arXiv:2303.04447v1 [stat.ME] 8 Mar 2023

时间序列条件极值

Graeme Auld [†*] 和**Ioannis Papastathopoulos**[††]

[†]朱拉隆功大学数学与计算机科学系，泰国曼谷
[††]爱丁堡大学数学学院及麦克斯韦研究所，英国爱丁堡 GraemeRoss.A@chula.ac.th
i.papastathopoulos@ed.ac.uk

摘要

准确刻画平稳时间序列中的联合极值依赖结构是一项具有挑战性的问题，在许多应用中都至关重要。此前针对这一问题的若干方法仅适用于时间序列中的某些类型极值依赖，例如渐近依赖，或有限阶马尔可夫时间序列。本文基于近期的概率结果，开发了用于时间序列极值的统计方法，使我们能够灵活地刻画平稳时间序列在经历极端事件后的衰减过程。有限阶马尔可夫序列自然适用于我们的方法；此外，我们还考虑了更广泛的设定，即基于条件极值模型，从而能够涵盖时间序列中多种可能的依赖结构。我们研究了基于蒙特卡洛模拟的推断方法，并推导出一种常用重要性抽样器的方差上界。我们通过对模拟数据中的簇泛函进行估计，以及对法国奥尔良的日最高气温时间序列进行分析，展示了该方法的应用。

关键词：条件　　极值、渐近独立性、时间序列极值、马尔可夫链

1 引言

自然界中的许多极端事件，其影响源于极值簇的出现，即在短时间内观测到多个极大值或极小值。洪水可能是连续数日强降雨的结果，而热浪则可能是连续数日高温的结果。这类事件可能导致重要基础设施受损、引发巨额保险理赔，并造成人员死亡。例如，由于2003年夏季毁灭性的欧洲热浪，据估计有40,000至70,000人死于高温相关原因（Fischer & Sch¨ar 2010，Robine 等人 2008），相关经济损失超过 €13 十亿（de Bono 等人 2004）。要开发能够准确再现自然过程极值聚集行为的统计模型，就必须理解该过程的极值依赖结构。

理解时间序列中极值依赖如何随滞后变化的一种方法，是使用尾部依赖函数 χ (Ledford & Tawn2003)。如果 $\{X_n\}_{n\in\mathbb{Z}}$ 是一个边缘分布函数为 F 的平稳时间序列，那么 χ 定义为

$$\chi_i = \lim_{x\to 1} \mathbb{P}(F(X_i) > x \mid F(X_0) > x), \quad i \in \mathbb{N}, \tag{1.1}$$

前提是该极限存在。当 $\chi_i = 0$时，我们称 X_0 和 X_i 渐近独立（AI）（Sibuya 1960），此时 X_0 和 X_i 不可能同时取到各自的最大值。若 $\chi_i = 0$ 对所有$i \geq 1$均成立，则称 $\{X_n\}_{n\in\mathbb{Z}}$ 为渐近独立时间序列。尽管渐近独立时间序列的极值会单独出现

*通讯作者

在极限下，但在中等极端水平上仍可能存在很强的依赖性；这一事实无法通过渐近测度 χ捕捉。当 $\chi_i > 0$ 对某个 $i \geq 1$成立时，我们称该时间序列渐近相依。此前的一些时间序列极值模型，例如Smith 等人（1997）、Perfekt（1997）以及Basrak 和 Segers（2009），仅适用于渐近相依的情形。正如 Winter 和 Tawn（2016年）在热浪建模背景下所展示的那样，如果错误地使用渐近相依过程来建模一个呈现渐近独立的过程，那么可能会严重高估与多个变量同时取较大值有关事件的概率。由于环境时间序列通常呈现渐近独立，因此在对此类过程建模时，指定能够适应该极值相关类别的模型非常重要。

一种强大的建模框架可用于刻画各种类型的极值依赖，即Heffernan & Tawn (2004) 提出的条件极值模型。为简洁起见，我们现在仅讨论二元情形。如果 X_0 和 X_1 是指数尾随机变量，Heffernan & Tawn (2004) 表明，对于（X_0, X_1）上的一大类依赖结构，存在位置和尺度标准化函数a_1： $\mathbb{R} \to \mathbb{R}$ 和 b_1： $\mathbb{R} \to \mathbb{R}_+$ 使得

$$\left(X_0 - u, \frac{X_1 - a_1(X_0)}{b_1(X_0)} \right) \Bigg| X_0 > u \xrightarrow{D} (E_0, Z^0), \quad u \to \infty, \tag{1.2}$$

其中， E_0 是与 Z^0独立的单位指数随机变量，且 $Z^0 \sim G^0$，其中 G^0 是一个非退化分布函数。此外，标准化函数 a_1 和 b_1 可确定为属于简单的参数族。特别地，当 X_0 和 X_1 是单位拉普拉斯随机变量时，则$a_1(x) = \alpha\, x$ 且 $b_1(x) = x^\beta$，其中 $\alpha \in [-1, 1]$ 和 $\beta \in (-\infty, 1)$（Keef 等人 2013）。 $\alpha < 1$和 $\alpha = 1$ 的情形分别对应于 X_0 和 X_1 渐近独立与渐近相依。

Winter & Tawn (2016年) 使用条件极值方法对日气温时间序列中的极值进行建模，并估计与热浪相关的极端事件发生概率。通过作出一阶马尔可夫假设，并假设 (X_t, X_{t+1}) 的依赖结构属于 Heffernan & Tawn (2004) 确定的类别，他们提出了一个模型，用于描述在大阈值 X_{t+1} 条件下的 $X_t > u$ u，形式如下：

$$X_{t+1} = \alpha\, X_t + X_t^\beta Z_{t+1}, \quad X_t > u, \tag{1.3}$$

其中待估参数为 α 和 β。模型 (1.3) 中的随机变量 Z_{t+1} 对应于收敛关系 (1.2) 中的极限随机变量 Z^0，其分布通过非参数方法估计。通过使用指数尾假设模拟一个初始超阈值 X_t，即超越阈值 u，并根据模型 (1.3) 向前迭代，他们得以模拟日气温时间序列进入极端状态后的行为。应用递推关系 (1.3) 时，残差 Z_{t+1}独立于初始超阈值 X_t进行模拟，而 Z_{t+i} 则独立于 Z_{t+j} 进行模拟，其中$i \neq j$。 Winter & Tawn (2017) 将这一模型推广至高阶马尔可夫时间序列。

本文还考虑了基于Heffernan & Tawn (2004)条件极值方法的时间序列模型。该方法能够模拟在超过大阈值后，渐近独立和渐近相依的平稳时间序列的行为。% 我们的框架比.....所采用的框架更广。我们考虑在观测到超过大阈值的超越后，对一大段连续观测值进行建模；也就是说，我们考虑在条件X_{t+1}, $X_{t+2}, \ldots, X_{t+k}$下的模型，其中 $X_t > u$ 对于某个大阈值 u，其中 k 是一个可能相当大的正整数。常数 k 的具体取值将视具体情境而定；在实际应用中，我们会选择足够大的值，使得 $k+1$ 个观测值（$X_t, X_{t+1}, \ldots, X_{t+k}$）涵盖极端事件的整个持续时间。例如，在第 6 节的数据应用中，我们考察了日温度时间序列在三周期间的行为，并以该期间开始时的高温为条件；此处我们选择 $k = 20$。采用这种方法的动机是，通过拟合输入数据由涵盖极端事件整个持续时间的一大段观测值构成的模型，我们希望能够

准确地模拟平稳时间序列进入极端状态后的后续行为，适用于广泛的依赖结构。

我们假设（X_t、 X_{t+1}、……、 X_{t+k}）的 Copula 属于由 Heffernan & Tawn （2004）确定的类别。这促使我们针对在 $X_t > u$给定条件下、以适当选择的大阈值 u 为条件的 X_{t+i} 条件分布建立如下

$$X_{t+i} = \alpha_i X_t + X_t^{\beta_i} Z_{t+i}, \quad 1 \le i \le k. \tag{1.4}$$

Eastoe & Tawn (2012) 也研究了这一模型框架，其与 Winter & Tawn (2017) 的模型类似；在后者中，k 被视为假定的马尔可夫过程的阶数。我们方法的新颖之处在于，考虑了对（α_1、 α_2、……、 α_k）和（β_1、 β_2、……、 β_k）施加结构约束的模型，而在 Eastoe & Tawn (2012) 和 Winter & Tawn (2017) 的模型中，当 $i \neq j$ 时， α_i 和 β_i 与 α_j 和β_j 完全相互独立。我们方法的另一项创新是，还可以在统计模型（1.4）中针对残差向量 $\boldsymbol{Z} =$（Z_{t+1}、……、 Z_{t+k}）考虑完全参数化的模型。

我们的方法受到Wadsworth 和 Tawn （2022）的研究思路启发。在该研究中，作者在空间情形下，考虑了在参考位置出现较大观测值这一条件下，对整个空间场进行模拟。我们沿用了他们对残差过程的形状参数施加结构约束的创新做法，同时也对残差的位置参数和尺度参数施加结构约束。在较简单的时间序列情境下开展研究的一个优势是：对于一大类平稳马尔可夫时间序列，近期已有重要的概率结果，使我们能够对模型参数施加具有理论依据的结构约束（Papastathopoulos 等人 2017，Papastathopoulos 等人 2023）。尽管我们考虑的示例侧重于渐近独立的马尔可夫时间序列，但该方法也可推广至既非马尔可夫、也非渐近独立的更一般过程。

本文结构如下。第2 节介绍全文将用到的符号和主要假设。第3 节探讨统计建模和统计推断方法。第4节介绍可与第3 节中的模型配合使用的模拟方法，以估计极端事件发生的概率；我们还证明了 Wadsworth 和 Tawn （2022）的5 节考察渐近独立的马尔可夫时间序列示例，并通过模拟研究评估第 3 节中模型的表现。第6节利用法国奥尔良的日最高气温时间序列展示我们的方法。

2 符号和假设

如果 i和 j是满足 $i \le j$的两个整数，我们将集合 $i{:}j$定义为从 i到 j（含首尾）的所有整数，即 $i{:}j = \{n \in \mathbb{Z}\colon i \le n \le j\}$。此外，连续随机变量区块 $\{X_n\colon i \le n \le j\}$ 记作 $\boldsymbol{X}_{i:j}$。更一般地，如果 A是一个正整数集合，则随机变量集合$\{X_i : i \in A\}$ 记作 $\boldsymbol{X}_A$。下文中，所有算术运算均为向量化运算，因此涉及两个向量的此类运算应按逐分量方式理解。此外，涉及不同长度向量的运算通过重复较短向量来匹配较长向量的长度。例如，如果 $\alpha \ge 0$和 $k \in \mathbb{N}$是常数，且 $\boldsymbol{y} = (y_1, \ldots, y_d) \in \mathbb{R}^d$，则 $\alpha^{1:k} = (\alpha, \alpha^2, \ldots, \alpha^k)$ 和 $\alpha + \boldsymbol{y} = (\alpha + y_1, \ldots, \alpha + y_d)$。

我们用符号 $\sim$ 表示“服从……分布”，其中 $\sim$ 右侧的符号可以是分布或密度函数。若 $\boldsymbol{X}$ 是一个随机向量，且 g 是一个可测函数，则诸如 $\mathbb{E}_\pi\{g(\boldsymbol{X})\}$ 这样的表达式表示当 $\boldsymbol{X} \sim \pi$时 $g(\boldsymbol{X})$的期望值。

除非另有说明， $\{X_n\}_{n\in\mathbb{Z}}$ 表示边际分布为拉普拉斯分布的平稳时间序列，因此

$$\mathbb{P}(X_n \le x) = \begin{cases} \exp(x)/2, & x \le 0, \\ 1 - \exp(-x)/2, & x > 0 \end{cases} \tag{2.1}$$

对于每个 $n \in \mathbb{Z}$ 和 $x \in \mathbb{R}$。假设边缘分布已给定并不失一般性，因为任何具有连续边缘分布的平稳时间序列，都可以通过概率积分变换转换为具有不同连续边缘分布的另一个平稳时间

通过概率积分变换处理具有不同连续边缘分布的序列。

We d将阈值 $u > 0$ 的超越时刻随机集定义为 se t

$$T_u = \{i \in \mathbb{N} : X_i > u\}. \tag{2.2}$$

我们的主要假设涉及在时间序列 $\{X_n\}_{n\in\mathbb{Z}}$ 的任意位置观测到极端状态（记为 X_0 ）时，该时间序列的极限行为。具体而言，我们假设对于每个有限集合 $A \subset \mathbb{Z} \setminus \{0\}$，都存在位置和尺度函数 $\boldsymbol{a}_A = \{a_i : \mathbb{R} \to \mathbb{R}\}_{i\in A}$ 和 $\boldsymbol{b}_A = \{b_i : \mathbb{R} \to \mathbb{R}_+\}_{i\in\mathbb{A}}$，使得

$$\left(X_0 - u, \frac{\boldsymbol{X}_A - \boldsymbol{a}_A(X_0)}{\boldsymbol{b}_A(X_0)}\right) \Bigg| X_0 > u \xrightarrow{D} (E_0, \boldsymbol{Z}_A) \tag{2.3}$$

当 $u \to \infty$，其中 E_0 是一个与随机向量 $\boldsymbol{Z}_A \sim G_A$ 相互独立的单位指数随机变量，其中 G_A 是定义在 $\mathbb{R}^{|A|}$ 上的联合分布函数，其边缘分布均非退化，且在 $+\infty$处没有质量。假设（2.3）仅仅说明，过程 $\{X_n\}_{n\in\mathbb{Z}}$ 的有限维 copula 属于 Heffernan & Tawn (2004) 所确立的类。Heffernan–Tawn 类范围广泛，包含 Joe (1997) 中考虑的所有 copula 示例。

在本文的大部分内容中，我们将考虑以下简单的特殊情形： $A = 1{:}\ k$ 对于某个$k \in \mathbb{N}$，即，我们考察过程进入极端状态后紧接着的行为。有时我们将这种方法称为 k步法，因为它可以让我们建立模型，描述从初始阈值超越开始向前 k 个时间步的时间序列。常数 k 将由统计学家根据具体情境选取；模型参数估计对 k的选择有多敏感，将在第5节讨论。我们认为这个简单的特殊情形最常见、最值得关注，不过在第4.2 节中，我们会介绍一些情形：在这些情形下，需要能够从极端事件出发，同时向前和向后模拟，因此需要采用更一般的集合 A 。

3 统计建模

3.1 边际模型

尽管在方程（2.1）中我们假设边缘分布为标准拉普拉斯分布，但实际应用中的数据通常并非处于这一尺度，因此需要进行边际变换，才能应用下文介绍的模型。如果 $\{Y_n\}_{n\in\mathbb{Z}}$ 是一个平稳时间序列，其边缘分布函数为 F，那么由下式定义的时间序列 $\{X_n\}_{n\in\mathbb{Z}}$ 为

$$X_n = \begin{cases} \log\{2F(Y_n)\}, & \text{if } F(Y_n) < 1/2, \\ -\log[2\{1 - F(Y_n)\}], & \text{if } F(Y_n) \geq 1/2, \end{cases} \tag{3.1}$$

具有平稳性，并且具有拉普拉斯边缘分布。在实践中， F 的确切形式通常未知，必须对其进行估计。我们采用标准的半参数方法（Coles&Tawn1991）来估计F ：在低于某个较高阈值 u^* 时使用经验分布函数，在高于 u^*时使用广义帕累托分布（GPD）。因此，我们假设 Y 的上尾模型形式为

$$\mathbb{P}(Y - u^* > y \mid Y > u^*) = (1 + \xi\, y/\sigma)_+^{-1/\xi}, \quad y > 0, \tag{3.2}$$

其中 $y_+ = \max\{y, 0\}$。参数 $\sigma > 0$ 和 ξ 分别控制 Y尾部的尺度和形状。我们通过最大似然法，使用每个 σ 的超阈值 ξ，对方程（3.2）中的 $Y_t - u^*$ 和 t 进行估计，其中 $Y_t > u^*$ (Davison & Smith 1990)。通过对一系列高阈值拟合广义帕累托分布（GPD）模型，并选取使 u^* 和 σ 的估计值趋于稳定的最小阈值，来确定阈值 ξ 。

ˆ 给定 σ 和 ξ 的最大似然估计量 ˆσ 和 ξ，基于样本 $\boldsymbol{Y}_{1:N}$ 的边缘分布 F 的估计量为

$$\hat{F}(y) = \begin{cases} N^{-1}\sum_{i=1}^{N} \mathbb{1}[Y_i \le y], & y \le u^*, \\ 1 - \hat{p}_{u^*}\{1 + \hat{\xi}\,(y - y^*\}/\hat{\sigma})_+^{-1/\hat{\xi}}, & y > u^*, \end{cases} \tag{3.3}$$

其中 ˆ$p_{u^*} = N^{-1}\sum_{i=1}^{N} \mathbb{1}[Y_i > u^*]$ 是超过阈值 u^*的 Y_i 经验估计量 .

3.2 标准化函数的参数模型

我们将以假设（2.3）为基础，对给定$X_t > u$时 $\boldsymbol{X}_{t+1:t+k}$的条件分布进行建模。我们假设正整数 k （它决定了我们希望对极端事件之后多少个滞后项进行建模）已固定，并将在第5节进一步讨论这个问题。我们还假设阈值 u 已选得足够大，因此收敛关系（2.3）在超过 u时恰好成立。根据平稳性，对于给定 $X_t > u$时的$_{t+k}$，可以使用与给定 $X_0 > u$时的 $\boldsymbol{X}_{1:k}$相同的函数 $\boldsymbol{a}_{1:k}$和 $\boldsymbol{b}_{1:k}$进行标准化: 因此，我们假设

$$\frac{\boldsymbol{X}_{t+1:t+k} - \boldsymbol{a}_{1:k}(X_t)}{\boldsymbol{b}_{1:k}(X_t)} = (Z_{t+1|t}, \ldots, Z_{t+k|t}) = \boldsymbol{Z}_{t+1:t+k|t}, \quad t \in T_u, \tag{3.4}$$

的边缘分布 $\boldsymbol{Z}_{t+1:t+k|t}$非退化。当 $u \to \infty$时， $\boldsymbol{Z}_{t+1:t+k|t}$对应于隐尾链中长度为 k 的区块；不过，与我们的应用情形相同，由于 u 将是有限的，我们将直接把 $\boldsymbol{Z}_{t+1:t+k|t}$称为残差向量。我们将考虑两种统计推断方法，它们在残差向量 $\boldsymbol{Z}_{t+1:t+k|t}$的建模方式上有所不同。第一种情形将在第3.3节讨论，$\boldsymbol{Z}_{t+1:t+k|t}$采用非参数方法建模；而第3.4节将讨论参数模型。

为了使统计推断易于处理，我们需要指定标准化函数的形式$\boldsymbol{a}_{1:\ k}$和 $\boldsymbol{b}_{1:k}$，并考虑两种可能性。我们的第一个模型受到Heffernan & Tawn (2004) 中所用标准化因子的启发，我们将其设定为

$$\textbf{Model 1:} \quad a_i(x) = \alpha_i x, \quad b_i(x) = x^{\beta}, \quad i \in 1\!:\!k,\ \alpha_i \in [-1, 1],\ \beta \in [0, 1). \tag{3.5}$$

一种更灵活的模型是指定 $b_i(x) = x^{\beta_i}$，即在每个滞后处采用不同的尺度标准化。我们选择简单的情形，即对每个 $i \in 1\!:\ k$，令 $\beta_i = \beta$；部分原因是缺乏可用于对 β_i 施加结构约束的有效理论结果，部分原因是 Papastathopoulos 等（2017）和 Papastathopoulos 等人（2023）的结果表明，对于马尔可夫时间序列，我们可以取 $\beta_i = \beta$。我们对 β 的参数限制比 Heffernan & Tawn（2004）中的限制更严格，其中 $\beta \in (-\infty, 1)$。当 β 取负值时，在给定 $X_t > u$ 的条件下， X_t 与 X_{t+i} 之间的关系会随着 $u \to \infty$ 而变成确定性的，且 $Z_{t+i|t}$ 的极限分布会退化，因此在应用中我们排除这种不现实的行为。为允许采用更灵活的尺度标准化因子，我们还考虑以下标准化函数形式，该形式受到 Wadsworth & Tawn（2022）中模型3的启发

$$\textbf{Model 2:} \quad a_i(x) = \alpha_i x, \quad b_i(x) = 1 + a_i(x)^{\beta}, \tag{3.6}$$

采用与模型 1 中相同的参数限制。未来也可以考虑更多尺度函数模型b_i ，例如 Wadsworth 和 Tawn（2022）提出的模型 1，但本文将仅关注模型（3.5）和（3.6）。

我们最感兴趣的是可以假设这些常数$\boldsymbol{\alpha}_{1:\ k}$具有某种结构的情形。特别是，如第 5 节所述，当时间序列为马尔可夫时间序列时， k参数 $\boldsymbol{\alpha}_{1:k}$可以简化为单个参数 α，从而大大简化推断。原则上，这些常数 $\boldsymbol{\alpha}_{1:k}$应受到约束，以确保它们是由平稳过程生成的；也就是说，它们不能取任意形式。与 Wadsworth 和 Tawn (2022) 一样，我们不探讨这个一般性问题，不过，未来研究若能进一步探究这一问题，将有助于提出理论依据更充分的非马尔可夫时间序列模型。

3.3 半参数建模

我们首先考虑条件极值文献中的一种标准方法（Heffernan & Tawn 2004，Keef 等人 2013），在对 $\boldsymbol{\alpha}_{1:k}$和 β的结构作出简化假设的前提下，构造复合对数似然。具体而言，我们暂且假设 $\boldsymbol{Z}_{t+1:t+k|t}$的 Copula 具有独立性，且对于每个 $t \in T_u$， i-th 个分量 $Z_{t+i|t}$，残差向量 $\boldsymbol{Z}_{t+1:t+k|t}$的密度为

$$f_i(z) = \frac{\delta_i}{2\sigma_i \Gamma(1/\delta_i)} \exp\left\{-|(z-\mu_i)/\sigma_i|^{\delta_i}\right\}, \quad \mu_i \in \mathbb{R},\ \sigma_i > 0,\ \delta_i > 0. \tag{3.7}$$

特殊情形 $(\mu_i, \sigma_i, \delta_i) = (0, 1, 2)$ 和 $(\mu_i, \sigma_i, \delta_i) = (0, 1, 1)$ 分别对应标准高斯分布和拉普拉斯分布。尽管在半参数推断方法中，残差的边际模型仅是一种临时工作假设，但纳入尾部参数 δ_i 后，模型便比标准高斯假设更准确，而这有望进一步帮助我们更好地识别其他感兴趣的参数。密度函数为 (3.7) 的随机变量早在统计学文献中就已出现（Subbotin 1923，Varanasi & Aazhang 1989），当时称其服从广义高斯分布。近来，这类分布也见于空间条件极值文献（Wadsworth & Tawn 2022，Shooter 等人 2021），并被称为 δ-拉普拉斯分布。我们沿用这一较新的约定；当随机变量Z 服从密度为 (3.7) 的 δ-拉普拉斯分布时，记作 $Z \sim \delta\text{Laplace}(\mu_i, \sigma_i, \delta_i)$。另一种工作假设是暂时假设 Copula 为独立性联结函数，并采用 $Z_{t+i|t} \sim N(\mu_i, \sigma_i^2)$，如 Heffernan & Tawn（2004）所做的那样。这样做的好处是可以加快计算；在第 6 节中，我们需要对许多自助抽样样本拟合多个模型，因此更倾向于采用这一假设。然而，根据第 5.1 节所述、用于估计次渐近极值指数的模拟结果， δ-拉普拉斯假设的效果略好，因此在下文其他部分我们将采用该假设。

在这些工作假设下，记作 $\boldsymbol{\theta} = (\boldsymbol{\alpha}_{1:k}, \beta, \boldsymbol{\mu}_{1:k}, \boldsymbol{\sigma}_{1:k}, \boldsymbol{\delta}_{1:k})$，即可得到采用模型 1 或模型 2 标准化因子时的复合对数似然：

$$l(\boldsymbol{\theta}) = \sum_{t \in T_u} \sum_{i=1}^{k} \left\{ \log \delta_i - \log b_i(x_t) - \log \sigma_i - \log \Gamma\left(1/\delta_i\right) - \left| \frac{x_{t+i} - a_i(x_t) - b_i(x_t)\mu_i}{b_i(x_t)\sigma_i} \right|^{\delta_i} \right\}. \tag{3.8}$$

当 $|t - t'| < k$且 $t,\ t' \in T_u$时，残差向量 $\boldsymbol{Z}_{t+1:t+k|t}$和 $\boldsymbol{Z}_{t'+1:\ t'+k|t'}$包含重复信息，即某些值同时出现在两个向量中。这一事实，加上我们对残差向量作出的错误工作假设，意味着表达式（ ℓ（$\boldsymbol{\theta}$）中的（3.8）并不对应一个适当的似然函数。在条件极值模型的其他应用中，使用复合似然是标准做法。我们关注的参数是 $\boldsymbol{\alpha}_{1:k}$和 β，而（$\boldsymbol{\mu}_{1:k}$， $\boldsymbol{\sigma}_{1:k}$， $\boldsymbol{\delta}_{1:k}$）对应干扰参数。我们通过最大化相应的剖面复合对数似然来估计（$\boldsymbol{\alpha}_{1:k}$， β）

$$l(\boldsymbol{\alpha}_{1:k}, \beta) = \sup_{(\boldsymbol{\mu}_{1:k}, \boldsymbol{\sigma}_{1:k}, \boldsymbol{\delta}_{1:k})} l(\boldsymbol{\theta}). \tag{3.9}$$

对于（，）中的固定提议值 (3.9)，使右侧达到最大值的值（:，:，:）可通过相应 δ-Laplace 样本的最大似然估计得到。具体而言，对于每个 $i = 1{:}\ k$，给定提议值（，），我们计算经验滞后 i残差，如下所示

$$Z_{t+i|t} = \frac{X_{t+i} - a_i(X_t)}{b_i(X_t)}, \quad t \in T_u. \tag{3.10}$$

在我们的工作假设下，残差 $\{Z_{t+i|t}\}_{t \in T_u}$ 是来自一个 δ-拉普拉斯分布的随机样本，其密度为 (3.7)。数值 $(\mu_i, \sigma_i, \delta_i)$ 是通过最大化 δ-拉普拉斯对数似然函数 $\sum_{t \in T_u} \log f_i(z_{i|t})$ 得到的，其中 f_i 如 (3.7) 所示。尽管这些数值没有闭式解，但通过数值方法求得它们相当直接。

ˆ 获得 $(\boldsymbol{\alpha}_{1:k}, \beta)$ 的点估计后，其中 $(\boldsymbol{\alpha}_{1:k}, \beta)$，可从拟合得到的经验残差中随机抽样，以获得残差向量 $\boldsymbol{Z}_{t+1:t+k|t}$ 的估计实现值

$$\hat{\boldsymbol{Z}}_{t+1:t+k|t} = \frac{\boldsymbol{X}_{t+1:t+k} - \hat{\boldsymbol{a}}_{1:k}(X_t)}{\hat{\boldsymbol{b}}_{1:k}(X_t)}, \quad t \in T_u. \tag{3.11}$$

如果我们改为采用工作高斯假设 $Z_{t+i|t} \sim N(\mu_i, \sigma_i^2)$，那么在剖面化消去 μ_i 和 σ_iˆ 后，我们可以将它们的闭式解分别表示为 $Z_{t+i|t}$ 的样本均值和标准差，$t \in T_u$对应于 (3.11)。因此，在这种情况下，剖面复合对数似然为

$$l(\boldsymbol{\alpha}_{1:k}, \beta) = \sum_{t \in T_u} \sum_{i=1}^{k} \left\{ \log \hat{\sigma}_i - \log b_i(x_t) - \frac{\left(x_{t+i} - a_i(x_t) - b_i(x_t)\hat{\mu}_i\right)^2}{2 b_i(x_t) \hat{\sigma}_i} \right\},$$

其中 ˆμ_ia　　ˆ 和 ˆσ_i 分别是 $\{Z_{t+i|t}\}_{t \in T_u}$ 的样本均值和标准差。

3.4 参数建模

现在我们考虑残差向量 $\boldsymbol{Z}_{t+1:t+k|t}$ 的参数建模。我们假设 $Z_{t+i|t}$ ∼δLaplace($\mu_i, \sigma_i, \delta_i$)，并关注如何将 δ-Laplace 参数化为 i 的函数。我们将这些参数可能采用的参数形式留待第5节确定。目前我们注意到，假设 $Z_{t+i|t} \sim \delta$Laplace($\mu_i, \sigma_i, \delta_i$) 意味着 $X_{t+i} \mid X_t > u \sim \delta$Laplace($a_i(X_t) + b_i(X_t)\mu_i, b_i(X_t)\sigma_i, \delta_i$)。对于较大的 i，我们预期 X_{t+i} 与 X_t 近似不相关，因此 $X_{t+i} \mid X_t > u$ 的分布应近似为单位 Laplace 分布，即 δLaplace(0, 1, 1)，这是 X_{t+i} 的无条件分布。只要该过程满足适当的混合条件，从而限制长程依赖，这一预期便有可能成立。对于第5节所考虑的马尔可夫过程，若干强混合条件均成立（Bradley 2005，第3节）。在这种情况下，当 i 较大时，对于模型 2 归一化，通过指定函数 μ_i、σ_i 和 δ_i，使得当 $i \to \infty$ 时 $\mu_i \to 0$、$\sigma_i \to 1$ 和 $\delta_i \to 1$，我们可以确保给定 $X_t > u$ 时 X_{t+i} 的条件分布近似为 δ Laplace(0, 1, 1)。模型 1 的尺度参数 σ_i 需要稍加谨慎地进行参数化，这将在第5节进一步讨论。

我们还假设，$\boldsymbol{Z}_{t+1:t+k|t}$的 copula 是具有正定相关矩阵 P 的高斯 copula。P的具体形式取决于上下文，但我们通常会将其建模为一个（$k+1$）维随机向量在已知其第一个分量的条件下的条件相关矩阵。在许多情况下，例如第5节所讨论的渐近独立马尔可夫过程，（$k+1$）维向量可以自然地取为平稳自回归过程的一个长度为 $k+1$ 的区块，该过程的阶数等于底层马尔可夫序列的阶数。对 P 进行条件设定的一种动机是，在（2.3）中也纳入 X_0 的位置和尺度标准化，即考虑以下极限：$\{\boldsymbol{X}_0 : {}_k - \boldsymbol{a}_0 : {}_k(X_0)\}/\boldsymbol{b}_0 : {}_k(X_0)$，条件为 $X_0 > u$，当 $u \to \infty$时。显然，我们可以取 $a_0(x) = x$ 和 $b_0(x) = 1$，使得 $\{\boldsymbol{X}_{0:k} - \boldsymbol{a}_{0:k}(X_0)\}/\boldsymbol{b}_{0:k}(X_0) \mid X_0 > u \to \boldsymbol{Z}_{0:ak}$，其中$Z_0 = 0$，而 $\boldsymbol{Z}_{1:k}$ 与（2.3）中相同。因此，我们对 P 的设定可以理解为对 $\boldsymbol{Z}_{1:k} \mid Z_0 = 0$ 的依赖结构进行设定。这种对 P 的条件设定基于 Wadsworth & Tawn（2022）提出的方法；他们设定了空间场残差在特定位置出现极端观测条件下的依赖结构。采用这种方法的另一个动机是，Papastathopoulos 等（2017）和 Papastathopoulos 等人（2023）中的隐尾链通常具有非平稳依赖结构。通过以这种条件方式对 P 建模，与直接将 P 设定为平稳过程长度为 k 的区块的相关矩阵相比，我们可以更准确地刻画残差向量中的依赖关系。

我们的假设意味着，随机向量 $(\Phi^{-1}\{F_{t+1}(Z_{t+1|t})\}$、…、$\Phi^{-1}\{F_{t+k}(Z_{t+k|t})\})$ 服从均值为零、相关矩阵为 P 的k维高斯分布，其中 Φ 是标准单变量高斯分布函数，F_{t+i} 是 $Z_{t+i|t}$ 的分布函数。尽管选择

高斯 Copula 来描述残差向量一般缺乏理论依据，Towe 等人(2019) 发现，在空间情形下，获得可靠的 Heffernan & Tawn(2004) 回归参数 α 和 β 比建立精确的残差分布模型更为重要。此外，他们发现，无论过程表现出渐近依赖还是独立性，高斯 Copula 模型都表现良好。虽然理论上可以考虑其他 Copula 模型，但其计算需求会将 k 限制在较小的取值范围内。

根据我们对 $\boldsymbol{Z}_{t+1:t+k|t}$结构的假设，如果$\pi_{1:\ k|t}(\boldsymbol{x})$表示给定 $X_t > u$时 $\boldsymbol{X}_{t+1:t+k}$的条件密度，那么

$$\log \pi_{1:k|t}(\boldsymbol{x}) = \sum_{i=1}^{k} \Bigg\{ \log \delta_i - \log b_i(x_t) - \log \sigma_i - \log \Gamma\left(1/\delta_i\right) - \left| \frac{x_{t+i} - a_i(x_t) - b_i(x_t)\mu_i}{b_i(x_t)\sigma_i} \right|^{\delta_i} \Bigg\} + 0.5 \log |Q| - 0.5\, \boldsymbol{w}^T Q \boldsymbol{w} + 0.5 \sum_{i=1}^{k} \left[\Phi^{-1}\{F_{t+i|t}(x_{t+i})\} \right]^2 \tag{3.12}$$

其中，$Q = P^{-1}$、$\boldsymbol{w} = (\Phi^{-1}\{F_{t+1|t}(x_{t+1})\}$，……，$\Phi^{-1}\{F_{t+k|t}(x_{t+k})\})$，$F_{t+i|t}$表示给定 $X_t > u$时 X_{t+i} 的条件分布函数。复合似然为 $\sum_{t\in T_u} \log \pi_{1:k|t}(\boldsymbol{x})$。

3.5 不确定性的量化与模型诊断

参数估计值和其他感兴趣量（例如第4节中讨论的那些量）的不确定性通过自助法获得。对于平稳时间序列，获取重复自助样本的标准方法包括区块自助法（Carlstein 1986），以及其推广方法，如移动区块自助法（Kunsch 1989）和平稳自助法（Politis & Romano 1994）。区块自助法要求我们从非重叠区块$\boldsymbol{X}_1$中进行有放回抽样：$_b$、$\boldsymbol{X}_{b+1:2b}$、$\boldsymbol{X}_{2b+1}$：$_{3b}$、……，每个区块的长度为 $b \in \mathbb{N}$。随后将依次抽取的区块连接起来，以生成一个自助样本。如果原始时间序列的长度不是 b 的整数倍，我们只需截短最后一个抽取的区块，使自助样本的长度与原序列一致。移动区块自助法通过从重叠区块$\boldsymbol{X}_1$中进行有放回抽样来推广这一过程：$_b$、$\boldsymbol{X}_2$：$_{b+1}$、$\boldsymbol{X}_{3:b+2}$、……。平稳自助法进一步通过抽取长度随机且服从几何分布的数据块来推广这一过程。平稳自助法具有一定吸引力，因为与区块自助法和移动区块自助法不同，它生成的样本是平稳的。然而，Lahiri（1999）的理论结果表明，使用非随机区块会带来更低的均方误差，而且重叠区块优于非重叠区块。因此，在第 6 节的数据应用中，我们选择移动区块自助法来量化参数估计值和其他感兴趣量的不确定性。

由于我们的拟合模型将用于对超出拟合阈值 u 的情况进行外推，甚至可能外推至大于任何观测值的阈值，因此参数 $\boldsymbol{\alpha}_{1:k}$ 和 β 的估计值在 u 以上保持稳定十分重要。可以通过在一系列阈值以上拟合条件模型，并通过图形评估参数估计何时趋于稳定，来检查这一点。为尽可能利用可用于拟合模型的数据，我们将 u 设为参数估计值开始保持稳定的最小阈值。在实际示例中，$\boldsymbol{\alpha}_{1:k}$ 包含的自由参数少于 k，因此只需检查少数几张图。此外，一项基本的建模假设是，残差向量 $\boldsymbol{Z}_{t+1:t:\ k|t}$ 在给定 $X_t > u$ 的条件下与 X_t 条件独立。可以通过选取若干分量 $i \in 1:k$，非正式地查看 $\hat{}(X_t,\ Z_{t+i|t})\ t \in T_u$，的散点图，或正式地进行假设检验，来检验这一假设是否成立。

选择不同模型的一种方法，是将模型得出的簇泛函估计值与经验估计值进行比较。可使用的两个此类泛函是

$$\theta(v, d) = \mathbb{P}\big(X_2 \le v, \ldots, X_d \le v \mid X_1 > v\big), \tag{3.13}$$

$$\chi(v, d) = \mathbb{P}(X_{d+1} > v \mid X_1 > v). \tag{3.14}$$

式（3.13）和（3.14）中的概率分别对应于极值指数的次渐近版本 θ（Lead-better 1983）和上尾依赖度量 χ，两者均在 Ledford & Tawn （2003）中进行了探讨。Winter & Tawn （3.13）和（3.14）均被 Winter & Tawn （2016年）和 Winter & Tawn （2017）用于区分模型，并作为诊断工具来确定极值马尔可夫模型的适当阶数。在中等偏高的阈值 v处，此时数据量较为充足，式（3.13）和（3.14）的经验估计值会相当准确。通过观察经验估计值随 v 增大时的变化，并与模型得出的估计值进行比较，我们可以选择与经验估计值所呈现的行为最为吻合的模型。

采用参数化方法后，我们还可以使用其他多种模型判别和诊断方法。例如，在空间情形下，Wadsworth 和 Tawn (2022) 建议使用复合似然下的赤池信息准则，而非真实似然，来选择模型。Shooter 等人 在贝叶斯情形下采用了类似的方法 (2019)。基于复合似然构造复合后验，并像使用真实后验一样，用它计算偏差信息准则，以判别模型。Varin & Vidoni (2005) 和Ng &Joe(2014) 也研究了专门针对复合似然校准的信息准则。

对于参数模型，我们可以分别检验残差向量 $\boldsymbol{Z}_{t+1:t+k|t}$ 的边缘分量和 Copula 的拟合优度。通过计算经验滞后 i残差

$$Z_{t+i|t} = \frac{X_{t+i} - \hat{a}_i(X_t)}{\hat{b}_i(X_t)}, \quad t \in T_u, \tag{3.15}$$

与拟合的 $\delta\text{Laplace}(\mu_i, \sigma_i, \delta_i)$ 分布的相容性，可通过分位数-分位数图等标准方法或更正式的假设检验进行评估。此外，残差向量 $\boldsymbol{Z}_{t+1:t+k|t}$的 Copula 拟合优度可通过 Genest 等人 (2009) 中介绍的检验方法进行评估。

4 稀有事件估计的模拟方法

现在，我们来讨论拟合前文所述条件极值模型的主要目的：当平稳时间序列处于极端状态时对其进行模拟。第 4.1节讨论如何从超过大阈值的观测值出发，沿时间向前进行模拟。这可用于估计以下形式的条件期望 $\mathbb{E}(g(\boldsymbol{X}_{1:d}) \mid X_1 > v)$ $v \geq u$， $d \in \mathbb{N}$，其中 g为某个感兴趣的函数 g。此外，还可用于模拟超阈值的重复簇，并据此计算各种感兴趣的泛函。典型例子包括簇最大值、平均簇大小，或者如 Winter & Tawn (2016年) 所研究的，簇内连续超越的最大数量。

第 4.2 节介绍了 Owen 等人(2019) 的重要性抽样方法。这可以估计以下形式的条件期望 $\mathbb{E}(g(\boldsymbol{X}_{1:d}) \mid$ 最大值 $\boldsymbol{X}_{1:d} > v)$ $v \geq u$， $d \in \mathbb{N}$，其中 g为某个感兴趣的函数 g。这里的条件是观测块内任意位置至少出现一次超越，而不是像第 4.1节中那样，要求在观测块的起始位置出现超越。

第 4.1 节和第 4.2 节中默认假设 $d - 1 \leq k$。也就是说，对于所涉及的观测值，其距阈值超越的滞后大于区块长度 k 的事件，我们不考虑对其进行估计；区块长度 ，用于拟合我们的条件模型。在第 5节所考虑的示例中，虽然我们可能通过将 $d - 1 > k$中的结构外推至 $\boldsymbol{\alpha}_{1:k}$来允许 α_i， $i > k$但在这种情况下，最好直接选择 k ，使其至少等于我们希望用于模拟的最大滞后。

4.1 前向模拟

根据拟合得到的条件模型 $\boldsymbol{X}_{t+1:t+k}$，我们可以通过重新排列（ k）从极端事件 $\{X_t > u\}$ 出发，向前模拟若干步，得到3.11

$$\boldsymbol{X}_{t+1:t+k} = \hat{\boldsymbol{a}}_{1:k}(X_t) + \hat{\boldsymbol{b}}_{1:k}(X_t)\hat{\boldsymbol{Z}}_{t+1:t+k|t}, \quad t \in T_u. \tag{4.1}$$

ˆ 残差向量 $\boldsymbol{Z}_{t+1:t+k|t}$ 既可以像第3.3 节那样通过经验值获得，也可以像第3.4 节那样从高斯 Copula 模型中模拟得到。从极端事件出发向前模拟，可以轻松估计各种感兴趣的量。除了 $\theta(v,\ d)$ 和 $\chi(v,\ d)$，它们在（3.13）和（3.14）中定义，其他可能感兴趣的量包括

$$e_1(v,d) = \mathbb{E}(\max \boldsymbol{X}_{1:d} \mid X_1 > v), \tag{4.2}$$

$$e_2(v,d) = \mathbb{E}(d^{-1}\sum_{i=1}^{d} X_i \mid X_1 > v) \tag{4.3}$$

$$e_3(v,d) = \mathbb{E}\Big(\sum_{i=1}^{d} \mathbb{1}[X_i > v] \,\Big|\, X_1 > v\Big) \tag{4.4}$$

$$p(v,d,r) = \mathbb{P}\Big(\sum_{i=1}^{d} \mathbb{1}[X_i > v] = r \,\Big|\, X_1 > v\Big), \tag{4.5}$$

其中 $v > u$。诸如（4.2）至（4.4）的量可以用作简单的汇总统计量，以帮助了解过程进入极端状态后的行为。（4.5）中的概率涉及在区块起始处发生超越的条件下，大小为 d 的区块中超越次数的分布。所有这些量都具有 $\mathbb{E}$（g（$\boldsymbol{X}_{1:d}$）$\mid X_1 > v$）的形式，其中函数 g取适当的形式。例如，对于（4.5），有 g（$\boldsymbol{x}$）$= \mathbb{1}\big[\sum_{i=1}^{d} \mathbb{1}[x_i > v] = r\big]$。

由于假定采用拉普拉斯边缘分布，（$X_t - u$）$\mid X_t > u$ 的分布为单位指数分布，因此我们可以使用算法 $\mathbb{E}$（g（$\boldsymbol{X}_{1:d}$）$\mid X_1 > v$）估计 8。该算法涉及从超越处反复进行向前模拟 X_1，并通过经验比例估计所关注的量。第 4 步的具体执行方式取决于采用半参数推断还是参数推断。对于半参数推断，第 4 步涉及从经验残差（3.11）中随机抽样；而对于参数建模，则涉及从高斯 copula 中进行模拟。

我们还可能希望估计簇大小分布。假设根据游程法（Smith 与 Weissman 1994）定义簇，游程长度为 r。因此，当超过阈值 v 时，簇即被视为开始；当连续出现 r 个未超越值时，簇结束。根据这一定义，Winter & Tawn （2017）通过从簇的首次超阈值开始，迭代地向前模拟，直到出现 r 个连续未超越值，从而模拟簇。在我们的情形下，我们从首次超阈值开始，联合模拟一个长度为 k 的完整区块，其中首次超阈值通过设定 $X_t = v + E$来模拟，其中 $E \sim$ 服从 exp(1) 分布。应选择常数 k，使得观测到长度大于 k 的簇的概率可以忽略。随后，模拟得到的区块 $\boldsymbol{X}_{t:t+k}$ 通常会在簇终止后包含若干个值，因此我们可以保留较小的区块$\boldsymbol{X}_{t:t+k'}$，其中 $k' < k$，对应于一个单独模拟的簇。

ˆ 输入：阈值 $v > u$, d, $n \in \mathbb{N}$ 以及拟合条件模型得到的常数 (ˆ$\boldsymbol{\alpha}_{1:k}$, β)。输出：$\mathbb{E}$(g($\boldsymbol{X}_{1:d}$) $\mid X_1 > v$) 的估计值 *1* 对于 $i \leftarrow 1$ 至 n，执行 *2* 模拟超越量 $E \sim$ exp(1)；*3* 设置 $X_1^i = v + E$；*4* 独立于 X_1，从拟合条件模型中模拟残差 ˆ$\boldsymbol{Z}^{(1),2:d}$；*5* 设置 $\boldsymbol{X}_2^{i}$: $_d$ = ˆ $\boldsymbol{\alpha}_1$: $_{d-1} X_1^i + (X_1^i)^{\hat{\beta}}$ ˆ$\boldsymbol{Z}^{(1),2:d}$；*6* 设置 $\boldsymbol{X}_1^i$: $_d$ = $(X_1^i, \boldsymbol{X}_{2:d}^i)$;

7 结束

8 返回ˆ$\mathbb{E}(g(\boldsymbol{X}_{1:d}) \mid X_1 > v) = n^{-1}\sum_{i=1}^{n} g(\boldsymbol{X}_1^i$: $_d)$。

算法1： 通过前向模拟估计 $\mathbb{E}(g(\boldsymbol{X}_{1:d}) \mid X_1 > v)$。

4.2 重要性抽样

本节介绍Owen 等人（2019）提出的估计量，用于估计形如$\mathbb{P}(\cup_{i=1}^{d}\{X_i > v_i\})$的概率，其中 $v_i \in (0, \infty)$， $i \in 1\colon d$。在略有不同的设定下，Wadsworth 和 Tawn （2022）和Adler 等人（2012）的模拟算法中也采用了相同的方法。尽管该估计量可用于估计任意事件并集的概率，但我们将注意力限制在特定形式的事件 $\{X_i > v_i\}$ 上，因为在时间序列背景下，这类事件对我们最为重要。此外，我们还将看到如何调整该估计量，使其能够估计其他我们感兴趣事件的概率。

考虑来自联合密度函数为 π 的时间序列的一个长度为 d 的区块 $\boldsymbol{X}_{1:d}$。对于 $i \in 1\colon d$，令$L_i \subseteq \mathbb{R}^d$ 为区域 $L_i = \{\boldsymbol{x} \in \mathbb{R}^d\colon x_i > v_i\}$，并令 $\mathcal{L} = \cup_{i=1}^{d} L_i$。我们最感兴趣的是每个 v_i， $1 \le i \le d$ 都是标准拉普拉斯分布的较大分位数的情况，因此，如果 $\boldsymbol{X}_{1:d}$ 落在$\mathcal{L}$，中，那么其至少一个分量较大。我们考虑对

$$p = \mathbb{P}\left(\bigcup_{i=1}^{d}\{X_i > v_i\}\right) = \mathbb{P}(\boldsymbol{X}_{1:d} \in \mathcal{L}). \tag{4.6}$$

一个常见且备受关注的特殊情况是所有阈值都相等，设为 $v_i = v, 1 \le i \le d$，此时$p = \mathbb{P}(\max \boldsymbol{X}_{1:d} > v)$是在包含 d 个观测值的区块中超过阈值 v 的概率。基于 n 个独立重复样本 $\{\boldsymbol{X}_{1:d}^i\}_{i=1}^n$（来自 $\boldsymbol{X}_{1:d}$），对 p 的显然的经验估计量 $n^{-1}\sum_{i=1}^{n}\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d}^i)$ 是无偏的，其方差为 $p(1-p)/n$。Owen 等（2019）表明，通过从适当的混合分布而不是直接从 $\boldsymbol{X}_{1:d}$ 抽样，可以改进该估计量，即降低其方差。具体而言，对于每个 $i \in 1\colon d$，定义 π_i^* 为给定 $X_i > v_i$ 时$\boldsymbol{X}_{1:d}$ 的条件密度，使得 $\pi_i^*(\boldsymbol{x}) = \pi(\boldsymbol{x})\mathbb{1}_{L_i}(\boldsymbol{x})/p_i, \boldsymbol{x} \in \mathbb{R}^d$，其中 $p_i = \mathbb{P}(X_i > v_i) > 0$。Owen 等（2019）提出的重要性抽样密度为 $\pi^* = \sum_{i=1}^{d} w_i\pi_i^*$，其中 $w_i = p_i/\bar{p}$，且 $\bar{p} = \sum_{i=1}^{d} p_i$ 是 p 的并集界。因此，混合分量 π_i^* 被抽取的概率与p_i成正比。由于

$$p = \mathbb{E}_{\pi}\{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d})\} = \mathbb{E}_{\pi^*}\left\{\frac{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d})\pi(\boldsymbol{X}_{1:d})}{\pi^*(\boldsymbol{X}_{1:d})}\right\}, \tag{4.7}$$

这启发我们构造以下针对 p 的估计量

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d}^i)\pi(\boldsymbol{X}_{1:d}^i)}{\pi^*(\boldsymbol{X}_{1:d}^i)} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d}^i)\pi(\boldsymbol{X}_{1:d}^i)}{\sum_{j=1}^{d}\mathbb{1}_{L_j}(\boldsymbol{X}_{1:d}^i)\pi(\boldsymbol{X}_{1:d}^i)\bar{p}^{-1}}, \quad \boldsymbol{X}_{1:d}^i \overset{iid}{\sim} \pi^*. \tag{4.8}$$

当 $\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d}^i) = 1$时， $\boldsymbol{X}_{1:d}^i \sim \pi^*$，估计量 (4.8) 可简化为

$$\hat{p} = \frac{\bar{p}}{n}\sum_{i=1}^{n}\frac{1}{S(\boldsymbol{X}_{1:d}^i)}, \quad \boldsymbol{X}_{1:d}^i \overset{iid}{\sim} \pi^*, \tag{4.9}$$

其中， $S(\boldsymbol{X}_{1:d}^i) = \sum_{j=1}^{d}\mathbb{1}_{L_j}(\boldsymbol{X}_{1:d}^i)$ 统计区块 $\boldsymbol{X}_{1:d}^i$（长度为 d）中发生的事件 $\{X_i > v_i\}$， $1 < i < d$ 的数量。当 $1 \le S(\boldsymbol{X}_{1:d}^i) \le d$在 $\boldsymbol{X}_{1:d}^i \sim \pi^*$ 时，$\hat{p}$ 总是定义良好，并满足理论界限 $\bar{p}/d \le p \le \bar{p}$。

表达式 (4.9) 中出现的并集界 $\bar{p}$，可轻易根据过程的边缘分布服从标准拉普拉斯分布这一假设推得。本节其余部分我们聚焦于共同阈值 $v_i = v, i \in 1\colon d$ 的情形，此时 $\bar{p} = de^{-v}/2$。为了能够通过 (4.9) 估计 p，我们需要能够反复模拟区块 $\boldsymbol{X}_{1:d}$ 的取值，条件是区块内至少有一个值超越 v，然后统计超阈值的总数。由于边缘分布相同，且我们假设阈值相同，因此有 $w_i = d^{-1}, i \in 1\colon d$，而分布 π_i^* 是 $\boldsymbol{X}_{1:d} \mid X_i > v$ 的条件分布。当 $2 \le i \le d-1$ 时，我们需要能够从 $\boldsymbol{X}_{1:i-1} \mid X_i > v$ 和 $\boldsymbol{X}_{i+1:d} \mid X_i > v$ 中模拟，也就是说，我们需要能够同时模拟

从事件发生时刻起沿时间正向和反向 $X_i > v$。 Janßen 和 Segers (2014) 研究了向量值马尔可夫时间序列在极端事件发生时刻前后沿时间正向和反向的行为，并借助所谓的前向尾链和后向尾链对其进行描述。在此背景下，极端事件对应于欧几里得范数较大的取值；研究表明，前向尾链决定后向尾链，反之亦然。然而，他们的结果不能直接应用于我们的情形，因为他们的多元正则变差（Resnick（1987））假设排除了并非相互独立的渐近独立过程。

根据（2.3），要从一次阈值超越向后模拟，存在位置和尺度标准化函数 $\{a_i : \mathbb{R} \to \mathbb{R}\}_{i=-k}^{k}$ 和 $\{b_i : \mathbb{R} \to \mathbb{R}_+\}_{i=-k}^{k}$，使得对任意 $k \in \mathbb{N}$

$$\left(X_0 - u, \frac{\boldsymbol{X}_{-k:k} - \boldsymbol{a}_{-k:k}(X_0)}{\boldsymbol{b}_{-k:k}(X_0)} \right) \Bigg| X_0 > u \xrightarrow{D} (E_0, \boldsymbol{Z}_{-k:k}), \tag{4.10}$$

当 $u \to \infty$ 时，其中 E_0 是一个与随机向量 $\boldsymbol{Z}_A \sim G_A$ 独立的单位指数随机变量，$A = -k{:}k$，其中 G_A 是定义在 $\mathbb{R}^{2k}$ 上的联合分布函数，其边缘分布 G_i、 $i \neq 0$ 均非退化，且在 $+\infty$ 处没有质量。在（4.10）中，我们采用如下约定： $a_0(x) = x$， $b_0(x) = 1$ 且 $Z_0 = 0$。随后，我们可以使用第3节的方法，为超越 $X_t > u$ 之前和之后 k 个步长的区块 $\boldsymbol{X}_{t-k:t+k}$ 拟合模型。例如，对于模型 1 归一化（3.5），我们可以拟合如下形式的模型

$$\boldsymbol{X}_{t-k:t+k} \overset{D}{=} \boldsymbol{\alpha}_{-k:k} X_t + X_t^{\boldsymbol{\beta}_{-k:k}} \boldsymbol{Z}_{t-k:t+k|t}, \quad t \in T_u. \tag{4.11}$$

其中 $\boldsymbol{\beta}_{-k:k} = (\beta^-, \ldots, \beta^-, 0, \beta^+, \ldots, \beta^+)$， β^- 和 β^+ 分别是与 X_t 处阈值超越之前和之后的观测值相关联的尺度参数。在采用半参数建模且需要估计参数 $\boldsymbol{\alpha}_{-k:k}, \beta^-$ 和 β^+ 的情况下，其中 $\alpha_0 = 1$，拟合模型（4.11）无需相较于第3.3节的方法引入新的创新项。与之前一样，我们作出工作假设：$Z_{t+i|t} \sim \delta$ 服从 Laplace$(\mu_i, \sigma_i, \delta_i)$，此时 $i \in -k{:}\, k \backslash \{0\}$。残差向量 $\boldsymbol{Z}_{t-k:t+k|t}$ 可通过重排（4.11），并以参数的估计值替换原参数，以经验方式模拟得到。

我们有兴趣了解前向链和后向链相关参数之间是否存在联系，即（$\boldsymbol{\alpha}_{-k:\,-1}$， β^-）和（$\boldsymbol{\alpha}_{1:k}$， β^+），因为在这种情况下，我们或许可以减少需要估计的参数数量，从而提高效率。例如，对于渐近相依过程，有 $\alpha_{-i} = \alpha_i = 1$ 且 $\beta^- = \beta^+ = 0$。对于我们在第5.1节中考虑的渐近独立马尔可夫过程，其联结函数为高斯和反向逻辑斯蒂 copula，也有 $\alpha_{-i} = \alpha_i$ 且 $\beta^- = \beta^+$。空间情境下与此类对称性相对应的概念是各向同性；对于一阶马尔可夫时间序列，这意味着（X_t， X_{t+1}）的 Copula 具有可交换性。实际上，这种对称性是否存在尚不可知；不过，我们始终可以拟合对称和非对称模型，并利用第3.5节所述的诊断方法等来比较拟合效果。

估计出 $\boldsymbol{\alpha}_{-k:\,k}, \beta^-$ 和 β^+ 后，使用算法 2 估计 $\mathbb{P}(\max \boldsymbol{X}_{1:d} > v)$ 就很简单。该算法假设使用了模型 1 归一化，也可以轻松修改为使用模型 2 归一化。此外，残差向量的具体模拟方式取决于残差是像第 3.3 节那样采用非参数方法建模，还是像第 3.4 节那样采用参数方法建模。Owen 等（2019）证明，$\hat{}p$ 是 p 的无偏估计量，且 $\text{var}(\hat{}p) \leq p(\bar{p} - p)/n$，由此可知 $\hat{}p$ 是 p 的相合估计量。尽管 $\hat{}p$ 可用于估计任意事件并集的概率，但它在稀有事件情形下效率最高；此时， p 和 $\bar{p}$ 较小，因此 $p(\bar{p} - p)$ 可能比 $p(1 - p)$ 小几个数量级。因此，在稀有事件情形下，我们通过从 π^* 而非 π 中抽样来提高估计精度。

现在，我们考虑估计 $\cup_{i=1}^{d}\{X_i > v\}$ 的子事件概率。我们感兴趣的典型例子包括 $\mathbb{P}(\sum_{i=1}^{d} \mathbb{1}[X_i > v] = r)$ 或 $\mathbb{P}(\sum_{i=1}^{d} \mathbb{1}[X_i > v] \geq r)$ $r \geq 1$，它们分别对应于区块 $\boldsymbol{X}_{1:d}$ 的 d 个观测值中“恰有 r 次超越”和“至少 r 次超越”阈值 v。设 g 为定义在 $\mathcal{L}$ 上的函数，使得对于 $\boldsymbol{x} \in \mathcal{L}^c$， $g(\boldsymbol{x}) = 0$。我们考虑估计 $\mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\} =$

$\int_{\mathcal{L}} g(\boldsymbol{x})\pi(\boldsymbol{x})\, d\boldsymbol{x}$。我们可以通过从 π^* 进行重要性抽样来估计 $\mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\}$，方法如下：

$$\widehat{\mathbb{E}}_\pi\{g(\boldsymbol{X}_{1:d})\} = \frac{\bar{p}}{n}\sum_{i=1}^{n}\frac{g(\boldsymbol{X}^i_{1:d})}{S(\boldsymbol{X}^i_{1:d})}, \quad \boldsymbol{X}^i_{1:d} \overset{iid}{\sim} \pi^*. \tag{4.12}$$

在 g 有界的情况下（例如当 g 为指示函数时），(4.12) 定义了 $\mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\}$的一个相合估计量。这是下文定理 4.1的主要内容，该定理在附录 A中证明。

定理 4.1。 如果 $\mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\}$如 *(4.12)* 所示，且 g的支撑集为 $\mathcal{L}$，那么

$$\mathbb{E}_{\pi^*}\big(\widehat{\mathbb{E}}_\pi\{g(\boldsymbol{X}_{1:d})\}\big) = \mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\} \tag{4.13}$$

and

$$\mathrm{var}\left(\widehat{\mathbb{E}}_\pi\{g(\boldsymbol{X}_{1:d})\}\right) = n^{-1}\left(\bar{p}\int_{\mathcal{L}}[\{g(\boldsymbol{x})^2\pi(\boldsymbol{x})\}/S(\boldsymbol{x})]d\boldsymbol{x} - \big(\mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\}\big)^2\right). \tag{4.14}$$

因此，若 g是有界函数，则 $\mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\}$是 $\mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\}$的相合估计量。此外，若 g 是指示函数，即对所有 $\boldsymbol{x}\in\mathbb{R}^d$，$g(\boldsymbol{x})\in\{0,1\}$，那么

$$\mathrm{var}\big(\widehat{\mathbb{E}}_\pi\{g(\boldsymbol{X}_{1:d})\}\big) \le n^{-1}\mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\}\big(\bar{p} - \mathbb{E}_\pi\{g(\boldsymbol{X}_{1:d})\}\big). \tag{4.15}$$

输入： 阈值 $v > u$，$n\in\mathbb{N}$以及来自拟合模型 ($\hat{}\boldsymbol{\alpha}_{-k:\,k}$、$\hat{}\beta^+$、$\hat{}\beta^-$) 的常数。**输出：** 对 $\mathbb{P}(\cup_{i=1}^d\{X_i > v\})$ 的估计值

for $i \leftarrow 1$ **to** n **do**
- *sample exceedance time* $j\in\{1,\dots,d\}$ *uniformly at random*;
- *simulate exceedance amount* $E\sim\exp(1)$;
- *set* $X^i_j = v + E$;
- *simulate the residual vector* $\hat{\boldsymbol{Z}}^{(j),i}_{1:d}$;
- **if** $j = 1$ **then**
  - *set* $\boldsymbol{X}^i_{2:d} = \hat{\boldsymbol{\alpha}}_{1:(d-1)}X^i_1 + (X^i_1)^{\hat{\beta}^+}\hat{\boldsymbol{Z}}^{(1),i}_{2:d}$;
  - *set* $\boldsymbol{X}^i_{1:d} = (X^i_1, \boldsymbol{X}^i_{2:d})$
- **else if** $j = d$ **then**
  - *set* $\boldsymbol{X}^i_{1:d-1} = \hat{\boldsymbol{\alpha}}_{-(d-1):-1}X^i_d + (X^i_d)^{\hat{\beta}^-}\hat{\boldsymbol{Z}}^{(d),i}_{1:d-1}$;
  - *set* $\boldsymbol{X}^i_{1:d} = (\boldsymbol{X}^i_{1:d-1}, X^i_d)$;
- **else**
  - *set* $\boldsymbol{X}^i_{1:j-1} = \hat{\boldsymbol{\alpha}}_{-(j-1):-1}X^i_j + (X^i_j)^{\hat{\beta}^-}\hat{\boldsymbol{Z}}^{(j),i}_{1:j-1}$;
  - *set* $\boldsymbol{X}^i_{j+1:d} = \hat{\boldsymbol{\alpha}}_{1:d-j}X^i_j + (X^i_1)^{\hat{\beta}^+}\hat{\boldsymbol{Z}}^{(j),i}_{j+1:d}$;
  - *set* $\boldsymbol{X}^i_{1:d} = (\boldsymbol{X}^i_{1:j-1}, X^i_j, \boldsymbol{X}^i_{j+1:d})$;
- **end**
- *calculate* $S(\boldsymbol{X}^i_{1:d}) = \sum_{k=1}^d \mathbb{1}[X^i_k > v]$;

end

return $\hat{p} = \frac{de^{-v}}{2}\frac{1}{n}\sum_{i=1}^n\frac{1}{S(\boldsymbol{X}^i_{1:d})}$.

算法2： 拉普拉斯边缘分布下平稳时间序列中长度为 d 的区块内至少出现一次超过阈值 v 的概率估计。

最后，我们考虑这样一种情况：希望估计在 $\cup_{i=1}^d\{X_i > v\}$发生的条件下的期望。设 π^{**}为给定 $\cup_{i=1}^d\{X_i > v\}$时 $\boldsymbol{X}_{1:d}$的条件密度，因此 $\pi^{**}(\boldsymbol{x}) = \pi(\boldsymbol{x})\mathbb{1}_{\mathcal{L}}(\boldsymbol{x})p^{-1}$，其中 $p = \mathbb{P}(\cup_{i=1}^d\{X_i > v\})$。我们希望估计

$$\mathbb{E}_{\pi^{**}}\{g(\boldsymbol{X}_{1:d})\} = \int g(\boldsymbol{x})\pi(\boldsymbol{x})\mathbb{1}_{\mathcal{L}}(\boldsymbol{x})p^{-1}d\boldsymbol{x} = p^{-1}\mathbb{E}_\pi\{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d})g(\boldsymbol{X}_{1:d})\}. \tag{4.16}$$

现在，由于 $\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d})g(\boldsymbol{X}_{1:d})$ 的支撑集为 $\mathcal{L}$，我们可以使用 (4.12) 中的估计量来估计 $\mathbb{E}_{\pi}\{\mathbb{1}_{\mathcal{L}}(\boldsymbol{X}_{1:d})g(\boldsymbol{X}_{1:d})\}$，并使用 (4.9) 的倒数来估计 p^{-1} 。因此，根据 (4.16)，我们可以使用$\mathbb{E}_{\pi^{**}}\{g(\boldsymbol{X}_{1:d})\}$来估计

$$\widehat{\mathbb{E}}_{\pi^{**}}\{g(\boldsymbol{X}_{1:d})\} = \frac{\sum_{i=1}^{n} g(\boldsymbol{X}_{1:d}^{i})/S(\boldsymbol{X}_{1:d}^{i})}{\sum_{i=1}^{n} 1/S(\boldsymbol{X}_{1:d}^{i})}, \quad \boldsymbol{X}_{1:d}^{i} \overset{iid}{\sim} \pi^{*} \tag{4.17}$$

当 g 是有界函数时，由于它是由相合估计量之比构成的，因此是相合估计量。估计量（4.17）出现在 Wadsworth 和 Tawn (2022）的算法3中。一个适合使用（4.17）进行估计的典型例子是

$$p^{*}(v,d,r) = \mathbb{P}\left(\sum_{i=1}^{d} \mathbb{1}[X_i > v] = r \,\middle|\, \max \boldsymbol{X}_{1:d} > v \right), \quad 1 \le r \le d, \tag{4.18}$$

当 r 从 1 变化到 d时，该量估计区块 $\boldsymbol{X}_{1:d}$内 d个观测值中超阈值v的数量分布，条件是至少发生一次超越。对于固定的 r，要估计（4.18），我们令 g为指示函数 $g(\boldsymbol{x}) = \mathbb{1}\big[\sum_{i=1}^{d} \mathbb{1}[x_i > v] = r\big]$。利用估计量（4.17），可以轻松调整算法2来估计（4.18），或者更一般地估计任意 g 对应的（4.16）。所需的唯一修改是在第19行：除计算 $S(\boldsymbol{X}_{1:d}^{i})$外，还要计算 $g(\boldsymbol{X}_{1:d}^{i})$；然后在第21行返回（4.17）的值。

5 示例与模拟研究

5.1 渐近独立马尔可夫时间序列的标准化函数

Papastathopoulos 等人 的结果(2017表明，对于一阶马尔可夫时间序列，我们可以采用一个大为简化的常数向量结构 $\boldsymbol{\alpha}_{1:k}$，而非第 3 节给出的一般形式。具体而言，若写作 $\alpha_1 = \alpha$，则可以取 $\alpha_i = \alpha^i$，其中 $i \in 1{:}\ k$。因此，给定 $\boldsymbol{X}_{t+1:t+k}$时， $X_t > u$的条件分布在模型1和模型2下分别表示为

$$\boldsymbol{X}_{t+1:t+k} \mid X_t > u \overset{D}{=} \alpha^{1:k} X_t + X_t^{\beta} \boldsymbol{Z}_{t+1:t+k|t}, \qquad \text{(Model 1)}, \tag{5.1}$$

and

$$\boldsymbol{X}_{t+1:t+k} \mid X_t > u \overset{D}{=} \alpha^{1:k} X_t + \{1 + (\alpha^{1:k} X_t)^{\beta}\} \boldsymbol{Z}_{t+1:t+k|t}, \qquad \text{(Model 2)}. \tag{5.2}$$

对于第 3.3 节所述的半参数建模，我们只需估计两个参数 α 和 β，。常数 $\boldsymbol{\alpha}_{1:k} = \alpha^{1:k}$也可以递归定义为 $\alpha_t = a(\alpha_{t-1})\ 1 \le t \le k$，初始条件为 $\alpha_0 = 1$，其中 $a{:}\ \mathbb{R} \to \mathbb{R}$是由 $a(x) = \alpha\, x$定义的函数。此外，数值 $\alpha^{1:k}$可视为一阶自回归模型（5.11）的自相关函数在滞后 1 至 k 处的取值，其中 $\rho = \alpha$。

Papastathopoulos 等人（2023）提出，对于高阶马尔可夫序列，常数 $\boldsymbol{\alpha}_{1:k}$的结构可以用类似的方法得到。具体而言，对于一个阶数为 $l < k$的马尔可夫序列，给定初始值 $\alpha_{0:l-1}$，其中 $\alpha_0 = 1$，我们可以通过递推关系 $\alpha_t = a(\boldsymbol{\alpha}_{t-l:t-1})$ 得到 $\boldsymbol{\alpha}_{l:k}$，其中 $a{:}\ \mathbb{R}^l \to \mathbb{R}$ 是一个可微的 1 次齐次函数，即 $a(t\boldsymbol{x}) = t a(\boldsymbol{x})$ $\boldsymbol{x} \in \mathbb{R}^l$。我们将考虑函数 a的两种可能函数形式。第一种情况下，我们假设 a 取平稳自回归模型（阶数为 l）的自相关函数形式；第二种情况下，则采用 Papastathopoulos 等人（2023）提出的一种函数形式。

目前，我们聚焦于时间序列为 $l = 2$阶马尔可夫序列的情形。考虑二阶平稳自回归模型 $Y_n = \theta_1 Y_{n-1} + \theta_2 Y_{n-2} + w_n$的自相关函数，其中 $\{w_n\}$ 是一个与 $\{Y_n\}$独立的零均值不相关序列。若 ρ_n 表示

序列 $\{Y_n\}_{n=0}^{\infty}$ 在滞后 n处的自相关函数值，则对所有 n ， ρ_n 均由递推关系 $\rho_n = \theta_1\rho_{n-1} + \theta_2\rho_{n-2}$， $n \geq 2$，确定，其初始条件为 $\rho_0 = 1$， $\rho_1 = \theta_1/(1-\theta_2)$。我们用这种递推关系形式来定义序列 $\boldsymbol{\alpha}_{1:k}$ 上的结构：

$$\alpha_t = \theta_1\alpha_{t-1} + \theta_2\alpha_{t-2}, \quad 2 \leq t \leq k, \quad \text{with } \alpha_0 = 1, \alpha_1 = \theta_1/(1-\theta_2). \tag{5.3}$$

(5.3)中的递推关系也可以改写为 $\alpha_t = a(\alpha_{t-1}, \alpha_{t-2})$，其中 $a: \mathbb{R}^2 \to \mathbb{R}$ 是函数 $a(x_1, x_2) = \theta_1 x_1 + \theta_2 x_2$。对于该模型，在第3.3节所述的半参数推断情形下， $k+1$需要在$\boldsymbol{\alpha}_{1:k}$、 β)中估计的参数，在（3.5）和（3.6）中减少为三个参数（θ_1， θ_2， β）。

实现该模型时会遇到一个困难：为了使（θ_1， θ_2）定义平稳过程的自相关函数，它们必须满足某些约束。这些约束要求（θ_1， θ_2）位于由以下不等式定义的三角形区域内部： $\theta_2 < 1+\theta_1$， $\theta_2 < 1-\theta_1$ 以及 $\theta_2 > -1$。为处理这些约束，我们考虑以偏自相关为参数的重新参数化。 Barndorff-Nielsen & Schou (1973)证明，阶数为 l 的平稳自回归过程可以用前 l 个偏自相关进行参数化，其中每个偏自相关都可以在（-1， 1）内自由取值。此外，研究表明，偏自相关与自回归参数之间存在一一连续可微对应关系。与直接使用自回归参数相比，这大大简化了推断，尤其是随着 l 增大，参数约束会变得更加复杂。对于二阶过程，若（r_1， r_2）是前两个偏自相关，则（r_1， r_2）与（θ_1， θ_2）之间的对应关系为

$$\theta_1 = r_1(1-r_2) \tag{5.4}$$

$$\theta_2 = r_2 \tag{5.5}$$

其中（r_1， r_2） $\in$ （-1， 1）^{2}。因此，在这种参数化下，我们需要推断参数（r_1， r_2， β）。根据 r_1 和 r_2的拟合值，我们可以利用方程（5.4）和（5.5）得到 θ_1 和 θ_2 的拟合值，进而利用方程（5.3）得到 $\boldsymbol{\alpha}_{1:k}$。

这种方法可以推广到高阶情形。例如，对于三阶马尔可夫序列， $\boldsymbol{\alpha}_{1:k}$上的结构由三阶自回归模型的自相关函数给出：

$$\alpha_t = \theta_1\alpha_{t-1} + \theta_2\alpha_{t-2} + \theta_3\alpha_{t-3}, \quad 3 \leq t \leq k, \tag{5.6}$$

with

$$\alpha_0 = 1, \quad \alpha_1 = \frac{\theta_1 + \theta_2\theta_3}{1-\theta_2-\theta_1\theta_3-\theta_3^2}, \quad \alpha_2 = \theta_2 + (\theta_1+\theta_3)\alpha_1. \tag{5.7}$$

参数（θ_1， θ_2， θ_3）需满足平稳性约束 $\theta_1+\theta_2+\theta_3 < 1$、 $-\theta_1+\theta_2-\theta_3 < 1$、$\theta_3$（$\theta_3-\theta_1$） $-\theta_2 < 1$ 和 $|\theta_3| < 1$。

随后，我们用（r_1、 r_2、 r_3）重新进行参数化，其中

$$\theta_1 = r_1 - r_1r_2 - r_2r_3$$

$$\theta_2 = r_2 - r_1r_3 + r_1r_2r_3$$

$$\theta_3 = r_3$$

其中（r_1、 r_2、 r_3） $\in$ （-1、 1）^{3}。尽管原则上该方法可用于任意阶数 l 的马尔可夫过程，但 $\boldsymbol{\alpha}_{0:l-1}$ 的初始条件由 l 阶自回归序列的相关函数的前 l 个值确定，随着 l 的增大，这些初始条件开始变得相当复杂。

Papastathopoulos 等人Papastathopouloset al.（2023）提出了一种略有不同的方法。对于阶数为 $l < k$ 的马尔可夫序列，给定初始 l 值 $\alpha_{0:l-1}$，其中 $\alpha_0 = 1$， $\boldsymbol{\alpha}_l$ 上的结构： k由以下递推关系给出

$$\alpha_t = c\left\{\sum_{i=1}^{d}\gamma_i(\gamma_i\alpha_{t-i})^{\delta}\right\}^{1/\delta}, \quad d \leq t \leq k, \tag{5.8}$$

其中 $0 < c^{-\delta} < \sum_{i=1}^{k} \gamma_i^{1+\delta}$、$\delta \in \mathbb{R}$、$\boldsymbol{\gamma}_{1:l} \in S_{l-1} = \{\boldsymbol{\gamma}_{1:l} \in [0,1]^l : \sum_{i=1}^{l} \gamma_i = 1\}$。对于二阶马尔可夫序列，相关方法中的递推关系 $\alpha_t = \theta_1 \alpha_{t-1} + \theta_2 \alpha_{t-2}$ 可写成式（5.8）的形式，其中 $\delta = 1$、$\gamma_1 = (\theta_1 - \sqrt{\theta_1 \theta_2}) / (\theta_1 - \theta_2)$ $\gamma_2 = 1 - \gamma_1$ 以及 $c = (\theta_1 - \theta_2)^2 / (\theta_1 - 2\sqrt{\theta_1 \theta_2} + \theta_2)$。与使用相关函数在 $\boldsymbol{\alpha}_{1:k}$ 上构建结构的方法相比，两者有两点不同。首先，现在初始条件 $\boldsymbol{\alpha}_{0:l-1}$ 被视为 $[-1,1]$ 中的自由参数，而不是用自回归模型的参数来表示。这种复杂度的降低，是以更复杂的递推关系（5.8）为代价的。对于二阶马尔可夫序列，我们有五个参数（α_1、β、c、δ、γ_1），而使用相关函数方法时只有三个参数。在实现该模型时，我们还需要满足参数约束 $\boldsymbol{\gamma}_{1:l} \in S_{l-1}$。当 $l = 2$ 时，这很容易做到：只需将 γ_1 限制在区间（0, 1）内，例如通过逻辑变换，然后令 $\gamma_2 = 1 - \gamma_1$。更一般地，我们可以考虑用 $\boldsymbol{\Gamma}_{1:l} \in \mathbb{R}^l$ 重新参数化，其中

$$\gamma_i = \exp(\Gamma_i) \Big/ \sum_{i=1}^{l} \exp(\Gamma_i), \quad 1 \le i \le l, \tag{5.9}$$

在此条件下，我们显然有 $\boldsymbol{\gamma}_{1:l} \in S_{l-1}$。然而，$\boldsymbol{\Gamma}_{1:l}$并不可识别，因为 $\boldsymbol{\Gamma}_{1:l}$和 $\boldsymbol{\Gamma}_{1:l} + x$对于任意 $x \in \mathbb{R}$都会产生相同的 $\boldsymbol{\gamma}_{1:l}$。对此，可以添加和为零的可识别性约束，$\sum_{i=1}^{l} \Gamma_i = 0$。于是，可识别的参数化形式为

$$\gamma_i = \begin{cases} \dfrac{\exp(\Gamma_i)}{\sum_{i=1}^{l-1} \exp(\Gamma_i) + \exp(-\sum_{i=1}^{l-1} \Gamma_i)}, & 1 \le i \le l-1, \\ \dfrac{\exp(-\sum_{i=1}^{l-1} \Gamma_i)}{\sum_{i=1}^{l-1} \exp(\Gamma_i) + \exp(-\sum_{i=1}^{l-1} \Gamma_i)}, & i = l, \end{cases} \tag{5.10}$$

其中 $\boldsymbol{\Gamma}_{1:l-1} \in \mathbb{R}^{l-1}$。

5.2 示例：半参数建模方法

本节将按照第 3.3 节所述，对两个一阶和一个二阶马尔可夫时间序列进行半参数建模。我们考虑的第一个示例是一个采用高斯 copula 的自回归模型。设 $Y_0 \sim N(0,1)$，并且

$$Y_{n+1} = \rho Y_n + \epsilon_n, \quad n \ge 0, \ |\rho| < 1, \tag{5.11}$$

其中 $\epsilon_n \sim N(0, 1-\rho^2)$，且 $\{\epsilon_n\}$ 与 $\{Y_n\}$ 相互独立。我们通过（3.1）将 $\{Y_n\}_{n=0}^{\infty}$ 转换为拉普拉斯边缘分布，其中 $F = \Phi$ 为一元高斯分布函数。与本节中的所有示例一样，我们利用 F 的已知形式来获得精确的拉普拉斯边缘分布，从而避免拟合边际模型（3.3）。因此，在下面的模拟中，我们忽略边际模型不确定性所产生的任何影响。

在第二个示例中，$(X_n,\ X_{n+1})$ 具有反向逻辑斯蒂 copula。令 Y_0 为单位指数随机变量，并令 $(Y_n,\ Y_{n+1})$ 的联合生存函数为

$$\bar{F}(y_n, y_{n+1}) = \exp\{-(y_n^{1/\gamma} + y_{n+1}^{1/\gamma})^{\gamma}\}, \quad y_n, y_{n+1} \ge 0, \tag{5.12}$$

对于 $\gamma \in (0,1]$ 和 $n \ge 0$。我们通过 $\{Y_n\}_{n=0}^{\infty}$ 经由（3.1）映射到拉普拉斯边缘分布，其中 $F(y) = 1 - e^{-y}$，$y > 0$是单位指数分布函数。

尽管这两个过程都是渐近独立的，但反向逻辑斯蒂 copula 对应的过程只需要进行尺度归一化，即该过程的 α 真实值等于零，而高斯 copula 对应的过程则为 $\alpha = \rho^2$。我们考虑这两个过程在 $\rho = 0.7$ 和 $\alpha = 0.5$时的特例，并研究模型 1和2中所用 k 的取值如何影响 α 和 β的估计值。我们针对每个过程进行了蒙特卡洛研究，每个过程包含长度为10^5 的1000次实现，并取了我们的

阈值 u 设为标准拉普拉斯分布的0.95分位数，以识别超阈值。对于每次实现，我们使用第3.3节中针对 $k=1$所述的半参数方法，估计 α 和 β 的取值，其中 $k=1$取值为5、10、15、20、30。图1展示了两种 copula 以及模型 1和模型 2归一化下的中位数估计值，以及估计值的0.025和0.975经验分位数。对于两种 copula 和不同的标准化因子选择， α 的估计值在区块长度 k 取值不同时都很稳定，且随 $k>1$变化很小。我们看到，在模型 1归一化下，随着 k 增加， β 的估计会出现明显偏差，尤其是估计值似乎趋近于零。在模型 1归一化下， β 的理论极限值对于高斯 copula 和反向逻辑斯蒂 copula 均为0.5。当 k 较大时， β 估计值中不断增大的偏差会由估计残差向量 $\boldsymbol{Z}_{1:k}$中较大的尺度参数所补偿，因此仍可能利用模型 1获得感兴趣量的合理估计值。模型 2归一化下的 β 估计值稳定得多，但代价是抽样变异性增大，不过这种变异性与 k并无明显关系。改变样本量和用于识别超阈值的阈值 u 时，也会发现类似结果。为了比较

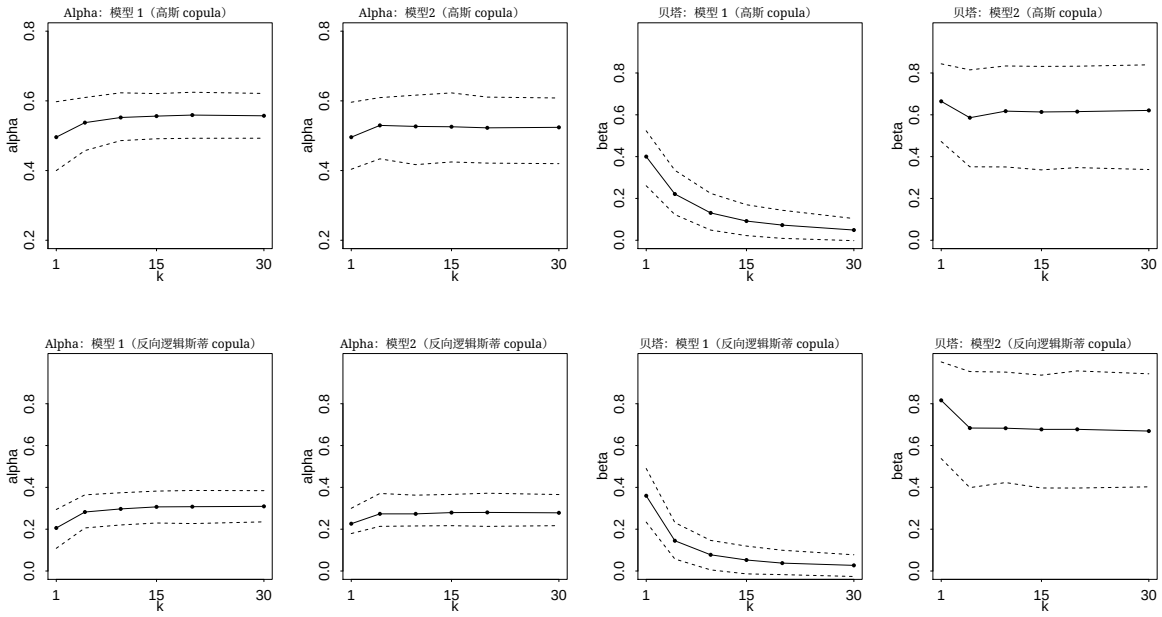

图1：展示估计值如何变化的图 α 以及 β 随区块长度取值 k 及不同标准化因子变化的图，针对高斯联结函数 AR(1)（5.11）（上排）和反向逻辑斯蒂 copula（5.12）（下排），基于每个过程的1000次实现。实线连接不同区块长度下所有1000次实现的中位数估计 k，虚线则表示估计值的0.025和0.975经验分位数。

这两种方法（5.3）和（5.8）用于在 $\boldsymbol{\alpha}_{1:k}$上引入结构；对于高阶马尔可夫序列，我们模拟了500个长度为 10^5 的二阶高斯自回归模型实现值

$$Y_n = 0.6Y_{n-1} + 0.3Y_{n-2} + w_n, \quad w_n \sim N(0, \sigma_w^2),\ n \geq 2, \tag{5.13}$$
$$(Y_0, Y_1) \sim \text{MVN}(\mathbf{0}, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix},$$

其中 $\{w_n\}_{n=2}^{\infty}$ 独立于 $\{Y_n\}_{n=0}^{\infty}$。我们取 $\sigma_w^2 = 1 - 0.6^2 - 0.3^2 - 2(0.6^2 \times 0.3)/(1-0.3) \approx 0.24$ 以确保 Y_n 对 $n \geq 0$服从标准正态分布。每个实现值都被转换为拉普拉斯边缘分布。我们只模拟了 500 个实现值，而非 1000 个（用于一阶马尔可夫序列），这是因为拟合 Papastathopoulos 等人（2023）在式（5.8）中的模型耗时更长。我们不展示两个模型中所有参数的结果，而只考察 $\boldsymbol{\alpha}_{1:k}$的拟合值，它们是

除 β外，所有模型参数均通过（5.3）和（5.8）成为其函数。中位数、0.025 和 0.975 经验分位数见图 2。使用区块长度 k 分别为 20 和 30 的情形重复了该实验。对于大于 k的 i 值，$\hat{}\alpha_i$ 的值通过式（5.3）和（5.8）的递推关系外推得到。我们仅使用式（3.5）中的模型 1 归一化。两种方法拟合出的曲线几乎完全相同；将它们绘制在同一图中时，中位数估计和分位数几乎无法区分。两种模型对 β 也得出了非常相似的估计值。当区块长度为 $k = 20$时，通过使用递推关系（5.3）和（5.8）得到的 β 估计值的 0.025 和 0.975 分位数分别为 $[0.158, 0.379]$ 和 $[0.157, 0.390]$ 。当 $k = 30$时，相应的区间在式（5.3）和（5.8）下分别为 $[0.095, 0.323]$和 $[0.102, 0.308]$。两种方法的主要区别在于，采用递推关系（5.8）的方法多出两个参数，因此拟合耗时显著更长。拟合阶数大于 2 的马尔可夫模型时，为加快拟合过程，可以考虑将残差向量的工作假设从 δ-拉普拉斯边缘分布替换为高斯边际分布。为了说明从一个极端值向前模拟一个较大区块可能带来的用途，

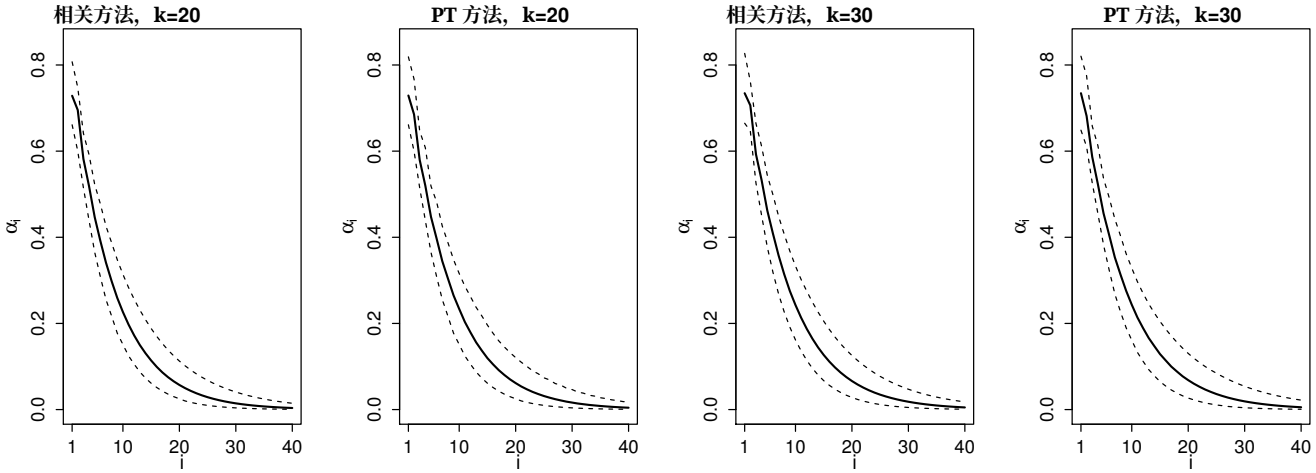

图 2：展示估计值 α_i 如何随滞后 i 变化的图，这些估计值针对二阶马尔可夫高斯 Copula 时间序列（5.13），分别采用带递推关系的相关函数方法（5.3）和带递推关系的 Papastathopoulos- Tawn (PT) 方法（5.8），基于该过程的 500 个实现值计算得出。实线连接全部 500 个实现值的中位数估计，虚线表示估计值的 0.025 和 0.975 经验分位数。

事件发生后，我们进行了实验，以估计式（3.13）中定义的次渐近极值指数 $\theta(v, d)$，其中 $d = 20$ ，并在高斯 copula AR(1) 模型（5.11）中考察一系列阈值 v 。所考察的阈值对应于边际拉普拉斯分位数从 0.9 到 0.99（间隔为 0.01），以及 0.999 分位数。由于我们需要能够从首次超阈值向前模拟 19 步，因此取 $k = 19$ ，并使用第（3.3）节的半参数方法和模型 2 归一化来估计 α 和 β 。我们将其与以下情形进行比较：其中 $k = 1$ ，并且在式（1.3）的残差 Z_{t+1} 上采用工作高斯假设；这对应于 Winter & Tawn (2016) 使用的方法。对于后一种方法，无法使用算法8估计 $\theta(v, 20)$，因此采用 Winter & Tawn (2016) 的算法1。这需要使用递推关系（1.3），从初始超阈值开始逐步向前模拟。我们模拟了 2×10^4 个长度为 2×10^4 的序列，并在 $k = 19$ 的情况下，于算法8中取 $n = 5 \times 10^4$；当 $k = 1$时也采用类似做法。我们采用等于 0.9 边际拉普拉斯分位数的阈值 u 来识别超阈值。对于每个序列，我们使用 α和 β 的拟合值来估计 $\theta(v, 20)$，并通过相对误差平方衡量估计质量。随后，将给定阈值下两种方法的整体表现定义为相对均方根误差（RMSRE）。也就是说，对于给定阈值 v，若某种方法得到的估计值

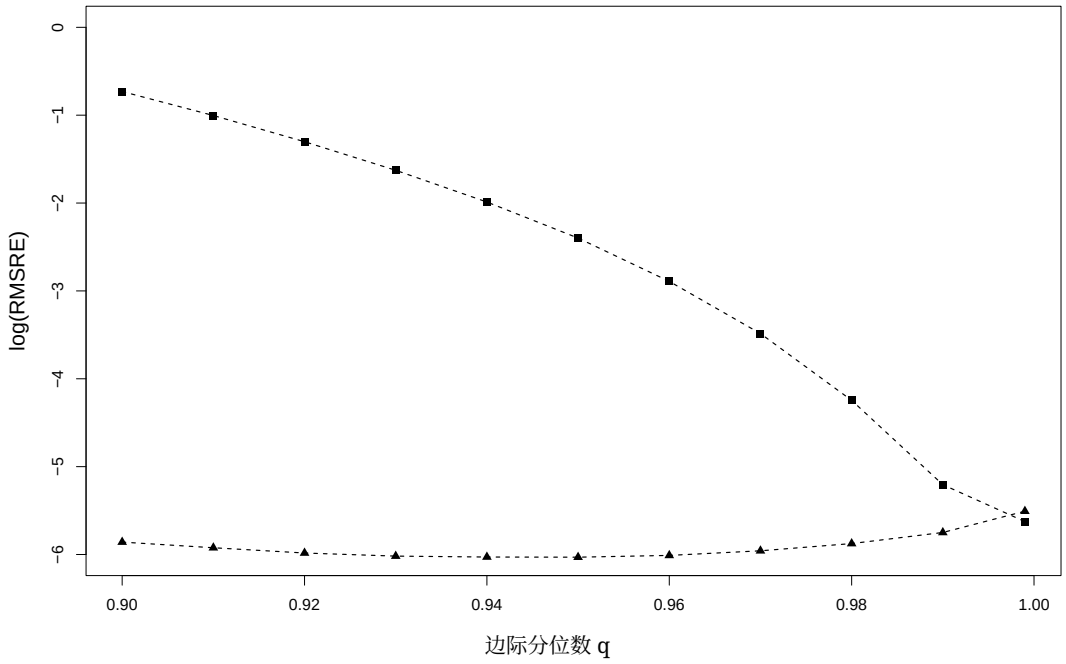

图3：展示估计值$\theta(v,\ 20)$的相对均方根误差（RMSRE）对数如何随单步方法（方块）和 k步方法（三角形）的阈值 $v = F^{-1}(q)$变化。在高斯 copula 时间序列模型中使用 $k = 19$ 和模型 2 归一化。估计值基于过程的2×10^4 个实现值。

$p_1, \ldots, p_M$，其中 $M = 2 \times 10^4$，我们报告数值 $\left[M^{-1} \sum_{i=1}^{M} (p_i/p - 1)^2\right]^{1/2}$，其中 $p = \theta(v, 20)$。图3展示了 RMSRE 的对数随阈值 v 的变化情况，涵盖了两种方法。对于不超过 0.98 边际分位数的所有阈值，从首次超阈值开始联合模拟长度为 19 的数据块并向前推进的 k 步方法，优于逐步向前模拟的方法。 k 步方法在所有阈值下都表现得极为稳定，RMSRE 值几乎没有明显差异。然而，随着阈值 v 升高，两种方法之间的差异逐渐缩小；达到 0.99 分位数时，两种方法几乎无法区分。这或许仅仅是由于该过程渐近独立，这保证了当 $v \to \infty$时， $\theta(v,20) \to 1$，因此在大阈值下， v、 $p_i/p \approx 1$ ，且两种方法的 RMSE ≈ 0 。解释 Winter–Tawn 模型在较低阈值下表现相对较差的另一个可能因素，是该模型所用模拟方案中的一个细节。对于该模型，我们使用递推关系（1.3），从初始阈值超越开始逐步向前模拟。如果对于某个 t， $X_t < 0$，那么根据式（1.3）， X_{t+1}将未定义，因为 $\beta < 1$。在这种情况下，Winter & Tawn (2016) 的算法1会将簇内所有后续值设为零。其思路是：由于零对应于拉普拉斯尺度上的边际中位数，如果某个时刻的值低于该值，那么后续值就不太可能是极端值，因此将它们设为零对估计簇泛函的影响应该很小。然而，在 0.9 分位数这类中等偏高的阈值下，即使值已降至该阈值以下，再次超越阈值的概率仍然虽小却不可忽略。因此，将所有后续值设为边际中位数，可能会导致估计值出现显著偏差。

5.3 示例：参数方法

对于第 5.2 节中的三种时间序列模型（5.11）、（5.12）和（5.13），我们现在考虑为以下参数识别参数形式的可能性： μ_i、σ_i 以及 δ_i，即与滞后 δ相关的 i -Laplace 参数和残差项 $Z_{t+i|t}$。

目前先关注一阶模型（5.11）和（5.12），我们取自第 5.2 节所述蒙特卡洛实验中 α 和 β的拟合值，分别记为α 和 β，并计算经验滞后i 残差如下：

$$\hat{Z}_i = \frac{X_{t+i|t} - \hat{\alpha}^i X_t}{X_t^{\hat{\beta}}}, \quad t \in T_u \tag{5.14}$$

对于模型 1，

$$\hat{Z}_i = \frac{X_{t+i|t} - \hat{\alpha}^i X_t}{1 + (\hat{\alpha}^i X_t)^{\hat{\beta}}}, \quad t \in T_u \tag{5.15}$$

对于模型 2 归一化。针对 $k = 1$ 的每个取值（5、10、15、20、30）分别重复上述过程。随后，我们根据这些残差，通过最大似然法估计残差 $Z_{t+i|t}$ 的 δ-Laplace 参数 μ_i、σ_i 和 δ_i。对滞后项 i 从 1 到 30（含 30）逐一重复此过程。图 4 展示了在使用 $k = 1$ 的情况下，对于高斯 copula 和反向逻辑斯蒂 copula，采用模型 1 和模型 2 归一化时，$Z_{t+i|t}$ 的拟合 δ-Laplace 参数随滞后 i 的变化情况。可以看到，两种 copula 均可采用相同的 μ_i、σ_i 和 δ_i 参数形式，但归一化方式似乎会产生一定影响，其中最明显的是尺度参数 σ_i。

目前先关注模型 1 归一化，对于每个 δ-拉普拉斯参数，采用指数衰减模型似乎是描述过程在尾部经历极端波动后如何回归主体部分的恰当方式。尽管高斯 copula 的 μ_i 曲线起始处似乎有一个转折点，但幅度足够小，或许可以忽略。正如预期，i 取较大值时，μ_i 和 δ_i的值分别稳定在 0 和 1；这正是边际拉普拉斯分布的取值。对于两个联结函数，σ_i 的值似乎都收敛到 0.6 至 0.7 之间的某个值。为了尝试理解较大滞后 i下 σ_i的行为，我们回顾基于模型的假设 $X_{t+i|t} \mid X_t > u \sim \delta\text{Laplace}(\alpha^i X_t + X_t^{\beta}\mu_i,\ X_t^{\beta}\sigma_i,\ \delta_i)$。因此，我们看到，只要满足 $|\alpha| < 1$，当 i 较大时，$X_{t+i|t} \mid X_t > u$，的分布的位置参数和形状参数就不会近似于标准拉普拉斯分布的相应参数，除非 μ_i 和 δ_i 满足这样的条件：当 i较大时，$\mu_i \approx 0$ 且 $\delta_i \approx 1$。然而，若要使 $X_{t+i|t} \mid X_t > u$ 的分布尺度参数与标准拉普拉斯分布的尺度参数相匹配，我们需要在 i 较大时满足 $\sigma_i \approx X_t^{-\beta}$，而这显然不可能对所有 $t \in T_u$都成立。一种可能的替代方案是为 σ_i 指定参数形式，使得当 i较大时，$\sigma_i \approx \mathbb{E}(X_t \mid X_t > u)^{-\beta} = (1+u)^{-\beta}$，这是由于拉普拉斯分布具有指数型右尾。因此，对于两个联结函数，在模型 1 归一化下，一种可能的参数化形式为

$$\begin{aligned} \mu_{i+1} &= Ae^{-Bi}, \\ \sigma_{i+1} &= (1+u)^{-\beta}(1 + Ce^{-Di}), \\ \delta_{i+1} &= 1 + Ee^{-Fi}, \end{aligned} \tag{5.16}$$

对于 $A, B, C, D, E, F > 0$和 $i \geq 0$。当 u是标准拉普拉斯分布的 0.95分位数，且$\beta = 0.35$（使用 $k = 1$ 作为两个 Copula 时，约为 β 的中位数估计）时，则 $(1+u)^{-\beta} \approx 0.66$，这解释了图 4 中 σ_i 收敛到约此值的现象。使用更大的 k重复这一实验时，μ_i, σ_i 和 δ_i 也会出现类似的曲线（未展示）。主要区别在于，当 k 较大时，β 的估计值会趋于零，因此 (5.16) 中的因子 $(1+u)^{-\beta}$ 约为 1。因此，我们发现 σ_i 收敛到约 1。

对于模型 2 归一化，我们看到 $\mu_i \to 0$，$\sigma_i \to 1$ 以及 $\delta_i \to 1$，因此我们不需要因子（$1+u$）$^{-\beta}$，如（5.16）中所示。这是由于模型 2 所采用的标准化确保了 δ的 -拉普拉斯尺度参数$X_{t+i} \mid X_t > u$ 为 $\{1 +$（$\alpha^i X_t$）$^{\beta}\}\sigma_i$，并且在 $\alpha \in (0,1)$的条件下，项（$\alpha^i X_t$）$^{\beta}$ 以概率趋于零

1. 使用模型 2 归一化时，另一个显著差异是，与模型 1 相比， σ_i 的变化似乎发生了反转。它现在单调递增，随后稳定于 1。对于两个联结函数，在模型 2 归一化下的一种可能参数化为

$$\begin{aligned}\mu_{i+1} &= Ae^{-Bi},\\ \sigma_{i+1} &= 1 + Ce^{-Di},\\ \delta_{i+1} &= 1 + Ee^{-Fi}\end{aligned} \tag{5.17}$$

对于 $i \geq 0$，其中现在 $C < 0$。曲线 μ_i 的起始部分附近似乎存在一个稍微更明显的众数。若要拟合这一特征，可以通过加入一个额外参数来实现，例如 $\mu_{i+1} = Ai^{B-1}e^{-Ci}$，它与伽马随机变量的密度函数成正比。我们考虑为高斯 copula 自回归拟合参数模型（5.16）

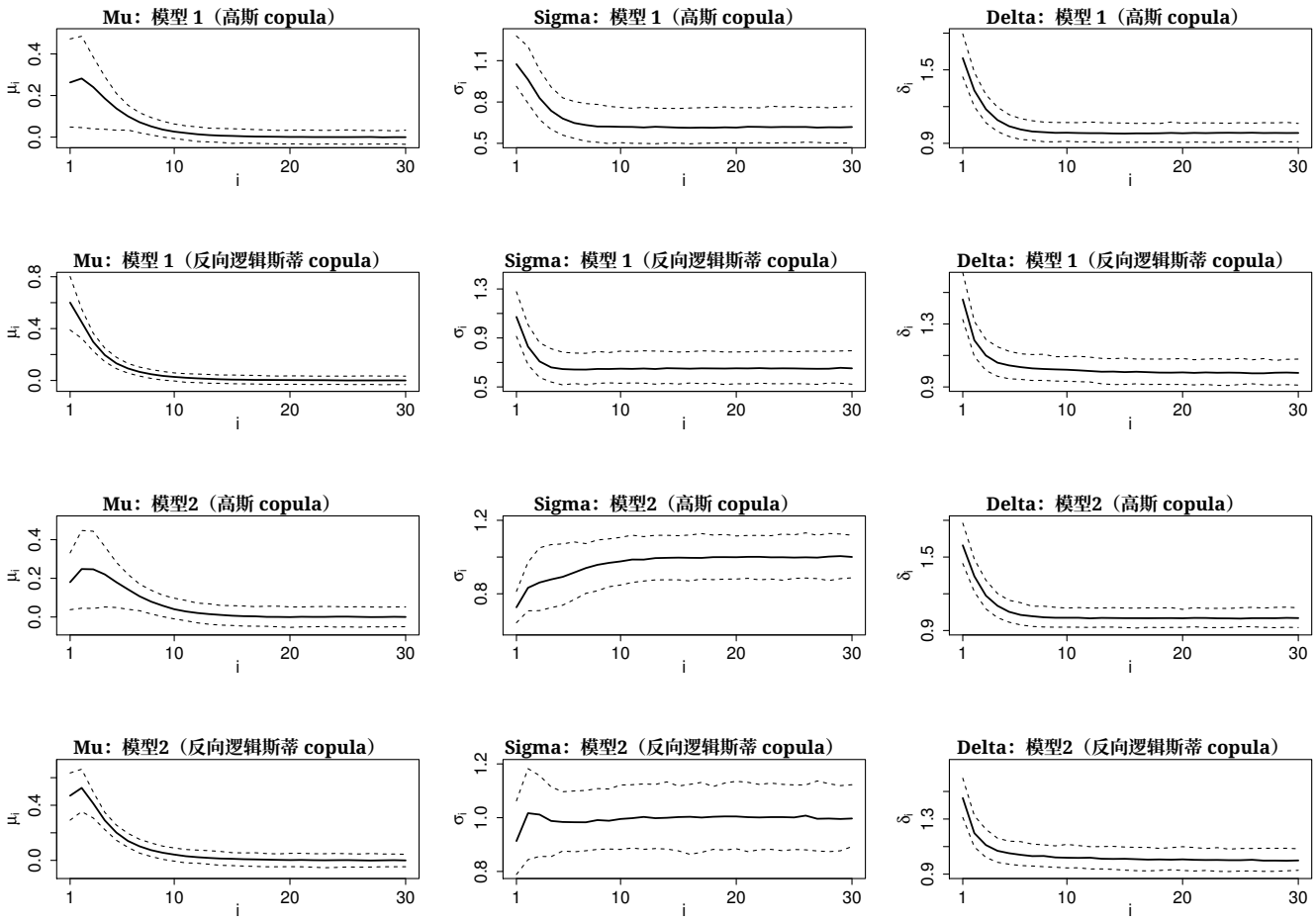

图 4：展示 δ-Laplace 参数 μ_i、σ_i 和 δ_i 如何随 i 变化的图。图中采用模型 1（上方两行）和模型2（下方两行）的高斯 copula 和反向逻辑斯蒂 copula 模型标准化因子，基于每个过程的1000次实现。实线连接所有1000次实现的中位数估计，虚线表示估计值的0.025和0.975经验分位数。

使用第 3.4 节的方法，针对半参数估计蒙特卡罗实验中使用的该过程的全部1000次实现。这需要我们指定残差向量 P 的高斯 copula 的相关矩阵 $\boldsymbol{Z}_{t+1:t+k|t}$，$t \in T_u$。如第 3.4 节所述，我们考虑在给定 Z_0条件下，（$k+1$）维随机向量 $\boldsymbol{Z}_{0:k}$的相关矩阵 Z_0。由于我们正在

表 1：参数模型（5.16）和高斯联结自回归模型中参数抽样分布的 0.025 和 0.975 分位数（保留两位小数）。这些结果基于过程在不同区块长度下的 1000次实现 k。

k	α	β	A	B	C	D	E	F	ρ
10	.54, .63	.08, .20	.00, .35	.10, 5.00	.63, 1.03	.35, 1.21	.45, .99	.44, 2.46	.65, .66
20	.54, .64	.03, .12	.00, .31	.13, 5.00	.65, 0.99	.35, .84	.48, .95	.44, 1.27	.67, .68
30	.54, .63	.01, .09	.00, .32	.15, 4.99	.63, 1.03	.35, 1.18	.46, .99	.42, 2.08	.68, .69

处理一个一阶马尔可夫序列，因此我们令 P 为自相关参数为 ρ的一阶平稳自回归模型的条件相关矩阵。直接用 $Q = P^{-1}$ 来表示会更方便；它是一个稀疏（带状）矩阵，并出现在（3.13）中。此外，与相关矩阵相比，在给定条件下更新精度矩阵要直接得多（Rue & Held 2005，定理 2.5）。我们首先构造（$k+1$）×（$k+1$）维 ~矩阵 Q如下：

$$\widetilde{Q}_{i,j} = \begin{cases} 1, & \text{if } (i,j) = (1,1) \text{ or } (i,j) = (k+1, k+1), \\ 1+\rho^2, & \text{if } i = j, \text{ and } 2 \le i \le k, \\ -\rho, & \text{if } |i-j| = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{5.18}$$

~随后，通过删除 Q 的首行和首列，得到式（3.13）中使用的矩阵 Q。虽然我们能够对所有参数（α、β、A、……、F、ρ）联合最大化复合似然，但计算速度太慢，无法进行模拟研究。因此，我们按照 Joe（1997）的建议，分两个阶段估计参数向量。首先，在残差向量服从独立性联结函数这一工作假设下，估计（α、β、A、……、F）；然后，在假定参数（α、β、A、……、F）固定为第一步所得估计值的条件下，最大化 ρ 的复合似然。与一次性估计所有参数相比，这大大加快了模型拟合速度。根据图 4，我们使用缩放 logit 变换将 A 和 E 限制在区间（0, 1），将 C 限制在（0, 3），并将 B、D 和 F 限制在（ 0, 5 ）。表 1 展示了基于 1000 次实现所得参数估计值的 0.025 和 0.975 经验分位数（保留两位小数）。我们在区块长度 k =10、20 和 30 下进行了估计。估计值并未显示出 k 存在明显依赖性；此前在使用模型 1 归一化因子的半参数方法中，我们已经注意到 β 存在这种依赖性。尤其是， B 的估计值具有很大的不确定性，其 0.025 和 0.975 分位数涵盖了区间（0, 5）的大部分范围。 $B = 5$ 取某个值时，意味着 μ_i 随 i 极快衰减，而图 4 中的图形似乎并不支持这一点。 B 估计值的不确定性较高，可能是因为 μ_i 的参数形式设定不佳；正如我们此前所指出的，图 4 中有证据显示存在一个转折点，但我们忽略了这一点。Shooter 等（2019）也报告了参数化 δ-Laplace 模型在基于似然的估计中遇到的困难，并发现使用贝叶斯方法能得到更好的结果。未来的研究或可在时间序列情境下采用类似方法。其他参数的估计值似乎合理；尤其是联结函数相关参数 ρ 的估计十分精确，并且随着区块长度 k 的变化而保持稳定。

为了检验二阶马尔可夫时间序列（ δ-Laplace 残差向量参数是否可以采用参数模型5.13进行建模，我们对一阶马尔可夫序列采用了相同的步骤。具体而言，基于第5.2节蒙特卡洛实验中给定的模型拟合结果，我们按照（ i 5.14）计算滞后残差，并通过最大似然法估计 δ-Laplace 参数 μ_i, σ_i 和 δ_i 。图5 展示了采用相关函数方法、递推关系（5.3）和区块长度 $k = 20$时，估计值的中位数估计以及 0.025 和 0.975 经验分位数。采用Papastathopoulos 等人(2023) 的递推关系（5.8）方法所得曲线与采用区块长度 $k = 30$时的曲线基本相同。和 σ_i 的曲线比一阶情形更加复杂，并且呈现出明显的众数，因此需要对其建模；不过， δ_i 与一阶情形中采用的相同模型

阶数情形似乎合适。

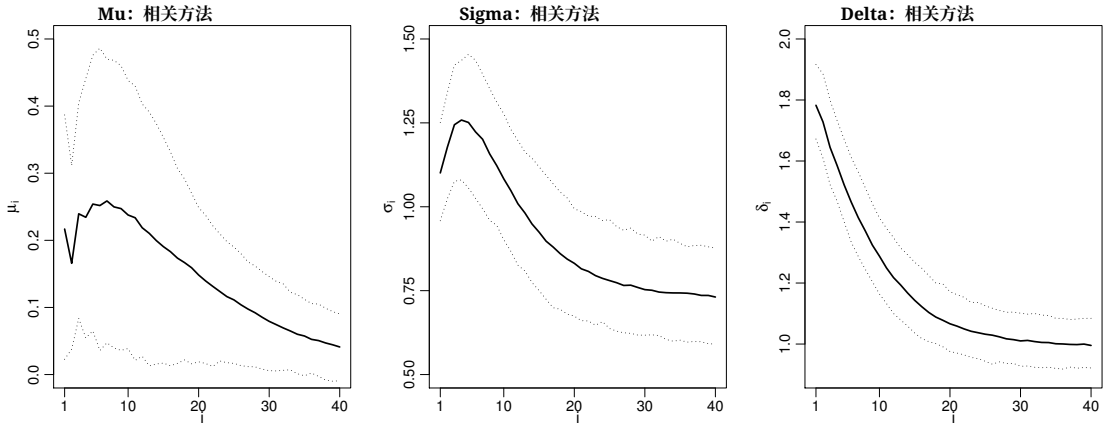

图 5：展示残差向量的 δ-Laplace 参数 μ_i、σ_i 和 δ_i 如何随滞后变化的图 i，针对二阶马尔可夫高斯 Copula 时间序列（5.13），使用相关函数方法和递推关系（5.3），区块长度为 $k = 20$，基于该过程的 500 个实现值。实线连接全部 500 个实现值的中位数估计，虚线表示估计值的 0.025 和 0.975 经验分位数。使用 Papastathopoulos-Tawn 模型时，也得到了几乎相同的曲线（未显示）。

6 数据应用

6.1 数据

本节将说明我们的方法如何应用于法国奥尔良 1946—2012 年的每日最高气温测量值时间序列。这与 Winter&Tawn (2016年) 和 Winter & Tawn (2017) 分析的数据集相同。我们仅使用 6 月至 8 月的夏季月份，并假设不同年份的这些月份构成平稳过程的近似独立实现值。共有四个缺失值，且分别出现在不同年份。我们使用 R（R核心团队 2021）软件包`imputeTS` (Moritz & Bartz-Beielstein 2017）中的卡尔曼滤波器对缺失值进行了插补。

6.2 方法和模型

Winter & Tawn (2016年) 根据滞后 1 处的偏自相关函数与零无显著差异这一事实，对奥尔良数据拟合了一阶马尔可夫模型。基于更精细的诊断方法（与第 3.5 节所述方法类似），Winter & Tawn (2017) 拟合了一个 3 阶马尔可夫模型。两篇论文均发现，渐近独立模型适用于这些数据。这里，我们考虑假设参数 $\boldsymbol{\alpha}_{1:k}$的结构为平稳自回归序列的自相关函数的渐近独立模型，如第 5.1 节所述。我们采用半参数方法和模型 1 归一化，拟合 1 阶、2 阶和 3 阶的自相关函数。为将其预测与我们自己的模型进行比较，我们还拟合了 Winter& Tawn(2016年) 的一阶马尔可夫模型。

我们将研究奥尔良在三周期间内，日最高气温若在期间开始时超越35°C时，日最高气温若干函数的模拟与估计；也就是说，我们只考虑第4.1节所述的前向模拟。35°C的气温大致相当于一年一遇的重现水平，即平均而言，这一气温每年会被超越

一次。正如Winter& Tawn(2016年)所讨论的，连续三天日平均最高气温超过35°C，可能会使奥尔良的超额死亡率增加17%至47%。

由于我们关注的是三周期间，因此需要能够从初始超阈值开始向前模拟20个步骤，所以在拟合我们的模型时，我们使用的区块长度为 $k = 20$ 。与第5节中模拟示例相比，由于数据的分段结构，模型拟合还会出现一个不太明显的额外复杂性。具体而言，如果 X_t 是在拉普拉斯尺度上超过我们的拟合阈值 u的超越，并且发生在8月12日之后，那么当年可用的连续值将少于20个，因此我们无法从 $\boldsymbol{X}_{t+1:t+20}$获得复合似然贡献。在这种情况下，我们仍然通过 $\boldsymbol{X}_{t+1:t+m}$ 允许复合似然贡献，其中 $m < 20$ 是可用的最大滞后。

为了量化参数估计值和其他感兴趣量的不确定性，我们采用区块长度为20的移动区块自助法，模拟生成1000个与原始数据集长度和结构相同的重复数据集。对于任何估计值，都可以通过计算所有自助样本中估计值的标准差来获得近似标准误。类似地，通过计算所有样本中估计值的0.025和0.975经验分位数，可以得到自助法95%的置信区间。在第6.4节中，由于计算簇函数概率的成本较高，我们使用较少数量的自助样本，即300个，来评估估计值的不确定性。

6.3 诊断方法

第3.1节的边际模型应用于奥尔良气温数据，所用阈值为 $u^* = 29.7$ ，对应于0.9边际经验分位数。对超过该阈值的超阈值拟合广义帕累托分布，得到参数估计值$\hat{}\sigma = 2.801$(0.185)和$\hat{}\xi = -0.192$(0.042)，括号内为估计标准误。Winter & Tawn (2016)给出了参数稳定性图，为这一阈值选择提供了依据。图6展示了一阶相关模型的参数 α 和 β ，以及二阶模型偏自相关参数化的参数 r_1 和 r_2 的参数稳定性图。所有阶数的相关模型中， β 的稳定性图几乎完全相同；三阶模型参数的图也与二阶模型的相应参数图十分相似。根据这些图，我们选择拉普拉斯尺度上对应于30°C的阈值 u 来拟合条件模型。对于一阶模型，我们得到估计值$\hat{}$ $\hat{}\alpha = 0.807$ (0.124)和 $\beta = 0.211$(0.095)；对于二阶模型，得到$\hat{}r_1 = 0.742$(0.157)、$\hat{}r_2 = 0.162$ (0.323)、$\beta = 0.210$ (0.095)；对于三阶模型，得到$\hat{}r_1 = 0.764$(0.137)、$\hat{}r_2 = -0.048$(0.286)、$\hat{}r_3 = 0.163$ (0.338)$\hat{}$以及 $\beta = 0.209$(0.095)。我们注意到，二阶模型中 $[-0.799, 0.352]$ 的 r_2 近似95%置信区间包含零。对于三阶模型， r_2和 r_3的$[-0.749, 0.494]$ 和 $[-0.747, 0.602]$ 的95%置信区间也包含零。因此，初步证据表明，二阶和三阶模型所提供的高阶结构可能是冗余的。对于 Winter–Tawn 一阶马尔可夫模型，我们得到估计值$\hat{}\alpha = 0.734$(0.237)和 $\beta = 0.546$(0.129)。表2(a)比较了经验估计与基于模型的估计：在一段三周期间内，已知该期间开始时超越阈值，至少出现 s 次0.9经验分位数超越的概率；表2(b)则给出了使用0.95经验分位数时的对应比较。表3比较了次渐近尾部依赖度量 $\chi(v, d)$的估计值，该度量定义于式(3.14)，阈值 v 设为经验0.9和0.95经验分位数，滞后 d 从1变化至10。请记住，进行此类比较的理由是，在0.9分位数这类中等偏高的阈值下，经验估计应当相当准确，因此我们可以将模型给出的估计与相应的经验估计进行比较，从而发现模型可能存在的不足。从表2(a)可以看出，对于每个 s ，各阶相关函数模型的估计值都与经验估计十分接近，尤其是始终处于经验估计的一个标准误范围内。表2(b)中较高阈值的情况也大致如此。与经验估计相比，Winter–Tawn 模型往往低估概率，并且随着 s 增大，概率衰减过快。其表现不佳，部分可能是由于第5.2节末尾指出的问题：在如此

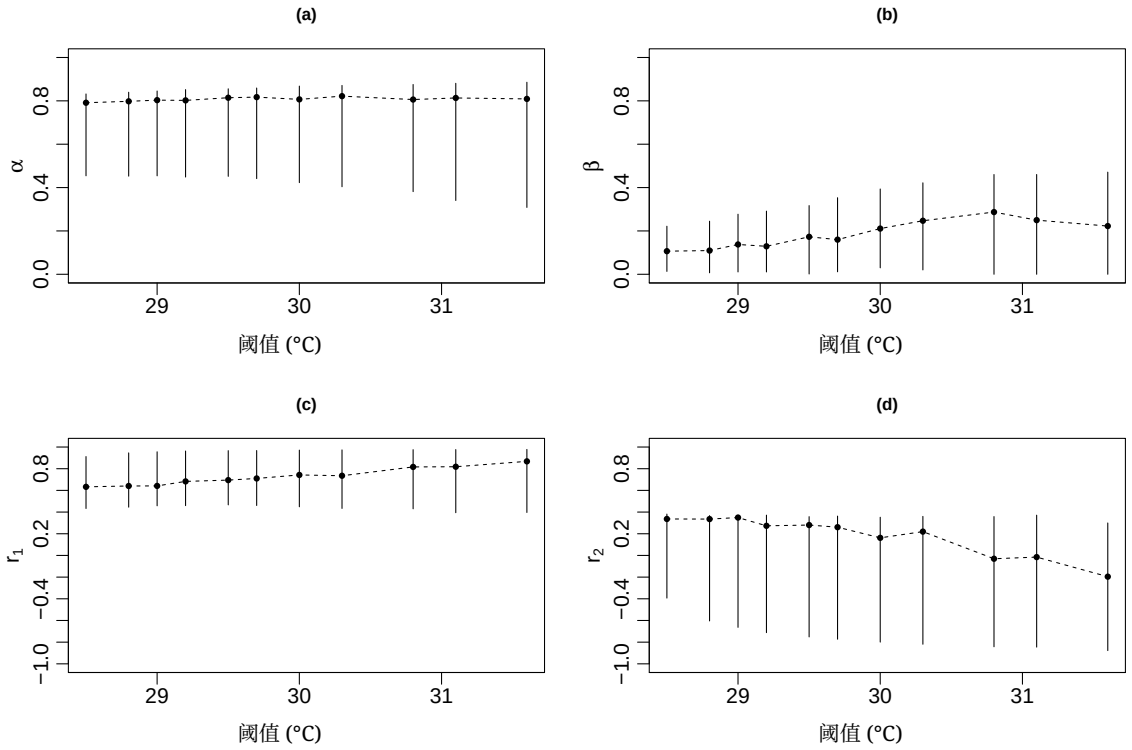

图6：图中展示了用于识别超阈值的不同阈值下的参数估计值及其 95% 自助法置信区间。第一行展示一阶马尔可夫相关模型的 (a) α 和 (b) β 的估计值。第二行展示二阶马尔可夫相关模型的 (c) r_1 和 (d) r_2 的估计值。

次渐近阈值下从该模型进行模拟。我们预计，在高得多的阈值下，这类问题不会那么明显。对于 $\chi(v,\ d)$，当 v 等于0.9经验分位数时，基于相关模型的估计值总体上与经验估计一致；尽管它们往往高估概率，但通常处于经验估计的一个标准误范围内。对于0.95经验分位数，整体情况也类似；虽然高估现象稍为明显，但基于相关模型的估计值通常处于经验估计的两个标准误范围内。基于 Winter–Tawn 模型的估计值在较短滞后 d 下与经验估计高度一致，但衰减过快。在较短滞后下，Winter–Tawn 模型前向模拟方案的问题并不十分明显；而当需要在长度大得多的区块上模拟并评估簇泛函时，该问题就明显得多。

对于第 6.4 节中估计的概率，表 2 是更相关的诊断方法，因为它考察三周期间所有观测值的联合行为，而 $\chi(v,\ d)$ 的估计值仅考察双变量行为。在其他应用中，要估计的簇函数概率的具体形式将决定哪些诊断方法最为相关。综合考虑所有证据，我们没有发现有力证据支持高阶相关模型优于一阶模型，因此我们选择一阶模型作为首选模型。不过，在第 6.4 节中，我们将给出所有模型产生的估计值。

我们在模拟中采用的一个重要建模假设是，残差向量 $\boldsymbol{Z}_{t+1:t+k|t}$在给定 X_t 的条件下与 $X_t > u$条件独立。我们通过散点图（未展示）评估了这一假设对于一阶相关函数模型是否合理，图中展示了 $Z_{t+i|t}$（即滞后 i处的残差向量分量）与 X_t、 $t \in T_u$之间的关系， $i \in 1$ 分别对应每个 $i \in 1:\ 20$。视觉上未见明显的依赖关系。更正式地说，我们计算了每个滞后 $i \in 1:\ 20$ 下各对 $\{(X_t,\ Z_{t+i|t})\}_{t \in T_u}$的Kendall关联度量 τ (Joe 1997，第2节2.1.9）的值。所有 τ的估计值都在 10^{-2} 量级，这表明残差向量各分量与阈值超越之间的任何依赖都非常弱。我们对所有自助抽样样本重复了这一过程。拟合值 τ 的自助法分布（即直方图）见附录 B。我们获得了95%的自助法置信

区间，并使用邦费罗尼校正对多重比较进行了校正。这20个区间均包含零，意味着在5%的显著性水平下不存在显著关联。尽管我们没有直接检验 $\boldsymbol{Z}_{t+1:t+k|t}$ 与 $X_t, t \in T_u$ 的独立性，但这些非正式诊断方法表明，这一假设是合理的。

(a)

模型	$s=2$	3	4	5	6	7	8	9	10	11
经验值	.880 (.012)	.749 (.021)	.639 (.027)	.527 (.030)	.431 (.032)	.355 (.031)	.273 (.030)	.218 (.027)	.159 (.024)	.110 (.020)
马尔可夫 1	.876 (.013)	.761 (.021)	.659 (.026)	.547 (.030)	.445 (.031)	.365 (.031)	.283 (.030)	.219 (.027)	.163 (.024)	.116 (.020)
马尔可夫 2	.877 (.013)	.761 (.021)	.658 (.026)	.546 (.030)	.444 (.032)	.365 (.031)	.283 (.030)	.219 (.027)	.164 (.024)	.116 (.020)
马尔可夫 3	.876 (.013)	.760 (.021)	.657 (.026)	.545 (.030)	.443 (.032)	.364 (.031)	.283 (.030)	.220 (.027)	.163 (.024)	.116 (.020)
WT	.623 (.033)	.406 (.040)	.269 (.037)	.180 (.031)	.122 (.025)	.083 (.020)	.057 (.015)	.038 (.012)	.026 (.009)	.018 (.007)

(b)

模型	$s=2$	3	4	5	6	7	8	9	10	11
经验值	.813 (.023)	.622 (.035)	.467 (.041)	.346 (.043)	.260 (.042)	.187 (.039)	.146 (.035)	.089 (.031)	.061 (.025)	.041 (.021)
马尔可夫 1	.810 (.013)	.644 (.021)	.512 (.026)	.395 (.030)	.292 (.031)	.216 (.031)	.158 (.029)	.113 (.027)	.077 (.024)	.052 (.020)
马尔可夫 2	.808 (.013)	.642 (.021)	.509 (.026)	.392 (.030)	.291 (.032)	.216 (.031)	.158 (.030)	.113 (.027)	.076 (.024)	.052 (.020)
马尔可夫 3	.808 (.013)	.642 (.021)	.511 (.026)	.393 (.030)	.290 (.032)	0.215 (.031)	.158 (.030)	.112 (.027)	.077 (.024)	.052 (.020)
WT	.582 (.031)	.356 (.037)	.224 (.034)	.143 (.028)	.092 (.023)	.059 (.018)	.039 (.014)	.025 (.011)	.016 (.008)	.011 (.006)

表2：在三周期间开始时超越相应阈值的条件下，至少发生 s 次超阈值的概率的经验值和基于模型的估计：(a) 0.9经验分位数；(b) 0.95分位数。Markov1-Markov3分别为指定阶数的相关函数模型，WT为Winter–Tawn一阶模型。括号内为估计标准误。

6.4 结果

在本节中，T 表示将奥尔良日最高气温从原始尺度（°C）转换为拉普拉斯尺度的函数，详见第3.1节。因此，如果 Y_t 表示奥尔良在第t日的气温，且 $X_t = T(Y_t)$，那么 X_t 服从标准拉普拉斯分布。

为了初步了解奥尔良日最高气温在超过35°C后如何变化，我们可以通过蒙特卡洛模拟计算一些简单的汇总统计量。例如，假设我们希望估计三周期间的最高气温期望值，条件是该期间开始时气温超过35°C。为此，我们从拟合的条件极值模型中模拟了 5×10^5 个 $\{\boldsymbol{X}^j_{1:21}\}_{j=1}^{5\times 10^5}$ 个 $\boldsymbol{X}_{1:21}$ 的实现值，条件为 $X_1 > T(35)$，方法是执行算法8 的步骤1至6，并使用 $d = 21$、$n = 5 \times 10^5$ 和 $v = T(35)$。然后进行反变换，得到5×10^5 个 $\{\boldsymbol{Y}^j_{1:21}\}_{j=1}^{5\times 10^5} = \{T^{-1}(\boldsymbol{X}^j_{1:21})\}_{j=1}^{5\times 10^5}$ 个 $\boldsymbol{Y}_{1:21}$ 的实现值，条件为 $Y_1 > 35$，并估计

(a)

模型	$d=1$	2	3	4	5	6	7	8	9	10
经验值	.542 (.021)	.357 (.025)	.285 (.023)	.265 (.022)	.235 (.021)	.222 (.019)	.190 (.019)	.187 (.018)	.188 (.017)	.181 (.018)
马尔可夫 1	.533 (.025)	.396 (.031)	.305 (.031)	.302 (.030)	.292 (.026)	.254 (.023)	.206 (.023)	.201 (.020)	.216 (.019)	.210 (.020)
马尔可夫 2	.590 (.024)	.410 (.031)	.318 (.032)	.298 (.031)	.269 (.027)	.278 (.023)	.216 (.023)	.220 (.020)	.265 (.019)	.228 (.020)
马尔可夫 3	0.557 (.025)	.399 (.032)	.298 (.032)	.263 (.031)	.254 (.026)	.233 (.023)	.218 (.023)	.210 (.020)	.206 (.019)	.212 (.020)
WT	.544 (.024)	.361 (.028)	.252 (.027)	.181 (.025)	.132 (.023)	.099 (.020)	.073 (.017)	.054 (.015)	.041 (.013)	.031 (.011)

(b)

模型	$d=1$	2	3	4	5	6	7	8	9	10
经验值	.517 (.030)	.316 (.036)	.233 (.037)	.199 (.036)	.175 (.036)	.135 (.030)	.126 (.027)	.131 (.024)	.121 (.021)	.117 (.019)
马尔可夫 1	.509 (.033)	.362 (.038)	.264 (.036)	.257 (.034)	.228 (.030)	.186 (.025)	.155 (.022)	.160 (.020)	.142 (.018)	.114 (.016)
马尔可夫 2	.513 (.035)	.364 (.042)	.258 (.040)	.236 (.037)	.205 (.032)	.206 (.026)	.170 (.022)	.178 (.020)	.165 (.018)	.120 (.016)
马尔可夫 3	.517 (.034)	.319 (.044)	.251 (.040)	.240 (.037)	.188 (.031)	.195 (.026)	.164 (.022)	.174 (.020)	.165 (.018)	.114 (.016)
WT	.501 (.032)	.315 (.034)	.212 (.031)	.147 (.026)	.105 (.022)	.075 (.018)	.055 (.014)	.040 (.011)	.030 (.009)	.022 (.008)

表3： $\chi(v,\ d)$，定义见（3.14），其中 d 从1到10。在（a）中，阈值 v 设为0.9经验分位数，而在（b）中使用0.95分位数。Markov1-Markov3为指定阶数的相关函数模型，WT为Winter–Tawn一阶模型。估计标准误见括号内。

$\mathbb{E}$(最大值 $\boldsymbol{Y}_{1:21} \mid Y_1 > 35$) 作为 $(5 \times 10^5)^{-1} \sum_{j=1}^{5\times 10^5}$ 最大值 $\boldsymbol{Y}_{1:21}^{j}$。对于阶数为 1、2 和 3 的相关函数模型，我们得到的估计值分别为 37.00°C（0.36°C）、36.98°C（0.38°C）和 36.98°C（0.37°C），括号内为估计标准误。对于 Winter–Tawn 模型，我们得到的估计值为 37.03°C（0.37°C）。其他汇总统计量也可以用类似方法计算。

如第 6.2 节所述，连续三天的日平均最高气温超过 35°C 可能导致奥尔良的超额死亡率增加 17% 至 47%。我们估计，在三周期间内发生这一事件的概率，条件是该期间开始时气温超过 35°C，估计结果为

$$(5 \times 10^5)^{-1} \sum_{j=1}^{5\times 10^5} \left(\mathbb{1}\left[\sum_{i=1}^{19} \mathbb{1}\left\{ \frac{1}{3} \sum_{m=0}^{2} Y_{i+m}^{j} > 35 \right\} \geq 1 \right] \right). \tag{6.1}$$

对于阶数为1、2和3的相关函数模型，该概率的估计值分别为0.457（0.095）、0.445（0.102）和0.437（0.105）。对于 Winter–Tawn 模型，我们得到的估计值为0.374（0.076）。

需要注意的是，迄今为止提到的示例中，感兴趣量的估计值都是在原始尺度上计算的，也就是说，使用 $\{\boldsymbol{Y}_{1:21}^{j}\}_{j=1}^{5\times 10^5}$ 而不是 $\{\boldsymbol{X}_{1:21}^{j}\}_{j=1}^{5\times 10^5}$。我们无法先在拉普拉斯尺度上计算估计值，再简单地进行反变换。不过，对于某些感兴趣量，可以直接在拉普拉斯尺度上计算估计值，无需进行反变换。例如，假设我们感兴趣的是：已知某三周期间开始时温度超过35°C，期间温度超过 s°C的概率。此时，由于 $\mathbb{P}(\max \boldsymbol{Y}_{1:21} > s \mid Y_1 > 35) = \mathbb{P}(\max \boldsymbol{X}_{1:21} > T(s) \mid X_1 > T(35))$，我们可以使用算法8直接估计该量，其中 $v = T(35)$ $d = 21$ 且 $g(\boldsymbol{x}) = \mathbb{1}[\max \boldsymbol{x}_{1:21} > T(s)]$。对于函数 g的以下取值，我们使用算法8，其中 $v = T(35)$ $d = 21$ 且 $n = 5 \times 10^5$，

$$g_s(\boldsymbol{x}) = \mathbb{1}\big[\max \boldsymbol{x}_{1:21} > T(s)\big], \tag{6.2}$$

$$g_s(\boldsymbol{x}) = \mathbb{1}\left[\sum_{i=1}^{21} \mathbb{1}[x_i > T(35)] \geq s\right], \tag{6.3}$$

$$g_s(\boldsymbol{x}) = \mathbb{1}\left[\max\left\{ i \in 1:20 \;\middle|\; \min \boldsymbol{x}_{t+1:t+i} > T(35) \text{ for some } t \in 0:(21-i) \right\} \geq s\right]. \tag{6.4}$$

(6.2)–(6.4) 中的所有函数 g 都取决于单个参数 s。以 (6.2) 中的 $s = 36$ 为例，通过算法8，可以估计在三周期间开始时温度已超越35°C的条件下，该期间温度超过36°C的概率。(6.3) 中的函数 g 用于估计三周期间内至少发生 s 次超阈值的概率，而 (6.4) 用于估计至少发生 s 次连续超越的概率。对于 (6.2)–(6.4) 中的每一个函数，$\mathbb{E}\{g_s(\boldsymbol{x}) \mid X_1 > T(35)\}$ 的估计值在参数 s取不同值时列于表4中。所有阶数的相关函数模型对于每个函数 g 以及参数 s的取值都给出了相近的估计值。对于估计超过温度 s的概率，即在 (6.2) 中令 g_s 时，Winter–Tawn 模型给出的估计值与相关模型相近，但对于其他函数，其估计值更小。不过，大多数 Winter–Tawn 估计值都处于我们的相关模型估计值的一个或两个标准误之内。因此，尽管第 6.3 节的诊断方法表明，我们的模型在较低阈值下表现更好，但对于本节所考虑的高得多的阈值，两者之间并不存在显著差异，这与第 5.2 节末尾报告的模拟结果一致。

在实际应用中，正如Winter & Tawn（2016年）所做的那样，可能更希望报告极值簇上泛函期望值的估计值，例如（6.2）-（6.4），而不是像本文这样报告固定区块长度上的估计值。正如第4.1节所指出的，这只需对本节的方法稍作修改即可。为进行这一修改，我们只需截断每个模拟块 $\{\boldsymbol{X}_{1:\,21}^{j}\}_{j=1}^{5\times 10^5}$ 使其对应一个簇，例如使用游程法（Smith 与 Weissman 1994），

然后利用这些簇计算任何感兴趣的量。区块长度为 21 已足够长，簇超过此长度的概率可以忽略不计；不过，如果情况并非如此，则可以采用更大的区块长度。

表4：基于模型的估计值，适用于 $\mathbb{E}\{g_s(\boldsymbol{X}_{1:21}) \mid X_1 > T(35)\}$，通过算法 8 得出。马尔可夫1–马尔可夫3分别表示指定阶数的相关函数模型，WT 表示 Winter–Tawn一阶模型。括号内为估计标准误。

函数	s	马尔可夫 1	马尔可夫 2	马尔可夫 3	WT
$g_s(\boldsymbol{x})$ 在（6.2）	36	.6998 (.0747)	.6966 (.0763)	.6970 (.0757)	.6804 (.0627)
	37	.4350 (.0968)	.4297 (.0998)	.4270 (.0992)	.4295 (.0877)
	38	.2259 (.0796)	.2211 (.0832)	.2206 (.0825)	.2483 (.0831)
	39	.1043 (.0538)	.1007 (.0571)	.1004 (.0566)	.1272 (.0622)
	40	.0464 (.0323)	.0443 (.0348)	.0445 (.0344)	.0542 (.0409)
	41	。0147 (。0181)	.0142 (.0197)	.0142 (.0194)	.0182 (.0243)
	42	.0014 (.0076)	.0013 (.0086)	.0013 (.0085)	.0040 (.0133)
$g_s(\boldsymbol{x})$（见6.3）	2	.6269 (.0949)	.6108 (.0984)	.6118 (.0959)	.5031 (.0632)
	3	.4040 (.0908)	。3967 (。0975)	.3904 (.0985)	.2716 (.0638)
	4	.2583 (.0716)	.2525 (.0767)	.2493 (.0764)	.1513 (.0517)
	5	.1650 (.0526)	。1626 (。0562)	.1605 (.0560)	.0859 (.0389)
	6	.1119 (.0379)	。1103（。0396）	.1093 (.0396)	.0494 (.0284)
	7	.0712 (.0265)	。0708 (。0271)	.0705 (.0272)	.0286 (.0205)
	8	.0482 (.0191)	.0481 (.0190)	.0487 (.0190)	.0164 (.0147)
	9	.0330 (.0137)	.0331 (.0136)	.0334 (.0135)	.0097 (.0105)
	10	.0197 (.0094)	.0192 (.0092)	.0199 (.0092)	.0058 (.0075)
	11	.0120 (.0068)	.0120 (.0066)	.0122 (.0066)	.0033 (.0054)
$g_s(\boldsymbol{x})$ 在（6.4）	2	.5310 (.0923)	.5115 (.0978)	.5163 (.0935)	.4629 (.0657)
	3	.2834 (.0755)	.2742 (.0840)	.2642 (.0872)	.2297 (.0612)
	4	.1513 (.0503)	.1470 (.0570)	.1417 (.0574)	.1187 (.0464)
	5	.0864 (.0320)	.0843 (.0365)	.0823 (.0371)	.0631 (.0331)
	6	.0466 (.0217)	.0457 (.0241)	.0453 (.0241)	.0340 (.0231)
	7	.0319 (.0161)	.0314 (.0169)	.0320 (.0170)	.0185 (.0160)
	8	.0222 (.0122)	.0220 (.0122)	.0225 (.0122)	.0101 (.0111)
	9	.0151 (.0092)	.0151 (.0092)	.0155 (.0092)	.0057 (.0078)
	10	.0101 (.0066)	.0100 (.0065)	.0105 (.0064)	.0032 (.0055)
	11	.0066 (.0049)	。0067（。0047）	.0068 (.0047)	.0017（.0039）

定理 3.1 的证明。

证明。 为证明无偏性，

$$\begin{aligned}\mathbb{E}_{\pi^*}\left\{\frac{g(\boldsymbol{X}_{1:d})}{S(\boldsymbol{X}_{1:d})}\right\} &= \int \frac{g(\boldsymbol{x})}{S(\boldsymbol{x})}\pi^*(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{\bar{p}}\int_{\mathcal{L}} \frac{g(\boldsymbol{x})}{S(\boldsymbol{x})}\left(\sum_{j=1}^{d}\mathbb{1}_{L_j}(\boldsymbol{x})\pi(\boldsymbol{x})\right)d\boldsymbol{x} \\ &= \frac{1}{\bar{p}}\int_{\mathcal{L}} \frac{g(\boldsymbol{x})}{S(\boldsymbol{x})}S(\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x} \\ &= \frac{1}{\bar{p}}\int_{\mathbb{R}^d} g(\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{\bar{p}}\mathbb{E}_{\pi}\{g(\boldsymbol{X}_{1:d})\}, \end{aligned} \tag{A.1}$$

其中（A.1）由 g 的支撑集为 $\mathcal{L}$可得。根据期望的线性性，可知 $\mathbb{E}_{\pi}\{g(\boldsymbol{X})\}$ 是无偏的。通过类似的计算，我们发现

$$\mathbb{E}_{\pi^*}\left\{\frac{g(\boldsymbol{X}_{1:d})^2}{S(\boldsymbol{X}_{1:d})^2}\right\} = \frac{1}{\bar{p}}\int_{\mathcal{L}} \frac{g(\boldsymbol{x})^2}{S(\boldsymbol{x})}\pi(\boldsymbol{x})d\boldsymbol{x} \tag{A.2}$$

因此

$$\mathrm{var}\left\{\frac{g(\boldsymbol{X}_{1:d})}{S(\boldsymbol{X}_{1:d})}\right\} = \frac{1}{\bar{p}}\int_{\mathcal{L}} \frac{g(\boldsymbol{x})^2}{S(\boldsymbol{x})}\pi(\boldsymbol{x})d\boldsymbol{x} - \left(\frac{\mathbb{E}_{\pi}\{g(\boldsymbol{X}_{1:d})\}}{\bar{p}}\right)^2. \tag{A.3}$$

利用独立性和（4.12），即可得到所述的方差公式。

为证明上界（4.15），注意到若 g 是一个指示函数，则 $g(\boldsymbol{x})^2 = g(\boldsymbol{x})$，因此

$$\int_{\mathcal{L}} \frac{g(\boldsymbol{x})^2}{S(\boldsymbol{x})}\pi(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathcal{L}} \frac{g(\boldsymbol{x})}{S(\boldsymbol{x})}\pi(\boldsymbol{x})d\boldsymbol{x} \le \int_{\mathcal{L}} g(\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathbb{R}^d} g(\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x} = \mathbb{E}_{\pi}\{g(\boldsymbol{X})\} \tag{A.4}$$

在 $S \ge 1$ 关于 $\mathcal{L}$ 上成立，并使用此上界即可得到所需结果。

最后，为了证明 $\widehat{\mathbb{E}}_{\pi}\{g(\boldsymbol{X})\}$ 在 g 有界时是相合估计量，我们只需验证

$$\left|\int_{\mathcal{L}} \frac{g(\boldsymbol{x})^2\pi(\boldsymbol{x})}{S(\boldsymbol{x})}d\boldsymbol{x}\right| < \infty \tag{A.5}$$

这很容易由以下事实得出：若 $|g(\boldsymbol{x})| < C < \infty$ $S(\boldsymbol{x}) \ge 1$ $\int_{\mathcal{L}} g(\boldsymbol{x})^2\pi(\boldsymbol{x})/S(\boldsymbol{x})\,d\boldsymbol{x}()$ $< C^2$ π 且 ()，则 () () ()，因为其是密度函数。

□

B 诊断图

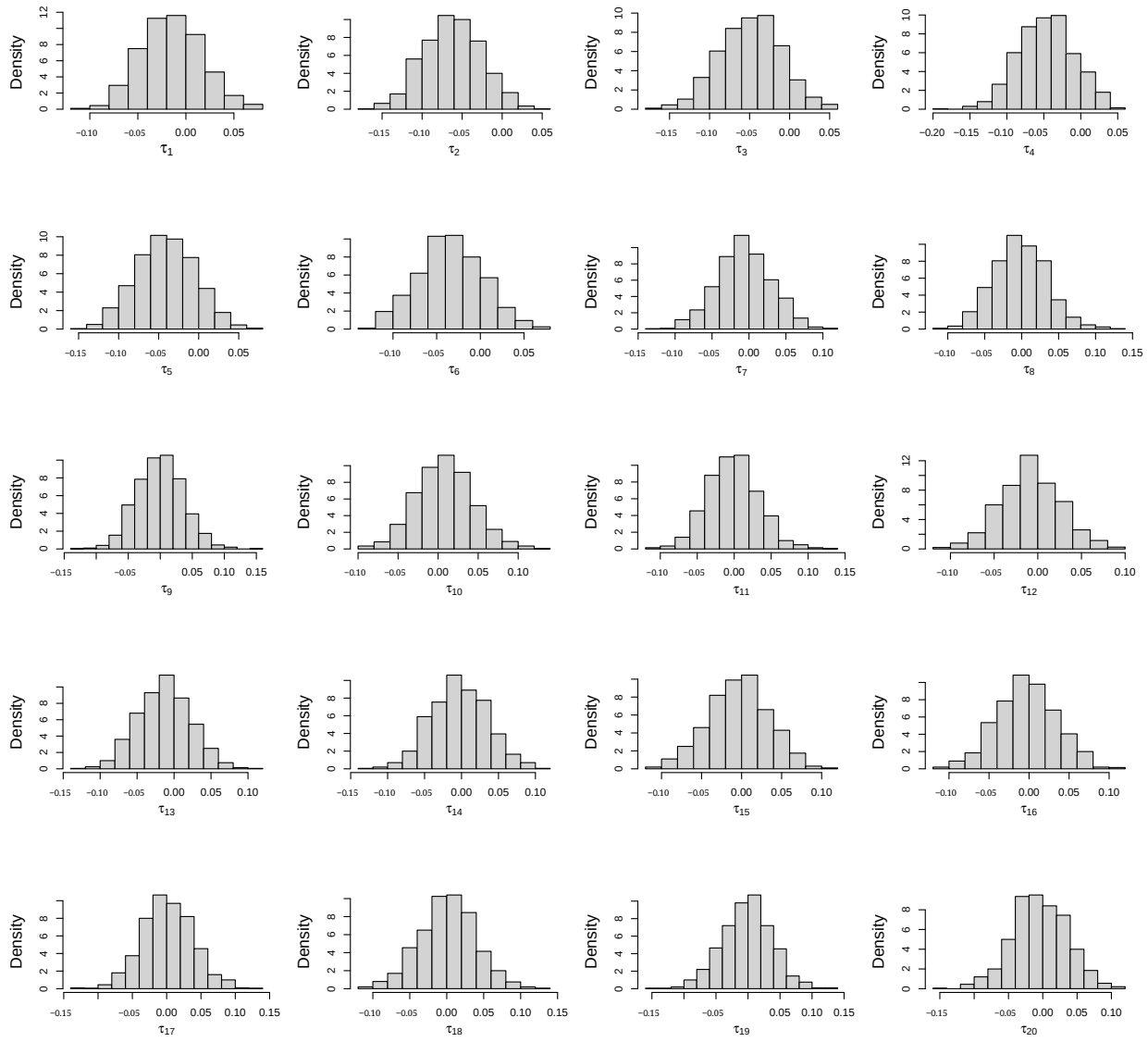

图7： τ_i、 $i \in 1$的自助法分布：对于一阶相关函数模型，其中 τ_iˆ是 $\{(X_t$、 $Z_{t+i|t})\}_{t \in T_u}$之间的Kendall关联度量。

致谢：本文基于Auld (2022）。作者感谢 Jonathan Tawn 和 Miguel de Carvalho 提供的宝贵意见，这些意见有助于改进本文。

参考文献

Adler, R., Blanchet, J. & Liu, J. (2012), '高斯随机场高水平超越的高效蒙特卡洛方法',《应用概率年刊》**22**, 1167–1214. Auld, G. (2022), 《环境极端事件统计建模》，博士论文，爱丁堡大学。**网址：** *https://era.ed.ac.uk/handle/1842/38730*Barndorff-Nielsen, O. & Schou, G. (1973), '基于偏自相关对自回归模型进行参数化',《多元分析杂志》**3**, 408–419. Basrak, B. & Segers, J. (2009), '正则变差多元时间序列', 随机过程及其应用 **119**, 1055–1080. Bradley, R. C. (2005), '强混合条件的基本性质：综述及若干未决问题',《概率调查》**2**, 107–144. Carlstein, E. (1986), '利用子序列值估计平稳序列一般统计量的方差',《统计年刊》**14**, 1171 – 1179. Coles, S. G. & Tawn, J. A. (1991), '多元极端事件建模',《皇家统计学会杂志：*B*辑（方法论）》**53**, 377–392. Davison, A. C. & Smith, R. L. (1990), '高阈值超越模型',《皇家统计学会杂志：*B*辑（方法论）》**52**, 393–442. de Bono, A., Giuliani, G., Kluser, S. & Peduzzi, P. (2004), '2003年夏季欧洲热浪的影响',*UNEP/DEWA/GRID Eur. Environ. Alert Bull.* **2**, 1–4. Eastoe, E. F. & Tawn, J. A. (2012), '次渐近阈值超越的簇最大值分布建模',《生物统计学》**99**, 43–55. Fischer, E. & Sch¨ar, C. (2010), '欧洲高影响热浪变化的一致地理格局',《自然·地球科学》**3**, 398–403. Genest, C., R´emillard, B. & Beaudoin, D. (2009), '联结函数拟合优度检验：综述与检验功效研究',《保险：数学与经济学》**44**, 199–213. Heffernan, J. E. & Tawn, J. A. (2004), '多元极值的条件方法（附讨论）',《皇家统计学会杂志：*B*辑（方法论）》**66**, 1–34. Janßen, A. & Segers, J. (2014), '马尔可夫尾链',《应用概率杂志》**51**, 1133–1153. Joe, H. (1997), 《多元模型与依赖概念》,《统计与应用概率专著》第73卷，Chapman & Hall, 伦敦。Keef, C., Papastathopoulos, I. & Tawn, J. A. (2013), '给定某个分量较大时多元变量条件分布的估计：Heffernan–Tawn模型的附加约束',《多元分析杂志》**115**, 396–404. Kunsch, H. R. (1989), '平稳观测的一般刀切法与自助法',《统计年刊》**17**, 1217 – 1241. Lahiri, S. N. (1999), '区块自助法的理论比较',《统计年刊》**27**, 386–404. Leadbetter, M. R. (1983), '平稳序列中的极值与局部依赖', *Z. Wahr-sch. verw. Gebiete***65**, 291 – 306.

Ledford, A. W. & Tawn, J. A. (2003), ‘时间序列极值中的依赖诊断方法’, 《皇家统计学会杂志：B辑（方法论）》65, 521–543. Moritz, S. & Bartz-Beielstein, T. (2017), ‘imputeTS：R中的时间序列缺失值插补’, 《R期刊》9, 207–218. Ng, C. T. & Joe, H. (2014), ‘基于复合似然信息准则的模型比较’, 《伯努利》20, 1738– 1764. Owen, A. B., Maximov, Y. & Chertkov, M. (2019), ‘对稀有事件并集进行重要性抽样及其在电力系统分析中的应用’, 《电子统计学期刊》13, 231–254. Papastathopoulos, I., Casey, A. & Tawn, J. A. (2023), ‘与高阶马尔可夫链极值相关的隐尾链及依赖参数递推方程’, arXiv:1903.04059 . Papastathopoulos, I., Strokorb, K., Tawn, J. A. & Butler, A. (2017), ‘马尔可夫链的极端事件’, 《应用概率进展》49, 134–161. Perfekt, R. (1997), ‘取值于 $\mathbb{R}^d$的一类马尔可夫链的极值理论’, 《应用概率进展》29, 138–164. Politis, D. N. & Romano, J. P. (1994), ‘平稳自助法’, 《美国统计学会会刊》89, 1303–1313. R核心团队 (2021), R：统计计算语言与环境，R统计计算基金会，维也纳，奥地利。Resnick, S. I. (1987), 《极值、正则变差与点过程》，Springer–Verlag，纽约。Robine, J.-M., Cheung, S. L. K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.-P. & Herrmann, F. R. (2008), ‘2003年夏季欧洲死亡人数超过70,000’, 《生物学报告》331, 171–178. Rue, H. & Held, L. (2005), 《高斯马尔可夫随机场：理论与应用》，《统计与应用概率专著》第104卷，Chapman & Hall/CRC，纽约。Shooter, R., Ross, E., Ribal, A., Young, I. R. & Jonathan, P. (2021), ‘基于卫星高度计测量的东北大西洋极端海况空间依赖’, 《环境计量学》32, e2674. Shooter, R., Ross, E., Tawn, J. & Jonathan, P. (2019), ‘海洋风暴强度的空间条件极值’, 《环境计量学》30, e2562. Sibuya, M. (1960), ‘双变量极值统计 I’, 《东京统计数学研究所年报》11, 195–210. Smith, R. L., Tawn, J. A. & Coles, S. G. (1997), ‘阈值超越的马尔可夫链模型’, 《生物统计学》84, 249–268. Smith, R. L. & Weissman, I. (1994), ‘极值指数的估计’, 《皇家统计学会杂志：B辑（方法论）》56, 515–528. Subbotin, M. F. (1923), ‘论误差频率定律’, 《数学文集》31, 296–301. Towe, R., Tawn, J., Lamb, R. & Sherlock, C. (2019), ‘基于模型的条件极值分布推断及其水文应用’, 《环境计量学》30, e2575. Varanasi, M. K. & Aazhang, B. (1989), ‘参数化广义高斯密度估计’, 《美国声学学会杂志》8 86, 1404–1415. Varin, C. & Vidoni, P. (2005), ‘关于复合似然推断与模型选择的说明’, 《生物统计学》92, 519– 528. Wadsworth, J. & Tawn, J. (2022), ‘通过单站点条件化研究高维空间极值’, Spatial Statistics 51, 100677.

Winter, H. C. & Tawn, J. A.（2016年），《法国中部热浪建模：极值依赖案例研究》，英国皇家统计学会学报*C*辑（应用统计）**65**，345–365。

Winter, H. & Tawn, J. A.（2017），《用于评估热浪风险的k阶马尔可夫极值模型》，《极值》**5**，393–415。