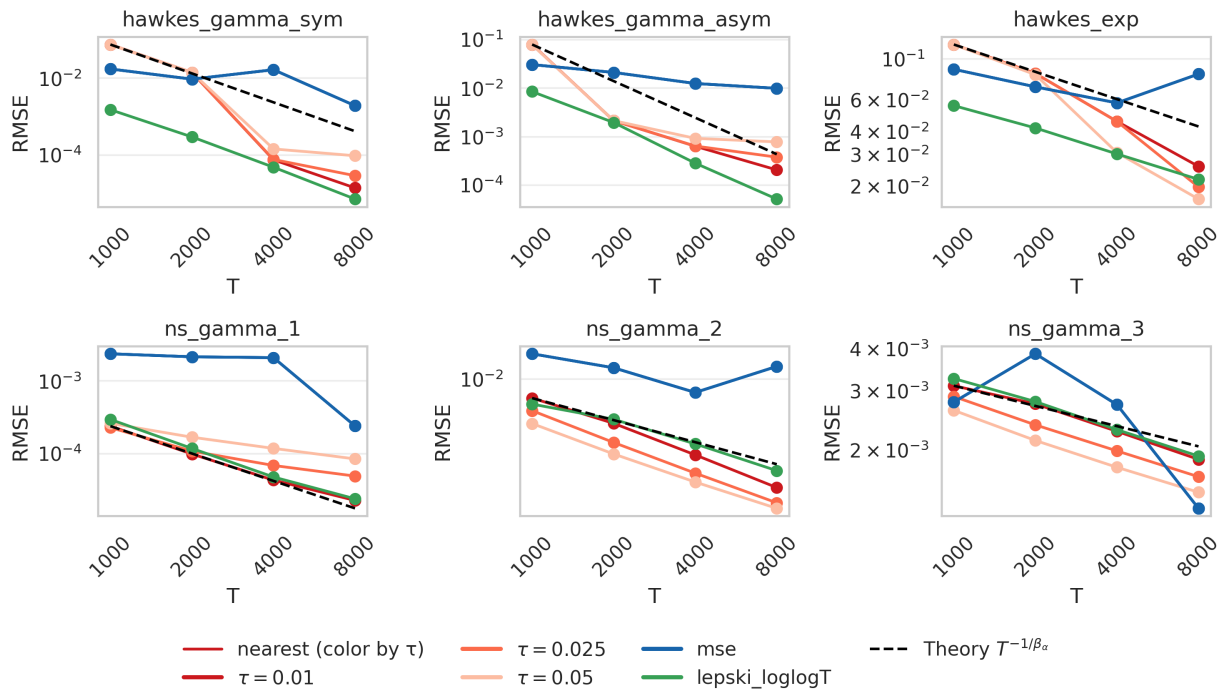

图 6：不同情景下带宽选择方法的性能比较：交叉验证使用 L_{nearest} （按 τ 呈红色渐变）、交叉验证使用 L_{mse} （蓝色）以及采用 $A_T = \log\log T$ 的 Lepski 方法（绿色）时，RMSE 随 T 的变化情况，绘于双对数坐标轴上；黑色虚线表示理论收敛速率 T^{-1/β_α}。

与 Lepski 型程序相比，最近范围交叉验证估计量在小 α 模型中的表现较差。事实上，对于 hawkes gamma sym、hawkes gamma asym 和 ns gamma 1，Lepski 估计量的均方根误差更低。然而，对于具有 $\alpha > 1$的更平滑模型，经过最佳调优的最近范围交叉验证估计量变得具有竞争力。总体而言，这些实验表明，基于 L_{nearest} 的交叉验证是一种很有前景的替代方法，但可能需要谨慎选择 τ。

综合来看，我们的研究结果给出了以下实用建议。在我们考察的带宽选择器中，Lepski 方法表现最为稳健。它只需要一个缓慢发散的阈值 A_T，在所有模型中都表现稳定，并且无论在理论上还是数值实验中，其表现都接近极小极大速率T^{-1/β_α}。最近邻范围交叉验证具有竞争力，尤其适用于更平滑的 CPCF；但除了带宽之外，它还需要选择修剪水平 τ，而目前从理论上看，最佳选择尚未得到充分理解。因此，我们目前建议将 Lepski 型带宽选择作为估计领先—滞后时间的默认方法。

致谢 我们感谢以下会议和研讨会的与会者提出的深刻见解和建议：“计量经济学、金融与统计中的大数据和人工智能”研讨会（芝加哥大学，2025年10月2日至4日）、新加坡国立大学数量金融研讨会（2025年10月24日）、筑波大学科研费研讨会（2025年10月30日至31日）、大阪大学中之岛研讨会（2025年12月5日至6日），以及伦敦大学 CFE-CMStatistics 2025（2025年12月13日至15日）。

并感谢他们对本研究提出的建设性建议。Takaaki Shiotani 的研究部分得到 JSPS研究员资助金（25KJ0933）和数学科学与物理学前沿世界领先创新研究生培养计划的支撑。Yuta Koike 的研究部分得到 JST CREST（项目编号：JPMJCR2115）和 JSPS科研费（项目编号：JP22H00834、JP22H01139）的支撑。

参考文献

[1] Abraham, C., Biau, G. & Cadre, B.（2003）。多元密度众数的简单估计。《加拿大统计学杂志》**31**，23–34。[2] Alsayed, H. & McGroarty, F.（2014）。国际指数期货市场的超高频算法套利。《预测学杂志》**33**，391–408。[3] Arias-Castro, E., Qiao, W. & Zheng, L.（2022）。密度全局众数的估计：极小极大性、自适应性与计算复杂度。《电子统计学杂志》**16**，2774–2795。[4] Bacry, E., Dayri, K. & Muzy, J.-F.（2012）。对称霍克斯过程的非参数核估计：在高频金融数据中的应用。《欧洲物理学杂志*B*辑》**85**，157。[5] Bacry, E., Delattre, S., Hoffmann, M. & Muzy, J.-F.（2013）。关于霍克斯过程的一些极限定理及其在金融统计中的应用。《随机过程及其应用》**123**，2475–2499。[6] Bacry, E., Mastromatteo, I. & Muzy, J.-F.（2015）。金融领域的霍克斯过程。《市场微观结构与流动性》**1**，1550005。[7] Bangsgaard, C. & Kokholm, T.（2024）。VIX期货与SPX期货之间的领先—滞后关系。《金融市场杂志》**67**，100851。[8] Bartlett III, R. P. & McCrary, J.（2019）。股票市场受到操纵的程度有多大？来自微秒时间戳的证据。《金融市场杂志》**45**，37–60。[9] Bercu, B., Gamboa, F. & Lavielle, M.（2002）。边际众数与谱众数的估计。《非参数统计学杂志》**14**，353–366。[10] Bollen, N. P., O'Neill, M. J. & Whaley, R. E.（2017）。尾部牵动整体：VIX市场的日内价格发现。《期货市场杂志》**37**，431–451。[11] Boly, O., Cheysson, F. & Nguyen, T. H.（2023）。多元霍克斯过程的混合性质。*arXiv*预印本 *arXiv:2311.11730* 。[12] Brillinger, D. R.（1972）。平稳区间函数的谱分析。载于第六届伯克利数学统计与概率研讨会论文集，第一卷：统计学理论，第6卷。加利福尼亚大学出版社，第483–514页。[13] Brillinger, D. R.（1975）。平稳点过程的统计推断。载于《随机过程及相关主题》。学术出版社，第55–99页。

[14] Brillinger, D. R.（1976）。二元平稳点过程二阶强度的估计。《英国皇家统计学会会刊B辑：统计方法学》38，60–66。[15] Bryant Jr, H., Marcos, A. R. & Segundo, J.（1973）。单突触连接和共同输入所产生的神经元尖峰放电之间的相关性。《神经生理学杂志》36，205–225。[16] Buccheri, G., Corsi, F. & Peluso, S.（2021）。高频领先—滞后效应与跨资产关联：一种多资产滞后调整模型。《商业与经济统计学杂志》39，605–621。[17] Chac´on, J. E.（2020）。统计学中的众数年龄。《国际统计评论》88，122–141。[18] Chen, Y.-C.（2018）。基于核密度估计的众数回归：综述。《Wiley跨学科评论：计算统计学》10，e1431。[19] Chen, Y.-C., Genovese, C. R., Tibshirani, R. J. & Wasserman, L.（2016）。非参数众数回归。《统计年刊》44，489–514。[20] Cox, D. R. & Lewis, P. A. W.（1972）。多元点过程。载于《第六届伯克利数学统计与概率研讨会论文集》第三卷。加利福尼亚大学出版社，伯克利，第401–448页。[21] Cucala, L.（2008）。对含噪声观测的空间点过程进行强度估计。《斯堪的纳维亚统计学杂志》35，322–334。[22] Da Fonseca, J. & Zaatour, R.（2017）。霍克斯微观结构模型中的相关性与领先—滞后关系。《期货市场杂志》37，260–285。[23] Daley, D. J. & Vere-Jones, D.（2003）。点过程理论导论，第一卷：基础理论与方法。施普林格，第2版。[24] Daley, D. J. & Vere-Jones, D.（2008）。点过程理论导论，第二卷：一般理论与结构。施普林格，第2版。[25] Dao, T. M., McGroarty, F. & Urquhart, A.（2018）。超高频领先—滞后关系与信息到达。《定量金融》18，725–735。[26] De Jong, F. & Nijman, T.（1997）。金融市场间领先—滞后关系的高频分析。《实证金融学杂志》4，259–277。[27] Dobrev, D. & Schaumburg, E.（2015）。高频跨市场交易：无模型测量及应用。工作论文。[28] Dobrev, D. & Schaumburg, E.（2023）。高频跨市场交易：无模型测量与可检验推论。工作论文。[29] Doukhan, P.（1994）。混合性：性质与实例。纽约：施普林格。

[30] Doukhan, P. & Louhichi, S.（1999）。一种新的弱依赖条件及其在矩不等式中的应用。《随机过程及其应用》84，313–342。[31] Eddy, W. F.（1980）。众数的最优核估计量。《统计年刊》8，870–882。[32] Ellis, S. P.（1991）。点过程的密度估计。《随机过程及其应用》39，345–358。[33] Gin´e, E. & Nickl, R.（2016）。无限维统计模型的数学基础，《剑桥统计与概率数学丛书》第40卷。剑桥大学出版社。[34] Guan, Y.（2007）。成对相关函数估计中的最小二乘交叉验证带宽选择方法。《统计与概率快报》77，1722–1729。[35] Hasbrouck, J.（2021）。高分辨率下的价格发现。《金融计量学杂志》19，395–430。[36] Has'minskii, R.（1979）。众数非参数估计的风险下界。载于《统计学文集》。D. Reidel 出版公司，第91–97页。[37] Hayashi, T.（2017）。高频限价订单簿数据的统计分析：日本股票市场跨市场单资产领先—滞后关系（日本文）。《统计数理研究所论文集》65，113–139。[38] Hayashi, T. & Koike, Y.（2018）。基于小波的方法进行高频领先—滞后分析。SIAM 金融数学杂志9，1208–1248。[39] Hayashi, T. & Koike, Y.（2020）。高频金融市场领先—滞后关系的多尺度分析。arXiv预印本 arXiv:1708.03992v4。[40] Herrmann, E. & Ziegler, K.（2004）。在缺乏光滑性假设时众数非参数估计的相合速度。《统计与概率快报》68，359–368。[41] Hessellund, K. B., Xu, G., Guan, Y. & Waagepetersen, R.（2022）。多元对数高斯 Cox 过程的二阶半参数推断。英国皇家统计学会杂志 C辑：应用统计71，244–268。[42] Hessellund, K. B., Xu, G., Guan, Y. & Waagepetersen, R.（2022）。多元点模式数据的半参数多项逻辑回归。《美国统计学会杂志》117，1500–1515。[43] Hoffmann, M., Rosenbaum, M. & Yoshida, N.（2013）。从非同步数据估计领先—滞后参数。Bernoulli 19，426–461。[44] Holden, C. W., Pierson, M. & Wu, J.（2023）。转瞬之间：交易所至 SIP 的延迟与交易分类准确率。可见于 SSRN 4441422。

[45] Huth, N. & Abergel, F.（2014）。高频领先—滞后关系——经验事实。《实证金融学杂志》26，41–58。[46] Ibragimov, I. & Has’minskii, R.（1981）。统计估计：渐近理论。施普林格。[47] Jalilian, A. & Waagepetersen, R.（2018）。成对相关函数估计的快速带宽选择。《统计计算与模拟杂志》88，2001–2011。[48] Jovanovi´c, S., Hertz, J. & Rotter, S.（2015）。霍克斯点过程的累积量。《物理评论E》91，042802。[49] Klemel¨a, J.（2005）。多元密度的众数自适应估计。《非参数统计》17，83–105。[50] Koike, Y.（2021）。基于超高频数据的时变领先—滞后关系推断。《日本统计与数据科学杂志》4，643–696。[51] K¨uchler, U. & Tappe, S.（2008）。双边伽马密度的形状。《统计与概率快报》78，2478–2484。[52] Leblanc, T.（2024）。最小假设下霍克斯过程的指数矩。《概率电子通讯》29，1–11。[53] Malliavin, P.（1995）。积分与概率，《数学研究生教材》第157卷。施普林格。[54] Meister, A.（2011）。关于反卷积众数估计的一般一致性。《统计规划与推断杂志》141，771–781。[55] Parzen, E.（1962）。概率密度函数与众数的估计。《数理统计年刊》33，1065–1076。[56] Pierson, M. & Wu, J.（2025）。延迟评述：为何动态RBBO优于固定延迟调整。可见于 SSRN 5185422。[57] Poinas, A., Delyon, B. & Lavancier, F.（2019）。相联点过程的混合性质与中心极限定理。Bernoulli 25，1724–1754。[58] Potiron, Y. & Volkov, V.（2025）。带延迟的相互激发点过程。《美国统计学会杂志》，1–22。[59] Poutr´e, C., Dionne, G. & Yergeau, G.（2024）。高频领先—滞后套利的盈利能力。《国际预测杂志》40，1002–1021。[60] Rambaldi, M., Filimonov, V. & Lillo, F.（2018）。利用霍克斯过程检测强度突发：高频金融数据中的应用。《物理评论E》97，032318。

[61] Saumard, A. & Wellner, J. A. (2014). 对数凹性与强对数凹性：综述。《统计调查》**8**, 45–114.[62] Schwenk-Nebbe, S. (2024). 大数据金融。博士论文，奥胡斯大学。[63] Shaw, T., Møller, J. & Waagepetersen, R. P. (2021). 适用于 K和成对相关函数的全局强度重加权估计量。《澳大利亚与新西兰统计学杂志》**63**, 93–118.[64] Sheather, S. J. & Jones, M. C. (1991). 一种用于核密度估计的可靠数据驱动带宽选择方法。《英国皇家统计学会杂志：*B*辑（方法论）》**53**, 683–690.[65] Shiotani, T. & Yoshida, N. (2024). 高度相关平稳点过程和含噪双变量 Neyman-Scott 过程的统计推断。*arXiv*预印本 *arXiv:2410.05732* 。[66] Strasser, H. (1985). 统计学数学理论。Walter de Gruyter & Co.[67] Tivnan, B. F., Dewhurst, D. R., Van Oort, C. M., Ring, J. H., Gray, T. J., Tivnan, B. F., Koehler, M. T. K., McMahon, M. T., Slater, D. M., Veneman, J. G. & Danforth, C. M. (2020). 美国股票市场的碎片化与低效率：来自道琼斯工业平均指数30只成分股的证据。《公共科学图书馆·综合》**15**, e0226968.[68] Tivnan, B. F., Slater, D., Thompson, J. R., Bergen-Hill, T. A., Burke, C. D., Brady, S. M., Koehler, M. T., McMahon, M. T., Tivnan, B. F. & Veneman, J. G. (2018). 美国股票市场中的价格发现与综合数据馈送的准确性。《风险与金融管理杂志》**11**, 73.[69] Tsybakov, A. B. (2009). 非参数估计导论。施普林格。[70] Vieu, P. (1996). 关于密度众数估计的简述。《统计与概率快报》**26**, 297–307.[71] Wegman, E. J. (1971). 关于众数估计的简述。《数理统计年刊》**42**, 1909–1915.[72] Zhou, H. & Huang, X. (2019). 非参数众数回归的带宽选择。《统计学中的通信——模拟与计算》**48**, 968–984.