arXiv:2303.10018v2 [math.ST] 5 Jan 2024

托普利茨协方差矩阵的高效非参数估计

Karolina Klockmann[1] Tatyana Krivobokova[1]

2024年1月8日

摘要

本文提出一种用于托普利茨协方差矩阵的新型高效非参数估计量。该估计量基于一种数据变换，将托普利茨协方差矩阵估计问题转化为近似高斯回归中的均值估计问题。由此得到的托普利茨协方差矩阵估计量按构造即为正定，完全由数据驱动，且计算速度非常快。此外，研究表明，对于一大类托普利茨矩阵，该估计量在谱范数意义下是极小极大最优的。这些结果可直接推广到托普利茨协方差矩阵逆矩阵的估计。此外，本文还提出一种基于离散余弦变换、用于谱密度的替代版 Whittle 似然。该方法已在随论文提供的 R 包 vstdct 中实现。

关键词：离散余弦变换；周期图；谱密度；方差稳定变换；Whittle似然。

1 引言

协方差矩阵和精度矩阵的估计是统计数据分析中的一个基本问题，在自然科学和社会科学中有无数应用。

[1]维也纳大学统计与运筹学系，奥斯卡-摩根斯特恩广场 1 号，1090 维也纳，奥地利

一种特殊的协方差矩阵，其每条向下的对角线上的元素均为常数，称为托普利兹矩阵，常见于平稳随机过程的研究中。平稳随机过程是许多应用领域中的重要建模工具，例如雷达目标检测、语音识别、互联网经济活动建模、脑电活动或晶体结构运动（Du 等人，2020；Roberts 和 Ephraim，2000；Quah，2000；Franaszczuk 等人，1985年；Grenander 和 Szeg¨o，1958，第232页）。

用于估计的数据是 n 个独立同分布的实现值，来自一个具有零均值和托普利茨协方差矩阵 $\Sigma = (\sigma_{|i-j|})_{i,\,j=1}^{p}$ 的 p 维向量。因此，假设 p 不断增大，而 n 可以等于 1，也可以趋于无穷大。对于 $p \to \infty$ 和 $p/n \to c \in (0, \infty]$，已知样本（自）协方差矩阵在谱范数下是 Σ 的不一致估计量（例如参见 Wu 和 Pourahmadi，2009；Pourahmadi，2013，第 2 章）。因此，人们提出对样本协方差矩阵进行渐消、带状化和阈值化，以正则化该估计量（参见 Wu 和 Xiao，2012；Cai 等，2013）。Cai 等（2013）确立了托普利茨协方差矩阵估计量的最优收敛速率，并特别指出，渐消和带状化估计量在某些托普利茨协方差矩阵空间上达到极小极大最优收敛速率。阈值化估计量的最优性仅在情形 $n = 1$ 下得到证明（参见 Wu 和 Xiao，2012）。然而，所有这些估计量都有若干实际缺点，会影响它们在小样本中的表现。首先，必须对估计量进行额外处理以确保正定性，例如参见 Cai 等（2013）的第 5 节。其次，在实践中，数据驱动地选择渐消、带状化或阈值化参数并非易事。对于 $n > 1$，Bickel 和 Levina（2008）提出了一种交叉验证准则，用于近似估计量的风险。Fang 等人（2016）在一项大规模模拟研究中，将该方法与基于自助法的风险近似方法进行了比较，并建议优先采用交叉验证而非自助法。然而，对于较小的 n，交叉验证选出的调优参数会有很大变异；而当 n 达到中等规模时，计算很快就会变得极其繁重。对于 $n = 1$，据我们所知，目前还没有完全由数据驱动的带状化／渐消／阈值化参数选择方法

可供使用。Wu 和 Pourahmadi（2009）建议先将时间序列拆分为互不重叠的子序列，然后应用 Bickel 和 Levina（2008）的交叉验证准则。然而，事实证明，子序列长度的恰当选择对这一方法至关重要，但无法通过数据驱动的方式完成。

本文提出了一种估计托普利茨协方差矩阵及其逆矩阵的替代方法。我们的方法利用了托普利茨协方差矩阵与其谱密度之间的一一对应关系。首先，将给定数据转换为近似高斯随机变量，其均值等于谱密度的对数。然后，使用带有数据驱动平滑参数的周期平滑样条估计对数谱密度。最后，将所得谱密度估计量转换为 Σ 或其逆矩阵的估计量。结果表明，该程序能得到一种完全由数据驱动、自动正定的估计量，并且在谱范数下，对一大类托普利茨协方差矩阵实现极小极大最优收敛速率。特别是，该类别包括与具有有界谱密度的长记忆过程相对应的托普利茨协方差矩阵。此外，该方法计算效率很高，无需迭代或重抽样方案，并且允许应用非参数高斯回归领域中开发的任意推断和自适应估计程序。

从单个平稳时间序列估计谱密度是一项历史悠久的研究课题。早期的非参数方法基于对（对数）周期图进行平滑，而周期图本身并不是一致估计量（Bartlett，1950；Welch，1967；Thomson，1982；Wahba，1980）。另一类非参数谱密度估计方法基于Whittle似然，它是对频域中时间序列精确似然的一种近似。例如，Pawitan和O'Sullivan（1994）通过惩罚Whittle似然估计谱密度，而Kooperberg等人（1995）则使用多项式样条，通过最大化Whittle似然来估计对数谱密度函数。近年来，有人提出了用于谱密度估计的贝叶斯方法（参见Choudhuri等人，2004；Edwards等人，2019；Maturana-Russel和Meyer，2021），但由于后验抽样，这些方法在大样本中可能会带来极高的计算开销，原因是

验

Bentkus（1985年）在 L_p 范数下得出了基于单个高斯平稳时间序列估计 H¨older 连续谱密度的非参数估计量的极小极大最优收敛速率， $1 \leq p \leq \infty$。关于谱密度估计的研究中，只有少数工作证明了相应估计量的最优性。具体而言，Kooperberg等人（1995）和Pawitan和O'Sullivan（1994）在 L_2 范数下推导了其对数谱密度估计量的收敛速率，但忽略了 Whittle 似然近似误差。

一般而言，大多数关于谱密度估计的研究并未进一步利用谱密度估计与相应托普利茨协方差矩阵估计之间的密切联系。特别是，谱密度估计量的 L_∞ 风险上界会自动给出相应托普利茨协方差矩阵估计量在谱范数下的风险上界。我们利用这一事实来证明，我们针对托普利茨协方差矩阵提出的非参数估计量具有极小极大最优性。本文的主要贡献在于证明，我们提出的谱密度估计量不仅数值计算效率很高，在小样本下表现出色，而且在 L_∞ 范数下达到极小极大最优速率，进而确保相应托普利茨协方差矩阵估计量具有极小极大最优性。

本文结构如下。第2节介绍模型，并讨论利用离散余弦变换对托普利茨协方差矩阵进行近似对角化。此外，本文提出了 Whittle 似然的一种替代形式。第3节推导了托普利茨协方差矩阵和精度矩阵的新估计量，第4节则介绍它们的理论性质。第5节讨论如何选择我们的方法中估计量的参数。第6节给出了高斯数据的模拟结果，第7节讨论我们的方法在非高斯数据上的适用性。第8节展示了一个真实数据示例。本文的证明见附录。文末补充材料还附有针对非高斯数据的进一步模拟研究。

2 设定 Toeplitz 矩阵的构造与对角化 s

设 $Y_1, \ldots, Y_n$ 独立同分布$\sim \mathcal{N}_p(0_p, \Sigma)$，其中 Σ 是一个具有托普利兹结构的 $p \times p$ 半正定协方差矩阵，即 $\Sigma = (\sigma_{|i-j|})_{i,\, j=1}^{p} \succeq 0$。样本量$n$ 可以趋于无穷大，也可以是常数。情形 $n = 1$ 对应于对平稳时间序列的一次观测，在这种情形下，数据记为 $Y \sim \mathcal{N}_p(0_p, \Sigma)$。假设维度 p 不断增长。与托普利茨协方差矩阵 Σ 相对应、且协方差序列 $(\sigma_k)_{k\in\mathbb{Z}}$绝对可和的谱密度函数 f为

$$f(x) = \sigma_0 + 2\sum_{k=1}^{\infty} \sigma_k \cos(kx), \quad x \in [-\pi, \pi].$$

傅里叶逆变换表明

$$\sigma_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\mathrm{d}x = \int_0^1 f(\pi x)\cos(k\pi x)\mathrm{d}x. \tag{1}$$

若 $\sum_{h=-\infty}^{\infty} |\sigma_h| = \infty$，即相应的随机过程是长记忆过程，则谱密度函数直接定义为一个以 a 2π 为周期的非负函数 $f:\ [-\pi, \pi] \to \mathbb{R}_{\geq 0}$ 该函数满足条件 (1)。因此，若存在情形f ，则 Σ 完全由 f刻画。此外，谱密度函数的非负性意味着协方差矩阵具有半正定性。此外，协方差 σ_k 的衰减与 f的光滑度直接相关。最后，Toeplitz 协方差估计量与相应谱密度估计量的收敛速率通过 $\|\Sigma\| \leq \|f\|_\infty = \sup_{x\in[-\pi,\pi]} |f(x)|$直接相关，其中 $\|\cdot\|$表示谱范数（参见 Grenander 和 Szeg¨o, 1958，第 5.2 章）。

与 Cai 等人（2013 年）一样，我们考虑具有 H¨older 连续谱密度的半正定 Toeplitz 协方差矩阵类。对于 $\beta = \gamma + \alpha > 0$，其中 $\gamma \in \mathbb{N} \cup \{0\}$、$0 < \alpha \leq 1$，且 $0 < M_0$、$M_1 < \infty$，令

$$\mathcal{P}_\beta(M_0, M_1) = \left\{ f \mid f : [-\pi, \pi] \to \mathbb{R}_{\geq 0},\, \|f\|_\infty \leqslant M_0,\, \|f^{(\gamma)}(\cdot + h) - f^{(\gamma)}(\cdot)\|_\infty \leqslant M_1 |h|^\alpha \right\}.$$

此外，对于 $f \in \mathcal{P}_\beta(M_0, M_1)$，我们用 $\Sigma(f)$ 表示 $\in \mathbb{R}^{p\times p}$相应的

$p \times p$ 通过傅里叶逆变换 (1) 得到的托普利茨协方差矩阵，并用 $\Sigma^{-1}(f)$ 表示精度矩阵。对于 $\Sigma(f)$，其中 $f \in \mathcal{P}_\beta(M_0, M_1)$，估计托普利茨协方差矩阵的最优收敛速率在很大程度上取决于 β。众所周知，某函数的第 k 阶傅里叶系数，若其第 γ 阶导数满足指数为 $\alpha \in (0,1]$ 的 H¨older 连续性，则其衰减阶至少为 $\mathcal{O}(k^{-\beta})$（见 Zygmund，2002年）。因此，β 决定了协方差 σ_k 的衰减速率；这些协方差是谱密度 f 的傅里叶系数，其衰减速率为 $k \to \infty$。对于 $\beta \in (0, 1/2]$，类 $\mathcal{P}_\beta(M_0, M_1)$ 包含某些长记忆过程的有界谱密度。

下面的引理进一步利用了托普利茨协方差矩阵与其谱密度之间的联系。

引理 1. 设 $\Sigma = \Sigma(f)$ 满足 $f \in \mathcal{P}_\beta(M_0, M_1)$，且 $x_j = (j-1)/(p-1)$ 对于 $j = 1, \ldots, p$。那么

$$\left(D^T \Sigma D\right)_{i,j} = f(\pi x_j)\delta_{i,j} + \frac{1 + (-1)^{|i-j|}}{2} \mathcal{O}\left(p^{-1} + p^{-\beta} \log p\right),$$

where δ_{ij} 是克罗内克δ，$\mathcal{O}(\cdot)$项在 $i, j = 1, \ldots, p$ 上一致，且

$$D = \left(\frac{2}{p-1}\right)^{1/2} \left[\cos\left\{\pi(i-1)\frac{j-1}{p-1}\right\}\right]_{i,j=1}^{p} \quad \textit{divided by } 2^{1/2} \textit{ when } i \textit{ or } j \textit{ is } 1 \textit{ or } p$$

是离散余弦变换I矩阵。

证明见附录A.1。该结果表明，离散余弦变换I矩阵能够近似对角化托普利茨协方差矩阵，且对角化误差在一定程度上取决于相应谱密度的光滑度。

在时间序列分析中，离散傅里叶变换矩阵 $F = p^{-1/2}\{\exp(2\pi \mathtt{i} ij/p)\}_{i,\,j=1}^{p}$（其中 $\mathtt{i}$ 为虚数单位）通常用于近似对角化托普利茨协方差矩阵。利用 $(F^T \Sigma F)_{i,\,i} = f(2\pi i/p) + o(1)$ 这一事实，Whittle（1957年）提出了对单个高斯

平稳时间序列（情形 $n = 1$）的似然进行近似的方法，即所谓的 Whittle似然

$$\mathcal{L}(Y|f) \propto \exp\left\{-\sum_{j=1}^{\lfloor p/2 \rfloor} \log f(2\pi j/p) + \frac{I_j}{f(2\pi j/p)}\right\}. \tag{2}$$

数量 $I_j = |F_j^T Y|^2$（其中 F_j表示矩阵 F 的第 j 列）称为第 j 个傅里叶频率处的周期图。由于周期图具有对称性，只有$\lfloor p/2 \rfloor$ 个数据点 I_1，...， $I_{\lfloor p/2 \rfloor}$ 可用于估计均值 $f(2\pi j/p)$（j =1，...， $\lfloor p/2 \rfloor$），其中 $\lfloor x \rfloor$ 表示严格小于 x 的最大整数。Whittle似然已成为平稳时间序列参数估计的常用工具，例如用于非参数和参数谱密度估计，或用于估计赫斯特指数，参见 Walker（1964年）；Taqqu 和 Teverovsky（1997）。

引理 1 给出了 Whittle似然的以下另一种形式

$$\mathcal{L}(Y|f) \propto \exp\left\{-\sum_{j=1}^{p} \log f(\pi x_j) + \frac{W_j}{f(\pi x_j)}\right\}, \tag{3}$$

其中 $W_j = (D_j^T Y)^2$, , D_j 表示矩阵 D 的第 j 列。请注意，这一似然近似基于的数据点 W_j 是标准 Whittle似然的两倍。因此，它能更有效地利用数据 Y 来估计感兴趣的参数，例如谱密度或赫斯特参数。这在小样本中尤其有利。

方程（2）或（3）提出了一种估计 f 的方法，即在某些线性空间（例如样条空间）上最大化（惩罚）似然；Kooperberg等人（1995）或Pawitan和O' Sullivan（1994）等人曾提出类似方法。然而，这种方法需要精心设计的数值方法来求解相应的优化问题，因为方程（2）或（3）第二项中的谱密度位于分母，这使得难以推导出估计量的闭式表达式，而且常常会导致数值不稳定性。此外，平滑参数的选择也颇具挑战性。

因此 ，我们提出一种替代方法，使谱密度能够

作为近似高斯回归中的均值进行估计。这类估计量具有闭式表达式，无需迭代优化算法，而且可以使用任何常规准则轻松确定平滑参数。首先注意，若 $Y \sim \mathcal{N}_p(0_p, \Sigma)$，且 $\Sigma = \Sigma(f)$ 和 $f \in \mathcal{P}_\beta(M_0, M_1)$，则 $D^T Y \sim \mathcal{N}_p(0_p, D^T \Sigma D)$。因此，对于 $W_j = (D_j^T Y)^2 \, (j = 1, \ldots, p)$，由引理 1 可得

$$W_j \sim \Gamma\left\{1/2, 2f(\pi x_j) + \mathcal{O}\left(p^{-1} + p^{-\beta} \log p\right)\right\}, \tag{4}$$

其中，$\Gamma(a, b)$表示形状参数为 a 、尺度参数为 b的伽马分布。随机变量 W_1，...，W_p 仅渐近独立。显然，对于 $j = 1$，...，p，$E(W_j) = f(\pi x_j) + o(1)$。要根据 W_1，...，W_p估计 f，可以采用以伽马分布为响应的广义非参数回归框架，例如参见 Hastie 和 Tibshirani（1990）的经典专著。然而，这种方法需要通过迭代过程进行估计，例如使用牛顿-拉夫森算法，并在每个迭代步骤中适当选择平滑参数。推导所得估计量的 L_∞ 速率也并非易事。为此，我们建议采用 Cai 和 Zhou（2010）的方差稳定变换，将伽马回归转换为近似高斯回归。在下一节中，我们将针对 $n \geq 1$的一般情形，更详细地介绍该方法论。

3 方法论

令 $L_\delta = \{f : \inf_x f(x) \geq \delta\}$，其中 $\delta > 0$，并设 $\mathcal{F}_\beta = \mathcal{P}_\beta(M_0, M_1) \cap L_\delta$。我们考虑根据样本 $Y_1, ..., Y_n \overset{\text{独立同分布}}{\sim} \mathcal{N}_p(0_p, \Sigma)$ 估计 Σ 和 $\Omega = \Sigma^{-1}$，其中 $\Sigma = \Sigma(f)$ 且 $f \in \mathcal{F}_\beta$。对于 $Y_i \sim \mathcal{N}_p(0_p, \Sigma) \, (i = 1, \ldots, n)$，上一节已证明，借助引理 1，数据可转换为伽马分布随机变量 $W_{i,\,j} = (D_j^T Y_i)^2 \, (i = 1, \ldots, n;\ j = 1, \ldots, p)$，其中，对每个固定的 i，随机变量 $W_{i,\,j}$ 与式 (4) 中给出的 W_j 具有相同的分布。现在，我们将 Cai 和 Zhou（2010）的方法应用于设定 $n \geq 1$。

首先，将变换后的数据点$W_{i,\,j}$ 进行分箱，也就是说，构造更少的新变量

$Q_k(k = 1, \ldots, T)$，其中 $T < p$，通过 $Q_k = \sum_{j=}^{kp/T}{}_{(k-1)p/T+1} \sum_{i=1}^{n} W_{i,\ j}$针对 $k = 1, \ldots, T$构造。请注意，每个分箱中的观测值数量为 $m = np/T$。在第4节的定理 1 中，我们证明，对于任意 $\upsilon \in (1 - \min\{1, \beta\}/3, 1)$，令 $T = \lfloor p^{\upsilon} \rfloor$ 可使谱密度估计量达到极小极大最优速率。为简化记号，将 m 视为整数（否则，可以舍弃最后一个分箱中的若干观测值）。接下来，对每个 Q_k 应用方差稳定变换 $G(x) = 2^{-1/2} \log(x/m)$，得到新的随机变量 $Y_k^* = 2^{-1/2} \log(Q_k/m)$；根据 Cai 和 Zhou（2010），这些随机变量近似服从高斯分布。由于谱密度是关于零对称且以 $[-\pi, \pi]$为周期的函数，因此可以镜像变换所得的观测值，以使用 $Y_T^*, \ldots, Y_2^*$, $Y_1^*, \ldots, Y_{T-1}^*$ 进行估计。重新编号观测值 Y_k^* 并将设计点缩放到区间$[0, 1)$，以便于处理，由此得到一个近似高斯回归问题

$$Y_k^* \overset{\text{approx.}}{\sim} \mathcal{N}\left[H\left\{f(x_k)\right\}, m^{-1}\right], \quad x_k = \frac{k-1}{2T-2} \quad (k = 1, ..., 2T-2),$$

其中 $H(y) = 2^{-1/2}\{\phi(m/2) + \log(2y/m)\}$，且 ϕ 是双伽马函数，参见 Cai 和 Zhou（2010）。现在，可以使用周期平滑样条估计经过缩放和平移的对数谱密度 $H(f)$

$$\widehat{H(f)}(x) = \underset{s \in S_{\text{per}}(2q-1)}{\arg\min} \left[\frac{1}{2T-2} \sum_{k=1}^{2T-2} \{Y_k^* - s(x_k)\}^2 + h^{2q} \int_0^1 \{s^{(q)}(x)\}^2 \, \mathrm{d}x \right], \quad (5)$$

其中 $h > 0$ 表示平滑参数，$q \in \mathbb{N}$ 是惩罚阶数，且 $S_{\text{per}}(2q-1)$ 是由 $2q-1$ 次多项式构成的周期样条空间。关于周期平滑样条的更多细节，参见附录 A.2。

一旦得到估计量 $H(f)$，应用逆变换函数 $H^{-1}(y) = m\exp\{2^{1/2}y - \phi(m/2)\}/2$，即可得到谱密度估计量 $\hat{f} = H^{-1}\{\widehat{H\ f}(\ )\}$。最后，通过傅里叶逆变换得到以下协方差矩阵估计量

$$\widehat{\Sigma} = (\hat{\sigma}_{|i-j|})_{i,j=1}^{p}, \text{ with } \hat{\sigma}_k = \int_0^1 \hat{f}(x) \cos(k\pi x) \, \mathrm{d}x \text{ for } k = 0, ..., p-1. \quad (6)$$

精度矩阵 Ω 通过对谱密度估计量的倒数进行傅里叶逆变换来估计

谱密度估计量的，即，

$$\widehat{\Omega} = (\hat{\omega}_{|i-j|})_{i,\,j=1}^{p}, \text{ 其中 } \hat{\omega}_k = \int_0^1 \hat{f}(x)^{-1} \text{余弦}\,(k\pi x)\,\mathrm{d}x \text{ 对于 } k = 0, ..., p-1. \tag{7}$$

$\widehat{\Sigma}$ 和 Ω 的估计程序可概括如下。

1. 数据变换：定义 $W_{i,j} = (D_j^T Y_i)^2\,(i = 1, \ldots, n;\ j = 1, \ldots, p)$，其中 D 是引理 1 中给出的离散余弦变换I矩阵，D_j 是其第 j 列。

2. 分箱：对于任意 $\upsilon \in (1 - \min\{1, \beta\}/3, 1)$，设定 $T = \lfloor p^{\upsilon} \rfloor$ 并计算

$$Q_k = \sum_{j=(k-1)p/T+1}^{kp/T} \sum_{i=1}^{n} W_{i,j}, \quad (k = 1, \ldots, T).$$

3. 方差稳定变换：对 $k = 1, ..., T$ 和 $m = np/T$，设 $Y_k^* = 2^{-1/2} \log\,(Q_k/m)$。对数据进行镜像处理，得到 $2T-2$ 个近似高斯随机变量 $Y_T^*, \ldots, Y_2^*, Y_1^*, \ldots, Y_{T-1}^*$。

4. 高斯回归：重新编号观测值 Y_k^*，将设计点缩放至 $[0, 1)$，并在近似高斯回归模型中，使用次数为 $2q-1$ 的周期平滑样条估计 $H(f)$

$$Y_k^* = H\{f(x_k)\} + \epsilon_k, \quad x_k = \frac{k-1}{2T-2} \quad (k = 1, \ldots, 2T-2),$$

其中 ϵ_k 是渐近独立同分布的高斯变量。

5. 逆变换：使用 $\hat{f} = H^{-1}\{\widehat{H\ f}\,(\,)\}$ 估计谱密度 f，其中

$$H^{-1}(y) = m \exp\left\{2^{1/2} y - \phi\,(m/2)\right\}/2, \text{ for a digamma function } \phi.$$

6. 估计量：设 $\widehat{\Sigma} = (\hat{\sigma}_{|i-j|})_{i,j=1}^{p}$，其中 $\hat{\sigma}_k = \int_0^1 \hat{f}(x) \cos\,(k\pi x)\,\mathrm{d}x$；以及 $\widehat{\Omega} = (\hat{\omega}_{|i-j|})_{i,j=1}^{p}$，其中 $\hat{\omega}_k = \int_0^1 \hat{f}(x)^{-1} \cos\,(k\pi x)\,\mathrm{d}x\,(k = 0, ..., p-1)$。

$\widehat{\Sigma}$ 和 $\widehat{\Omega}$根据构造，估计量 Σ 和 Ω 是正定矩阵，因为谱密度估计量 f 按定义为非负。关于获得这些估计量所需所有参数的选择，详见第 5 节。

4 理论性质

ˆ ˆ ˆ本节研究估计量 f、 Σ 和 Ω 的渐近性质。设ˆ $f = m\exp\{2^{1/2}\widehat{H\ f}() - \phi(m/2)\}/2$ 为第3节定义的谱密度估计量⌒，其中 $H(f)$由式 (5) 给出， $m = np/T$ ，而 ϕ 是 digamma 函数 ˆ ˆ。此外，设 Σ 为托普利兹协方差矩阵估计量，Ω 为分别由式 (6) 和 (7) 定义的相应精度矩阵。以下定理表明， Σ和 Ω 在满足 $f \in \mathcal{F}_\beta$、 $\beta > 0$的托普利兹矩阵类 $\Sigma(f)$ 上，均达到极小极大最优收敛速率。

定理 1. 设 $Y_1, ..., Y_n \overset{i.i.d.}{\sim} \mathcal{N}_p(0_p, \Sigma)$ $n \geq 1$，其中 $\Sigma = \Sigma(f)$ 满足 $f \in \mathcal{F}_\beta$且 $\beta = \gamma + \alpha > 0$。若 $h > 0$ 满足 $h \to 0$ 且 $hT \to \infty$，则对于任意 $\upsilon \in (1 - \min\{1, \beta\}/3, 1)$ 和 $q = \max\{1, \gamma\}$，谱密度估计量 f,ˆ、相应的协方差矩阵估计量 Σ 以及精度矩阵估计量ˆΩ 满足：对于 $p \to \infty$ 和 n ，存在 $p^{\min\{1,\beta\}}/n \to c \in (0, \infty]$

$$\sup_{f\in\mathcal{F}_\beta} E_f\|\widehat{\Sigma} - \Sigma(f)\|^2 \leq \sup_{f\in\mathcal{F}_\beta} E_f\|\hat{f} - f\|_\infty^2 = \mathcal{O}\left\{\frac{\log(np)}{nph}\right\} + \mathcal{O}(h^{2\beta})$$
$$\sup_{f\in\mathcal{F}_\beta} E_f\|\widehat{\Omega} - \Sigma^{-1}(f)\|^2 = \mathcal{O}\left\{\frac{\log(np)}{nph}\right\} + \mathcal{O}(h^{2\beta}).$$

对于 $h \asymp \{\log(np)/(np)\}^{\frac{1}{2\beta+1}}$ ，可得

$$\sup_{f\in\mathcal{F}_\beta} E_f\|\widehat{\Sigma} - \Sigma(f)\|^2 \leq \sup_{f\in\mathcal{F}_\beta} E_f\|\hat{f} - f\|_\infty^2 = \mathcal{O}\left[\left\{\frac{\log(np)}{np}\right\}^{\frac{2\beta}{2\beta+1}}\right]$$
$$\sup_{f\in\mathcal{F}_\beta} E_f\|\widehat{\Omega} - \Sigma^{-1}(f)\|^2 = \mathcal{O}\left[\left\{\frac{\log(np)}{np}\right\}^{\frac{2\beta}{2\beta+1}}\right].$$

定理 1 是在如下渐近情形下建立的： $p \to \infty$ 和 n 满足$p^{\min\{1,\beta\}}/n \to c \in (0, \infty]$, 即维度 p 不断增大，而样本量 n 要么保持不变，要么也随之增大，但增长速度不超过 $p^{\min\{1,\beta\}}$。请注意，这种渐近情形涵盖了样本协方差矩阵不一致的情况。特别是，对于 $\beta \geq 1$ ，样本量 n 的增长速度不超过 p ，增加样本可以提高收敛速率。如果 $\beta \in (0, 1)$，那么以快于 p^β 的速率增加 n 并不会带来更快的收敛速率，这是由于对角化

误差所致。该误差来自引理 1，只有对谱密度作出额外假设才能改善。

从 n 个独立同分布的高斯向量 Y_1、...、Y_n 中估计 Σ 和 Σ^{-1} 的极小极大最优收敛速率，已由 Cai 等（2013）在均值为零且具有托普利茨协方差矩阵 Σ（f）并满足 $f \in \mathcal{F}_\beta$ 的情形下确立。由于 Cai 等（2013）的定理 5 和定理 7 给出的下界速率与我们定理 1 得出的上界速率相匹配，因此我们得出结论：我们的估计量是极小极大最优的。对于非高斯数据，Σ 和 Σ^{-1} 的极小极大最优收敛速率尚不明确。注意，满足 $\beta > 0$ 的 $\mathcal{F}_\beta$ 包含长记忆过程的有界谱密度。

定理 1 的证明见附录 A.3，这是我们工作的主要结果。该证明中最重要的部分，是推导谱密度估计量 f 在 L_∞ 范数下的收敛速率。在原始研究中，Cai 和 Zhou（2010）针对伽马回归中的小波非参数均值估计量，确立了一个 L_2 速率；该研究假设数据相互独立。在我们的研究中，谱密度估计量 f 基于伽马分布数据 $W_{i,\,1}$、...、$W_{i,\,p}$，这些数据仅在渐近意义下相互独立。此外，这些数据的均值并不恰好是 f（πx_1）...、f（πx_p），而是受到引理 1 所述对角化误差的影响。该误差会叠加到分箱和方差稳定变换带来的误差上；正如 Cai 和 Zhou（2010）所推导的，这种误差刻画了数据偏离高斯分布的程度。最后，我们需要为谱密度估计量推导一个 L_∞ 速率，而非 L_2 速率。总的来说，该证明需要采用不同于 Cai 和 Zhou（2010）所用方法、且部分新颖的工具。一项特别的挑战是处理 $W_{i,\,1}$、...、$W_{i,\,p}$ 的依赖性。

ˆ 为了得到 L_∞ 的速率，我们首先推导对数谱密度的周期平滑样条估计量 H(*f*) 的速率。为此，我们使用 Schwarz 和 Krivobokova（2016）得到的该估计量有效核的闭式表达式，并仔细处理各种（相关）误差；这些误差描述了与均值为 $f(\pi x_i)$ 且误差独立的高斯非参数回归之间的偏离。还需注意，尽管周期平滑样条估计量是基于分箱后的点得到的，但

其速率是以向量维数 p 和样本量 n 表示的。接下来，利用柯西-施瓦茨不等式和中值论证，将该速率转化为谱密度估计量 f 的 L_∞ 速率。要得到托普利兹协方差矩阵估计量的速率，只需注意 $E\|\widehat{\Sigma}-\Sigma\|^2 \leq E\|\hat{}f-f\|_\infty^2$。

5 实际问题

实际应用中，要得到我们的估计量，必须作出若干选择。首先，将数据$Y_i (i = 1, \ldots, n)$ 变换为 $W_{i,\ j} = (D_j^T Y_i)^2 (j = 1, \ldots, p)$，随后进行分箱。根据定理 1，分箱数 $T = \lfloor p^v \rfloor$ 取任意$v \in (1 - \min\{\beta, 1\}/3, 1)$，都能得到极小极大最优估计量。也就是说，对于 $\beta \geq 1$，可以取任意 $v \in (2/3, 1)$；而对于 $\beta < 1$，取值区间取决于 β；例如，对于 $\beta = 1/2$，区间为 $v \in (5/6, 1)$。在模拟研究中，我们发现，对于 v的各种取值，结果都相当稳健。如果不了解 β，可以采取以下做法：进行任何一种标准的长程依赖检验；如果拒绝长程依赖的零假设，即 $\beta > 1/2$，则可以取任意 $v \in (5/6, 1)$。否则，应考虑较小的 v取值区间。此外，还可以通过生成 Y_k^*的QQ图，来检验所选的 T 值是否合适；理想情况下，QQ图应显示其与正态性的偏离很小。补充材料中的几幅图展示了高斯数据的QQ图如何随 T变化。

数据变换后，使用周期平滑样条估计 $Y_k^* (k = 1, \ldots, T)$ 的均值，即对数谱密度。为此，需要选择周期样条空间的基函数、惩罚阶数 $q \in \mathbb{N}$ 以及平滑参数 $h > 0$。

节点设在观测值处的周期样条空间，其基为傅里叶基 $\{2^{1/2} \cos(2\pi x)\ 2^{1/2} \sin(2\pi x)\}$，在 $x_k = (k - 1)/(2T - 2)$ 处取值，其中 $k = 1, \ldots, 2T - 2$。

平滑参数 h 可以使用任何数据驱动方法来选择，例如（广义）交叉验证或经验贝叶斯方法，参见 Wahba（1985年）。

根据定理 1，为了获得极小极大最优估计量， q 的选择应与谱密度的真实光滑度 β相关。假设用于估计的 q 大于真实光滑度 β。那么，收敛速率由真实的 β 决定，并且是极小极大的，与 q 有多大无关。特别地，如果 $\gamma = 0$，那么 $\beta = \alpha \in (0, 1)$，且取 $q = \max\{1, \gamma\} = 1$将得到极小极大最优估计量。如果用于估计的 q 小于β，那么收敛速率将取决于 q 以及平滑参数 h的选择。假设 $\beta > q$ ，且平滑参数 h 通过（广义）交叉验证估计。那么，已有研究表明（Wahba，1985年），周期样条估计量能够适应未知的光滑度，直至 $2q$。也就是说，如果 $\beta < 2q$，那么收敛速率将是极小极大最优；而对于 $\beta \geq 2q$ ，收敛速率将由q决定。如果 $\beta > q$ ，且平滑参数 h 通过经验贝叶斯方法估计，那么 Wahba（1985年）表明，由此得到的估计量无法适应额外的光滑度。然而，在小样本下，由于风险中的常数较小，经验贝叶斯平滑样条的表现可能与交叉验证平滑样条相近，甚至更好，参见 Krivobokova（2013 年）。

由于实际中 β 并非完全已知，因此也无法知道在选定 q时估计量将达到何种速率。不过，通过观察样本协方差函数的衰减情况，可以大致判断 β 是较小（衰减非常缓慢时）还是较大（衰减非常迅速时），进而决定如何选择 q。更具吸引力的方法是采用自适应估计方法，无需事先了解 β，便能得到尽可能好的估计量。例如，在对函数空间作出某些假设的前提下，经验贝叶斯框架能够估计未知的 γ （Serra 等人，2017）。其他用于自适应非参数估计的方法包括聚合和 Lepski 方法，概述及相关参考文献见例如（Chagny，2016）。对自适应谱密度估计量的详细研究超出了本文的范围。

在接下来的两节中，我们将通过模拟研究和真实数据示例，明确讨论参数的选择。

6 模拟研究

本节中，我们将比较所提出的托普利兹协方差估计量与 Cai 等人（2013 年）的渐消估计量，以及样本协方差矩阵的性能。我们使用 R（版本 4.1.2，随机种子 42）进行蒙特卡罗模拟，共生成 100 个样本。我们考虑高斯向量 $Y_1, ..., Y_n \overset{\text{独立同分布}}{\sim} \mathcal{N}_p(0_p, \Sigma)$，其中 (A) $p = 5\,000, n = 1$，(B) $p = 1\,000, n = 50$ 和 (C) $p = 5\,000, n = 10$，协方差函数为 $\sigma : \mathbb{Z} \to \mathbb{R}, k \mapsto \sigma_k$

(1) 多项式衰减，即 $\sigma_k = 1.44(1 + |k|)^{-5.1}$；(2) 自回归过程 $y_t = \epsilon_t + 0.1y_{t-1} - 0.1y_{t-2}$，其中 ϵ_t 是独立同分布的高斯噪声，且 $\text{var}(\epsilon_t) = 1.44$；(3) 相应的谱密度利普希茨连续但不可微：$f(x) = 1.44\{|\sin(x + 0.5\pi)|^{1.7} + 0.45\}$。

特别地，$\sigma_0 = 1.44$ 在全部三个例子中成立。图 1 展示了这三个例子的谱密度及相应的协方差函数。设定参数时请注意，过程（1）的谱密度属于 $\mathcal{F}_4$，过程（2）的谱密度是解析函数，而过程（3）的谱密度来自 $\mathcal{F}_1$。

由于对于所有三个过程 $\beta \geq 1$，根据定理 1，$\upsilon \in (2/3, 1)$ 的任意取值都应能得到最优估计。我们在所有情景中均将分箱数设为 $T = 500$ 个分箱，对应于情景（A）和（C）的 $\upsilon \approx 0.73$，以及情景（B）的 $\upsilon \approx 0.90$。

设置 $q = 2$，若平滑参数通过（广义）交叉验证选取，则过程（1）和过程（3）可达到极小极大最优速率，详见第 5 节的讨论。根据上一节的论证，过程（2）的收敛速率与过程（1）相同，而过程（3）的收敛速率较慢。因此，对于给定情景，我们应观察到过程（1）和过程（2）的平均范数大小相近，而过程（3）的平均范数则稍大

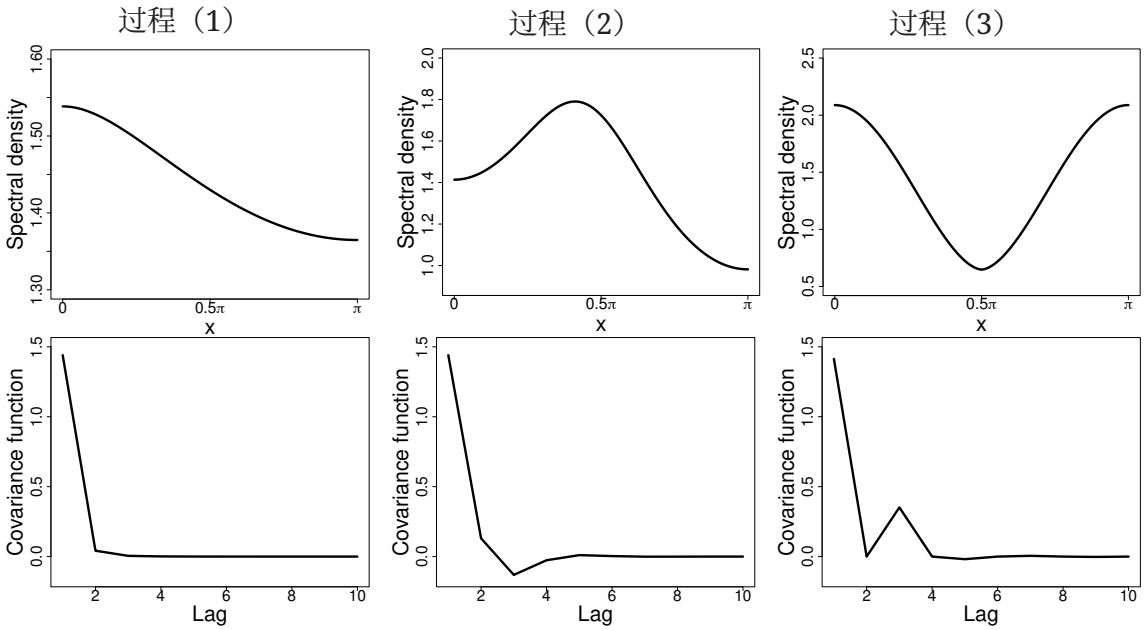


图 1：示例（1）–（3）的谱密度函数（第一行）和协方差函数（第二行）。

。在不同情景之间，我们预计情景（A）的平均范数值较大，因为其中仅使用了 $p = 5\,000$ 个观测值，而情景（B）和（C）中使用了 $np = 50\,000$ 个数据点。

为了为我们的估计量选择正则化参数，我们采用了限制最大似然法、广义交叉验证及相应的预言机版本，即假设已知 Σ 。

~样本协方差矩阵 $\Sigma = (\tilde{\sigma}_{|i-j|})_{i,\ j=1}^{p}$ 定义为 $n^{-1}\sum_{i=1}^{n} Y_i Y_i^T$，通过对角线取平均以获得托普利兹结构。锥化参数为 ta-~pering parameter $k \leq p/2$ 的锥化估计量定义为 $\mathrm{Tap}_k(\Sigma) = (\tilde{\sigma}_{|i-j|} w_{|i-j|})_{i,\ j=1}^{p}$，其中 $w_m = 1$当 $m = 0$ 时，...，$k/2$，当 $k/2 < m \leq k$ 时为 $w_m = 2 - 2m/k$，其他情况为 $w_m = 0$，参见 Cai 等人（2013 年）。只有在 $n > 1$ 时，才能使用交叉验证选择参数 k （参见 Bickel 和 Levina，2008），即在情形 (B) 和 (C) 下。为此，将 n个观测值通过 30 次随机划分，分为大小为$n_1 = 2n/3$ 的训练集和大小为 $n_2 = n/3$的测试集。令 Σ_1^{ν} 和 Σ_2^{ν} 为样本协方差

m矩阵，来自第 ν 次划分。然后将锥化参数 k 估计为

$$\hat{k}_{\mathrm{cv}} = \arg \min_{k=2,3,\ldots,p/2} \frac{1}{30} \sum_{\nu=1}^{30} \|\mathrm{Tap}_k(\Sigma_1^\nu) - \Sigma_2^\nu\|,$$

其中，$\mathrm{Tap}_k(\Sigma_2^\nu)$ 是参数为 k 的渐消估计量。若 $n = 1$，即在情景（A）下，Wu 和 Pourahmadi（2009）建议将时间序列 Y 划分为l 个长度为 p/l 的非重叠子序列，然后按照前述方法选择调优参数 k。据我们所知，目前尚无用于选择 l 的数据驱动方法。利用真实协方差矩阵 Σ，我们为示例（1）预先选取了 $l = 30$ 个子序列的预言值，为示例（2）和（3）预先选取了 $l = 15$ 个子序列的预言值。随后可像上文一样通过交叉验证选择参数 k 。在情景（A）下，我们采用这种方法，而非无法获得的完全数据驱动准则，并将其称为半预言法。最后，对于情景（A）、（B）和（C），通过网格搜索，针对每个蒙特卡洛 ˆ ~ ~样本计算预言渐消参数：$k_{\mathrm{or}} = \arg\min_{k=2,\ 3,\ldots,p/2} \|\mathrm{Tap}_k(\Sigma) - \Sigma\|$，其中 Σ 是样本协方差矩阵。为加快计算，可按 Bickel 和 Levina（2008）的建议，用 ℓ_1 范数替代谱范数。

表 1：(A) $p = 5000, n = 1$：托普利茨协方差矩阵和谱密度估计量相对于谱范数和 L_2 范数的误差；最后一列为每个蒙特卡洛样本的协方差估计量平均计算时间（秒）；除最后一列外，所有数值均乘以 100

	过程（1）		过程（2）		过程（3）		time
	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	秒
我们的方法（广义交叉验证）	0.688	0.255	1.591	0.439	3.401	0.606	4.235
我们的方法（最大似然）	0.591	0.224	1.559	0.417	3.747	0.628	4.224
渐消（半预言）	0.558	0.216	2.325	0.674	3.551	0.979	4.617
样本协方差	16240.895	3810.680	20291.809	3694.392	24438.036	3809.486	0.342
我们的方法（广义交叉验证预言机）	0.421	0.175	1.373	0.378	3.321	0.575	
我们的方法（最大似然预言机）	0.464	0.186	1.487	0.391	3.781	0.624	
渐消（预言机）	0.418	0.171	1.045	0.288	1.547	0.371	

表1、表2和表3分别报告了情景（A）、（B）和（C）下，三个过程的托普利兹协方差估计量相对于谱范数的误差，以及每个蒙特卡洛样本的平均计算时间。为说明谱密度的拟合优度，还计算了 L_2 范数 $\|f - f\|_2$ 。

表 2：(B) $p = 1000, n = 50$：托普利茨协方差矩阵和谱密度估计量相对于谱范数和 L_2 范数的误差；最后一列为每个蒙特卡洛样本的协方差估计量平均计算时间（秒）；除最后一列外，所有数值均乘以 100

	过程（1）		过程（2）		过程（3）		time
	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	秒
我们的方法（广义交叉验证）	0.100	0.028	0.205	0.050	0.531	0.082	27.194
我们的方法（最大似然）	0.076	0.024	0.230	0.051	0.611	0.089	27.098
截断（交叉验证）	0.110	0.031	0.218	0.055	0.348	0.073	23.908
样本协方差	79.603	56.262	95.090	61.000	127.528	61.001	0.141
我们的方法（广义交叉验证预言机）	0.062	0.020	0.163	0.043	0.462	0.074	
我们的方法（最大似然预言机）	0.067	0.021	0.221	0.050	0.603	0.088	
渐消（预言机）	0.057	0.020	0.133	0.037	0.265	0.055	

表3：（C） $p = 5000,\ n = 10$：托普利茨协方差矩阵和谱密度估计量相对于谱范数和 L_2 范数的误差；最后一列为协方差估计量针对一个蒙特卡洛样本的平均计算时间（秒）；除最后一列外，所有数值均乘以100

	过程（1）		过程（2）		过程（3）		time
	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	以秒计
我们的方法（广义交叉验证）	0.088	0.026	0.217	0.050	0.593	0.079	4.260
我们的方法（最大似然）	0.078	0.023	0.231	0.050	0.677	0.086	4.251
截断（交叉验证）	0.143	0.034	0.217	0.051	0.422	0.070	635.345
样本协方差	673.122	370.946	792.714	360.687	1014071.587	375.728	1.189
我们的方法（广义交叉验证预言机）	0.062	0.020	0.172	0.043	0.500	0.071	
我们的方法（最大似然预言机）	0.069	0.021	0.224	0.048	0.663	0.085	
渐消（预言机）	0.055	0.018	0.147	0.039	0.257	0.051	

我们的估计量整体表现完全符合预期。此外，就谱范数风险而言，渐消方法与我们的估计量表现相近。这并不令人意外，因为已证明这两种估计量都是速率最优的。预言机估计量表现相似，但与数据驱动估计量相比，波动略小。显然，渐消方法和我们的估计量都优于不一致的样本协方差矩阵。就计算时间而言，在情景 (A) 和 (B) 中，两种方法的速度相近。在情景 (C) 中，由于交叉验证方法涉及多次高维矩阵乘法，渐消方法慢得多。预计对于更大的 p，与我们的方法相比，锥化估计量的计算开销会大得多。

7 非高斯数据的应用

虽然我们认为，对非高斯数据情形下我们估计量进行严谨的理论研究超出了本文的范围，但本节将讨论我们的方法在实践中应用于非高斯数据的情况。

要应用我们的方法，需要确保变换后的数据 Y_k^*近似服从高斯分布，且均值为 $H\{f(x_k)\}$。如果 Y_k^* 偏离高斯分布的程度足够小，我们的估计量将达到极小极大最优。一般来说，需要确保

(i) $(D_j^T Y_i)^2$的均值与谱密度相差可忽略的误差 $f(\pi x_j)\{1+o(1)\}$；(ii) 经过适当的中心化和缩放后，中心极限定理适用于 Q_k；(iii) 对 $(D_j^T Y_i)^2$而言，对数变换具有方差稳定化作用。

只要 $D_j^T Y_i$ 的二阶矩存在，第一点总是成立。事实上，假设 Y_i 具有托普利茨协方差矩阵 Σ，且其边缘分布为非高斯、零均值分布。那么， $E\{(D_j^T Y_i)^2\} = (D^T \Sigma D)_{jj} = f(\pi x_j)\{1+o(1)\}$.如果已知 Y_i 的分布，最后两点可以明确验证。例如，对于伽马分布数据，第二点显然成立，而第三点已在补充材料中得到证明。

当然，在实践中， Y_i 的分布鲜为人知。由于我们的方法依赖于 Y_k^*的渐近正态性，因此只需检查数据是否明显偏离正态分布即可。请注意，在小样本中对 Y_k^* 进行正态性检验可能会产生误导，因为即使Y_i 服从高斯分布，Y_k^* 也仅渐近服从高斯分布。因此，我们建议绘制 Y_k^*的QQ图。如果该QQ图显示其与高斯分布的偏离很小，那么在实践中就可以放心地应用我们的方法。如果数据明显偏离正态分布，这可能表明对数变换不适用于这些数据。为找到合适的方差稳定化

变换，可以采用Box-Cox 变换。请注意，为了估计Box-Cox 变换参数，必须考虑分箱数据的相关性，例如，可以使用Guerrero（1993）针对时间序列提出的方法。

在补充材料中，我们提供了一项小规模模拟研究，以及若干关于 Y_k^* 的QQ图，其中 $\Sigma^{-1/2}Y_i$ 的分布设为伽马分布和均匀分布。可以观察到，在所有示例中，QQ图与高斯分布的偏离都很小；无论 Y_i 服从何种分布，所有模拟结果都非常相似。

8 蛋白质动力学的应用

我们重新分析了 Krivobokova 等人（2012）和 Singer 等人（2016）对蛋白质动力学所做的数据分析。我们考察了通过分子动力学模拟生成的酵母水通道蛋白数据；该蛋白是酵母毕赤酵母的门控水通道。分子动力学模拟是研究生物系统在纳秒至微秒时间尺度上原子级行为的成熟工具。数据包含在 100 纳秒时间段内观测到的水通道蛋白全部 783 个原子的欧几里得坐标，并分为 20 000 个等间隔观测值。此外，还给出了通道直径 y_t 在时间 t 的数值，该直径通过蛋白质特定残基的两个质心之间的距离测得。分析的目的是识别导致通道打开的原子集体运动。为了根据蛋白质原子的运动对响应变量 y_t（即一个距离）进行建模，我们选择用原子与某些固定基准点之间的距离来表示蛋白质结构，而非使用欧几里得坐标。也就是说，我们计算了

$$X_t = \{d(A_{t,1}, B_1), \ldots, d(A_{t,783}, B_1), d(A_{t,1}, B_2), \ldots, d(A_{t,783}, B_y)\} \in \mathbb{R}^{4 \cdot 783},$$

其中，$A_{t,\,i} \in \mathbb{R}^3$，$i = 1, \ldots, 783$ 表示时间 t,$B_j \in \mathbb{R}^3$ ($j = 1, ..., 4$) 时蛋白质的第 i 个原子，j 是第 j 个基点，而 $d(\cdot, \cdot)$ 表示欧几里得距离。图2展示了直径 y_t 以及第一个原子与第一个质心之间的距离。

可以是 得出结论：线性模型 $Y = Xb+\epsilon$ 成立，其中 $Y = (y_1, \ldots, y_{20\,000})^T$，

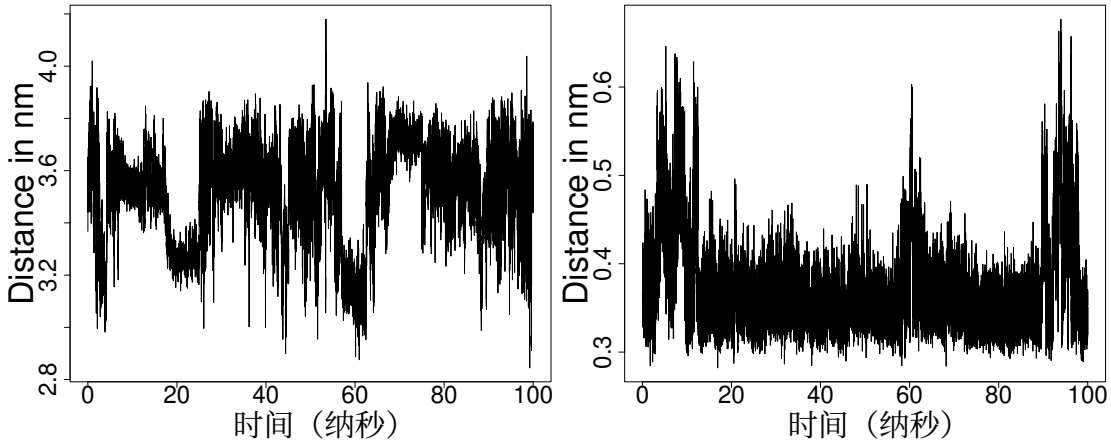


图2：水通道蛋白中第一个原子与第一个质心之间的距离（左），以及开口直径 y_t 随时间 t 的变化（右）。

$X = (X_1^T, \ldots, X_{20\,000}^T)^T$, $b \in \mathbb{R}^{4\cdot 783}$, $\epsilon \in \mathbb{R}^{20\,000}$。该线性模型具有两个问题本身固有的特征：首先，观测值随时间并非相互独立；其次，在每个 t 上， X_t 都是高维的，而 X 中只有少数列与 Y相关。Krivobokova 等人（2012）表明，偏最小二乘（PLS）算法在此类数据上表现出色，能够以低维且稳健的方式表示蛋白质，并识别与 Y相关的原子动力学。Singer 等人（2016）研究了偏最小二乘（PLS）算法在观测值相关时的收敛速率，并表明在运行偏最小二乘（PLS）算法之前对数据进行去相关处理，可以提高算法性能。由于 Y 是 X各列的线性组合，因此可以假定 Y 与 X的所有列具有相同的相关结构。因此，只需估计 $\Sigma = \text{cov}(Y)$，即可对数据进行去相关处理，以便应用偏最小二乘（PLS）算法；也就是说，$\Sigma^{-1/2}Y = \Sigma^{-1/2}Xb + \Sigma^{-1/2}\epsilon$会得到一个误差相互独立的标准线性回归。

我们现在的目标是估计 Σ，并比较偏最小二乘算法在原始数据和去相关数据上的表现。为此，我们将数据集划分为训练集和测试集，每个数据集包含 $p = 10\,000$ 个观测值。首先，我们检验数据是否平稳。增广迪基-富勒检验确认 Y 平稳， p 值为 < 0.01。 Y 的赫斯特指数为 0.85，表明存在中等程度的长程依赖，这一点也得到样本协方差衰减相当缓慢的支持，

见图3左图中的灰线。因此，我们为估计量设定 $q = 1$，以匹配相应谱密度较低的光滑度。应用 R 软件包 `forecast`（该软件包实现了 Guerrero（1993）提出的 Box-Cox 变换参数估计方法）后，确认对这些数据进行对数变换是合适的。分箱数设为 $T = 2500$，即 $\upsilon \approx 0.85$。通过广义交叉验证选择对数谱密度的平滑参数。图3左图中的黑线表明，使用我们的方法估计的协方差矩阵几乎完全消除了训练数据集上通道直径 Y 的相关性。接下来，我们使用通常的偏最小二乘算法估计回归系数 b，不考虑数据中的依赖性。最后，我们使用偏最小二乘算法和协方差估计量 Σ 估计 b，并通过该算法考虑数据中的依-$\widehat{}$赖性。基于这些回归系数估计量，我们计算了测试集上的预测值。图2右侧的图展示了测试集上真实通道直径与同一测试集上预测值之间的皮尔逊相关系数，其中预测值分别基于原始数据（灰色）和去相关数据（黑色）。显然，当成分数较少时，偏最小二乘算法在去相关数据上的表现明显更好。尤其是，仅使用一个成分时，考虑数据依赖性的预测值与测试集上真实开口直径之间的相关系数已达到 0.45；而忽略数据依赖性的偏最小二乘方法所得相关系数则接近于零。Krivobokova 等人（2012）指出，基于一个偏最小二乘成分得到的 b 估计量恰好是集成加权的最大相关众数，它定义为原子的一种集体模态，具有实现响应 Y 特定变化的最高概率。因此，要解释结果，准确估计这一量至关重要，而只有考虑数据中的依赖性才能做到这一点。

使用锥化协方差估计量估计 Σ 存在两个实际问题。首先，由于我们只有时间序列 Y 的单次观测，即 $n = 1$，因此没有基于数据的方法可用于选择锥化参数。其次，对于当前数据，锥化估计量并非正定。为解决第二个问题，我们将相应的谱密度估计量 f_{tap} 截断至一个较小的正值，即

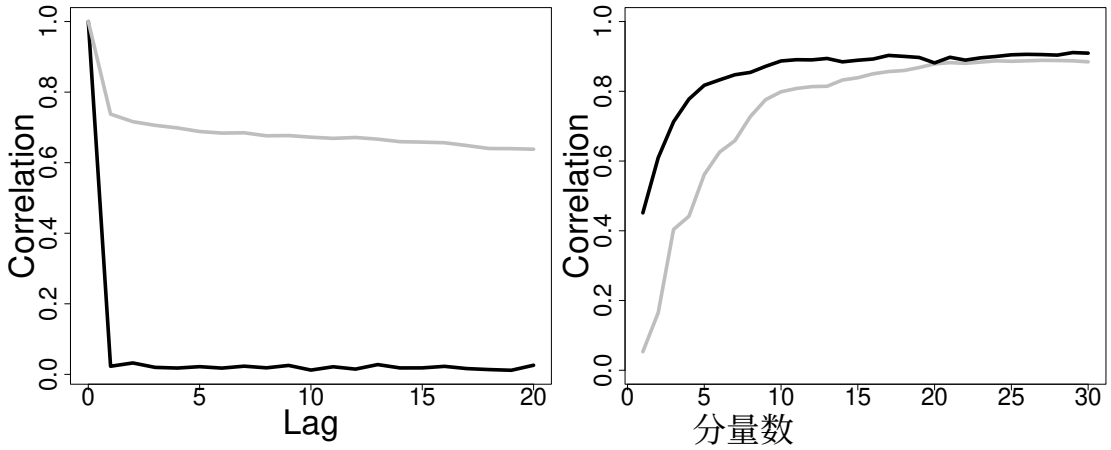


图 3：左侧为 Y（灰色）和 $\widehat{\Sigma}^{-1/2}Y$（黑色）的相关函数，$\widehat{\ }$其中 Σ由我们的方法估计；右侧为测试数据集上的真实值与基于偏最小二乘（灰色）和校正后的偏最小二乘（黑色）所得预测值之间的相关性。

$\hat{f}_{\text{tap}}^{+} = \max\{\hat{f}_{\text{tap}}, 1/\log\ p\}$（参见 Cai 等，2013；McMurry 和 Politis，2010）。为通过交叉验证选择锥化参数，我们尝试了不同的子序列长度，发现锥化估计量对这一选择非常敏感。例如，基于长度为 $8/15/30$ 的子序列估计渐缩协方差矩阵，得到的相关系数为 $0.42/0.53/0.34$，分别对应真实直径与第一分量。

总而言之，我们提出的估计量完全由数据驱动，即使样本量很大也能快速计算，自动满足正定性，并且能够处理某些长记忆过程。蛋白质数据示例以及补充材料中的进一步模拟表明，即使数据不服从正态分布，我们的方法仍能得到稳健的估计量。相比之下，锥化估计量并非在所有数据情形下都由数据驱动，而且必须经过调整才能变为正定。据我们所知，尚未有研究探讨锥化估计量用于非高斯数据的情况。我们的方法已在 R 软件包 `vstdct` 中实现。

A 附录

在整个附录中，我们用 c、c_1、C、C_1 等表示与 n 和 p 无关的通用常数。为简化记号，有时会省略这些常数，并用 . 表示在相差一个常数因子的意义下小于或等于。

A.1 引理1的证明

我们将 p维托普利兹矩阵 $\Sigma = \mathsf{toep}(\sigma_0, \ldots, \sigma_{p-1})$嵌入到一个$(2p-2)$-$\widetilde{\Sigma} = \mathsf{toep}(\sigma_0, \ldots, \sigma_{p-1}, \sigma_{p-2}, \ldots, \sigma_1)$维循环矩阵 $\widetilde{\Sigma} = U\Lambda U^*$，其中 $U = (2p-2)^{-1/2}[\exp\{\pi \mathtt{i}(k-1)(l-1)/(p-1)\}]_{k\,l=1}^{2p-2}$，</style>其共轭转置为 U^*，且 Λ 是一个对角矩阵，其第 k 个对角元素（$k = 1, ..., p$）由下式给出

$$\lambda_k = \sum_{l=1}^{p} \sigma_{l-1} \exp\{\pi \mathtt{i} x_k (l-1)\} + \sum_{l=p+1}^{2p-2} \sigma_{2p-1-l} \exp\{\pi \mathtt{i} x_k (l-1)\},$$

其中 $x_k = (k-1)/(p-1)$。由对称性可得，对于 $k = p+1, ..., 2p-2$，有 $\lambda_k = \lambda_{2p-k}$。此外，$\widetilde{\Sigma} = V^* \Lambda V$，其中 $V \in \mathbb{C}^{(2p-2)\times p}$ 包含 U的前 p 列。于是，对于 $i, j = 1, \ldots, p$,

$$(D\Sigma D)_{i,j} = (DV^*\Lambda V D)_{i,j} = \sum_{r=1}^{2p-2} \lambda_r \sum_{q=1}^{p} D_{iq} V_{qr}^* \sum_{s=1}^{p} V_{rs} D_{sj}.$$

定义缩放因子 $a_s = 2^{-1/2}$，用于 $s = 1$、p，以及用于 $s = 2$、...、$p-2$ 的 $a_s = 1$；这些因子出现在离散余弦变换I矩阵的定义中。然后,

$$\begin{aligned}
\sum_{s=1}^{p} V_{rs} D_{sj} &= (p-1)^{-1} \sum_{s=1}^{p} a_s a_j \exp\{\pi \mathtt{i}(r-1)x_s\} \cos\{\pi(j-1)x_s\} \\
&\quad \pm \frac{a_j}{p-1} \exp(0)\cos(0) \pm \frac{a_j}{p-1} \exp\{\pi \mathtt{i}(r-1)\}) \cos\{\pi(j-1)\} \\
&= a_j \frac{2^{1/2}-2}{p-1} \mathbb{I}^{\{|j-r| \text{ even}\}} + \frac{a_j}{p-1} \sum_{s=1}^{p} \exp\{\pi \mathtt{i}(r-1)x_s\} \cos\{\pi(j-1)x_s\} \\
&= a_j b(j,r) + a_j \mathtt{i} c(j,r),
\end{aligned}$$

其中 $\mathbb{I}$ 是示性函数，并且

$$b(j,r) = \frac{2^{1/2}-2}{p-1}\mathbb{I}^{\{|j-r| \text{ even}\}} + \frac{1}{p-1}\sum_{s=1}^{p}\cos\{\pi(r-1)x_s\}\cos\{\pi(j-1)x_s\},$$

$$c(j,r) = \frac{1}{p-1}\sum_{s=1}^{p}\sin\{\pi(r-1)x_s\}\cos\{\pi(j-1)x_s\}.$$

根据余弦和正弦的对称性，$b(j,r) = b(j,2p-r)$ 且 $c(j,r) = -c(j,2p-r)$，对于 $r = p+1,\ldots,2p-2$。利用 $(\sum_{s=1}^{p} V_{rs}D_{sj})^* = \sum_{q=1}^{p} D_{jq}V_{qr}^*$，可得

$$(D\Sigma D)_{i,j} = \begin{cases} a_j^2 \sum_{r=1}^{2p-2}\lambda_r\left\{b(j,r)^2 + c(j,r)^2\right\}, & i=j \\ a_i a_j \sum_{r=1}^{2p-2}\lambda_r(b(i,r)+\mathtt{i}c(i,r))(b(j,r)-\mathtt{i}c(j,r)), & i\neq j. \end{cases} \tag{A.1}$$

经过一些计算可得，对于 $r = 1,\ldots p$,

$$\frac{1}{p-1}\sum_{s=1}^{p}\cos\{\pi(r-1)x_s\}\cos\{\pi(j-1)x_s\} = \begin{cases} \frac{1}{p-1}, & r\neq j, r-j \text{ even} \\ 0, & r-j \text{ odd} \\ \frac{p+1}{2p-2}, & r=j\neq 1,p \\ \frac{p}{p-1}, & r=j=1,p \end{cases} \tag{A.2}$$

$$\frac{1}{p-1}\sum_{s=1}^{p}\sin\{\pi(r-1)x_s\}\cos\{\pi(j-1)x_s\}$$
$$= \frac{\{1-(-1)^{r-j}\}}{4(p-1)}\left[\cot\{\pi(r-j)/(2p-2)\} + \cot\{\pi(j+r-2)/(2p-2)\}\right]. \tag{A.3}$$

利用余切的泰勒展开，即对于 $0<|x|<\pi$，$\cot(x) = x^{-1} - x/3 - x^3/45 + ...$，可得对于 $r = 1,\ldots p$

$$c(j,r) = \frac{(r-1)\{1-(-1)^{r-j}\}}{\pi\{(r-1)-(j-1)\}\{(r-1)+(j-1)\}} + \mathcal{O}\left\{\frac{2\pi(2r-2)}{3(2p-2)^2}\right\}, \tag{A.4}$$

其中，$\mathcal{O}$ 项与 j 无关，且常数与 r, p无关。

情形 $i = j$

方程 (A.1) – (A.3) 表明

$$
\begin{aligned}
(\boldsymbol{D\Sigma D})_{j,j} &= a_j^2 \sum_{\substack{r=1\\ r\neq j,\\ r\neq 2p-j}}^{2p-2} \lambda_r \left[\frac{(2^{1/2}-2)^2+1+2(2^{1/2}-2)(p-1)}{(p-1)^2} \mathbb{I}^{\{|r-j| \text{ even}\}} + c(j,r)^2 \right] \\
&\quad + a_j^2 \lambda_j \cdot \begin{cases} 2\left(\frac{2^{1/2}-2}{p-1}+\frac{p+1}{2p-2}\right)^2, & j\neq 1,p \\ \left(\frac{2^{1/2}-2}{p-1}+\frac{p}{p-1}\right)^2, & j=1,p \end{cases} \\
&= a_j^2 \sum_{\substack{r=1\\ r\neq j,\\ r\neq 2p-j}}^{2p-2} \lambda_r \left[\frac{(2^{1/2}-1)^2}{(p-1)^2} \mathbb{I}^{\{|r-j| \text{ even}\}} + c(j,r)^2 \right] + \lambda_j \cdot \begin{cases} \frac{2(2^{1/2}-3/2+p/2)^2}{(p-1)^2}, & j\neq 1,p \\ \frac{(2^{1/2}-2+p)^2}{2(p-1)^2}, & j=1,p \end{cases} \\
&= \mathcal{O}(p^{-1}) + 2a_j^2 \sum_{r=1}^{p-1} \lambda_r c(j,r)^2 + \lambda_j\{0.5+\mathcal{O}(p^{-1})\},
\end{aligned}
$$

这里用到了 λ_r 的对称性，且根据假设 $\|f\| < M_0$。注意， $\mathcal{O}$ 项不依赖于 j。由于复指数函数是利普希茨连续的，且利普希茨常数为 $L=1$，因此有$\lambda_r = \lambda_j + L_{r,\ j}|r-j|(p-1)^{-1}$ 其中$-1 \le L_{r,\ j} \le 1$ 是一个依赖于 r、 j的常数。于是，

$$
\sum_{r=1}^{p-1} \lambda_r c(j,r)^2 = \lambda_j \sum_{r=1}^{p-1} c(j,r)^2 + p^{-1} \sum_{r=1}^{p-1} L_{r,j}|r-j| c(r,j)^2.
$$

由于 $\sum_{r=1}^{p-1} \lambda_r c(1,r)^2 = -\sum_{r=1}^{p-1} \lambda_r c(p,r)^2$,， 只需考虑 $j=1,...,p-1$。我们先处理第一项求和。为简洁起见，以下使用 $k=r-1$和 $l=j-1$。于是，对 $l=0,...,p-2$，将式 (A.4) 中第一项的平方求和可得

$$
\frac{1}{\pi^2} \sum_{k=0}^{p-2} \frac{k^2\{1-(-1)^{k-l}\}^2}{(k-l)^2(k+l)^2} = \frac{4}{\pi^2} \begin{cases} \sum_{x=1}^{\infty} \frac{1}{(2x-1)^2} - R_1(l,p) = \frac{\pi^2}{8} - R_1(l,p), & l=0 \\ \sum_{x=1}^{\infty} \frac{(2x)^2}{(2x-l)^2(2x+l)^2} - R_2(l,p) = \frac{\pi^2}{16} - R_2(l,p), & l \text{ odd} \\ \sum_{x=1}^{\infty} \frac{(2x-1)^2}{(2x-1-l)^2(2x-1+l)^2} - R_2(l,p) = \frac{\pi^2}{16} - R_3(l,p), & \text{else} \end{cases}
$$

参见 Abramowitz and Stegun（1964年）第 23.2 章中关于倒数幂次求和的内容。若 p 为偶数，则残差项为

$$
R_1(l,p) = \sum_{x=(p-2)/2}^{\infty} \frac{1}{(2x-1)^2} = \frac{\phi^{(1)}\{(p-3)/2\}}{4},
$$

$$
\begin{aligned}
R_2(l,p) &= \sum_{x=p/2}^{\infty} \frac{(2x)^2}{(2x-l)^2(2x+l)^2} \\
&= \frac{2\phi\{(p+l)/2\}-2\phi\{(p-l)/2\}+l[\phi^{(1)}\{(p+l)/2\}+\phi^{(1)}\{(p-l)/2\}]}{16l}, \\
R_3(l,p) &= \sum_{x=p/2}^{\infty} \frac{(2x-1)^2}{(2x-1-l)^2(2x-1+l)^2} \\
&= \frac{2\phi\{(p-1+l)/2\}-2\phi\{(p-1-l)/2\}+l[\phi^{(1)}\{(p-1+l)/2\}+\phi^{(1)}\{(p-1-l)/2\}]}{16l},
\end{aligned}
$$

其中，ϕ 和 $\phi^{(1)}$ 分别表示双伽马函数及其导数。若 p 为奇数，也可以推导出类似的余项。为证明当 i =为 1、2、3 且对 l 一致时，$R_i(l,p) = \mathcal{O}(p^{-1})$，我们利用以下渐近关系：$\phi(x)\sim\log(x)-1/(2x)$，以及$\phi^{(1)}(x)\sim 1/x+1/(2x^2)$。混合项 $4/3\sum_{k=0}^{p-2}\ {}^{k^2\{1-}_{k-l}\ (-1)^{\ k-l}\}^2\ \ ()\ \ (k+l)\ \ (2p-2)^{\ 2}$ 和平方$\underline{\mathcal{O}\ 项\ \ 16\pi^2/9\sum_{k=0}^{p-2}}\ \ {}^{k^2}\ \ {}^{(2p-2)}_{\ 4}$ 的阶均为 p^{-1}。此外，

$$
\frac{1}{p-1}\sum_{r=1}^{p-1} L_{r,j}|r-j|c(r,j)^2 = \frac{1}{p-1}\sum_{k=0}^{p-2}\frac{k^2\{1-(-1)^{k-l}\}^2 L_{k,l}|k-l|}{\pi^2(k-l)^2(k+l)^2} = \mathcal{O}\{p^{-1}\log p\},
$$

因为调和和以 log p 的速度发散。最后，由 Hölder 连续函数的离散傅里叶级数的一致逼近性质（Zygmund，2002年），$\lambda_j = f(x_j)+\mathcal{O}\{p^{-\beta}\log\ p\}$。综上所述，我们证明了 $(D\Sigma D)_{j,j} = f(x_j)\ +\mathcal{O}\{p^{-1}+p^{-\beta}\log\ p\}$，其中 $\mathcal{O}$ 项在 $j=1,...,p$ 上一致。

情形 $i\neq j$ 且 $|i-j|$ 为偶数

在这种情形下，$(D\Sigma D)_{i,j} = a_i a_j\sum_{r=1}^{2p-2}\lambda_r\{b(i,r)b(j,r)\ +c(i,r)c(j,r)\}$，因为 $|r-j|$，且$|r-i|$ 两者均为奇数或偶数。注意，对于 $r\neq i$，j，若 $i=r$ 或 $j=r$，则 $b(i,r)b(j,r)\ =\mathcal{O}(p^{-2})$，而 $\mathcal{O}(\cdot)$ 项在 i,j 上一致。由于 $b(i,r)\ \geq 0$ 且 $\|f\|_\infty < M_0$，可得 $a_i a_j\sum_{r=1}^{2p-2}\lambda_r b(i,r)b(j,r)\ =\mathcal{O}(p^{-1})$ 在 i,j 上一致。我们采用与之前类似的方法，证明 $a_i a_j\sum_{r=1}^{2p-2}\lambda_r c(i,r)c(j,r)\ =\mathcal{O}(p^{-1})$。令 $k=r-1, l=j-1, m=i-1$，并利用 $l\neq m$ 和 $|l-m|$ 均为偶数，可得

$$
\frac{1}{\pi^2}\sum_{k=0}^{p-2}\frac{k^2\{1-(-1)^{k-l}\}\{(-1)^{k-m}-1\}}{(k-l)(k+l)(k-m)(k+m)}
$$

$$= \frac{4}{\pi^2} \begin{cases} \sum_{x=1}^{\infty} \frac{(2x)^2}{(2x-l)(2x+l)(2x-m)(2x+m)} - R_1(l,m,p) = -R_1(l,m,p), & l,m \text{ odd} \\ \sum_{x=1}^{\infty} \frac{(2x-1)^2}{(2x-1-l)(2x-1+l)(2x-1-m)(2x-1+m)} - R_2(l,m,p) = -R_2(l,m,p), & l,m \text{ even}, \end{cases}$$

其中，对于偶数 p，残差项为

$$R_1(l,m,p) = \sum_{x=p/2}^{\infty} \frac{(2x)^2}{(2x-l)(2x+l)(2x-m)(2x+m)}$$

$$R_2(l,m,p) = \sum_{x=p/2}^{\infty} \frac{(2x-1)^2}{(2x-1-l)(2x-1+l)(2x-1-m)(2x-1+m)}.$$

与之前类似，可以证明 $R_1(l,m,p)$ 和 $R_2(l,m,p)$ 的阶为 $\mathcal{O}(p^{-1})$，其中隐藏常数不依赖于 l,m。若 p 为奇数，则可以推导出类似的残差项。运用与之前类似的技巧，可以证明这两个残差项以及其余的混合项和平方项以 $\mathcal{O}(p^{-1})$ 的阶趋于零，并且对 i,j 一致。

情形 $i \neq j$ 和 $|i-j|$ 为奇数

$|r-i|$ 和 $|r-j|$ 要么一奇一偶，要么一偶一奇。不失一般性，假设 $|r-i|$ 为偶数。那么，$(D\Sigma D)_{i,j} = a_i a_j \sum_{r=1}^{2p-2} \lambda_r b(i,r) c(j,r)$。由于 $b(i,\cdot)$ 是偶函数，$c(j,\cdot)$ 是奇函数，且 $\lambda_r = \lambda_{2p-r}$，因此有 $(D\Sigma D)_{i,j} = 0$。

□

A.2 周期平滑样条

本节回顾 Schwarz 和 Krivobokova（2016）关于周期平滑样条估计量等价核的研究结果。给定数据点 $\{(x_i, Y_i)\}_{i=1}^N$，其真实关系为 $Y_i = f(x_i) + \epsilon_i$，其中 $x_i = i/N$，且残差 ϵ_i 满足标准假设，周期平滑样条估计量 f 是以下问题的解：

$$\min_{s \in S_{\text{per}}(2q-1, \underline{x}_N)} \left[\frac{1}{N} \sum_{i=1}^{N} \{Y_i - s(x_i)\}^2 + h^{2q} \int_0^1 \{s^{(q)}(x)\}^2 \, \mathrm{d}x \right],$$

其中，$h>0$为平滑参数，$q\in\mathbb{N}$为惩罚阶数，$S_{\mathrm{per}}(2q-1,\underline{x}_N)$为基于节点$\underline{x}_N=(x_i)_{i=0}^N=(i/N)_{i=0}^N$、次数为$2q-1$的周期样条空间。Schwarz 和 Krivobokova（2016）证明，$\hat{f}(x)=N^{-1}\sum_{i=1}^N W(x,x_i)Y_i$，其中$W(x,x_i)$称为周期平滑样条估计量的有效核，其显式表达式为

$$W(x,x_i)=\sum_{j=1}^{N}\frac{\Phi\{Nx+q,\exp(-2\pi\mathrm{i}j/N)\}\overline{\Phi\{Nx_i+q,\exp(-2\pi\mathrm{i}j/N)\}}}{Q_{2q-2}(j/N)+h^{2q}(2\pi j)^{2q}\mathrm{sinc}(\pi j/N)^{2q}},$$

其中，$\Phi(\cdot,\cdot)$表示指数样条（Schoenberg，1973），其具有复共轭$\overline{\Phi}(\cdot,\cdot)$。多项式$Q_{2q-2}(x)$的形式定义为$Q_{2q-2}=\sum_{l=-\infty}^{\infty}\mathrm{sinc}\{\pi(x+l)\}^{2q}$；显式表达式见 Schwarz（2012）中的第 4.1.1 节。以带宽参数h表示的$W(x,t)$的缩放版本$K(x,t)$定义为$K(x,t)=hW(hx,ht)$。我们用$K_h(x,t)$表示$=K(x/h,t/h)=hW(x,t)$。此外，我们还利用$W(x,t)=\sum_{l=-\infty}^{\infty}\mathcal{W}(x,t+l)$（见 Schwarz（2012）中的引理 14），其中

$$\mathcal{W}(x,t)=\int_0^N\frac{\Phi\{Nx+q,\exp(-2\pi\mathrm{i}u)\}\overline{\Phi\{Nx_i+q,\exp(-2\pi\mathrm{i}u)\}}}{Q_{2q-2}(u)+h^{2q}(2\pi j)^{2q}\mathrm{sinc}(\pi u)^{2q}}\,\mathrm{d}u$$

是平滑样条估计量在$\mathbb{R}$上的渐近等效核。最后，我们记$\mathcal{K}_h(x,t)=\mathcal{K}(x/h,t/h)=h\mathcal{W}(x,t)$。

A.3 定理1的证明

证明的结构如下。首先，我们推导周期性平滑样条估计量$H(f)$的L_∞收敛速率。然后，利用柯西-施瓦茨不等式和中值论证，建立谱密度估计量f的收敛速率。结合关系$E\|\widehat{\Sigma}-\Sigma\|\le E\|\hat{f}-f\|_\infty^2$，即可得出定理的第一个结论。最后，我们证明关于精度矩阵的第二个结论。为清晰起见，辅助引理 2、3 和 4 的证明单独列于 B 节。

A.3.1 关于 $E\|H(f) - H(f)\|_\infty^2$的上界

命题 1. 设 $\Sigma = \Sigma(f)$ 满足 $f \in \mathcal{F}_\beta$，使得 $\beta > 0$。若 $h > 0$ 满足$h \to 0$且 $hT \to \infty$，则对于任意 $\upsilon \in (1 - \min\{1, \beta\}/3, 1)$，在 $T = \lfloor p^\upsilon \rfloor$ 条件下，第3节中描述的估计量 $H(f)$ 满足 $q = \max\{1, \gamma\}$，其中 $p \to \infty$ 且n，使得 $p^{\min\{1,\beta\}}/n \to c \in (0, \infty]$

$$E\|\widehat{H(f)} - H(f)\|_\infty^2 = \mathcal{O}\left\{\log(np)/(nph)\right\} + \mathcal{O}(h^{2\beta}).$$

证明。应用三角不等式可得偏差-方差分解

$$E\|\widehat{H(f)} - H(f)\|_\infty^2 \leq 2E\|\widehat{H(f)} - E\{\widehat{H(f)}\}\|_\infty^2 + 2\|E\{\widehat{H(f)}\} - H(f)\|_\infty^2. \quad \text{(A.5)}$$

~ ~ ~设 $T = 2T - 2$ 且 $x_k = (k-1)/T$，其中 $k = 1, ..., T$. 周期平滑样条估计量 $H(f)$ 可表示为关于核函数 $W(x, y)$ 的核估计量，也可表示为其缩放形式 $K_h(x,t) = hW(x,t)$。核的显式表示见 Schwarz 和 Krivobokova，2016。附录 A.2 总结了他们的结果。B 节列出的引理 4 给出了 $x \in [0,1]$ 时 $H(f)(x)$ 的如下分解：将其分为确定性部分、高斯部分和非高斯部分

$$\widehat{H(f)}(x) = \frac{1}{\tilde{T}} \sum_{k=1}^{\tilde{T}} W(x, x_k) Y_k^* = \frac{1}{\tilde{T}h} \sum_{k=1}^{\tilde{T}} K_h(x, x_k) \left[H\{f(x_k)\} + \epsilon_k + \zeta_k + \xi_k\right],$$

其中 $|\epsilon_k| \lesssim (np)^{-1} + (np)^{-\beta} \log p$，$\zeta_k \sim \mathcal{N}(0, m^{-1})$，且 $\text{cov}(\zeta_k, \zeta_l) = \mathcal{O}\{p^{-2} + p^{-2\beta}(\log p)^2\}$，其中 $k \neq l$。随机变量 ξ_k 满足 $E|\xi_k|^\ell \lesssim (\log m)^{2\ell}\{m^{-\ell} + (T^{-1} + T^{-1}p^{1-\beta} \log p)^\ell\}$，对每个整数 $\ell > 1$ 均成立，且其均值为零。对 ζ_k进行镜像并重新编号，与处理 $Y_k^*(k = 1, ..., T)$ 的方式类似。

方差上界

利用上述表示，可以写出

$$E\|\widehat{H(f)} - E\{\widehat{H(f)}\}\|_\infty^2 = E\left\|\frac{1}{\tilde{T}} \sum_{k=1}^{\tilde{T}} W(x, x_k)(\zeta_k + \xi_k)\right\|_\infty^2. \quad \text{(A.6)}$$

首先，我们用有限个点上的最大值来界定上确界。如果$q > 1$，那么 $W(\cdot, x_k)$是利普希茨连续的，常数为 $L > 0$。在这种情形下，几乎必然有

$$\begin{aligned}
&\sup_{x\in[0,1]}\left|\frac{1}{\tilde{T}}\sum_{k=1}^{\tilde{T}}W(x,x_k)(\zeta_k+\xi_k)\right|^2 \\
&\leq \left[\max_{1\leqslant j\leqslant\tilde{T}}\left|\frac{1}{\tilde{T}}\sum_{k=1}^{\tilde{T}}W(x_j,x_k)(\zeta_k+\xi_k)\right| + \sup_{\substack{x,x'\in[0,1],\\ |x-x'|<1/\tilde{T}}}\left|\frac{1}{\tilde{T}}\sum_{k=1}^{\tilde{T}}\{W(x,x_k)-W(x',x_k)\}(\zeta_k+\xi_k)\right|\right]^2 \\
&\leq 2\max_{1\leqslant j\leqslant\tilde{T}}\left|\frac{1}{\tilde{T}}\sum_{k=1}^{\tilde{T}}W(x_j,x_k)(\zeta_k+\xi_k)\right|^2 + \frac{2L^2}{\tilde{T}^4}\left(\sum_{k=1}^{\tilde{T}}|\zeta_k+\xi_k|\right)^2.
\end{aligned}$$

利用 $E\{(\sum_{k=1}^{T}|\zeta_k+\xi_k|)^2\} \lesssim E\{(\sum_{k=1}^{T}|\zeta_k|)^2\} + E\{(\sum_{k=1}^{T}|\xi_k|)^2\} \lesssim \tilde{T}^2 m^{-1}$，可得

$$(A.6) \lesssim E\left\{\max_{1\leqslant j\leqslant\tilde{T}}\left|\frac{1}{\tilde{T}}\sum_{k=1}^{\tilde{T}}W(x_j,x_k)(\zeta_k+\xi_k)\right|^2\right\} + o\{(np)^{-1}\}. \qquad \text{(A.7)}$$

若 $q = 1$，则 $\sum_{k=1}^{T}W(\cdot, x_k)$ 是一个分段线性函数，其节点位于 $x_j = j/\tilde{T}$。因子 $(\zeta_k + \xi_k)$ 可视为随机权重，不会影响分段线性这一性质。因此，上确界在节点 $x_j = j/\tilde{T}$之一处取得，其中 $j = 1, ..., \tilde{T}$，且 (A.7) 对 $q = 1$ 也成立。再次利用 $(a+b)^2 \leq 2a^2+2b^2$，可得

$$(A.6) \lesssim E\left\{\max_{1\leqslant j\leqslant\tilde{T}}\left|\frac{1}{\tilde{T}}\sum_{k=1}^{\tilde{T}}W(x_j,x_k)\zeta_k\right|^2\right\} + E\left\{\max_{1\leqslant j\leqslant\tilde{T}}\left|\frac{1}{\tilde{T}}\sum_{k=1}^{\tilde{T}}W(x_j,x_k)\xi_k\right|^2\right\} + o\{(np)^{-1}\}.$$

我们首先利用 Tsybakov（2009）的引理 1.6 对 $T_1 = E\{\max_{1\leq j\leq\tilde{T}}|(\tilde{T}h)^{-1}\sum_{k=1}^{T}K_h(x_j,x_k)\zeta_k|^2\}$ 进行界定。这需要对 $\|(\tilde{T}h)^{-1}\sum_{k=1}^{T}K_h(x_j,x_k)\zeta_k\|_{\psi_2}^2$ 给出一个界，其中$\|\cdot\|_{\psi_2}$ 表示次高斯范数。对于高斯随机变量，该范数等于方差。有关次高斯分布的更多详情，请参见 Vershynin（2018）。

利用补充材料引理 2 中给出的核函数 K_h 的性质，我们得到

$$\begin{aligned}
&\left\| \frac{1}{\tilde{T}h} \sum_{k=1}^{\tilde{T}} K_h(x_j, x_k)\zeta_k \right\|_{\psi_2}^2 = \mathrm{var}\left\{ \frac{1}{\tilde{T}h} \sum_{k=1}^{\tilde{T}} K_h(x_j, x_k)\zeta_k \right\} \\
&= \frac{1}{\tilde{T}^2h^2} \sum_{k=1}^{\tilde{T}} K_h(x_j, x_k)^2 \mathrm{var}(\zeta_k) + \frac{1}{\tilde{T}^2h^2} \sum_{k=1}^{\tilde{T}} K_h(x_j, x_k) \sum_{l=1}^{\tilde{T}} K_h(x_j, x_l)\mathrm{cov}(\zeta_k, \zeta_l) \\
&\leq C(Thm)^{-1} + C'\{p^{-2} + p^{-2\beta}(\log p)^2\}.
\end{aligned}$$

由（Tsybakov，2009）的引理 1.6 可得

$$T_1 \lesssim \log(2\tilde{T})[C(Thm)^{-1} + C'\{p^{-2} + (\log p)^2 p^{-2\beta}\}] = \mathcal{O}\{(Thm)^{-1}\log(2\tilde{T}) + h^{2\beta}\}. \tag{A.8}$$

为说明这一点，注意到 $p^{\min\{1,\beta\}}/n \to c \in (0, \infty]$ 意味着 $p^{-1} = \mathcal{O}(n^{-1})$。因此，若$\beta > 1$，则

$$\log(2\tilde{T})\{p^{-2} + p^{-2\beta}(\log p)^2\} \lesssim \log(2\tilde{T})p^{-2} \lesssim \log(2\tilde{T})(Tm)^{-1}.$$

现在考虑 $0 < \beta \leq 1$。回想一下，对某个固定的 $\upsilon \in (1 - \min\{1, \beta\}/3, 1)$，有 $T = \lfloor p^\upsilon \rfloor$。可以找到一个常数 $a = C_\upsilon$，它依赖于 υ，但不依赖于 n, p，使得不等式 $\log(x) \leq x^a/a$ 蕴含 $\log(2\tilde{T})\log(p)^2 p^{-2\beta}T^{2\beta} = \mathcal{O}(1)$。因此，由于根据假设 $hT \to \infty$，$\log(2\tilde{T})\log(p)^2 p^{-2\beta} = \mathcal{O}(h^{2\beta})$。

接下来，我们为第二项 $T_2 = E\{\max_{1\leq j\leq \tilde{T}} |(\tilde{T}h)^{-1} \sum_{k=1}^{T} K_h(x_j, x_k)\xi_k|^2\}$推导一个界。核 K_h的指数衰减性质（见 B 节中的引理 2）给出

$$T_2 \lesssim E\left\{ \max_{1\leq j\leq \tilde{T}} \left| \frac{1}{\tilde{T}h} \sum_{k=1}^{\tilde{T}} \gamma_h(x_j, x_k)|\xi_k| \right|^2 \right\},$$

其中 $\gamma_h(x, t) = \gamma^{|x-t|/h} + \gamma^{1/h}\{\gamma^{(x-t)/h} + \gamma^{(t-x)/h}\}(1 - \gamma^{1/h})^{-1}$ 且 $\gamma \in (0, 1)$ 是一个常数。对于稍后指定的某个阈值 $R > 0$，定义 $\xi_k^- = |\xi_k|_{\mathbb{I}}\{|\xi_k| \leq R\}$ 并且

$\xi_k^+ = |\xi_k|_{\mathbb{I}}\{|\xi_k| > R\}$。于是,

$$T_2 \lesssim E\left\{\max_{1\le j\le \tilde{T}}\left|\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)\xi_k^-\right|^2\right\} + E\left\{\max_{1\le j\le \tilde{T}}\left|\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)\xi_k^+\right|^2\right\}. \tag{A.9}$$

（A.9）中的第一项可再次利用 Tsybakov（2009）的引理 1.6 进行界定。我们利用这样一个事实：对于不一定相互独立的随机变量 $X_1, ..., X_N$，若满足 $|X_i| < R$ $(i = 1, ..., N)$，则有 $\|\sum_{i=1}^N a_i X_i\|_{\psi_2}^2 \le 4R^2\sum_{i=1}^N a_i^2$，其中 $a_1, ..., a_N \in \mathbb{R}$ 和 $R > 0$ 均为常数。这是 Azuma （1967）的引理 1 的推论，该引理给出 $E\{\exp(\lambda\sum_{i=1}^N a_i X_i)\} \le \exp(\lambda^2 R^2\sum_{i=1}^N a_i^2)$，对所有 $\lambda \in \mathbb{R}$ 均成立。因此，对所有 $t > 0$，均有 $\text{pr}(|\sum_{i=1}^N a_i X_i| > t) \le 2\exp(-\lambda t + \lambda^2 R^2\sum_{i=1}^N a_i^2)$。令 $\lambda = t/2\cdot(R^2\sum_{i=1}^N a_i^2)^{-1}$，可得 $\sum_{i=1}^N a_i X_i$ 服从次高斯分布，且其次高斯范数有界于 $2R(\sum_{i=1}^N a_i^2)^{1/2}$。

综上可得 $\|(\tilde{T}h)^{-1}\sum_{k=1}^T \gamma_h(x_j,x_k)\xi_k^-\|_{\psi_2}^2 \le 4R^2(\tilde{T}h)^{-2}\sum_{k=1}^T \gamma_h(x_j,x_k)^2 \lesssim R^2(\tilde{T}h)^{-1}$，其中我们使用了引理 2 中关于 γ_h 的界。接着应用 Tsybakov（2009）的引理 1.6，得到

$$E\left\{\max_{1\le j\le \tilde{T}}\left|\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)\xi_k^-\right|^2\right\} \lesssim \frac{R^2}{\tilde{T}h}\log(2\tilde{T}). \tag{A.10}$$

为了界定 (A.9) 中的第二项，我们使用引理 4 中推导出的 ξ_k 矩界。于是，对于所有整数 $\ell > 1$

$$\begin{aligned}
&E\left\{\max_{1\le j\le \tilde{T}}\left|\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)\xi_k^+\right|^2\right\} = E\left\{\max_{1\le j\le \tilde{T}}\left|\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)(\xi_k^+)^\ell(\xi_k^+)^{1-\ell}\right|^2\right\} \\
&\le R^{2-2\ell}E\left\{\max_{1\le j\le \tilde{T}}\frac{1}{\tilde{T}^2h^2}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)^2\cdot\sum_{k=1}^{\tilde{T}}(\xi_k^+)^{2\ell}\right\} \le R^{2-2\ell}E\left(\frac{C_1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}\xi_k^{2\ell}\right) \\
&\lesssim h^{-1}R^{2-2\ell}(\log m)^{4\ell}\{m^{-2\ell} + (T^{-1} + T^{-1}p^{1-\beta}\log p)^{2\ell}\}.
\end{aligned} \tag{A.11}$$

Com合并误差界 (A.10) 和 (A.11)，并选择 $R = m^{-1/2}$，可得

$$T_2 \lesssim \frac{\log T}{mTh} + \frac{(\log m)^{4\ell}}{hm^{1-\ell}} \cdot \begin{cases} \frac{(\log p)^{2\ell} p^{2\ell(1-\beta)}}{T^{2\ell}} + m^{-2\ell}, & 0 < \beta \le 1, \\ T^{-2\ell} + m^{-2\ell}, & 1 < \beta. \end{cases}$$

根据假设， $T = \lfloor p^{\upsilon} \rfloor$ 和 $m = \lfloor np^{(1-\upsilon)} \rfloor$ 对某个固定的 $\upsilon \in (1 - \min\{1, \beta\}/3, 1)$成立。若 ℓ 是满足 $\ell \ge 1/(1-\upsilon)$的整数，则

$$m^{\ell-1}(\log m)^{4\ell} m^{-2\ell} \lesssim m^{-l} \lesssim n^{-\ell} p^{-(1-\upsilon)\ell} = \mathcal{O}\{(np)^{-1}\},$$

其中我们使用了 $\log x \le x^a/a$ ，且 $a = 1/(4\ell)$。现在考虑 $0 < \beta \le 1$。于是，根据假设，$n \le cp^{\beta}$ 。令 $0 < \chi < 1$ 为一个较小的常数。将不等式 $\log x \le x^a/a$ 在 $a = \chi/(2\ell)$下使用两次，可得

$$\begin{aligned} & m^{\ell-1}(\log m)^{4\ell}(\log p)^{2\ell} p^{2\ell(1-\beta)} T^{-2\ell} \lesssim n^{\ell-1+\chi} p^{(1-\upsilon)(\ell-1+\chi)-2\ell\upsilon+2\ell(1-\beta)+\chi} \\ & \lesssim p^{\ell(3-3\upsilon-\beta)+(\beta+\chi+1)(1-\upsilon)+\chi}. \end{aligned} \tag{A.12}$$

注意，对于 $\upsilon \in (1 - \beta/3, 1)$，条件 $\ell(3 - 3\upsilon - \beta) + (\beta + \chi + 1)(1 - \upsilon) + \chi < -2$ 等价于 $\ell > \{-2 - (\beta + \chi + 1)(1 - \upsilon) - \chi\}/(3 - 3\upsilon - \beta)$。因此，对任意固定的 υ ，都可以找到一个与 n、p 无关的整数 ℓ ，使得(A.12)右侧的阶为 $\mathcal{O}(p^{-2})$。对于同样选取的 ℓ ，有$m^{\ell-1}(\log\ m)^{4\ell} T^{-2\ell} = \mathcal{O}(p^{-2})$。

总之，选择整数 $\ell \ge \max[1/(1-\upsilon)\ \{-2 - (\beta + \chi + 1)(1 - \upsilon)\ - \chi\}/(3 - 3\upsilon - \beta)]$ 可得

$$\mathbb{E}\|\widehat{H(f)} - \mathbb{E}\{\widehat{H(f)}\}\|_{\infty}^2 = \mathcal{O}\left\{\frac{\log(np)}{nph} + h^{2\beta}\right\}. \tag{A.13}$$

偏差上界

再次利用引理 4 中的表示，可得对每个 $x \in [0, 1]$

$$E\{\widehat{H(f)}(x)\} - H\{f(x)\} = \frac{1}{\tilde{T}h} \sum_{k=1}^{\tilde{T}} K_h(x, x_k) \left[H\{f(x_k)\} + \epsilon_k\right] - H\{f(x)\}.$$

关于 ϵ_k 的界意味着

$$\left|\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}K_h(x,x_k)\epsilon_k\right| \lesssim \frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)|\epsilon_k| \lesssim (np)^{-1}+(np)^{-\beta}\log p.$$

考虑 $\beta \geq 1$ 的情形。特别地， $q=\gamma$ 和 $f^{(q)}$ 是 α-H¨older 连续的。由于 f 是周期函数，且 $f(x) \in [\delta, M_0]$，并且 $H(y) \propto \phi(m/2) + \log(2y/m)$，因此 $\{H(f)\}^{(q)}$ 也是 α-H¨older 连续的。将 $g = H(f)$ 延拓到整个实数轴上，可得

$$\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}K_h(x,x_k)g(x_k) = \int_{-\infty}^{\infty}h^{-1}\mathcal{K}_h(x,t)g(t)\,\mathrm{d}t + \mathcal{O}(\tilde{T}^{-\beta})$$

其中， $\mathcal{K}_h$ 是将 K_h 延拓到整个实数轴上的函数（参见 Schwarz 和 Krivobokova，2016）。在 x 处对 $g(t)$ 展开泰勒级数，并利用 $h^{-1}\mathcal{K}_h$ 是一个 $2q$ 阶核这一事实（参见引理 2（iii）），可知对于任意 $x \in [0,1]$

$$\begin{aligned}
&\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}K_h(x,x_k)g(x_k) = g(x) + \int_{-\infty}^{\infty}h^{-1}\mathcal{K}_h(x,t)(x-t)^q\frac{g^{(q)}(\xi_{x,t})}{q!}\,\mathrm{d}t \; + \mathcal{O}(\tilde{T}^{-\beta})\\
&= g(x) + \int_{-\infty}^{\infty}h^{-1}\mathcal{K}_h(x,t)(x-t)^q\frac{g^{(q)}(\xi_{x,t})-g^{(q)}(x)}{q!}\,\mathrm{d}t + \mathcal{O}(\tilde{T}^{-\beta})\\
&= g(x) + \sum_{l=-\infty}^{\infty}\int_{x+(l-1)h}^{x+lh}\mathcal{K}_h(x,t)(x-t)^q\frac{g^{(q)}(\xi_{x,t})-g^{(q)}(x)}{hq!}\,\mathrm{d}t + \mathcal{O}(\tilde{T}^{-\beta}),
\end{aligned}$$

其中 $\xi_{x,\,t}$ 是位于 x 和 t 之间的一个点。利用核 $\mathcal{K}_h$ 指数衰减这一事实，以及 $g^{(q)}$ 在 $[\delta, M_0]$ 上是 α-H¨older 连续的（其中常数为 L），可得

$$\begin{aligned}
&\left|\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}K_h(x,x_k)g(x_k) - g(x)\right| \leq \frac{CL}{q!}\sum_{l=-\infty}^{\infty}\int_{x+(l-1)h}^{x+lh}\gamma^{\frac{|x-t|}{h}}|x-t|^q\frac{|\xi_{x,t}-x|^\alpha}{hq!}\,\mathrm{d}t + \mathcal{O}(\tilde{T}^{-\beta})\\
&\leq \frac{CL}{q!}\sum_{l=-\infty}^{\infty}\int_{x+(l-1)h}^{x+lh}\gamma^{\frac{|x-t|}{h}}\frac{|x-t|^\beta}{h}\,\mathrm{d}t + \mathcal{O}(\tilde{T}^{-\beta}) \leq h^\beta\frac{CL}{q!}\sum_{l=-\infty}^{\infty}\gamma^{|l-1|}|l|^\beta + \mathcal{O}(\tilde{T}^{-\beta})\\
&= \mathcal{O}(h^\beta) + \mathcal{O}(\tilde{T}^{-\beta}).
\end{aligned}$$

若 $0 < \beta \leq 1$，则 $q=1$ 和 g 是 β-H¨older 连续的。由于 $f(x) \in [\delta, M_0]$，且对数函数在紧区间上是利普希茨连续的，因此可得 $g = H(f)$ 是

β-H¨older 连续。将 g 扩展到整个实数轴，并用$m = 0$ 应用引理 2(iii)，可得

$$\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}} K_h(x, x_k)g(x_k) - g(x) = \int_{-\infty}^{\infty} \mathcal{K}_h(x, t)\frac{g(t) - g(x)}{h}\,\mathrm{d}t + \mathcal{O}(\tilde{T}^{-\beta}).$$

与之前类似，可以得到

$$\left|\frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}} K_h(x, x_k)g(x_k) - g(x)\right| = \mathcal{O}(h^{\beta}) + \mathcal{O}(\tilde{T}^{-\beta}).$$

注意，当 $hT \to \infty$ 时，$\tilde{T}^{-\beta} = o(h^{\beta})$，并且根据假设有 $h \to 0$ 。由于所得界对 $x \in [0, 1]$ 一致，因此有

$$\left\|E\{\widehat{H(f)}(x)\} - H\{f(x)\}\right\|_{\infty} = \mathcal{O}(h^{\beta}) + \mathcal{O}\{(np)^{-1} + (np)^{-\beta}\log p\}. \tag{A.14}$$

将界 (A.13) 和 (A.14) 合并可得

$$E[\|\widehat{H(f)} - H(f)\|_{\infty}^2] = \mathcal{O}\left\{\log(np)/(nph)\right\} + \mathcal{O}(h^{2\beta}).$$

□

A.3.2 关于 $E\|\hat{}f - f\|_{\infty}^2$ 的上界

命题 2. 设 $\Sigma = \Sigma(f)$，其中 $f \in \mathcal{F}_{\beta}$ 满足 $\beta > 0$。若 $h > 0$ 满足$h \to 0$ 且 $hT \to \infty$，则对于任意 $\upsilon \in$ （$1 - \min\{1, \beta\}/3, 1$），当 $T = \lfloor p^{\upsilon} \rfloor$ 时，第3节中描述的 ˆ估计量 f 在 $q = \max\{1, \gamma\}$ 的条件下，对 $p \to \infty$ 以及满足 $p^{\min\{1,\beta\}}/n \to c \in (0, \infty]$的 n 有

$$E\|\hat{f} - f\|_{\infty}^2 = \mathcal{O}\left\{\log(np)/(nph)\right\} + \mathcal{O}(h^{2\beta}).$$

证明。由中值定理，对于某个函数 g between，有 $\widehat{H(f)}$ 和$H(f)$，满足

$$\begin{aligned} E\|\hat{f} - f\|_{\infty}^2 &= E\|H^{-1}\{\widehat{H(f)}\} - H^{-1}\{H(f)\}\|_{\infty}^2 = E\|(H^{-1})'(g)(\widehat{H(f)} - H(f))\|_{\infty}^2 \\ &\lesssim E\|H^{-1}(g)\{\widehat{H(f)} - H(f)\}\|_{\infty}^2. \end{aligned}$$

A应用带有常数 $C>0$ 的柯西-施瓦茨不等式可得

$$E\|\hat{f}-f\|_\infty^2 \le C^2 E\|\widehat{H(f)}-H(f)\|_\infty^2 + (E\|\hat{f}-f\|_\infty^4)^{1/2}\cdot \mathrm{pr}(\|H^{-1}(g)\|_\infty > C)^{1/2}. \tag{A.15}$$

为了证明 (A.15) 右侧第二项可以忽略，我们使用马尔可夫不等式。为此，我们首先推导 $\frown H(f)$ 的矩母函数在 $n,p\to\infty$ 下的渐近阶。

$\frown \|H(f)$ 的矩母函数$\|_\infty$

根据引理 2 中核 K 的指数衰减性质，几乎必然有

$$\|\widehat{H(f)}\|_\infty \lesssim \sup_{x\in[0,1]} \frac{1}{\tilde{T}h}\sum_{k=1}^{\tilde{T}} \gamma_h(x,x_k)\cdot|Y_k^*|\,,$$

其中 $\gamma_h(x,t) = \gamma^{|x-t|/h} + \gamma^{1/h}\{\gamma^{(x-t)/h}+\gamma^{(t-x)/h}\}/(1-\gamma^{1/h})$，某个常数$\frown\gamma\in(0,1)$。首先，$\|H(f)\|_\infty$ 由有限个 x_k 点上的最大值限定。计算 s 的导数：$[0,1]\to\mathbb{R}, x\mapsto s(x) = \sum_{k=1}^T \gamma_h(x,x_k)\,|Y_k^*|$，对于 $x\neq x_k$，几乎必然有

$$\frac{\partial}{\partial x}s(x) = h^{-1}\log\gamma\sum_{k=1}^{\tilde{T}}\left[\gamma^{|x-x_k|/h}|x-x_k| + \frac{\gamma^{1/h}}{1-\gamma^{1/h}}\left\{\gamma^{(x-x_k)/h}+\gamma^{(x_k-x)/h}\right\}\right]|Y_k^*|\,.$$

˜由于对于 $x\neq x_k$，$\partial s(x)/\partial x>0$ 几乎必然成立，因此极值出现在 $x_k(k=1,...,T)$。$\frown$因此，对于 $\lambda>0$，$\|H(f)\|_\infty$ 的矩母函数至多相差常数地有界于

$$E[\exp\{\lambda\|\widehat{H(f)}\|_\infty\}] \lesssim \sum_{j=1}^{\tilde{T}} E\left[\exp\left\{\frac{\lambda}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)\,|Y_k^*|\right\}\right].$$

˜令 $M_j = (\tilde{T}h)^{-1}\sum_{k=1}^T\gamma_h(x_j,x_k)$，根据引理 2，它在 j上一致有界，且上界为某个全局常数 $M>0$。由指数函数的凸性，我们

得到

$$E\left[\exp\left\{\frac{\lambda}{\tilde{T}h}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)|Y_k^*|\right\}\right]\leq\frac{1}{M_j\tilde{T}h}\sum_{k=1}^{\tilde{T}}\gamma_h(x_j,x_k)E\left[\exp\left\{\lambda M_j|Y_k^*|\right\}\right].$$

回顾一下，$Y_k^*=2^{-1/2}\log(Q_k/m)$，且根据假设 $0\leq\delta\leq f\leq M_0$。根据引理 3，$Q_k$ 可写成 $m=np/T$ 个独立伽马随机变量之和，即 $Q_k=\sum_{i=1}^m A_i$，其中 $A_i\sim\Gamma(1/2,2a_i)$，且 $a_i\in(c_1\delta,c_2M_0)$，其中 c_1、$c_2>0$为某些常数。因此，我们可以找到一些（完全相关的）随机变量$Q_k^-\sim\Gamma(m/2,\tilde{c}_1\delta)$和 $Q_k^+\sim\Gamma(m/2,\tilde{c}_2M_0)$，使得 $Q_k^-\leq Q_k\leq Q_k^+$ 几乎必然成立，且 $\tilde{c}_1$、$\tilde{c}_2>0$ 为某些常数。于是，

$$E[\exp\{\lambda M_j|Y_k^*|\}]\leq E\left[\exp\left\{\lambda M_j 2^{-1/2}\max\left\{|\log(Q_k^+/m)|,|\log(Q_k^-/m)|\right\}\right\}\right].$$

$|\log X|$的矩母函数（其中 $X\sim\Gamma(a,b)$）为

$$E\left\{\exp(t\left|\log X\right|)\right\}=b^{-t}\frac{\Gamma(a-t)-\gamma(a-t,1/b)+b^{2t}\gamma(a+t,1/b)}{\Gamma(a)},\quad |t|<a,$$

其中，$\Gamma(a)$ 是伽马函数，$\gamma(a,b)$ 是下不完全伽马函数。特别地，

$$E\{\exp(t|\log X|)\}\leq\frac{b^{-t}\Gamma(a-t)+b^t\Gamma(a+t)}{\Gamma(a)}.\tag{A.16}$$

对于 $0<t<a.$，在我们的设定中，$X=Q_k^-/m\sim\Gamma(m/2,\tilde{c}_1\delta/m)$ 或 $X=Q_k^+/m\sim\Gamma(m/2,\tilde{c}_2M_0/m)$，分别如此。为了推导 $E[\exp\{\lambda\|H(f)\|_\infty\}]$ 在 $p/n\to c\in(0,\infty]$ 时的渐近阶，其中 $\lambda>0$，我们首先建立比值 $\Gamma(a+t)/\Gamma(a)$ 在 $a\to\infty$ 时的渐近阶。我们区分两种情形：t 与 a 无关，以及 t 线性依赖于 a。基于斯特林近似，可对 $a>1$ 和 $t>0$ 推导出

$$\begin{aligned}\frac{\Gamma(a+t)}{\Gamma(a)}&=\frac{\{2\pi(a+t)^{-1}\}^{1/2}(a+t)^{a+t}\exp(-a-t)\left[1+\mathcal{O}\{(a+t)^{-1}\}\right]}{(2\pi/a)^{1/2}a^a\exp(-a)\left\{1+\mathcal{O}(a^{-1})\right\}}\\&=\{a(a+t)^{-1}\}^{1/2}\left\{(a+t)a^{-1}\right\}^a(a+t)^t\exp(-t)\left[\frac{1+\mathcal{O}\{(a+t)^{-1}\}}{1+\mathcal{O}(a^{-1})}\right].\end{aligned}\tag{A.17}$$

因此，对于与 a独立的 $0<t<a$ 和 t，由公式 (A.17) 可得，对于 $a\to\infty$

$\Gamma(a+t)/\Gamma(a) = \mathcal{O}(a^t)$。同理，可以看出 $\Gamma(a-t)/\Gamma(a) = \mathcal{O}(a^{-t})$。若 $0 < t < a$ 和 t 关于 a呈线性依赖关系，即对于某个常数 $c \in \in(0,1)$，有 $t = ca$，则可得 $\Gamma(a \pm t)/\Gamma(a) = \mathcal{O}\{a^{\pm t}\exp(a)\}$，对于 $a \to \infty$。因此，对于一个不依赖于 n 的固定 λ，且满足 $0 < \lambda < m2^{-1/2}M_j^{-1}$ 的 p，当n充分大时，可得 p

$$E\left[\exp\left\{\lambda M_j|Y_k^*|\right\}\right] = E\left[\exp\left\{\lambda M_j 2^{-1/2}\left|\log\left(Q_k/m\right)\right|\right\}\right]$$
$$\leq c_3 \sum_{b\in\{\tilde{c}_1\delta/m, \tilde{c}_2 M_0/m\}} (bm/2)^{-\lambda M_j 2^{-1/2}} + (bm/2)^{\lambda M_j 2^{-1/2}} = \mathcal{O}(1).$$

若 $\lambda = cm$ 使得 $0 < \lambda < m2^{-1/2}M_j^{-1}$，则对充分大的 n，p

$$E\left[\exp\left\{\lambda M_j|Y_k^*|\right\}\right] \leq c_4 \exp(m/2) \sum_{b\in\{\tilde{c}_1\delta/m, \tilde{c}_2 M_0/m\}} (bm/2)^{-\lambda M_j 2^{-1/2}} + (bm/2)^{\lambda M_j 2^{-1/2}}$$
$$= \mathcal{O}(L^m),$$

对于某个常数 $L > \exp(1/2)$。令 $K = \min_{j=1,\ldots,\tilde{T}} 2^{-1/2}M_j^{-1}$，它是一个与 n、p 无关的常数。综上，我们证明了对于 $0 < \lambda < Km$ 和 n，$p \to \infty$

$$E[\exp\{\lambda\|\widehat{H(f)}\|_\infty\}] = \mathcal{O}(T), \qquad \text{if } \lambda \text{ does not depend on } n, p, \tag{A.18}$$
$$E[\exp\{\lambda\|\widehat{H(f)}\|_\infty\}] = \mathcal{O}(TL^m), \qquad \text{if } \lambda = cm \text{ for } c \in (0, K). \tag{A.19}$$

对 (A.15) 的右侧进行界定

注意到对于 $m \geq 1$，$\|f\|_\infty \leq M_0$ 和 $m/2\exp\{\phi(-m/2)\} \in [1,4]$，由 (A.18) 可得，对于某些常数 c_0、$c_1 > 0$ 和 n，$p \to \infty$

$$E\|\widehat{f} - f\|_\infty^4 \leq c_0 E\|\widehat{f}\|_\infty^4 + c_1 = c_0 E\|(m/2)\exp\{2^{1/2}\widehat{H(f)} - \phi(m/2)\}\|_\infty^4 + c_1$$
$$\leq c_0 \left|(m/2)^4 \exp\{-4\phi(m/2)\}\right| \cdot E[\exp\{4\cdot 2^{1/2}\|\widehat{H(f)}\|_\infty\}] + c_1 = \mathcal{O}(T).$$

$\widehat{H^{-1}} \propto \exp$，因此可得 $\widehat{H^{-1}(g)} \leq H^{-1}\{H(f)\} + f$ 几乎处处逐点成立。因此，对于 $C > \|f\|_\infty = M_0$，有

$$\text{pr}\left\{\|H^{-1}(g)\|_\infty > C\right\} \leq \text{pr}\left[\|H^{-1}\{\widehat{H(f)}\}\|_\infty > C - M_0\right] \leq \text{pr}\left[\|\widehat{H(f)}\|_\infty > c_1\right],$$

其中 $c_1 = H(C - M_0)$。对 $t = cm$ 应用马尔可夫不等式，其中 $c \in (0, K)$

以及 $C = 2L^{4/c} + M_0$，其中 c、K、L 是 (A.19) 中的常数，可得

$$\mathrm{pr}\{\|H^{-1}(g)\|_\infty > C\} \le \exp(-c_1 t) E[\exp\{t\|\widehat{H(f)}\|_\infty\}] = \mathcal{O}(TL^{-m}).$$

因此，(A.15) 的右侧相对于 (A.15) 的第一项渐近可忽略。结合命题 1，可得

$$E\|\hat{f} - f\|_\infty^2 = \mathcal{O}\left\{\log(np)/(nph)\right\} + \mathcal{O}(h^{2\beta}).$$

□

A.3.3 $E\|\widehat{\Sigma} - \Sigma\|^2$ 的上界

利用托普利兹矩阵的谱范数不超过其谱密度的上确界范数这一事实，并注意到命题 2 对每个 $f \in \mathcal{F}_\beta$ 都成立，可得

$$\sup_{f\in\mathcal{F}_\beta} E_f\|\widehat{\Sigma} - \Sigma(f)\|^2 \le \sup_{f\in\mathcal{F}_\beta} E_f\|\hat{f} - f\|_\infty^2 = \mathcal{O}\left\{\log(np)/(nph)\right\} + \mathcal{O}(h^{2\beta}). \quad \text{(A.20)}$$

A.3.4 $E\|\hat{\Omega} - \Sigma^{-1}\|^2$ 的上界

根据中值定理，对于介于 $H(f)$ 和 $H(f)$ 之间的函数 g，有

$$\begin{aligned} E\|\hat{\Omega} - \Omega\|^2 &\le E\|1/\hat{f} - 1/f\|_\infty^2 = E\|1/H^{-1}\{\widehat{H(f)}\} - 1/H^{-1}\{H(f)\}\|_\infty^2 \\ &= E\|(1/H^{-1})'(g)\{\widehat{H(f)} - H(f)\}\|_\infty^2 \lesssim E\|1/H^{-1}(g)\{\widehat{H(f)} - H(f)\}\|_\infty^2. \end{aligned}$$

应用常数 $C > 0$ 的柯西-施瓦茨不等式，可得

$$E\|1/\hat{f} - 1/f\|_\infty^2 \le C^2 E\|\widehat{H(f)} - H(f)\|_\infty^2 + (E\|1/\hat{f} - 1/f\|_\infty^4)^{1/2} \mathrm{pr}\left[\|1/\{H^{-1}(g)\}\|_\infty > C\right]^{1/2}.$$

注意，对某个不依赖于 n, p 的常数 $c_1 > 0$，$\|1/H^{-1}(g)\|_\infty \le 2/m\exp(2^{1/2}\|g\|_\infty)\exp\{\phi(m/2)\} \le c_1\|H^{-1}(g)\|_\infty$。取与命题 2 的证明中相同的常数 C，即可得到

$$\mathrm{pr}\left[\left\|1/\{H^{-1}(g)\}\right\|_\infty > C\right] = \mathcal{O}(TL^{-m}).$$

注意到，对于 $m \geq 1$，$\|1/f\|_\infty \leq 1/\delta$ 和 $2/m \exp\{\phi(m/2)\} \in [0.25, 1]$，由 (A.18) 可知，存在常数 $c_2, c_3 > 0$ 和 $n, p \to \infty$

$$\begin{aligned} E\|1/\hat{f} - 1/f\|_\infty^4 &\leq c_2|(2/m)^4 \exp\{4\phi(m/2)\}| \cdot E\| \exp\{-42^{1/2}\widehat{H(f)}\}\|_\infty + c_3 \\ &\leq c_2|(2/m)^4 \exp\{4\phi(m/2)\}| \cdot E \exp\{42^{1/2}\|\widehat{H(f)}\|_\infty\} + c_3 = \mathcal{O}(T). \end{aligned}$$

由于推导出的界对每个满足 $f \in \mathcal{F}_\beta$ 的 $\Sigma(f)$ 都成立，因此综合可得

$$\sup_{f \in \mathcal{F}_\beta} E_f[\|\widehat{\Omega} - \Sigma^{-1}(f)\|^2] = \mathcal{O}\{\log(np)/(nph))\} + \mathcal{O}(h^{2\beta}). \tag{A.21}$$

A.3.5 最优带宽参数 h

针对带宽参数 h 最小化 (A.20) 的右侧，得到 $h \asymp \{\log(np)/(np)\}^{\frac{1}{2\beta+1}}$。特别地，$h \to 0$。此外，

$$hT \geq cp^{1-\min\{\beta,1\}/3-(1+\min\{1,\beta\})/(2\beta+1)}\{\log(np)\}^{1/(2\beta+1)} \to \infty$$

由于根据假设，$n \lesssim p^{\min\{1,\beta\}}$ 且 $T = \lfloor p^v \rfloor$，其中 $v \in (1 - \min\{1, \beta\}/3, 1)$。因此，将 $h \asymp \{\log(np)/(np)\}^{1/(2\beta+1)}$ 代入 (A.20)，即可得到第二个结果。 □

B 定理 1 的辅助引理证明

本节给出证明定理 1 所需的三个技术引理。第一个引理列出核 K_h 及其在实数轴上的扩展 $\mathcal{K}_h$ 的一些性质。证明基于 Schwarz 和 Krivobokova（2016）以及（Schwarz，2012）。

引理 2. 设 $h > 0$ 为依赖于 N 的带宽参数。

(i) Th 存在常数 $0 < C < \infty$ 和 $0 < \gamma < 1$，使得对所有 $x, t \in [0, 1$]

$$\begin{aligned} &|\mathcal{K}_h(x,t)| < C\gamma^{|x-t|/h}, \\ &|K_h(x,t)| < C[\gamma^{|x-t|/h} + \gamma^{1/h}\{\gamma^{(x-t)/h} + \gamma^{(t-x)/h}\}/(1-\gamma^{1/h})]. \end{aligned}$$

(ii) 设 $t_i = (i-1)/(N-1)$ $i = 1, ..., N$。如果对 $N \to \infty$，有 $Nh \to \infty$，那么

$$\frac{1}{Nh}\sum_{i=1}^{N}\gamma_h(x,t_i) = \mathcal{O}(1), \quad \frac{1}{Nh}\sum_{i=1}^{N}\gamma_h^2(x,t_i) = \mathcal{O}(1)$$

对 $x \in [0,1]$ 一致成立，其中 $\gamma_h(x) = \gamma^{|x-t|/h} + \gamma^{1/h}\{\gamma^{(x-t)/h} + \gamma^{(t-x)/h}\}/(1-\gamma^{1/h})$。

(iii) 成立

$$h^{-1}\int_{-\infty}^{\infty}(x-t)^m\mathcal{K}_h(x,t)\,dt = \delta_{m,0}, \qquad for\ m = 0, ..., 2q-1,$$
$$h^{-1}\int_{-\infty}^{\infty}(x-t)^m\mathcal{K}_h(x,t)\,dt \neq 0, \qquad for\ m = 2q.$$

证明。(*i*) 第一条结论的证明见 Schwarz（2012）的引理 16。此外，对于 x, $t \in [0,1]$ 有

$$\begin{aligned}|K_h(x,t)| &\leq \sum_{l=-\infty}^{\infty}|\mathcal{K}_h(x,t+l)| \\ &\leq C\sum_{l=-\infty}^{\infty}\gamma^{|x-t-l|/h} = C\left\{\gamma^{|x-t|/h} + \sum_{l>0}^{\infty}\gamma^{(x-t+l)/h} + \sum_{l>0}^{\infty}\gamma^{(-x+t+l)/h}\right\} \\ &= C[\gamma^{|x-t|/h} + \gamma^{1/h}\{\gamma^{(x-t)/h} + \gamma^{(t-x)/h}\}/(1-\gamma^{1/h})].\end{aligned}$$

(*ii*) 由于 $\gamma_h(x,\cdot)$ 除了在 x 处外处处可微，因此有

$$\begin{aligned}&\frac{1}{Nh}\sum_{i=1}^{N}\gamma_h(x,t_i) = h^{-1}\int_0^1\gamma_h(x,t)\,\mathrm{d}t + \mathcal{O}\{(Nh)^{-1}\} \\ &= h^{-1}\int_0^1\gamma^{|x-t|/h}\,\mathrm{d}t + h^{-1}\frac{\gamma^{1/h}}{1-\gamma^{1/h}}\int_0^1\gamma^{(x-t)/h} + \gamma^{(t-x)/h}\,\mathrm{d}t + \mathcal{O}\{(Nh)^{-1}\}.\end{aligned}$$

注意到

$$h^{-1}\int_0^1\gamma^{|x-t|/h}\,\mathrm{d}t = \{\gamma^{(1-x)/h} + \gamma^{x/h} - 2\}/\log\gamma$$
$$\text{and } h^{-1}\int_0^1\gamma^{(x-t)/h} + \gamma^{(t-x)/h}\,\mathrm{d}x = \gamma^{-t/h}\{\gamma^{1/h}-1\}/\log\gamma - \gamma^{t/h}\{\gamma^{-1/h}-1\}/\log\gamma.$$

特别地，对于某些常数 $C_1, C_2 > 0$，它们取决于 $\gamma \in (0,1)$，但不取决于 h 和 x，有 $h^{-1}\int_0^1\gamma^{|x-t|/h}\,\mathrm{d}t \leq C_1$ 以及 $h^{-1}\gamma^{1/h}(1-\gamma^{1/h})^{-1}\int_0^1\gamma^{(x-t)/h} + \gamma^{(t-x)/h}\,\mathrm{d}x \leq$

C_2。证明 $h^{-1}\int_0^1 \gamma_h(x,t)^2\,\mathrm{d}t \le C_3$ 对某个常数 $C_3>0$ 成立的方法类似。根据假设 $(Nh)^{-1}=o(1)$ 对 $N\to\infty$ 成立，至此证明完成。

(iii) 参见 Schwarz（2012）的引理 15，其中 $p=2q-1$。 □

下面的引理指出，每个分箱中的相关伽马随机变量之和可以改写为独立伽马随机变量之和。

引理 3. 设 $\Sigma=\Sigma(f)$，其中 $f\in\mathcal{F}_\beta$ 且 $\beta>0$。那么，对于 n，$p\to\infty$ 满足 $p/n\to c\in(0,\infty]$，且对于 $p\to\infty$ 和常数 n，$Q_k=\sum_{j=}^{kp/T}{}_{(k-1)p/T+1}\sum_{i=1}^{n}\tilde{W}_{i,}$ $_j(k=1,...,T)$，其中

$$\tilde{W}_{i,j}\overset{indep.}{\sim}\Gamma\left\{1/2,\,2f(x_j)+\mathcal{O}\left(T^{-1}+T^{-1}p^{1-\beta}\log p\right)\right\}$$

对于 $i=1$，...，n 和 $j=(k-1)m+1$，...，km，以及 $x_j=(j-1)/(2p-2)$。

证明。由 Y_i的独立性，只需证明 $n=1$ 时的结论。然后，每个分箱中的点数为 $m=p/T$。为简单起见，以下省略下标 i。首先，我们将 Q_k 写成矩阵向量乘积，并对其重新分解，使其对应于独立缩放卡方随机变量之和。第二步，我们计算缩放因子。设 $E^{(km)}$ 为一个对角矩阵，其第（$k-1$）$m+1$、...、km个对角元素为 1，其余对角元素为 0。那么，

$$Q_k=\sum_{j=(k-1)m+1}^{km}W_j=Y^TDE^{(km)}DY.$$

定义 $X=\Sigma^{-1/2}Y\sim\mathcal{N}_p(0_p,I_p)$，其中 I_p是 $p\times p$单位矩阵。设 $\Sigma^{1/2}DE^{(jm)}D\Sigma^{1/2}$ 的特征值分解为 $U^T\Gamma U$，其特征值为 $(\lambda_j)_{j=1}^p$。矩阵$\Sigma^{1/2}$ 可逆，因为 Σ 可逆。于是，

$$Q_k=X^T\Sigma^{1/2}DE^{(km)}D\Sigma^{1/2}X=X^TU^T\Gamma UX=\tilde{X}^T\Gamma\tilde{X}=\sum_{j=1}^{p}\lambda_j\tilde{X}_j^2,$$

~ ~其中 $X=UX$。由于 U是一个正交矩阵，因此可得 $X\sim\mathcal{N}_p(0_p,I_p)$。特别是，$Q_k$ 是独立缩放卡方随机变量之和 $W_j=\lambda_j\tilde{X}_j^2\sim\Gamma(1/2,2\lambda_j)$，其中缩放因子为特征值 λ_j。接下来需要

计算 λ_j。为此，我们利用矩阵 $\Sigma^{1/2}DE^{(km)}D\Sigma^{1/2}$ 与该矩阵相似这一事实

$$A = D\Sigma^{-1/2}\left(\Sigma^{1/2}DE^{(km)}D\Sigma^{1/2}\right)\Sigma^{1/2}D = E^{(km)}D\Sigma D.$$

特别地， $\Sigma^{1/2}DE^{(km)}D\Sigma^{1/2}$ 与 A具有相同的特征值。 A 的行除第（$k-1$）$m+1$、...、第 km行外均为零，而这些行分别等于 $D\Sigma D$的第（$k-1$）$m+1$、...、第 km行。根据引理 1，

$$(D\Sigma D)_{i,j} = f(x_i)\delta_{i,j} + \frac{1+(-1)^{|i-j|}}{2}\mathcal{O}\left(p^{-1}+p^{-\beta}\log p\right).$$

定义 $B = (E^{(km)}D\Sigma DE^{(km)})_{r,\ s=(k-1)m+1}^{km}$。于是， $\lambda(A) = \lambda(B)\cup\{0\}$，且 A 至少有一个重数为 $p-m$ 的零特征值。由于 B是对称的，因此所有特征值均为实数。对 B 应用格尔施戈林圆盘定理可得，其余 m 个特征值 $\lambda_{(k-1)m+1}$、...、 λ_{km}均包含在以下集合中

$$\bigcup_{j=(k-1)m+1}^{km}\left\{z\in\mathbb{R}:\ |f(x_j)+\mathcal{O}\left(p^{-1}+p^{-\beta}\log p\right)-z| = \mathcal{O}\left(mp^{-1}+mp^{-\beta}\log p\right)\right\}.$$

由于 f是 $\min\{\beta,1\}$-H¨older 连续的，因此对于 $j=(k-1)m+1$、...、 km，有 $\lambda_j = f(x_j)+\mathcal{O}(mp^{-1}+mp^{-\beta}\log\ p)$。利用 $T^{-1}=mp^{-1}$即可完成证明。 □

最后，引理 4 给出了方差稳定变换的随机误差和确定性误差的显式界。

引理 4. 若 $\Sigma=\Sigma(f)$ 满足 $f\in\mathcal{F}_\beta$和 $\beta>0$，则 $Y_k^* = 2^{-1/2}\log(Q_k/m)$ 可写为

$$Y_k^* = H\left\{f\left(x_k\right)\right\}+\epsilon_k+m^{-1/2}Z_k+\xi_k, \qquad (k=1,...,T)$$

其中 $x_k=(k-1)/(2T-2)$ $|\epsilon_k|\lesssim(np)^{-1}+(np)^{-\beta}\log\ p$， $Z_k\sim\mathcal{N}(0,1)$，且 $E(\xi_k)=0$。此外，

$$E|\xi_k|^\ell \lesssim (\log m)^{2\ell}\left\{m^{-\ell}+\left(T^{-1}+T^{-1}p^{1-\beta}\log p\right)^\ell\right\} \qquad (\ell\in\mathbb{N}_{>1}),$$
$$cov(Z_k,Z_l) = \mathcal{O}\{mp^{-2}+mp^{-2\beta}(\log p)^2\} \qquad (k\neq l=1,...,T).$$

结果适用于 $n,p\to\infty$使得 $p/n\to c\in(0,\infty]$，并且适用于 $p\to\infty$，其中 n为常数。

证明。在 Cai 和 Zhou（2010）以及 Brown 等人（2010）的回归设定中，每个分箱内的伽马随机变量相互独立。利用引理 3，我们可以将当前设定改写为每个分箱内观测值相互独立的设定。因此，证明的第一步是应用 Cai 和 Zhou（2010）以及 Brown 等人（2010）关于将 Y_k^* 近似为高斯随机变量 $m^{-1/2}Z_k$ 和非高斯余项 ξ_k的结果。由于在我们的设定中，不同分箱的观测值 $i \neq j$ 相关，因此证明的第二部分将推导高斯随机变量之间的协方差。

根据引理 3，Q_k 可写为 $m = np/T$ 个相互独立的伽马随机变量之和，其均值函数为 $f(x) = f(x) + S(x)$，其中 $S(x) = \mathcal{O}(T^{-1} + T^{-1}p^{1-\beta}\log p)$。由引理 1 和引理 3 可知，$S$ 项关于 $x \in [0,1]$ 连续，且 $\mathcal{O}$ 项中存在一个与 x无关的全局常数~。此外，根据假设，$f \in [\delta, M_0] \subset \mathbb{R}_{>0}$。由于 f 是伽马随机变量的均值，因此 $f > 0$成立。由此，f 连续，且其取值位于伽马分布支撑集的紧子集 $[\epsilon, \nu] \subset \mathbb{R}_{>0}$ 中。根据 Cai 和 Zhou（2010）关于伽马分布的定理 1，可得

$$Y_k^* = H\left\{\tilde{f}(x_k^*)\right\} + m^{-1/2}Z_k + \xi_k \quad (k = 1, ..., T),$$

其中 $(k-1)/(2T-2) \leq x_k^* \leq k/(2T-2)$，$Z_k \sim \mathcal{N}(0,1)$，且 ξ_k是一个零均值随机变量。Z_k 通过分位数耦合定义。它是 Q_k 标准化版本的高斯近似，其中假设第 k个分箱中的 W_{ij} 独立同分布。特别地，Z_k 近似于

$$\bar{Q}_k = \frac{\sum_{j=(k-1)p/T+1}^{kp/T}\sum_{i=1}^{n}\bar{W}_{i,j} - m\tilde{f}(x_k^*)}{(2m)^{\frac{1}{2}}\tilde{f}(x_k^*)},$$

其中 $\bar{W}_{i,\ j} \overset{\text{独立同分布}}{\sim} \Gamma(1/2, 2\tilde{f}(x_k^*))$，并且满足 $\text{cov}(\bar{W}_{i,\ j}, \bar{W}_{i,\ h}) = \text{cov}(W_{i,\ j}, W_{i,\ h})$，对于$j = (k-1)p/T + 1$，...，$kp/T$以及 $h \in \{1, ..., p\} \setminus \{(k-1)p/T + 1, ..., kp/T\}$。

令 θ 为每个分箱中观测值均值的最大差异。那么，

$$\theta = \max_{k=1,...,T}\ \max_{j,l=(k-1)p/T+1,...,kp/T}\left|E[\tilde{W}_{1,j}] - E[\tilde{W}_{1,l}]\right| \leq c_1\left(T^{-1} + T^{-1}p^{1-\beta}\log p\right).$$

应用 Brown 等人（2010）的引理 4 和定理 1 得出，ξ_k 满足

$$E[|\xi_k|^\ell] \lesssim (\log m)^{2\ell}\{m^{-\ell} + (T^{-1} + T^{-1}p^{1-\beta}\log p)^\ell\}$$

对于任意常数整数 $\ell > 0$，且隐含常数依赖于 ℓ。接下来,

$$Y_k^* = H\left\{f\left(\frac{k-1}{2T-2}\right)\right\} + \epsilon_k + m^{-1/2}Z_k + \xi_k \quad (k = 1, ..., T),$$

其中 $\epsilon_k = H[f\{(k-1)/(2T-2)\}] - H\{\tilde{f}(x_k^*)\}$，且 $|\epsilon_k| \lesssim (np)^{-1} + (np)^{-\beta}\log p$, 这是由对数的 Lipschitz 连续性以及 f 的 α-H¨older 连续性得出的。

为计算 $k \neq l$ 的 cov(Z_k, Z_l)，我们使用 Hoeffding 协方差恒等式。具体而言, 我们证明 cov(Z_k, Z_l) . cov(Q_k, Q_l)。由此可得

$$0 \leq \operatorname{cov}(\bar{Q}_k, \bar{Q}_l) = \frac{m\mathcal{O}\{\operatorname{cov}(W_{i,j}, W_{i,h})\}}{2\tilde{f}(x_k^*)\tilde{f}(x_l^*)} = \mathcal{O}(mp^{-2} + mp^{-2\beta}\log p),$$

其中 $j = (k-1)p/T + 1$, ..., kp/T, 且 $h = (l-1)p/T + 1$, ..., lp/T。

令 Φ 和 $F_{\bar{Q}}$ 分别表示 Z_k、Z_l、Q_k 和 Q_l 的累积分布函数。由于 Z_k 是通过分位数耦合定义的，因此有 $Z_k = \Phi^{-1}\{F_{\bar{Q}}(\bar{Q}_k)\}$，参见 Koml´os 等（1975）；Mason 和 Zhou（2012）。此外，定义具有累积分布函数 F_U 的一致随机变量 $U_k = \Phi(Z_k) = F_{\bar{Q}}(Q_k)$ 和 $U_l = \Phi(Z_l) = F_{\bar{Q}}(Q_l)$。相应的联合累积分布函数记为 $F_{Z,\,Z}$, $F_{\bar{Q},\,\bar{Q}}$, $F_{U,\,U}$。根据 Hoeffding 协方差恒等式,

$$\begin{aligned}\operatorname{cov}(\bar{Q}_k, \bar{Q}_l) &= \int_{\mathbb{R}}\int_{\mathbb{R}} F_{\bar{Q},\bar{Q}}(q,r) - F_{\bar{Q}}(q)F_{\bar{Q}}(r)\,\mathrm{d}q\,\mathrm{d}r \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \{F_{Z,Z}(x,y) - \Phi(x)\Phi(y)\}\,\frac{\phi(x)}{f_{\bar{Q}}\{F_{\bar{Q}}^{-1}(x)\}}\frac{\phi(y)}{f_{\bar{Q}}\{F_{\bar{Q}}^{-1}(y)\}}\,\mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

令 $\rho = \text{cov}(Z_k, Z_l)$。则有恒等式

$$F_{ZZ}(x,y) = \frac{1}{2\pi}\int_0^\rho \frac{1}{(1-r^2)^{1/2}}\exp\left\{-\frac{x^2 - 2rxy + y^2}{2(1-r^2)}\right\}\mathrm{d}r + \Phi(x)\Phi(y)$$

意味着 $F_{Z,\,Z}(x,y) - \Phi(x)\Phi(y) \geq 0$ 对所有 $x, y \in \mathbb{R}$，当且仅当 $\rho \geq 0$ 时成立, 参见 Patel 和 Read（1996）的公式 9.3.1.3。由于 cov(Q_k, Q_l) ≥ 0，且两个

密度的比值非负，因此 $\mathrm{cov}(Z_k, Z_l) \geq 0$。此外，

$$\begin{aligned}
\mathrm{cov}(Z_l, Z_k) &= \int_{\mathbb{R}}\int_{\mathbb{R}} F_{Z,Z}(x,y) - \Phi(x)\Phi(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^1\int_0^1 \{F_{UU}(u,v) - F_U(u)F_U(v)\}\,\frac{1}{\phi\{\Phi^{-1}(u)\}}\frac{1}{\phi\{\Phi^{-1}(v)\}}\,\mathrm{d}u\,\mathrm{d}v \\
&\leq \int_{1/2}^1\int_{1/2}^1 \{F_{UU}(u,v) - F_U(u)F_U(v)\}\,\frac{1}{\phi\{\Phi^{-1}(u)\}}\frac{1}{\phi\{\Phi^{-1}(v)\}}\,\mathrm{d}u\,\mathrm{d}v \\
&\quad + \int_0^1\int_0^{1/2} \{F_{UU}(u,v) - F_U(u)F_U(v)\}\,\frac{2}{\phi\{\Phi^{-1}(u)\}}\frac{1}{\phi\{\Phi^{-1}(v)\}}\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned}$$

最后一个积分对应于 $E(Z_k Z_l^-) = \mathrm{cov}(Z_k, Z_l)/2$，其中$Z_l^- = Z_l \mathbb{I}\{Z_l \leq 0\}$。接下来需要证明，存在一个 $c > 0$ 使得 $f_{\bar{Q}}\{F_{\bar{Q}}^{-1}(u)\}/\phi\{\Phi^{-1}(u)\} \leq c$ 对所有 $u \in (1/2, 1)$ 均成立。于是可得

$$\begin{aligned}
\mathrm{cov}(Z_l, Z_k) &\leq 2c^2 \int_0^1\int_0^1 \{F_{UU}(u,v) - F_U(u)F_U(v)\}\,\frac{1}{f_{\bar{Q}}\{F_{\bar{Q}}^{-1}(u)\}}\frac{1}{f_{\bar{Q}}\{F_{\bar{Q}}^{-1}(v)\}}\,\mathrm{d}u\,\mathrm{d}v \\
&= 2c^2 \mathrm{cov}(\bar{Q}_k, \bar{Q}_l).
\end{aligned}$$

此外，对于 $u \in (1/2, 1)$，我们有

$$\frac{\partial}{\partial u}\frac{f_{\bar{Q}}\{F_{\bar{Q}}^{-1}(u)\}}{\phi\{\Phi^{-1}(u)\}} = \frac{f'_{\bar{Q}}\{F_{\bar{Q}}^{-1}(u)\}}{f_{\bar{Q}}\{F_{\bar{Q}}^{-1}(u)\}\phi\{\Phi^{-1}(u)\}} - \frac{\Phi^{-1}(u) f_{\bar{Q}}\{F_{\bar{Q}}^{-1}(u)\}}{\phi\{\Phi^{-1}(u)\}^2}. \tag{B.1}$$

由于对于 $u \in (1/2, 1)$，$\Phi^{-1}(u) > 0$，第二项为正。为了说明当 $u \in (1/2, 1)$ 时第一项为负，我们将 $f_{\bar{Q}}$ 和 $F_{\bar{Q}}^{-1}$ 改写为关于未归一化伽马随机变量 $X = \sum_{j=}^{kp/T}{}_{(k-1)p/T+1} \sum_{i=1}^n W_{i,j} \sim \Gamma(m/2, 2f(x_k^*))$ 的形式。因此，

$$f_{\bar{Q}}(x) = \sigma f_X(x\sigma + \mu), \qquad F_{\bar{Q}}^{-1}(u) = \sigma^{-1}\{F_X^{-1}(u) - \mu\},$$

其中，σ 和 μ 分别是 X的标准差和均值。由于 $A \sim \Gamma(a, b)$ 的众数位于 $b(a-1)$，且 $x\sigma + \mu = 2\tilde{f}(x_k)(m/2 - 1)$ 意味着 $x = -(2/m)^{1/2}$，因此 $f_{\bar{Q}}(x)$ 关于 $x \geq -(2/m)^{1/2}$单调递减。此外，由于 $f_{\bar{Q}}(x)$呈右偏分布，对所有 $m \in \mathbb{N}$ 都有 $F_{\bar{Q}}(-(m/2)^{1/2}) \leq 0.5$。特别地，对所有 $m \in \mathbb{N}$，$-(m/2)^{1/2} \leq F_{\bar{Q}}^{-1}(1/2)$。最后，由于 $f_{\bar{Q}}(-(2/m)^{1/2}) \to \phi(0)$，对于

$m \to \infty$ 存在一个常数 $c > 0$，它不依赖于 m，使得

$$\frac{f_{\bar{Q}}\{F_{\bar{Q}}^{-1}(u)\}}{\phi\{\Phi^{-1}(u)\}} \leq \frac{f_{\bar{Q}}\{-(2/m)^{1/2}\}}{\phi(0)} \leq c.$$

□

C 补充材料

C.1 伽马分布模拟研究

C.1.1 数据模拟

观测值服从示例1、2、3的中心化伽马分布，其协方差矩阵为Σ_1、 Σ_2、 Σ_3，即 $Y_i = \Sigma_j^{1/2} Z_i (j = 1, 3;\ i = 1, ..., n)$，其中 $Z_i = (z_{i,\ 1},\ ....,z_{i,\ p})^T$，且 $z_{i,\ 1} + 2^{1/2}$，...，$z_{i,\ p} + 2^{1/2}$ 独立同分布$\sim$ $\Gamma(2, 2^{-1/2})$。对于示例2，R 函数 `arima.sim` 使用参数 `innov` 来传入创新项 $z_{i,\ 1} + 2^{1/2}$，...，$z_{i,\ p} + 2^{1/2}$ 独立同分布$\sim$ $\Gamma(2, 2^{-1/2})$。所得过程乘以 $(1.2)^{1/2}$。

C.1.2 方差稳定变换函数的推导

DY_i 的协方差为 $\mathrm{cov}(DY_i) = D\Sigma D$，而 $D_j^T Y_i$ 的边缘分布服从方差为 $f(x_j) + o(1)$ 的中心化伽马分布。因此，$D_j^T Y_i + E(D_j^T Y_i) \sim \Gamma(2, \theta)$，其中 $\theta = [\{f(\pi x_j) + o(1)\}/2]^{1/2}$。回忆一下，$A \sim \Gamma(k, \theta)$ 的概率密度及其中心化版本 $B = A - E[A]$ 为

$$g_A(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta} \mathbb{I}\{x \geq 0\}$$

$$g_B(x) = \frac{1}{\Gamma(k)\theta^k} (x + k\theta)^{k-1} e^{-(x+k\theta)/\theta} \mathbb{I}\{x \geq -k\theta\}.$$

由此可得，B^2 的概率密度为

$$\begin{aligned} g_{B^2}(x) &= \frac{g_B(x^{1/2}) + g_B(-x^{1/2})}{2x^{1/2}} \mathbb{I}\{x \geq 0\} \\ &= \frac{1}{\Gamma(k)\theta^k} \frac{(x^{1/2} + k\theta)^{k-1} e^{-(x^{1/2}+k\theta)/\theta}}{2x^{1/2}} \mathbb{I}\{x \geq 0\} \\ &\quad + \frac{1}{\Gamma(k)\theta^k} \frac{(-x^{1/2} + k\theta)^{k-1} e^{-(-x^{1/2}+k\theta)/\theta}}{2x^{1/2}} \mathbb{I}\{-(k\theta) < -x^{1/2} < 0\} \end{aligned}$$

利用该概率密度可得 $E(B^2) = k\theta^2 =$，且 $\mathrm{var}(B^2) = c_k(k\theta)^2$，其中 c_k是依赖于 k的某个显式常数。因此，$E(W_{i,\ j}) = E\{(D_j^T Y_i)^2\} = f(\pi x_j) + o(1)$

且 $\mathrm{var}\{(D_j^T Y_i)^2\} = 5\{f(\pi x_j) + o(1)\}^2$。相应的方差稳定化函数 $G: \mathbb{R} \to \mathbb{R}$ 满足 $G'(\mu) = V^{-1/2}(\mu)$，其中 $V(\mu)$ 表示方差关于均值 μ的函数。特别地，对于 $W_{i,\ j}$ 的分布，可得 $G(\mu) = 5^{-1/2}\log(\mu)$，其中 $\mu = f(x_j) + o(1)$。图 5 中的 QQ 图证实，对数变换确实是方差稳定化函数。三个示例中的变换后的数据Y_k^*，($k = 1, ..., T$) 看起来都与具有齐次方差的正态分布随机变量非常相似。

C.1.3 结果

由于 $W_{i,j}$ 的方差稳定化函数是缩放对数，我们的方法可用于根据（中心化的）伽马分布数据 $Y_1, ..., Y_n$ 估计 Σ。锥化估计量和我们的方法所用的调优参数与高斯数据情形相同，见第6节。分箱数 $T = 500$ 根据 QQ图选定，另见下一节的图6。表4–6分别展示了（A） $p = 5000, n = 1$、（B） $p = 1000, n = 50$ 和（C） $p = 5000, n = 10$的结果。

表4：(A) $p = 5000, n = 1$：托普利茨协方差矩阵和谱密度估计量在谱范数和 L_2 范数下的误差；最后一列为每个蒙特卡洛样本的协方差估计量平均计算时间（秒）；除最后一列外，所有数值均乘以100

	过程（1）		过程（2）		过程（3）		time
	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	秒内
我们的方法（广义交叉验证）	1.181	0.434	2.285	0.609	5.182	0.884	4.238
我们的方法（最大似然）	0.973	0.391	2.121	0.581	4.787	0.819	4.241
截断（交叉验证）	0.951	0.372	2.626	0.803	4.156	1.151	4.682
样本协方差	16750.719	3380.642	20195.498	3631.043	24828.431	3597.585	0.350
我们的方法（广义交叉验证预言机）	0.794	0.346	1.701	0.516	4.172	0.757	
我们的方法（最大似然预言机）	0.820	0.356	1.965	0.550	4.789	0.812	
渐消（预言机）	0.759	0.335	1.313	0.443	1.991	0.539	

表5：(B) $p = 1000, n = 50$：托普利茨协方差矩阵和谱密度估计量在谱范数和 L_2 范数下的误差；最后一列为每个蒙特卡洛样本的协方差估计量平均计算时间（秒）；除最后一列外，所有数值均乘以100

	过程（1）		过程（2）		过程（3）		time
	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	以秒计
我们的方法（广义交叉验证）	0.125	0.042	0.457	0.142	0.590	0.095	23.169
我们的方法（最大似然）	0.106	0.038	0.525	0.151	0.736	0.106	23.093
截断（交叉验证）	0.122	0.040	0.497	0.154	0.410	0.087	24.444
样本协方差	80.845	57.839	103.125	61.602	126.620	60.425	0.155
我们的方法（广义交叉验证预言机）	0.087	0.033	0.390	0.134	0.552	0.090	
我们的方法（最大似然预言机）	0.095	0.035	0.515	0.150	0.731	0.105	
渐消（预言机）	0.074	0.032	0.371	0.137	0.307	0.068	

表6：(C) $p = 5000, n = 10$：托普利茨协方差矩阵和谱密度估计量在谱范数和 L_2 范数下的误差；最后一列为每个蒙特卡洛样本的协方差估计量平均计算时间（秒）；除最后一列外，所有数值均乘以100

	过程（1）		过程（2）		过程（3）		time
	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	$\|\widehat{\Sigma}-\Sigma\|^2$	$\|\hat{f}-f\|_2^2$	秒
我们的方法（广义交叉验证）	0.138	0.042	0.386	0.129	0.626	0.099	4.347
我们的方法（最大似然）	0.109	0.037	0.441	0.137	0.700	0.105	4.352
截断（交叉验证）	0.175	0.047	0.485	0.142	0.508	0.092	680.471
样本协方差	651.720	352.037	846.298	380.690	1061033.779	381.164	1.202
我们的方法（广义交叉验证预言机）	0.091	0.033	0.330	0.121	0.557	0.091	
我们的方法（最大似然预言机）	0.097	0.034	0.434	0.136	0.697	0.105	
渐消（预言机）	0.076	0.031	0.307	0.120	0.305	0.068	

C.2 均匀分布模拟研究

C.2.1 数据模拟

观测值按如下方式生成：设定 $Y_i = \Sigma_j^{1/2} Z_i (i = 1, ..., n;\ j = 1, 3)$，其中 Σ_j 是示例 j 的托普利茨协方差矩阵，而 $Z_i = (z_{i,\ 1}, ..., z_{i,\ p})^T$ 的 $z_{i,\ 1}, ..., z_{i,\ p}$ 均服从 $\sim$ 均匀分布 $[-3^{1/2}, 3^{1/2}]$ 独立同分布。对于示例2，使用 R 函数 `arima.sim` 的参数 `innov` 来传入创新项 $X_1, ..., X_n$ 独立同分布 $\sim$ 均匀分布 $[-3^{1/2}, 3^{1/2}]$。所得过程再乘以 $(1.2)^{1/2}$。

C.2.2 方差稳定变换函数

对于这个例子，无法显式推导方差稳定变换函数。图6中的QQ图表明，对数变换具有方差稳定化作用

对于 $(D_j^t Y_i)^2$，所有三个例子中的变换后的数据 $Y_k^*(k = 1, ..., T)$ with$T = 500$ 看起来都与具有齐性方差的正态分布随机变量非常相似。因此，我们的方法可用于根据数据$Y_1, ..., Y_n$估计 Σ 。

C.2.3 结果

表7–9分别展示了（A） $p = 5000, n = 1$、（B） $p = 1000, n = 50$ 和（C） $p = 5000, n = 10$的结果。锥化估计量和我们的方法所用的调优参数与高斯数据情形相同，见第6节。

表7：（A） $p = 5000, n = 1$：托普利茨协方差矩阵和谱密度估计量在谱范数和 L_2 范数下的误差；最后一列为每个蒙特卡洛样本的协方差估计量平均计算时间（秒）；除最后一列外，所有数值均乘以100

	过程（1）		过程（2）		过程（3）		time
	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	以秒计
我们的方法（广义交叉验证）	0.737	0.230	1.952	0.488	4.010	0.616	4.282
我们的方法（最大似然）	0.572	0.190	1.560	0.422	3.991	0.610	4.279
截断（交叉验证）	0.490	0.163	2.282	0.703	2.843	0.769	4.873
样本协方差	17131.995	3599.128	19077.308	4031.457	26960.006	3770.437	0.355
我们的方法（广义交叉验证预言机）	0.392	0.139	1.271	0.367	3.280	0.543	
我们的方法（最大似然预言机）	0.417	0.147	1.461	0.398	3.903	0.599	
渐消（预言机）	0.371	0.130	1.022	0.314	1.463	0.342	

表 8：(B) $p = 1000, n = 50$：托普利茨协方差矩阵和谱密度估计量相对于谱范数和 L_2 范数的误差；最后一列为每个蒙特卡洛样本的协方差估计量平均计算时间（秒）；除最后一列外，所有数值均乘以 100

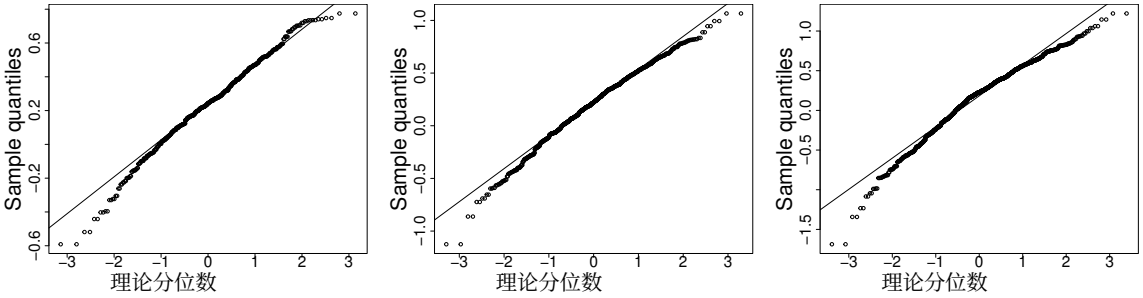
	过程（1）		过程（2）		过程（3）		time
	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	在秒
我们的方法（广义交叉验证）	0.090	0.024	0.397	0.116	0.682	0.092	31.574
我们的方法（最大似然）	0.079	0.021	0.452	0.123	0.766	0.096	31.473
截断（交叉验证）	0.134	0.030	0.424	0.126	0.423	0.074	23.444
样本协方差	80.230	56.741	102.424	61.282	123.505	58.847	0.136
我们的方法（广义交叉验证预言机）	0.057	0.016	0.332	0.106	0.567	0.078	
我们的方法（最大似然预言机）	0.066	0.018	0.443	0.122	0.757	0.094	
渐消（预言机）	0.047	0.015	0.318	0.110	0.260	0.052	

表 9：(C) $p = 5000, n = 10$：托普利茨协方差矩阵和谱密度估计量相对于谱范数和 L_2 范数的误差；最后一列为每个蒙特卡洛样本的协方差估计量平均计算时间（秒）；除最后一列外，所有数值均乘以 100

	过程（1）		过程（2）		过程（3）		time
	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	$\|\widehat{\Sigma} - \Sigma\|^2$	$\|\hat{f} - f\|_2^2$	秒
我们的方法（广义交叉验证）	0.090	0.023	0.353	0.112	0.568	0.079	4.292
我们的方法（最大似然）	0.074	0.019	0.412	0.119	0.635	0.085	4.285
截断（交叉验证）	0.127	0.025	0.437	0.124	0.427	0.066	638.830
样本协方差	662.955	365.878	824.441	379.398	1049580.980	386.305	1.205
我们的方法（广义交叉验证预言机）	0.060	0.016	0.314	0.105	0.478	0.070	
我们的方法（最大似然预言机）	0.066	0.017	0.408	0.118	0.626	0.085	
渐消（预言机）	0.043	0.013	0.299	0.105	0.230	0.046	

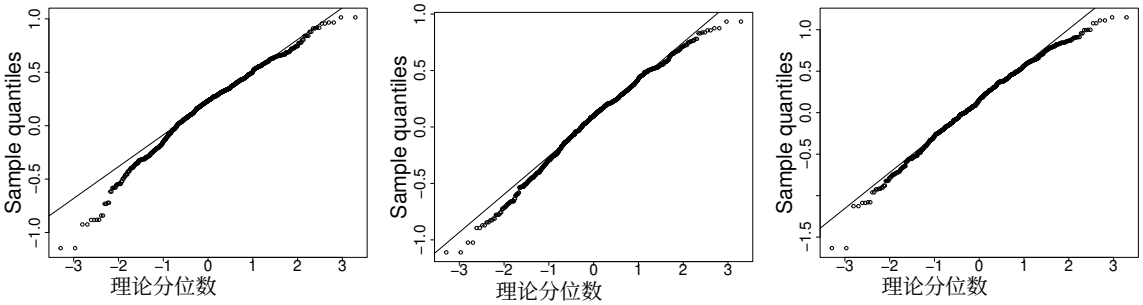
C.3 QQ图

本节展示模拟数据中变换后的数据 $Y_k^*(k = 1, ..., T)$ 的 QQ图。图中展示了情景（A） $p = 5000, n = 1$ 的 QQ图，因为其数据点少于情景（B）和（C），因此变换后的数据 Y_k^* 的高斯近似误差最大。图4展示了分箱数 T 对例（1）中变换后的数据 Y_k^* 的影响，其中使用的是高斯数据Y（参见主文第 6 节的模拟研究）。这些图验证了在模拟研究中选择 $T = 500$ 的合理性，因为在这种情况下， Y_k^* 与高斯分布随机变量非常接近。图5和图6展示了分箱数为 $T = 500$ 时，例 1–3 中 Y_k^* 的 QQ图；其中原始数据 Y 分别服从伽马分布或均匀分布。这些图与高斯数据 Y 的图4中的 QQ图相似，表明我们的方法也适用于这些非高斯数据的托普利兹协方差估计。



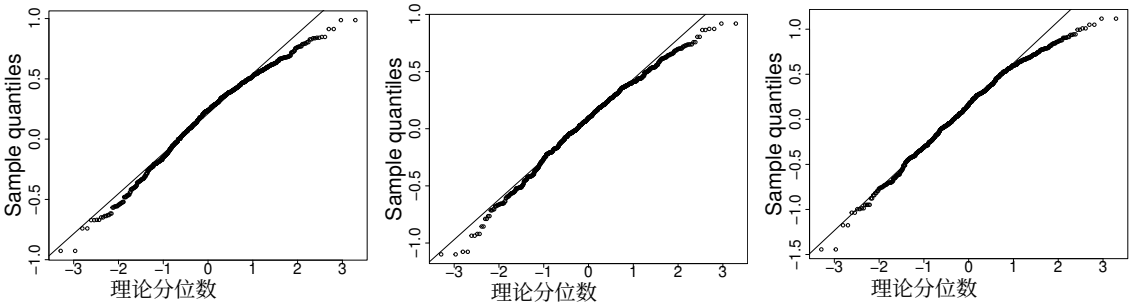
示例（1），$T = 300$　　示例 (1)，$T = 500$　　示例 (1)，$T = 700$

图4：高斯数据的模拟研究；分箱并变换后的数据的QQ图 $Y_k^*(k = 1, ..., 2T - 2)$，使用不同的分箱数 T例如 (1)，其中$p = 5000, n = 1$。



示例 (1)，$T = 500$　　示例 (2)，$T = 500$　　示例 (3)，$T = 500$

图5：伽马数据的模拟研究；示例 (1)–(3) 中分箱并变换后的数据的 QQ图 $Y_k^*(k = 1, ..., 2T - 2)$，其中 $T = 500$ $p = 5000, n = 1$。



示例 (1)，$T = 500$　　示例 (2)，$T = 500$　　示例 (3)，$T = 500$

图6：均匀数据的模拟研究；示例 (1)–(3) 中分箱并变换后的数据 $Y_k^*(k = 1, ..., 2T - 2)$ 的 QQ图，其中 $T = 500$ for $p = 5000, n = 1$。

参考文献

Abramowitz, M. 和 Stegun, I. A.（1964年）。《数学函数手册：公式、图表与数学表》，第55卷。美国政府印刷局。Azuma, K.（1967）。某些相关随机变量的加权和。东北数学杂志（第二辑），19(3)：357–367。Bartlett, M. S.（1950）。周期图分析与连续谱。Biometrika，37(1/2)：1–16。Bentkus, R.（1985年）。可微函数空间中谱密度统计估计量的一致收敛速率。立陶宛数学杂志，25(3)：209–219。Bickel, P. J. 和 Levina, E.（2008）。大协方差矩阵的正则化估计。《统计年刊》，36(1)：199–227。Brown, L. D., Cai, T. T., Zhou, H. H. 等（2010 年）。指数族中的非参数回归。《统计年刊》，38(4)：2005–2046。Cai, T. T., Ren, Z. 和 Zhou, H. H.（2013 年）。托普利茨协方差矩阵估计的最优收敛速率。《概率论及相关领域》，156(1-2)：101–143。Cai, T. T. 和 Zhou, H. H.（2010 年）。自然指数族中的非参数回归。载于《借力：驱动应用的理论——劳伦斯·D·布朗纪念文集》，第199–215页。数理统计学会。Chagny, G.（2016）。非参数自适应估计导论。《数学研究生学刊》，1(2)：105–120。Choudhuri, N., Ghosal, S. 和 Roy, A.（2004）。时间序列谱密度的贝叶斯估计。《美国统计学会会刊》，99(468)：1050–1059。Du, X., Aubry, A., De Maio, A. 和 Cui, G.（2020）。面向雷达应用的托普利茨结构协方差矩阵估计。IEEE信号处理快报，27：595–599。Edwards, M. C., Meyer, R. 和 Christensen, N.（2019）。使用B样条先验的贝叶斯非参数谱密度估计。统计与计算，

29(1)：67–78。Fang, Y., Wang, B. 和 Feng, Y.（2016）。大协方差矩阵正则化估计中的调参选择。统计计算与模拟杂志，86(3)：494–509。Franaszczuk, P. J., Blinowska, K. J. 和 Kowalczyk, M.（1985年）。参数化多通道谱估计在脑电活动研究中的应用。生物控制论，51(4)：239–247。Grenander, U. 和 Szeg¨o, G.（1958）。托普利兹型及其应用。加利福尼亚大学出版社。Guerrero, V. M.（1993）。幂变换辅助的时间序列分析。预测杂志，12(1)：37–48。Hastie, T. 和 Tibshirani, R.（1990）。广义加性模型。Wiley在线图书馆。Koml´os, J., Major, P. 和 Tusn´ady, G.（1975）。独立随机变量部分和的近似及样本分布函数。I.《概率论与相关领域》，32：111–131。Kooperberg, C., Stone, C. J. 和 Truong, Y. K.（1995）。对数样条谱密度估计的收敛速率。时间序列分析杂志，16(4)：389–401。Krivobokova, T.（2013 年）。惩罚样条两种框架中的平滑参数选择。《英国皇家统计学会会刊B辑：统计方法论》，75(4)：725–741。Krivobokova, T., Briones, R., Hub, J. S., Munk, A. 和 de Groot, B. L.（2012）。偏最小二乘函数众数分析：在膜蛋白水通道蛋白1、水通道蛋白Y1和ClC-ec1中的应用。《生物物理学杂志》，103(4)：786–796。Mason, D. M. 和 Zhou, H. H.（2012）。分位数耦合不等式及其应用。Maturana-Russel, P. 和 Meyer, R.（2021）。采用基于分位数的节点放置的P样条贝叶斯谱密度估计。《计算统计学》，36(3)：2055–2077。McMurry, T. L. 和 Politis, D. N.（2010）。用于自

托普利茨协方差矩阵与线性过程自助法。《时间序列分析杂志》，31(6):471–482。Patel, J. K. 和 Read, C. B. (1996)。正态分布手册，第150卷。CRC出版社。Pawitan, Y. 和 O'Sullivan, F. (1994)。使用惩罚Whittle似然进行非参数谱密度估计。《美国统计学会会刊》，89(426):600–610。Pourahmadi, M. (2013)。高维协方差估计：针对高维数据，第882卷。约翰·威立父子出版公司。Quah, D. (2000)。互联网集群的兴起。《欧洲经济评论》，44(4-6):1032–1044。Roberts, W. J. 和 Ephraim, Y. (2000)。使用托普利茨协方差矩阵进行语音隐马尔可夫建模。《语音通信》，31(1):1–14。Schoenberg, I. (1973)。基数样条插值。费城：SIAM。Schwarz, K. (2012)。样条估计量的统一框架。博士论文，G¨ öttingen大学。Schwarz, K. 和 Krivobokova, T. (2016)。样条估计量的统一框架。Biometrika，103(1):121–131。Serra, P.、Krivobokova, T. 等 (2017)。自适应经验贝叶斯平滑样条。《贝叶斯分析》，12(1):219–238。Singer, M.、Krivobokova, T.、Munk, A. 和 De Groot, B. (2016)。相关数据的偏最小二乘。Biometrika，103(2):351–362。Taqqu, M. S. 和 Teverovsky, V. (1997)。长程依赖时间序列的Whittle型估计量的稳健性。《统计学通讯：随机模型》，13(4):723–757。Thomson, D. J. (1982)。谱估计与谐波分析。《IEEE会刊》，70(9):1055–1096。Tsybakov, A. B. (2009)。非参数估计导论。施普林格统计学丛书。施普林格。Vershynin, R. (2018)。高维概率：数据科学中的导论与应用，第47卷。剑桥大学出版社。

Wahba, G. (1980)。对数周期图的自动平滑。《美国统计学会会刊》，75(369):122–132。Wahba, G. (1985)。在广义样条平滑问题中，使用GCV和GML选择平滑参数的比较。《统计年刊》，13(4)。Walker, A. (1964)。平稳非确定性时间序列谱参数最小二乘估计的渐近性质。《澳大利亚数学会杂志》，4(3):363–384。Welch, P. (1967)。快速傅里叶变换在功率谱估计中的应用：一种基于对短时修正周期图进行时间平均的方法。IEEE 音频与电声学汇刊，15(2):70–73。Whittle, P. (1957年)。曲线与周期图平滑。《英国皇家统计学会杂志：B辑（方法论）》，19(1):38–47。Wu, W. B. 和 Pourahmadi, M. (2009)。平稳过程样本自协方差矩阵的带状化。《中华统计学志》，页码1755–1768。Wu, W. B. 和 Xiao, H. (2012)。时间序列中的协方差矩阵估计。载于《统计学手册》，第30卷，第187–209页。爱思唯尔。Zygmund, A. (2002)。三角级数，第1卷。剑桥大学出版社。