TI 2026-002/III
Tinbergen Institute Discussion Paper

Estimation and inference for the persistence of extremely high temperatures

Juan Juan Cai[1]
Yicong Lin[2]
Julia Schaumburg[3]
Chenhui Wang[4]

1 Vrije Universiteit Amsterdam, Tinbergen Institute
2 Vrije Universiteit Amsterdam, Tinbergen Institute
3 Vrije Universiteit Amsterdam, Tinbergen Institute
4 Vrije Universiteit Amsterdam

Estimation and inference for the persistence of extremely high temperatures[1]

Juan-Juan Cai[†,‡], Yicong Lin[†,‡], Julia Schaumburg[†,‡] and Chenhui Wang[†]

[†] *Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands.*
E-mail: {j.cai, yc.lin, j.schaumburg, c.h.wang}@vu.nl
[‡] *Tinbergen Institute, Valkenierstraat 59, 1018 XE, Amsterdam, the Netherlands.*

Summary We propose a nonparametric framework for estimating the extremal index that captures the persistence of extreme observations. The framework provides unified and simple procedures for verifying the well-known local dependence condition $D^{(d)}(u_n)$, which characterizes the extremal index yet is often assessed through heuristic checks, and for selecting d (a key parameter for estimation) when the condition holds. Under a general ϕ-mixing condition, we establish the asymptotic normality of the proposed estimator and prove the consistency of both the tuning parameter selection and the verification procedure for the $D^{(d)}(u_n)$ condition. Simulation studies show improved performance relative to two commonly used methods in terms of empirical mean squared errors. We analyze summer apparent temperature data for nine European cities from 1940 to 2025. The results show strong evidence of persistence in extreme temperatures for all cities, with such extremes typically lasting at least two days. The probability of two-day extreme-temperature events is two to four times higher in the most recent three decades relative to 1940–1974.

Keywords: *Extremal index, extremal serial dependence, nonparametric, heatwaves.*

1. INTRODUCTION

Heatwaves can have severe impacts on human health, infrastructure, and economic activity, making their modeling and prediction a matter of scientific and policy interest. Definitions often combine local high-temperature thresholds with a minimum duration,

[1] An earlier version of this paper, focusing primarily on the methodological aspects, was circulated under the title "Statistical inference on $D^{(d)}(u_n)$ condition and estimation of the extremal index."

typically two or more consecutive days.[2] From a statistical perspective, estimating heatwave occurrence probabilities is challenging due to scarce tail observations and the need for a dependence notion suited to extremes.

We propose a nonparametric approach to estimate the probability of severe heatwaves, using extreme value theory that accounts for persistence in extremes. Our framework centers around the so-called $D^{(d)}(u_n)$ condition (Chernick et al., 1991), a local dependence condition which restricts the occurrence of multiple exceedances above a high threshold (u_n) to lie within a time window (d) and thereby captures the dependence structure of extreme events. In particular, we propose a procedure that unifies the validation of the $D^{(d)}(u_n)$ condition, the data-driven selection for the time-window parameter d when the condition holds, and nonparametric estimation of the extremal index θ (Leadbetter, 1983). We show empirical evidence that daily summer temperature extremes satisfy the $D^{(2)}(u_n)$ condition, yielding a simple estimator of the heatwave probability.

Several key theoretical results are established. In particular, we show that under the $D^{(d)}(u_n)$ condition, the minimal value exists at which a distinct switching behavior arises in the asymptotic order of the extremal index estimator: the order changes at this value while remaining identical both below and above it. We exploit this dichotomous pattern, reminiscent of the classical $I(0)/I(1)$ distinction in trend analysis (e.g., Canjels and Watson, 1997; Harvey et al., 2007; Perron and Yabu, 2009), to jointly validate the $D^{(d)}(u_n)$ condition and select d, and we establish consistency of both procedures together with asymptotic normality of the estimator under a flexible ϕ-mixing condition for strictly stationary time series. Simulation results support the theoretical findings.

We adopt the framework to investigate extreme summer apparent temperatures for nine European cities. Comparing an early period (1940–1974) with a recent one (1991–2025), we document that extreme temperatures have become substantially higher, which is in line with the Intergovernmental Panel on Climate Change's findings on global warming in Europe (IPCC, 2021). Furthermore, we find statistically significant evidence for both periods that temperature extremes exhibit persistence, typically lasting at least two days when they occur. On the other hand, we do not find evidence that the per-

[2]The World Health Organization, for instance, adopts a minimum duration of two days in its Heat–Health Action Plan (Matthies, 2008). Definitions used by national meteorological institutes usually range from three to six days, see European Centre for Medium-Range Weather Forecasts (2023).

sistence levels differ across the two periods. Under the $D^{(2)}(u_n)$ condition, which holds for all periods and cities, we estimate heatwave severity as the probability that extreme temperatures persist for at least two consecutive days. We find statistically significant increases for all cities in 1991–2025 relative to 1940–1974, which are driven mainly by changes in marginal exceedance probabilities.

Our work relates to two strands of literature. First, changes in heatwave occurrence attributable to anthropogenic climate change have been widely studied in the extreme event attribution literature in climate science, initiated by Stott et al. (2004). Most existing statistical approaches to heatwave attribution rely on univariate extreme value models, applied either to daily temperatures or to temporally aggregated indices such as block maxima or rolling multi-day averages (e.g., Naveau et al., 2020; Vautard et al., 2020). Aggregation provides a pragmatic way to reflect duration, but it compresses the temporal structure of extreme events into a single scalar quantity and does not explicitly model tail persistence. Although we do not make causal claims in this paper, our heatwave severity estimator can be readily incorporated into attribution studies. More closely related are the hierarchical model by Reich et al. (2014) and the Markov-switching model for heatwaves by Shaby et al. (2016). While flexible, both of these Bayesian frameworks are fully parametric and computationally intensive. In contrast, our framework is nonparametric and straightforward to implement.

The second strand of literature we relate to is the statistical literature on dependence in time series extremes. The $D^{(d)}(u_n)$ condition underpins a substantial body of work (see, e.g., Ferreira and Ferreira, 2018; Holešovskỳ and Fusek, 2020), yet its verification commonly relies only on diagnostic plots proposed by Süveges (2007) and Ferreira and Ferreira (2018) that lack theoretical justification. Regarding the estimation of the nonparametric estimation of the extremal index θ, the work most closely related to ours is Hsing (1993), where asymptotic results are obtained under the assumptions that the window parameter d is known and that the data generating process is m-dependent. Both requirements are more restrictive than those imposed in our framework. Similarly to the verification of the $D^{(d)}(u_n)$ condition, Holešovskỳ and Fusek (2025) propose graphical diagnostics and several tests building on Süveges and Davison (2010) for the censored estimator of θ, which differs from our nonparametric formulation. A related strand of literature studies extremal index estimation without imposing the $D^{(d)}(u_n)$ condition.

Classical approaches include the blocks and runs estimators (Smith and Weissman, 1994; Weissman and Novak, 1998). Their asymptotic approximation is developed under a deterministic threshold u_n, which implicitly assumes knowledge of the unknown stationary distribution (Robert, 2009). In contrast, our thresholds are data-dependent and therefore random, which renders the theoretical analysis considerably more involved, as uniform convergence results are required to accommodate threshold randomness. Further approaches include the inter-exceedance times estimator of Ferro and Segers (2003), the (pseudo) maximum likelihood estimators studied in Northrop (2015) and Berghaus and Bücher (2018), and the recent moment estimators developed by Bücher and Jennessen (2020). We find in our extensive simulations that our estimator, which explicitly exploits the $D^{(d)}(u_n)$ condition, outperforms the widely used inter-exceedance times and maximum likelihood estimators in terms of empirical mean squared errors.

The rest of the paper is organized as follows. Section 2 introduces the proposed nonparametric framework and its associated asymptotic theory. Simulation results are reported in Section 3, and Section 4 examines the persistence of extreme summer temperatures across nine European cities. Section 5 concludes. The main proofs are collected in the Appendix, while additional proofs and supplementary simulation and empirical results are provided in Online Supplement. All code used for the simulation study and empirical application is available at `https://github.com/ohhwangch/persistence_ts`.

2. THE NONPARAMETRIC FRAMEWORK

Let $\{X_t,\, t \in \mathbb{Z}\}$ be a strictly stationary sequence of random variables with a continuous marginal distribution function F. Whenever no confusion arises, we write $\{X_t\}$ to denote $\{X_t,\, t \in \mathbb{Z}\}$. We observe a subset of this sequence with sample size $n \in \mathbb{Z}^+$. In our empirical application, the data $\{X_1, \ldots, X_n\}$ correspond to daily maximum apparent temperatures during summer seasons. Since the persistence level of extreme events can have severe consequences, for instance in the climate system, it is important to assess the strength of extremal serial dependence in $\{X_t\}$, which is the aim of this paper.

To this end, we define $U_t = F(X_t)$, which maps the marginal distribution to the standard uniform distribution on $[0, 1]$. This standardization isolates the extremal serial dependence structure in the sequence, as the marginal distribution F itself carries no information about extremal temporal dependence. That is, the two processes $\{X_t\}$

and $\{U_t\}$ share the same extremal serial dependence structure. We say that $\{X_t\}$ (or equivalently $\{U_t\}$) has an extremal index $\theta \in (0, 1]$, if

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{1\le t\le n} U_t \le 1 - \frac{u}{n}\right) = e^{-\theta u}, \qquad u > 0, \tag{2.1}$$

see, for instance, Leadbetter (1983). Clearly, when there is no extremal serial dependence in $\{X_t\}$, the left-hand side of (2.1) reduces to $\lim_{n\to\infty} \mathbb{P}\big(\max_{1\le t\le n} U_t \le 1 - u/n\big) = \lim_{n\to\infty} (1 - u/n)^n = e^{-u}$. Accordingly, $\theta = 1$ characterizes the case of no extremal serial dependence, with smaller values of θ reflecting stronger dependence.

However, the existence of the extremal index θ is not immediate. To guarantee its existence, the literature commonly imposes the following two "mixing" conditions on the dependence structure of the sequence (see, e.g., Leadbetter et al., 1983; Smith and Weissman, 1994). These conditions underpin our approach.

Condition $D(u_n)$ Let $\{u_n\}$ be a sequence of constants. For any integers $1 \le i_1 < \cdots < i_q < j_1 < \cdots < j_{q'} \le n$ such that $j_1 - i_q \ge l \ge 1$, the following condition is assumed to hold:

$$\left|\mathbb{P}\Big(\max_{1\le t\le q} U_{i_t} \le u_n, \max_{1\le t\le q'} U_{j_t} \le u_n\Big) - \mathbb{P}\Big(\max_{1\le t\le q} U_{i_t} \le u_n\Big)\,\mathbb{P}\Big(\max_{1\le t\le q'} U_{j_t} \le u_n\Big)\right| \le \alpha_{n,l}, \tag{2.2}$$

where $\lim_{n\to\infty} \alpha_{n,l_n} = 0$ for some sequence $l_n = o(n)$ and $l_n \to \infty$.

Condition $D(u_n)$ is reminiscent of the common α-mixing condition (see, e.g., Davidson, 1994, Chapter 14) and describes the long-range dependence of extremes. This condition is mild and widely used, and it is implied by several other mixing assumptions, for instance by the uniform mixing property introduced in Section 2.1 below. Next, we present the second key condition, often referred to as the local dependence condition.

Condition $D^{(d)}(u_n)$ For α_{n,l_n} specified in the $D(u_n)$ condition, there exists a sequence of integers r_n such that $r_n \to \infty$, with $n\alpha_{n,l_n}/r_n \to 0$ and $l_n/r_n \to 0$ as $n \to \infty$. Moreover, for a positive integer $d \ge 1$,

$$\lim_{n\to\infty} n\,\mathbb{P}\Big(U_1 > u_n \ge U^{\max}_{2,d},\, U^{\max}_{d+1,r_n} > u_n\Big) = 0, \tag{2.3}$$

where $U^{\max}_{i,j} := -\infty$ for $i > j$ and $U^{\max}_{i,j} := \max_{i\le t\le j} U_t$ for $i \le j$.

Intuitively, $D^{(d)}(u_n)$ limits the local occurrence of multiple exceedances over a threshold, thereby constraining the dependence among extremes; the extremes separated by a gap of d steps are approximately independent. We shall discuss a formal approach for verifying this local dependence condition in Section 2.3. The following lemma, adapted from Chernick et al. (1991, Corollary 1.3), provides the necessary and sufficient conditions for the existence of an extremal index θ for $\{X_t\}$ under Conditions $D(u_n)$ and $D^{(d)}(u_n)$.

Lemma 2.1. (Existence of θ) *Let $\{X_t\}$ be a strictly stationary sequence of random variables such that for some $d \geq 1$ the conditions $D(u_n)$ and $D^{(d)}(u_n)$ hold for $u_n = u_n(u) = 1 - u/n$ for all $u > 0$. Then the extremal index of $\{X_t\}$ exists and is equal to θ if and only if*

$$\lim_{n\to\infty} \mathbb{P}\big(U_{2,d}^{\max} \leq u_n \,\big|\, U_1 > u_n\big) = \theta, \qquad \forall u > 0. \tag{2.4}$$

The condition (2.4) further illustrates the role of θ. For example, setting $d = 2$ and $\theta = 1$ in (2.4) implies that, conditional on U_1 exceeding a high threshold u_n, the probability that the subsequent variable U_2 also exceeds the same threshold is approximately zero. Hence, the extremal behavior can be regarded as independent.

It is important to note that the value of d satisfying both (2.3) and (2.4) is not unique. First, observe that if the $D^{(d)}(u_n)$ condition holds, then $D^{(s)}(u_n)$ also holds for any finite $s \geq d$. Indeed, given (2.3), we have, as $n \to \infty$,

$$n\,\mathbb{P}\big(U_1 > u_n \geq U_{2,s}^{\max},\, U_{s+1,r_n}^{\max} > u_n\big) \leq n\,\mathbb{P}\big(U_1 > u_n \geq U_{2,d}^{\max},\, U_{d+1,r_n}^{\max} > u_n\big) \to 0.$$

Second, for $d \in \mathbb{Z}^+$, if the limit of $\mathbb{P}\big(U_{2,d}^{\max} \leq u_n \,\big|\, U_1 > u_n\big)$ exists, we denote it by

$$\Delta(d) := \lim_{n\to\infty} \mathbb{P}\big(U_{2,d}^{\max} \leq u_n \,\big|\, U_1 > u_n\big). \tag{2.5}$$

Lemma 2.1 implies that $\Delta(s) = \theta$ for all $s \geq d$. Since the choice of d is not unique, we assume for now that there exists a minimal value d_L such that

$$d_L = \min\big\{d \in \mathbb{Z}^+ : \Delta(d) = \theta\big\}. \tag{2.6}$$

Then any estimator of $\Delta(d)$ with $d \geq d_L$ can be used as an estimator of θ. We establish the conditions that ensure the existence of d_L in Theorem 5.1 in Appendix A. Without loss of generality, we assume throughout the paper that there also exists an upper bound $d_U \geq d_L$ such that $\Delta(d)$ exists for all $d \leq d_U$.

The remainder of this section proceeds as follows. First, under the $D^{(d)}(u_n)$ condition for some d (the $D(u_n)$ is imposed throughout), we propose a nonparametric estimator of $\Delta(d)$ and establish its asymptotic properties. Building upon these results, we show that the first difference of the estimator of $\Delta(d)$ exhibits a change in asymptotic order at $d = d_L$ under the $D^{(d)}(u_n)$ condition. By exploiting this property, we construct a selector of d_L. If such a change in asymptotic order is not observed, then the $D^{(d)}(u_n)$ condition can be considered violated, which provides a way to validate Condition $D^{(d)}(u_n)$.

2.1. Nonparametric estimation of $\Delta(d)$

We first approximate $\Delta(d)$ in (2.5) by taking $u_n = u_n(k) = 1 - k/n$, where $k = k(n)$ is an intermediate sequence satisfying $k \to \infty$ and $n/k \to \infty$ as $n \to \infty$. Since $\mathbb{P}(U_1 > u_n) = k/n$, an approximation of $\Delta(d)$ can be obtained as follows:

$$\Delta(d) \approx \frac{n}{k}\,\mathbb{P}\Big(U_{2,d}^{\max} \le 1 - \frac{k}{n} < U_1\Big) \approx \frac{n}{k}\cdot\frac{1}{n}\sum_{t=1}^{n-d+1} \mathbb{1}\Big\{U_{t+1,t+d-1}^{\max} \le 1 - \frac{k}{n} < U_t\Big\}$$
$$= \frac{1}{k}\sum_{t=1}^{n-d+1} \mathbb{1}\Big\{X_{t+1,t+d-1}^{\max} \le F^{-1}\Big(1 - \frac{k}{n}\Big) < X_t\Big\}, \tag{2.7}$$

where $X_{i,j}^{\max} = \max_{i \le t \le j} X_t$, and $\mathbb{1}\{\cdot\}$ denotes an indicator function. The quantity on the right-hand side above is not yet an estimator, as it depends on the unknown marginal cumulative distribution function (CDF) F, which can be naturally estimated via the *empirical* CDF. This leads to our estimator of $\Delta(d)$ given by

$$\hat{\Delta}_n(d) := \frac{1}{k}\sum_{t=1}^{n-d+1} \mathbb{1}\big\{X_{t+1,t+d-1}^{\max} \le X_{n-k,n} < X_t\big\}, \tag{2.8}$$

where $X_{1,n} \le X_{2,n} \le \ldots \le X_{n,n}$ denote the order statistics of the sample. It is worth highlighting a key distinction: (2.7) uses a *deterministic* threshold $F^{-1}(1 - k/n)$, while (2.8) relies on the *random* threshold $X_{n-k,n}$, which introduces additional complications in the proofs as discussed later.

In what follows, $\hat{\Delta}_n(d)$ is used as an estimator of the persistence level θ for extreme observations. However, as previously discussed, any estimator of $\Delta(d)$ with $d \ge d_L$ can be used to estimate θ. Before developing a selector of d_L, we first establish the asymptotic properties of $\hat{\Delta}_n(d)$. Our theory builds on a ϕ-mixing condition that characterizes the extremal dependence structure of the process $\{U_t\}$, or equivalently $\{X_t\}$. Let $\mathcal{H}_l^s =$

$\sigma\big(\mathbb{1}\{U_t \geq 1 - k/n\}, l \leq t \leq s\big)$ and define the uniform mixing coefficient

$$\phi_n(l) = \max_{s\geq 1} \sup_{A\in\mathcal{H}_1^s, B\in\mathcal{H}_{s+l}^n, \mathbb{P}(A)>0} \big|\,\mathbb{P}(B \mid A) - \mathbb{P}(B)\big|. \tag{2.9}$$

We now introduce the required assumptions.

Assumption 2.1. *Let $x, y \in [1/2, 3/2]$ be some constants.*

A1 There exist positive sequences r_n and l_n such that $r_n \to \infty$, $r_n/n \to 0$, $l_n/r_n \to 0$, and $n\,\phi_n(l_n)/r_n \to 0$, as $n \to \infty$.

A2 For r_n satisfying Condition A1,

$$\lim_{n\to\infty} \frac{n}{k} \sum_{t=d+1}^{r_n} \mathbb{P}\left(U_{2,d}^{\max} < 1 - \frac{kx}{n} < U_1,\, U_t > 1 - \frac{ky}{n}\right) = 0.$$

A3 For r_n satisfying Condition A1,

$$\lim_{n\to\infty} \frac{n}{k} \sum_{t=1}^{r_n} \mathbb{P}\left(U_1 > 1 - \frac{kx}{n},\, U_{t+1} > 1 - \frac{ky}{n}\right) = \Lambda_1(x, y) \in [0, \infty),$$

$$\lim_{n\to\infty} \frac{n}{k} \sum_{t=1}^{r_n} \mathbb{P}\left(U_1 > 1 - \frac{k}{n},\, U_{t+2,t+d}^{\max} < 1 - \frac{k}{n} < U_{t+1}\right) = \lambda_1 \in [0, \infty).$$

A4 For $j = d$ and $j = d - 1$, there exist constants $\ell_j > 0$ and $\rho > 0$ such that as $t \to 0$,

$$t^{-1}\mathbb{P}\Big(U_{1,j}^{\max} > 1 - t\Big) - \ell_j = O\big(t^{\rho}\big). \tag{2.10}$$

Note that Assumption A1 implies Condition $D(u_n)$ and also ensures the absolute regularity of the sequence; see Bradley (2005). This assumption arises from the use of the common Bernstein blocking technique in the time series literature to establish a central limit theorem (see, e.g., Davidson, 1994, Chapter 24.5). Assumption A2 is a strengthened version of $D^{(d)}(u_n)$, with a similar assumption found in Chernick et al. (1991, Eq. (1.2)). When $d = 1$, Assumption A2 reduces to the so-called $D'(u_n)$ condition in Leadbetter et al. (1983, Chapter 3), which implies $\theta = 1$, and when $d = 2$, it coincides with the $D''(u_n)$ condition in Leadbetter and Nandagopalan (1989). Assumption A3 imposes technical conditions on the tail dependence structure of $(X_1, \ldots, X_{r_n})$ for deriving the asymptotic long-run variance of $\hat{\Delta}_n(d)$. In practice, both limiting quantities in A3 can be estimated by substituting the probabilities with their empirical counterparts. Assumption A4 ensures the existence of θ (see (A.2) in Appendix A) and removes the asymptotic bias of the estimator.

Theorem 2.1 below establishes the asymptotic normality of $\hat{\Delta}_n(d)$ for any $d \in \mathbb{Z}^+$.

Theorem 2.1. *Suppose that* $\sum_{t=r_n}^{n}(1-t/n)\,\phi_n(t)=o(1)$, $r_n\,k/n=o(1)$, *and that* $k=o\left(n^{2\rho/(2\rho+1)}\right)$. *Under Assumption 2.1,*

$$\sqrt{k}\left(\hat{\Delta}_n(d)-\theta\right)\xrightarrow{d}\mathcal{N}(0,\sigma^2),\qquad n\to\infty, \tag{2.11}$$

for any $d\in\mathbb{Z}^+$, *where* $\sigma^2=\theta(1-2\lambda_1)+\theta^2(2\Lambda_1(1,1)-1)$.

Building on Theorem 2.1, we construct a consistent selector of d_L in the following section, which subsequently leads to our final estimator of the extremal index θ; see (2.21) below. It is worth highlighting that Hsing (1993) derives a related result under the more restrictive assumption of m-dependence (and assuming d_L is known), whereas we work under a general mixing condition. Weissman and Novak (1998) establish the asymptotic normality of the runs and blocks estimators using a deterministic threshold, namely $F^{-1}(1-k/n)$, rather than the data-dependent threshold $X_{n-k,n}$ considered here, as seen by the contrast between (2.7) and (2.8). Employing $F^{-1}(1-k/n)$ substantially simplifies the proofs, as it uses additional unknown information in F, whereas deriving asymptotic results under a data-dependent threshold requires us to derive some uniform convergence results; see Proposition 5.1–5.2 in Appendix A.

2.2. Determination of the lower bound d_L

Estimating $\Delta(d)$ for $d\geq d_L$ necessitates selecting an appropriate value of d_L. We exploit the properties of $\Delta(\cdot)$ to estimate the minimal value d_L such that $\Delta(d)=\theta$ for $d\in[d_L,d_U]$, as defined in (2.6). By construction, the function $d\mapsto\Delta(d)$ is non-increasing. Then it is immediate that $\Delta(d)>\theta$ for $1\leq d<d_L$, and $\Delta(d)=\theta$ for $d\in[d_L,d_U]$. This dichotomy provides a straightforward approach to select d_L. Let $\delta(\bullet):=\Delta(\bullet)-\Delta(\bullet+1)$ be a difference operator. Then $\delta(d_L-1)>0$ and $\delta(d)=\theta-\theta=0$ for any $d_L\leq d\leq d_U-1$. We can select d_L once $\delta(\bullet)$ can be estimated. Following the plug-in principle, we define

$$\hat{\delta}_n(\bullet):=\hat{\Delta}_n(\bullet)-\hat{\Delta}_n(\bullet+1), \tag{2.12}$$

where $\hat{\Delta}_n(\bullet)$ is provided in (2.8). Next, we examine the asymptotic behavior of $\hat{\delta}_n(d)$ for a given $d\geq 1$. To this end, we employ techniques similar to those used in the proof of Theorem 2.1, and therefore require that the conditions of Theorem 2.1, or analogous ones, hold for both d and $d+1$.

Assumption 2.2. *For some $d \in \mathbb{Z}^+$, the following conditions are assumed to hold jointly for the pair $(d, d+1)$.*

B1 Assumption A4 holds for $j = d$, and for $x, y \in [1/2, 3/2]$, there exist nonnegative constants $\Lambda_2(x,y)$, λ_2, $\tilde{\lambda}_1$, and λ_3 such that the following limits exist:

$$\lim_{n\to\infty} \frac{n}{k} \sum_{t=d+1}^{r_n} \mathbb{P}\Big(U_{2,d}^{\max} < 1 - \frac{kx}{n} < U_1,\, U_{t+1,t+d-1}^{\max} < 1 - \frac{ky}{n} < U_t\Big) = \Lambda_2(x,y), \quad (2.13)$$

$$\lim_{n\to\infty} \frac{n}{k} \sum_{t=d+1}^{r_n} \mathbb{P}\Big(U_{2,d}^{\max} < 1 - \frac{k}{n} < U_1,\, U_{t+1,t+d}^{\max} < 1 - \frac{k}{n} < U_t\Big) = \lambda_2, \quad (2.14)$$

$$\lim_{n\to\infty} \frac{n}{k} \sum_{t=1}^{r_n} \mathbb{P}\Big(U_1 > 1 - \frac{k}{n},\, U_{t+2,t+d}^{\max} < 1 - \frac{k}{n} < U_{t+1}\Big) = \tilde{\lambda}_1, \quad (2.15)$$

$$\lim_{n\to\infty} \frac{n}{k} \sum_{t=d+1}^{r_n} \mathbb{P}\Big(U_{2,d}^{\max} < 1 - \frac{k}{n} < U_1,\, U_t > 1 - \frac{k}{n}\Big) = \lambda_3. \quad (2.16)$$

B2 Assumption 2.1 also holds when d is replaced by $d+1$.

For a given pair $(d, d+1)$ that satisfies Assumption 2.2, we immediately obtain that the result of Theorem 2.1 holds with d replaced by $d+1$, owing to Assumption B2. Meanwhile, Assumption B1 ensures the existence of the asymptotic long-run variance for this given d by constraining the probabilities of threshold exceedances, in the spirit of Condition $D^{(d)}(u_n)$, thereby yielding an asymptotic approximation. Taken together, these two assumptions imply the following limiting distribution of $\hat{\delta}_n(d)$, which is subsequently used to establish the consistency of our selection of d_L.

Theorem 2.2. *For some $d \in \mathbb{Z}^+$, if Assumption 2.2 holds for the pair $(d, d+1)$, then*

$$\sqrt{k}\,\big(\hat{\delta}_n(d) - \delta(d)\big) \xrightarrow{d} \mathcal{N}\big(0, \kappa^2(d)\big), \qquad n \to \infty,$$

where $\kappa^2(d) = \delta^2(d)[2\Lambda_1(1,1) - 1] - 2\delta(d)\big(\tilde{\lambda}_1 - \lambda_1 + \lambda_3 - 1/2\big) + 2\Lambda_2(1,1) - 2\lambda_2$.

Building on Theorem 2.2, we immediately obtain that for any $d \geq d_L$ such that the pair $(d, d+1)$ satisfies Assumption 2.2, the limiting distribution in Theorem 2.2 becomes degenerate. This observation underlies the construction of our selector for d_L, which will be formalized later.

Corollary 2.1. *Suppose Assumption 2.2 holds for the pair $(d, d+1)$ for some $d \geq$*

$d_L \geq 1$, and Assumption A2 holds with d replaced by d_L. Then we have

$$\sqrt{k}\,\hat{\delta}_n(d) = o_{\mathbb{P}}(1). \tag{2.17}$$

If $d_L \geq 2$ and Assumption 2.2 holds for the pair $(d_L - 1, d_L)$, then

$$\hat{\delta}_n(d_L - 1) = \delta(d_L - 1) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right), \tag{2.18}$$

where $\delta(d_L - 1) > 0$.

The proof of Corollary 2.1 is straightforward. For $d \geq d_L$, $\delta(d) = 0$ as noted earlier. If Assumption A2 holds with d replaced by d_L, then $\Lambda_2(1,1) = 0$ and $\lambda_2 = 0$, implying $\kappa^2(d) = 0$, hence (2.17). Eq. (2.18) follows directly from Theorem 2.2 and the definition of d_L in (2.6).

In view of Corollary 2.1, $\hat{\delta}_n(d)$ exhibits a change in asymptotic order at $d = d_L$. Specifically, we have $\hat{\delta}_n(d) = o_{\mathbb{P}}(k^{-1/2})$ for $d \geq d_L$ and $\hat{\delta}_n(d) = O_{\mathbb{P}}(1)$ for $d = d_L - 1$. This is similar to the unit root literature, where the asymptotic orders of estimators, such as trend coefficient estimators in regression models, differ depending on whether the error process is $I(0)$ or $I(1)$ (see, e.g., Harvey et al., 2007; Perron and Yabu, 2009). Exploiting this change in asymptotic orders, we propose the following selector for d_L based on the intermediate sequence k:

$$\hat{d}_L(k) = \min\left\{d \in \mathbb{Z}^+ : \max_{d \leq s \leq d_U} \hat{\delta}_n(s) < \frac{1}{\sqrt{k}}\right\}. \tag{2.19}$$

If the assumptions in Corollary 2.1 hold, the consistency of $\hat{d}_L(k)$ follows.

Proposition 2.1. *Let $d_U \geq d_L$. Suppose the conditions of Corollary 2.1 hold, then*

$$\lim_{n\to\infty} \mathbb{P}\big(\hat{d}_L(k) = d_L\big) = 1. \tag{2.20}$$

where $\hat{d}_L(k)$ is defined in (2.19).

Substituting $\hat{d}_L = \hat{d}_L(k)$ into (2.8) gives our final estimator $\hat{\theta}_n = \hat{\theta}_n(\hat{d}_L)$ of θ, where

$$\hat{\theta}_n(\hat{d}_L) := \hat{\Delta}_n(\hat{d}_L) = \frac{1}{k}\sum_{t=1}^{n-\hat{d}_L+1} \mathbb{1}\left\{X^{\max}_{t+1,t+\hat{d}_L-1} \leq X_{n-k,n} < X_t\right\}. \tag{2.21}$$

Given the consistency of $\hat{d}_L$, it follows that $\hat{\theta}_n(\hat{d}_L)$ preserves the asymptotic normality established in Theorem 2.1 with $d = d_L$.

A remark on practical implementation is in order. Two parameters, k and d_U, must be specified in advance when computing $\hat{d}_L(k)$ in (2.19). The parameter d_U defines the upper bound of the search range, and its specific choice is not sensitive to the procedure, as d_L is typically small in empirical applications. Regarding the parameter k, different values may be used for selecting $\hat{d}_L(k)$ and for estimating $\hat{\theta}_n$, respectively. In practice, we recommend first selecting d_L and then substituting its estimate into (2.21). The optimal choice of k depends on the convergence rate of the underlying model; nevertheless, our simulation study below shows that the proposed procedure performs robustly over a broad range of k values.

2.3. The verification of Condition $D^{(d)}(u_n)$

Corollary 2.1 also suggests a way to conduct inference for validating the $D^{(d)}(u_n)$ condition. This condition is fundamental not only to our estimation of the persistence parameter θ, but also to several existing methods in the literature. For instance, Süveges (2007) studies a likelihood-based estimator of θ under Condition $D^{(2)}(u_n)$. Ferreira and Ferreira (2018) propose an estimator of θ by linking a stationary sequence satisfying Condition $D^{(d)}(u_n)$ to a regenerative process satisfying Conditions $D^{(1)}(u_n)$ or $D^{(2)}(u_n)$. Finally, Holešovskỳ and Fusek (2020) consider an estimator of θ based on censoring inter-exceedance times under Condition $D^{(d)}(u_n)$. In view of (2.17), for any given integer $d_0 \geq 1$, Condition $D^{(d_0)}(u_n)$ can easily be validated by checking whether

$$\max_{d_0 \leq d \leq d_U} \hat{\delta}_n(d) < \frac{1}{\sqrt{k}}. \tag{2.22}$$

This procedure is consistent in the following sense: if Condition $D^{(d_0)}(u_n)$ holds, then (2.22) will hold with probability tending to one. If Condition $D^{(d_0)}(u_n)$ is violated, two scenarios may arise. First, d_L exists but $d_L > d_0$, in which case Condition $D^{(d_L)}(u_n)$ holds; then, by (2.18), we have

$$\sqrt{k} \max_{d_0 \leq d \leq d_U} \hat{\delta}_n(d) \geq \sqrt{k}\hat{\delta}_n(d_L - 1) \gg 1, \qquad n \to \infty.$$

Thus, the procedure remains consistent under this scenario. In this case, one proceeds to check $d_0 + 1$, and so on. The second scenario arises when Condition $D^{(d)}(u_n)$ fails for all $d \in \mathbb{Z}^+$. In this case, the assumptions required for Corollary 2.1 are no longer satisfied, implying that all existing methods relying on Condition $D^{(d)}(u_n)$ are not applicable.

Table 1: Simulation DGPs and their corresponding theoretical properties

Model	DGP	Parameter	θ	d_L
AR-N	$X_t = \varrho X_{t-1} + \epsilon_t$, $t \geq 1$, $X_0 \sim \mathcal{N}(0, 1/(1-\varrho^2))$, $\epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$	$\lvert\varrho\rvert < 1$	1	1
Moving Maxima	$X_t = \max_{0 \leq i \leq m} \epsilon_{t+i}$, $t \geq 1$, $\epsilon_t \overset{\text{i.i.d.}}{\sim} F_\epsilon$, $F_\epsilon(x) = \mathbb{P}(\epsilon_t \leq x) = \exp\big(-1/(mx)\big)$	$m \geq 2$	$1/m$	2
Max AR	$X_t = \max\{\varrho X_{t-1}, \epsilon_t\}$, $t \geq 2$, $X_1 = \epsilon_1/(1-\varrho)$, $\epsilon_t \overset{\text{i.i.d.}}{\sim} F_\epsilon$, $F_\epsilon(x) = \mathbb{P}(\epsilon_t \leq x) = \exp(-1/x)$	$\varrho \in [0,1)$	$1-\varrho$	2
AR-C	$X_t = \varrho X_{t-1} + \epsilon_t$, $t \geq 1$, $X_0 \sim \text{Cauchy}(0,1)$, $\epsilon_t \overset{\text{i.i.d.}}{\sim} \text{Cauchy}(0, 1-\lvert\varrho\rvert)$	$\lvert\varrho\rvert < 1$	$1-\varrho$ if $\varrho \geq 0$ $1-\varrho^2$ if $\varrho < 0$	2 if $\varrho \geq 0$ 3 if $\varrho < 0$
ARCH	$X_t = \big(2 \times 10^{-5} + 0.7X_{t-1}^2\big)^{1/2} \epsilon_t$, $t \geq 1$, $\epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$	—	0.721	not exist

We further examine the finite sample performance of the procedure above under both scenarios in the simulation study.

3. SIMULATIONS

We first evaluate the finite sample performance of the procedure for verifying Condition $D^{(d)}(u_n)$ in Section 2.3, and then assess the performance of our estimator $\hat{\theta}_n(\hat{d}_L)$ in (2.21) using empirical mean squared errors (MSE). The data-generating processes (DGPs) and their corresponding properties are summarized in Table 1. Derivations of these theoretical properties are provided in Online Appendix S2.

Given the DGPs in Table 1, we further distinguish several cases based on the parameter settings used in the simulations. Setting $\varrho = 0$ in the AR-N model yields the IID case. For the AR-N and MAX-AR models, we set $\varrho = 0.5$, and for the Moving Maxima model we take $m = 3$. For the AR-C model, we use $\varrho = -0.5$. We note that the marginal distributions of the DGPs do not affect the simulation results. Any continuous transformation that alters the marginal distribution, and thereby changes the max-domain of attraction, leaves the outcomes unchanged. Throughout, we fix the sample size at $n = 5000$ and $d_U = 10$, and all reported results are based on 1000 Monte Carlo repetitions.

Table 2 reports the empirical acceptance rates obtained by applying the verification procedure in Section 2.3. Specifically, we sequentially verify whether $D^{(d_0)}(u_n)$ holds for $d_0 = 1, 2, 3$. If (2.22) is satisfied for a given d_0, we say that $D^{(d_0)}(u_n)$ is accepted. For the IID and AR-N models, $D^{(d_0)}(u_n)$ holds theoretically for all three values of d_0, since

Table 2: Empirical acceptance rate (in percent) for condition $D^{(d_0)}(u_n)$.

Model	d_L	$k = 50$			$k = 100$		
		$d_0 = 1$	$d_0 = 2$	$d_0 = 3$	$d_0 = 1$	$d_0 = 2$	$d_0 = 3$
IID	1	100	100	100	100	100	100
AR-N	1	56	100	100	2.5	99.8	100
Moving Maxima	2	0	100	100	0	100	100
MAX-AR	2	0	100	100	0	100	100
AR-C	3	3.5	3.6	100	0	0	100
Transformed AR-C	3	4.5	4.9	100	0	0	100
ARCH	not exist	16	99.1	100	0.4	91.6	100

the minimal theoretical value is $d_L = 1$. The procedure performs well for the IID case, whereas for the AR-N model it often fails when $d_0 = 1$: the failure rates are 44% for $k = 50$ and 97.5% for $k = 100$. However, for $d_0 = 2$ and 3, the failure rate is (nearly) zero for both choices of k. For the Moving Maxima and MAX-AR models, the procedure reaches optimal accuracy: it correctly fails to support $d_0 = 1$ and succeeds when $d_0 \geq 2$. For the AR-C model, since the transformation keep the extremal dependence structure unchanged, the procedure performs optimally for $k = 100$ for both versions, and the failure rates remain below 5% for $k = 50$. Finally, for the ARCH model, which does not satisfy $D^{(d)}(u_n)$ for any finite d (see Proposition S.2 in the supplement), and for which all existing methods relying on $D^{(d)}(u_n)$ therefore break down, we observe that $D^{(2)}(u_n)$ and $D^{(3)}(u_n)$ are frequently accepted for both choices of k. This is mainly due to the somewhat limited sample size. For example, when $n = 50,000$ and $k/n = 0.1$, the acceptance rate is zero for $d_0 = 1, \ldots, 9$ in simulation results not reported here.

Next, we evaluate the performance of $\hat{\theta}_n(\hat{d}_L)$, using the same threshold k for constructing $\hat{d}_L$. For comparison, we include two benchmark estimators: $\hat{\theta}_n^{\text{int}}$, the interval estimator of Ferro and Segers (2003, p. 549), and $\hat{\theta}_n^{\text{B,sl}}$, the sliding-block pseudo–maximum likelihood estimator of Berghaus and Bücher (2018, p. 2314). The block length for $\hat{\theta}_n^{\text{B,sl}}$ is set to n/k. The results are displayed in Figure 1. The sharp increase in the empirical MSE of our estimator $\hat{\theta}_n(\hat{d}_L)$ for the IID model occurs because $\hat{d}_L(k)$ tends to overestimate the true d_L when k is relatively large. This issue can be largely mitigated by a two-step procedure: first estimate d_L using a smaller k, and then use this estimate as an input for estimating θ without relying on the same k in both steps. For the other four models that

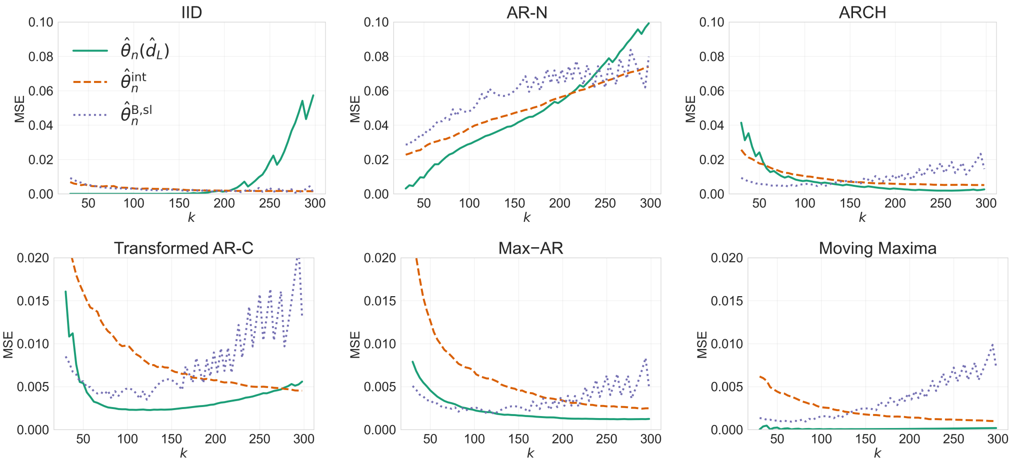

Figure 1: Empirical MSE of three estimators across six models, where $\hat{\theta}_n(\hat{d}_L)$ is defined in (2.21), $\hat{\theta}_n^{\text{int}}$ denotes the interval estimator of Ferro and Segers (2003, p. 549) and $\hat{\theta}_n^{\text{B,sl}}$ is the pseudo maximum likelihood estimators based on the sliding blocks proposed by Berghaus and Bücher (2018, p. 2314).

satisfy $D^{(d)}(u_n)$ condition, our estimator outperforms the two alternatives, as it has the smallest MSE over a sufficiently wide range of k.

Even for the ARCH model, the minimum MSE of our estimator remains smaller than that of the other two methods. For the ARCH model, we remark that θ can be well estimated by the runs estimator $\hat{\Delta}_n(r_n)$, where $\hat{\Delta}_n(\bullet)$ is given in (2.8) and r_n denotes a block length. Our procedure produces an estimator $\hat{d}_L$ such that the difference between $\hat{\Delta}_n(\hat{d}_L)$ and $\hat{\Delta}_n(d_U)$ is very small. Therefore, in finite samples, $\hat{d}_L$ can be used as a data-driven method for selecting the block length r_n.

4. EMPIRICAL STUDY

Our empirical analysis addresses two questions. First, we ask whether extreme summer temperatures in Europe exhibit dependence consistent with the $D^{(d)}(u_n)$ condition and, if so, whether the strength of this persistence has changed over time. Second, we assess whether the severity of heatwaves, measured as the probability of observing multi-day exceedances of high thresholds, has increased in recent decades.

We analyze daily maximum apparent temperature (X_t) during the summer months (June, July, and August) for nine European cities (London, Paris, Munich, Budapest, Milan, Barcelona, Rome, Valencia, and Athens), following D'Ippoliti et al. (2010). These

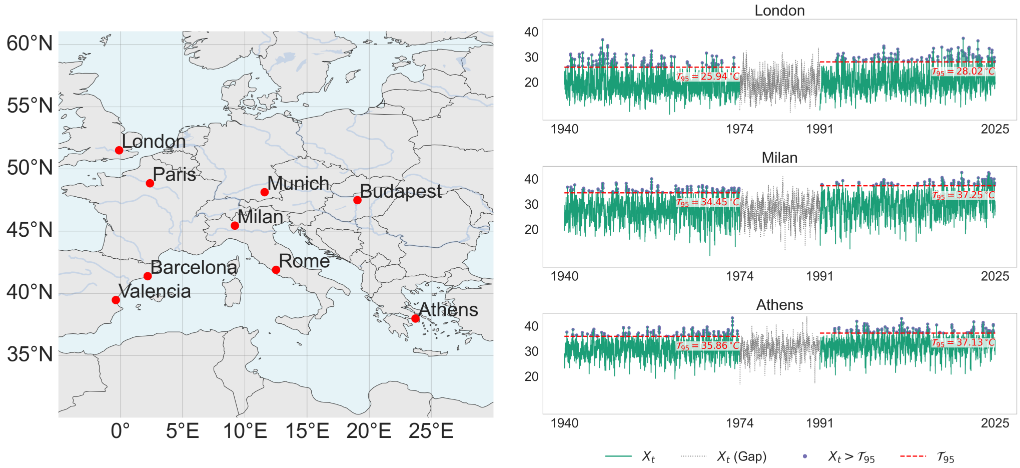

Figure 2: Left panel: Map of the nine selected cities. Right panel: Time series of X_t during summer days from 1940 to 2025 for three representative cities in the north, central, and southern regions of Europe: London, Milan, and Athens. The two subsamples covering 1940–1974 and 1991–2025 are shown as solid green lines, while observations in the intervening gap, denoted by X_t(Gap), are plotted as a gray dotted line. The red dashed lines labeled $\mathcal{T}_{95}$ represent the 95th percentiles within each subsample, and the purple dots mark observations exceeding these respective thresholds.

cities span a broad range of climatic environments across Europe, enabling meaningful comparisons between northern, central, and Mediterranean conditions. A map is shown in the left panel of Figure 2. Apparent temperature combines air temperature and dew-point temperature and thus reflects both thermal intensity and humidity-induced heat stress, which is the key determinant of adverse health impacts. Since our interest is the persistence of hazardous heat exposure rather than dry-bulb temperature alone, apparent temperature is more appropriate for defining extreme events (Steadman, 1979).

We obtain daily observations spanning 1940–2025 from the ERA5 reanalysis.[3] To satisfy the stationarity requirement underlying the $D^{(d)}(u_n)$ framework, we split the sample into two locally stationary subsamples: an early period from 1940 to 1974 (Period 1) and a recent period from 1991 to 2025 (Period 2), leaving 1975-1990 as a transition gap. Each period comprises 35 summers, corresponding to $n = 3220$ daily observations per

[3]See the link https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview. We use ERA5 hourly single-level data, including the 2 m dewpoint temperature T_{dewpt} and the 2 m air temperature T_{air}, covering the period from 1940 to 2025 and accessed on November 25, 2025. The apparent temperature X_t is computed as $X_t = -2.653 + 0.994\,T_{\text{air}} + 0.0153\,T_{\text{dewpt}}^2$.

city. Augmented Dickey-Fuller tests, both with and without a deterministic trend, reject the null hypothesis of unit-root nonstationarity for all time series in both periods at the 1% significance level.

The right panel of Figure 2 displays the time series X_t for three representative cities: London from northern Europe, Milan from central-southern Europe, and Athens from southern Europe. The green solid lines indicate the observations in the two subperiods, while the gray dotted lines correspond to the temporal gap. The clustering of exceedances above extreme percentiles in the two subperiods provides visual insight into the strength of extremal serial dependence within each period. The red dashed lines mark the 95% percentiles of X_t computed separately for the two subperiods. Two observations emerge. First, the 95% percentile in Period 2 is clearly higher than in Period 1, indicating an overall increase in temperature levels. In other words, extreme temperatures have become more extreme in recent decades. This pattern holds across all nine cities for extreme thresholds, see also Table 3. Second, although the cities exhibit higher temperatures in the later period, there is no visual evidence that the dependence structure of the most extreme 5% of observations within each subperiod undergoes a marked change across the two periods.

To analyze these observed patterns formally, we estimate θ in each period using the estimator $\hat{\theta}_n$ introduced in Section 2, which relies on a data-driven selection of d_L. For each city and period, we determine $\hat{d}_L$ by applying the procedure in (2.22). Figure 3 plots $\hat{\delta}_n(s)$ for $s = 1, \ldots, 4$, as defined in (2.12), over a range of $k/n \in [0.02, 0.08]$ for the three representative cities London, Milan, and Athens. The curve for $s = 1$ (green solid) stands out clearly from those for $s = 2, 3, 4$. According to our selection procedure in (2.19), and by comparing the curves of $\hat{\delta}_n(s)$ with $1/\sqrt{k}$ (dashed lines), we conclude that $\hat{d}_L = 2$ for both periods, all three cities, and all considered values of k/n. Additional results for all cities also support $\hat{d}_L = 2$ in almost all cases, with the exceptions of Barcelona in Period 1 and Valencia for $k/n \geq 6\%$ (see Figure S.3 in the Online Supplement). We conclude that Condition $D^{(d)}(u_n)$ holds with $d_L = 2$.

We subsequently compute $\hat{\theta}_n$ using the top 3%, 5%, and 7% observations of X_t within each period as the threshold. Table 3 reports the resulting estimates of the extremal index θ along with the corresponding temperature thresholds. As noted earlier, all cities exhibit a pronounced rise in high temperatures: the top 3% threshold in Period 1 is approximately

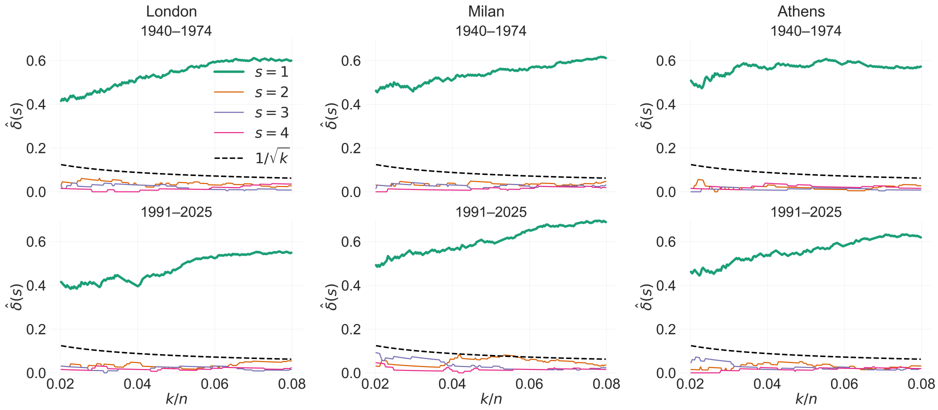

Figure 3: $\hat{\delta}_n(s)$ for $s = 1, \ldots, 4$, as defined in (2.12), over a range of $k/n \in [0.02, 0.08]$ for the three representative cities London, Milan, and Athens. The fact that $\hat{\delta}_n(1)$ lies above $1/\sqrt{k}$ while $\hat{\delta}_n(s)$ for $s = 2, 3, 4$ lie below $1/\sqrt{k}$ indicates that $\hat{d}_L = 2$ according to the selection procedure in (2.19).

Table 3: Estimates of θ from Section 2 for $k/n \in \{3\%, 5\%, 7\%\}$, together with corresponding thresholds $X_{n-k,n}$. The cities are ordered from the northernmost to the southernmost latitude when read from top to bottom and from left to right.

Period	City	$\hat{\theta}$			$X_{n-k,n}$ (in °C)			City	$\hat{\theta}$			$X_{n-k,n}$ (in °C)		
		3%	5%	7%	3%	5%	7%		3%	5%	7%	3%	5%	7%
1940–1974	London	0.54	0.45	0.39	27.30	25.94	25.04	Rome	0.52	0.44	0.42	34.92	34.32	33.84
1991–2025		0.60	0.53	0.45	29.34	28.04	27.12		0.43	0.36	0.35	38.14	37.49	37.06
1940–1974	Paris	0.55	0.48	0.45	31.52	30.15	29.01	Barcelona	0.60	0.46	0.40	31.38	30.86	30.56
1991–2025		0.61	0.58	0.53	33.90	32.29	31.36		0.51	0.48	0.40	34.46	33.81	33.36
1940–1974	Munich	0.54	0.48	0.48	31.32	30.32	29.56	Valencia	0.69	0.65	0.55	34.27	33.64	33.25
1991–2025		0.62	0.56	0.51	32.88	31.81	31.17		0.69	0.57	0.47	37.14	36.51	36.07
1940–1974	Budapest	0.58	0.52	0.50	33.44	32.49	31.79	Athens	0.46	0.43	0.43	36.63	35.86	35.37
1991–2025		0.54	0.48	0.44	35.56	34.77	34.14		0.51	0.45	0.38	38.00	37.14	36.68
1940–1974	Milan	0.54	0.46	0.41	35.14	34.46	33.89							
1991–2025		0.46	0.40	0.33	38.08	37.25	36.62							

at the same level as, and for some cities even lower than, the top 7% threshold in Period 2. We do not, however, observe substantial changes in θ across periods in Table 3. The variation that does appear is loosely related to latitude: for northwestern and central European cities, from London to Munich, persistence levels decline slightly, whereas a tendency toward higher persistence is evident for eastern cities (e.g., Budapest) and southern cities (from Milan to Valencia), with Athens as an exception. These patterns are not unexpected, given that the relevant percentiles in Period 2 are substantially

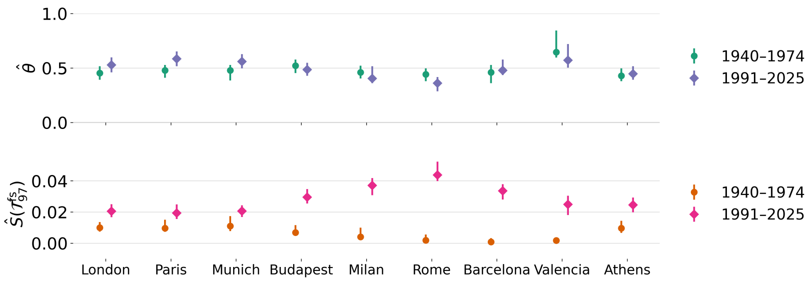

Figure 4: Estimates and 90%-level bootstrap confidence intervals of θ (top panel) and $S(\mathcal{T}_{97}^{\text{fs}})$ (bottom panel). For both panels, $k/n = 5\%$ and confidence intervals are based on 999 bootstrap replications.

higher than in Period 1. At such extreme temperature levels in Period 2, large shifts in persistence would be less likely to occur.

To further quantify the estimation uncertainty of θ, we adapt the block-type bootstrap procedure of Ferro and Segers (2003). Note that this bootstrap scheme also allows us to construct confidence intervals for the severity probability introduced later, within the same unified framework. Specifically, for a given time period, blocks (or clusters) are constructed using the estimated dependence parameter $\hat{d}_L$. The key idea of this bootstrap scheme is to preserve extremal temporal dependence within each block, while generating a sufficient number of approximately independent blocks to induce randomness. This approach differs from the conventional block bootstrap, where blocks typically have (approximately) equal lengths. Instead, block construction in our resampling scheme depends on the timing of exceedances, with exceedances assigned to the same extremal cluster according to the estimate $\hat{d}_L$. Full details of the resampling procedure and an accompanying simulation study are provided in Online Supplement, where we find satisfactory finite sample performance in terms of empirical coverage. Figure 4 presents the point estimates $\hat{\theta}$ for $k/n = 5\%$ in both periods, together with the corresponding 90% bootstrap confidence intervals. The intervals overlap across periods for all cities, indicating no evidence of changes in extremal serial dependence.

We can further estimate the heatwave severity probability $S(\mathcal{T})$, defined as the joint exceedance probability at a high temperature threshold $\mathcal{T}$. Given $\hat{d}_L = 2$, it is reasonable to define, for a high threshold $\mathcal{T}$, $S(\mathcal{T}) := \mathbb{P}(X_1 \geq \mathcal{T}, X_2 \geq \mathcal{T}) = \mathbb{P}(X_1 \geq \mathcal{T})\,\mathbb{P}(X_2 \geq \mathcal{T} \mid X_1 \geq \mathcal{T})$, i.e., the probability that two consecutive days exceed the threshold $\mathcal{T}$. The second term is related to the tail copula of (X_1, X_2). However, classical estimators of tail copulas are developed under the assumption of i.i.d. observations and are therefore not directly applicable to time series data. In the context of heatwave analysis, several parametric approaches have been proposed to model and estimate this quantity. For instance, Reich et al. (2014) employ a multivariate logistic copula, while Shaby et al. (2016) introduce a Markov-switching model to capture the tail dependence of (X_1, X_2).

We take a fully nonparametric approach, linking this quantity to the extremal index. Given that the $D^{(2)}(u_n)$ condition is supported by our data, we have $\lim_{\mathcal{T}\to x^*} \mathbb{P}(X_2 \geq \mathcal{T} \mid X_1 \geq \mathcal{T}) = 1-\theta$, where x^* denotes the right endpoint of the temperature distribution. Let $P_{\mathcal{T}} = \mathbb{P}(X_1 \geq \mathcal{T})$ denote the marginal exceedance probability, and let $\hat{P}_{\mathcal{T}}$ be its estimator; see Online Supplement for details. A plug-in estimator of $S(\mathcal{T})$ is then given by $\hat{S}(\mathcal{T}) = \hat{P}_{\mathcal{T}}(1-\hat{\theta})$. Let $\mathcal{T}_p^{\mathrm{fs}}$ denote the $p\%$ percentile of the full sample from 1940 to 2025. We take $k/n = 5\%$ and use $\mathcal{T}_{97}^{\mathrm{fs}}$ as $\mathcal{T}$ for evaluating the probability of joint occurrence above this extremal level. The bottom panel of Figure 4 displays the estimated values of $\hat{S}(\mathcal{T}_{97}^{\mathrm{fs}})$ for the two periods, together with bootstrap 90%-level confidence intervals constructed using the same resampling scheme as for θ. Across all cities, the probability of a two-day exceedance above $\mathcal{T}_{97}^{\mathrm{fs}}$ increases significantly in recent decades. In several cases, the point estimates for the later period are two to four times larger than those in the earlier period. This indicates a substantial increase in the probability of short heatwave spells in the last three decades compared with 1940–1974. Importantly, these increases are primarily due to changes in the marginal exceedance probability, largely through shifts in the location parameter of the marginal distributions; see Table S.1 in Online Supplement. Note also that the supplement provides a comparison with an alternative estimator of $S(\mathcal{T})$ based on the empirical distribution function. For a very high $\mathcal{T}$, this estimator frequently yields zero, as the sample may contain no exceedances.

5. CONCLUSION

We proposed a unified nonparametric framework that enables validation of the core local dependence condition $D^{(d)}(u_n)$ and provides procedures for tuning parameter selection and extremal index estimation. We established the asymptotic normality of the proposed extremal index estimator by deriving fundamental weak convergence results for empirical processes to accommodate random thresholds in our construction of the estimator. We showed the consistency of the validation procedure for the $D^{(d)}(u_n)$ condition and, conditional on this condition holding, proved the consistency of the tuning parameter selection procedure for the parameter that determines the lower bound required for the $D^{(d)}(u_n)$ condition to hold. An extensive simulation study supported our asymptotic results and showed that the proposed estimator substantially outperformed two commonly used existing estimators in terms of empirical MSE. We studied summer apparent temperature data for nine European cities over the period 1940–2025. The results reveal two important takeaways. First, there is clear evidence of extremal serial dependence: the $D^{(d)}(u_n)$ condition holds with $d \geq 2$ for all countries in both subperiods (1940–1974 and 1991–2025). This implies that when an extreme temperature event occurs, it persists for at least two days. Second, the heatwave severity probability increases substantially, with two-day exceedance probabilities typically two to four times larger in the more recent period compared with 1940–1974. These increases, however, are primarily due to shifts in the marginal exceedance probability, meaning that extreme temperatures become more extreme in the recent period rather than changes in the extremal index.

REFERENCES

Aldous, D. (1978). Stopping times and tightness. *Annals of Probability 6*(2), 335–340.

Berghaus, B. and A. Bücher (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Annals of Statistics 46*(5), 2307–2335.

Billingsley, P. (1999). *Convergence of Probability Measures.* John Wiley & Sons.

Bradley, R. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys 2*, 107–144.

Bücher, A. and T. Jennessen (2020). Method of moments estimators for the extremal index of a stationary time series. *Electronic Journal of Statistics 14*, 3103–3156.

Canjels, E. and M. W. Watson (1997). Estimating deterministic trends in the presence of serially correlated errors. *Review of Economics and Statistics 79*(2), 184–200.

Chernick, M., T. Hsing, and W. McCormick (1991). Calculating the extremal index for a class of stationary sequences. *Advances in Applied Probability 23*(4), 835–850.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford: Oxford University Press.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction.* Springer Verlag.

D'Ippoliti, D., P. Michelozzi, C. Marino, F. de'Donato, B. Menne, K. Katsouyanni,

U. Kirchmayer, A. Analitis, M. Medina-Ramón, A. Paldy, et al. (2010). The impact of heat waves on mortality in 9 European cities: results from the EuroHEAT project. *Environmental Health 9*(1), 37.

European Centre for Medium-Range Weather Forecasts (2023). Heat waves and cold spells in Europe derived from climate projections.

Ferreira, H. and M. Ferreira (2018). Estimating the extremal index through local dependence. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 54*(2), 587–605.

Ferro, C. A. T. and J. Segers (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 65*(2), 545–556.

Harvey, D. I., S. J. Leybourne, and A. R. Taylor (2007). A simple, robust and powerful test of the trend hypothesis. *Journal of Econometrics 141*(2), 1302–1330.

Holešovskỳ, J. and M. Fusek (2020). Estimation of the extremal index using censored distributions. *Extremes 23*(2), 197–213.

Holešovskỳ, J. and M. Fusek (2025). Statistical inference on the local dependence condition of extreme values in a stationary sequence. *Extremes 28*, 557–578.

Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *Annals of Statistics 21*(4), 2043–2071.

IPCC (2021). *Climate Change 2021: The Physical Science Basis.* Cambridge University Press.

Leadbetter, M. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 65*(2), 291–306.

Leadbetter, M., G. Lindgren, and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes.* New York: Springer-Verlag.

Leadbetter, M. and S. Nandagopalan (1989). On exceedance point processes for stationary sequences under mild oscillation restrictions. *In: Hüsler J., Reiss RD. (eds) Extreme Value Theory. Lecture Notes in Statistics 51*, 69–80.

Matthies, F. (2008). *Heat-health action plans: guidance.* World Health Organization.

Naveau, P., A. Hannart, and A. Ribes (2020). Statistical methods for extreme event attribution in climate science. *Annual Review of Statistics and Its Application 7*(1), 89–110.

Northrop, P. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes 18*(4), 585–603.

Perron, P. and T. Yabu (2009). Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics 151*(1), 56–69.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Reich, B. J., B. A. Shaby, and D. Cooley (2014). A hierarchical model for serially-dependent extremes: A study of heat waves in the western US. *Journal of Agricultural, Biological, and Environmental Statistics 19*(1), 119–135.

Robert, C. (2009). Inference for the limiting cluster size distribution of extreme values. *Annals of Statistics 37*(1), 271–310.

Shaby, B. A., B. J. Reich, D. Cooley, and C. G. Kaufman (2016). A Markov-switching model for heat waves. *Annals of Applied Statistics*, 74–93.

Smith, R. and I. Weissman (1994). Estimating the extremal index. *Journal of the Royal Statistical Society. Series B (Methodological) 56*(3), 515–528.

Steadman, R. G. (1979). The assessment of sultriness. part i: A temperature-humidity index based on human physiology and clothing science. *Journal of Applied Meteorology and Climatology 18*(7), 861–873.

Stott, P. A., D. A. Stone, and M. R. Allen (2004). Human contribution to the European heatwave of 2003. *Nature 432*(7017), 610–614.

Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes 10*, 41–55.

Süveges, M. and A. C. Davison (2010). Model misspecification in peaks over threshold analysis. *Annals of Applied Statistics 4*(1), 203 – 221.

Utev, S. (1991). On the central limit theorem for φ-mixing arrays of random variables. *Theory of Probability & Its Applications 35*(1), 131–139.

Vautard, R., M. van Aalst, O. Boucher, A. Drouin, K. Haustein, F. Kreienkamp, G. J. Van Oldenborgh, F. E. Otto, A. Ribes, Y. Robin, et al. (2020). Human contribution to the record-breaking June and July 2019 heatwaves in Western Europe. *Environmental Research Letters 15*(9), 094077.

Weissman, I. and S. Novak (1998). On blocks and runs estimators of the extremal index. *Journal of Statistical Planning and Inference 66*(2), 281–288.

APPENDIX A: MAIN PROOFS

We provide the proofs of Theorems 2.1 and 2.2. We first present several auxiliary results. Theorem 5.1 establishes a useful identity and ensures the existence of a minimal value as defined in (2.6). The auxiliary results in Propositions 5.1 and 5.2 establish weak convergence properties that allow our asymptotic approximations to accommodate random thresholds $X_{n-k,n}$, in contrast to, for example, Weissman and Novak (1998). Detailed proofs of these auxiliary results are provided in the supplement.

THEOREM 5.1. *Suppose that for all $1 \leq s \leq d_U$, the limit of $t^{-1}\mathbb{P}\big(U_{1,s}^{\max} > 1-t\big)$ exists as $t \to 0$, and define $\ell_s := \lim_{t\to 0} t^{-1}\mathbb{P}\big(U_{1,s}^{\max} > 1-t\big)$. Let $\ell_0 := 0$. Then*

$$\Delta(s) = \ell_s - \ell_{s-1}, \qquad 1 \leq s \leq d_U. \tag{A.1}$$

Moreover, if the $D(u_n)$ and $D^{(d)}(u_n)$ conditions hold for some $d \geq 1$, then there exists a $d_L \leq d$ such that

$$\theta = \Delta(s) = \ell_s - \ell_{s-1}, \qquad d_L \leq s \leq d_U, \tag{A.2}$$

and $\Delta(s) > \theta$ for $1 \leq s < d_L$.

PROOF. The proof is provided in the online supplement. □

To proceed, define, for $x \in [1/2, 3/2] =: \mathscr{X}$,

$$\tilde{\Delta}_n(x,d) = \frac{1}{k}\sum_{t=1}^{n-d+1} \mathbb{1}\left\{U_{t+1,t+d-1}^{\max} < 1 - \frac{kx}{n} < U_t\right\}. \tag{A.3}$$

Note that $\tilde{\Delta}_n(\cdot)$ is a *pseudo* estimator because the U_t's are unobservable when F is unknown. By the strict stationarity of the U_t's, one has

$$\mathbb{E}\big(\tilde{\Delta}_n(x,d)\big) = \frac{n-d+1}{k}\,\mathbb{P}\left(U_{2,d}^{\max} < 1 - \frac{kx}{n} < U_1\right) \to x\theta,$$

by Assumption A4 and (A.1). By (2.8), we also have

$$\hat{\Delta}_n(d) = \frac{1}{k}\sum_{t=1}^{n-d+1} \mathbb{1}\big\{U_{t+1,t+d-1}^{\max} \leq U_{n-k,n} < U_t\big\} = \tilde{\Delta}_n\left(\frac{n}{k}\left(1 - U_{n-k,n}\right), d\right).$$

Since $(n/k)\left(1 - U_{n-k,n}\right) - 1 = o_{\mathbb{P}}(1)$, we first derive the asymptotic properties of $\tilde{\Delta}_n(x,d)$ for $x \in \mathscr{X}$. Specifically, for $d \in \mathbb{Z}^+$, we aim to establish the weak convergence of $\nu_n(x,d)$ as a process indexed by $x \in \mathscr{X}$, where

$$\nu_n(x,d) := \sqrt{k}\big(\tilde{\Delta}_n(x,d) - \tilde{\Delta}_{0,n}(x,d)\big), \tag{A.4}$$

and $\tilde{\Delta}_{0,n}(x,d) = (n/k)\,\mathbb{P}\big(U_{2,d}^{\max} < 1 - kx/n < U_1\big)$. To this end, let $\Rightarrow$ denote weak convergence in the space of functions on the compact interval $\mathscr{X}$ having at most discontinuities of the first kind and endowed with the Skorokhod J_1 topology (see Aldous, 1978, for details), denoted by $\mathcal{D}(\mathscr{X})$.

PROPOSITION 5.1. *Under the assumptions of Theorem 2.1, for $d \in \mathbb{Z}^+$, we obtain the weak convergence of $\nu_n(x,d)$ as a process indexed by $x \in \mathscr{X}$:*

$$\big\{\nu_n(x,d)\big\}_{x\in\mathscr{X}} \Rightarrow \big\{W_d(x)\big\}_{x\in\mathscr{X}}, \qquad n \to \infty, \tag{A.5}$$

where W_d denotes a zero-mean continuous Gaussian process with covariance function

$$\mathbb{E}\Big(W_d(x)W_d(y)\Big) = \lim_{n\to\infty} \frac{n}{k}\sum_{t=0}^{d-1}\mathbb{P}\Big(U_{2,d}^{\max} < 1 - \frac{kx}{n} < U_1,\, U_{t+2,t+d}^{\max} < 1 - \frac{ky}{n} < U_{t+1}\Big),$$

for $x \le y$. In particular, $\mathbb{E}\big(W_d^2(x)\big) = x\theta$.

PROOF. The proof is provided in Online Supplement. □

By a similar but simpler proof, we can obtain the weak convergence for the empirical tail process: $\nu_n(x,1) = \sqrt{k}\big(k^{-1}\sum_{t=1}^{n}\mathbb{1}\big\{U_t > 1 - kx/n\big\} - x\big)$.

PROPOSITION 5.2. *Under the conditions of Theorem 2.1, we have $\{\nu_n(x,1)\}_{x\in\mathscr{X}} \Rightarrow \{\tilde{W}(x)\}_{x\in\mathscr{X}}$, as $n \to \infty$, where $\tilde{W}$ is a zero-mean continuous Gaussian process with covariance function $\mathbb{E}\big(\tilde{W}(x)\tilde{W}(y)\big) = \min(x,y) + \Lambda_1(x,y) + \Lambda_1(y,x)$.*

PROOF. The proof is similar to that of Proposition 5.1 and is therefore omitted. □

Note that this result is generally different from Proposition 5.1 with $d = 1$. The difference lies in the covariance structure. The two coincide with each other only when Condition A2 holds with $d = 1$, which implies $D^{(1)}(u_n)$ condition holds. Recall that $\hat{\theta}_n = \tilde{\Delta}_n(e_n, d)$, where $e_n := (n/k)\,(1 - U_{n-k,n})$. We suppress the dependence of e_n on k, as k itself depends on n. We now derive some asymptotic properties of e_n. First, by de Haan and Ferreira (2006, Theorem A.0.1 and Lemma A.0.2), Proposition 5.2 implies that $\sqrt{k}\,(e_n - 1) \xrightarrow{d} -\tilde{W}(1)$, as $n \to \infty$. In particular, one has $e_n - 1 = o_{\mathbb{P}}(1)$. Note that convergence in the Skorokhod topology implies uniform convergence when the limiting process is continuous (see, e.g., Billingsley, 1999, Chapter 12). Using the continuity of $\tilde{W}$

together with the triangle inequality, we obtain

$$\left|\sqrt{k}\,(e_n-1)+\nu_n(1,1)\right| \le \left|\nu_n(e_n,1)-\tilde{W}(e_n)\right| + \left|\nu_n(1,1)-\tilde{W}(1)\right| + \left|\tilde{W}(1)-\tilde{W}(e_n)\right| = o_{\mathbb{P}}(1). \tag{A.6}$$

Proof. (**Proof of Theorem 2.1**) For clarity of presentation, all processes involved in the proof are defined on a common probability space using the Skorohod representation theorem (see, e.g., Pollard, 1984). We use the same notation for these versions, although they are only equal in distribution to the original processes. We begin with the following decomposition, which follows from the definitions of $\tilde{\Delta}_n$ and $\nu_n(x,d)$: $\sqrt{k}\big(\hat{\Delta}_n(d)-\theta\big) = \sqrt{k}\big(\tilde{\Delta}_n(e_n,d)-\theta\big) = \nu_n(e_n,d)+\sqrt{k}\big(\tilde{\Delta}_{0,n}(e_n,d)-\theta\big)$. Let $R_n := \sqrt{k}\big(\tilde{\Delta}_{0,n}(e_n,d)-\theta\big)$. It suffices to show: *(a)* $|\nu_n(e_n,d)-\nu_n(1,d)| \xrightarrow{p} 0$; *(b)* $|R_n+\theta\nu_n(1,1)| \xrightarrow{p} 0$; *(c)* $\nu_n(1,d)-\theta\nu_n(1,1) \xrightarrow{d} \mathcal{N}(0,\sigma^2)$, where σ^2 is defined in Theorem 2.1.

Proof of (a) By applying Proposition 5.1 and Billingsley (1999, Chapter 12), we have

$$\begin{aligned}|\nu_n(e_n,d)-\nu_n(1,d)| &= \Big|\big(\nu_n(e_n,d)-W_d\,(e_n)\,\big)+\big(W_d(1)-\nu_n(1,d)\big)+\big(W_d\,(e_n)-W_d(1)\big)\Big| \\ &\le 2\sup_{x\in\mathscr{X}}\big|\nu_n(x,d)-W_d(x)\big| + \big|W_d(e_n)-W_d(1)\big| \xrightarrow{p} 0,\end{aligned}$$

by the continuous mapping theorem, the continuity of W_d, and the fact that $e_n \xrightarrow{p} 1$.

Proof of (b) We first consider R_n. Define $\ell_{d,n}(x) := (n/k)\mathbb{P}\big(U^{\max}_{1,d} > 1-kx/n\big)$. Then by Assumption A4 , we have $\lim_{n\to\infty}\ell_{d,n}(x) = x\ell_d$ and also $\tilde{\Delta}_{0,n}(x,d) = \ell_{d,n}(x)-\ell_{d-1,n}(x)$. Therefore, $R_n = \sqrt{k}\big(\ell_{d,n}(e_n)-\ell_{d-1,n}(e_n)-\theta\big) =: R_{n,1}+R_{n,2}$, where $R_{n,1} = \sqrt{k}\big[\big(\ell_{d,n}(e_n)-e_n\ell_d\big)-\big(\ell_{d-1,n}(e_n)-e_n\ell_{d-1}\big)\big]$ and $R_{n,2} = \sqrt{k}\big[e_n(\ell_d-\ell_{d-1})-\theta\big]$. By Assumption A4, $k = o\left(n^{2\rho/(2\rho+1)}\right)$, and $e_n \xrightarrow{p} 1$, we have $R_{n,1} = o_{\mathbb{P}}(1)$. Moreover, by (A.2) and (A.6), we obtain $R_{n,2} = \theta\sqrt{k}\,(e_n-1) = -\theta\nu_n(1,1)+o_{\mathbb{P}}(1)$.

Proof of (c) We denote $V_t = \mathbb{1}\big\{U^{\max}_{t+1,t+d-1} < 1-k/n < U_t\big\} - \theta\mathbb{1}\big\{U_t > 1-k/n\big\}$, then $\nu_n(1,d)-\theta\nu_n(1,1) = k^{-1/2}\sum_{t=1}^{n-d+1}\big(V_t-\mathbb{E}(V_1)\big)+O_{\mathbb{P}}\big(k^{-1/2}\big)$. Analogous to the proof of Proposition 5.1, we apply the main theorem of Utev (1991) to establish the CLT for the partial sum $\sum_{t=1}^{n-d+1}\big(V_t-\mathbb{E}(V_1)\big)$. We begin with the asymptotic variance:

$$\mathbb{V}\mathrm{ar}\left(\sum_{t=1}^{n-d+1}\big(V_t-\mathbb{E}(V_1)\big)\right) = \frac{n-d+1}{k}\,\mathbb{V}\mathrm{ar}(V_t)+\frac{2}{k}\sum_{t=1}^{n-d+1}(n-t)\,\mathrm{Cov}(V_1,V_{1+t}).$$

Note that $(n/k)\,\mathbb{E}(V_1) \to \theta - \theta = 0$. By the $D^{(d)}(u_n)$ condition, we have

$$\frac{n-d+1}{k}\,\mathbb{V}\mathrm{ar}(V_t) = \frac{n}{k}\Bigg[\mathbb{P}\Big(U_{2,d}^{\max} < 1-\frac{k}{n} < U_1\Big) + \theta^2\mathbb{P}\Big(U_1 > 1-\frac{k}{n}\Big) - 2\theta\mathbb{P}\Big(U_{2,d}^{\max} < 1-\frac{k}{n} < U_1\Big)\Bigg] + o(1) \to \theta - \theta^2.$$

Furthermore, under the mixing condition, it is clear that the tail sum $(2/k)\sum_{t=r_n+1}^{n-d+1}(n-t)\,\mathrm{Cov}(V_1, V_{1+t}) = o(1)$. Note that

$$\frac{2}{k}\sum_{t=1}^{r_n}(n-t)\,\mathrm{Cov}(V_1, V_{1+t}) =: 2\big(Q_{1,n} - Q_{2,n} - Q_{3,n} + Q_{4,n}\big) + o\Big(\frac{r_n k}{n}\Big),$$

where $Q_{1,n} = \theta^2\,(n/k)\sum_{t=1}^{r_n}\mathbb{P}\big(U_1 > 1-k/n,\, U_{t+1} > 1-k/n\big)$, $Q_{2,n} = \theta\,(n/k)\sum_{t=1}^{r_n}\mathbb{P}\big(U_1 > 1-k/n,\, U_{t+2,t+d}^{\max} < 1-k/n < U_{t+1}\big)$, $Q_{3,n} = \theta\,(n/k)\sum_{t=1}^{r_n}\mathbb{P}\big(U_{2,d}^{\max} < 1-k/n < U_1,\, U_{t+1} > 1-k/n\big)$, $Q_{4,n} = (n/k)\sum_{t=1}^{r_n}\mathbb{P}\big(U_{2,d}^{\max} < 1-k/n < U_1,\, U_{t+2,t+d}^{\max} < 1-k/n < U_{t+1}\big)$. By Assumptions A2–A3, $\lim_{n\to\infty}(2/k)\sum_{t=1}^{r_n}(n-t)\,\mathrm{Cov}(V_1, V_{1+t}) = 2(\theta^2\Lambda_1(1,1) - \theta\lambda_1)$. The Condition (2) in Utev (1991) follows from the same argument as that for (S.4). This completes the proof of Part (c). □

PROOF. (**Proof of Theorem 2.2**) For some given $d \in \mathbb{Z}^+$, since Assumption 2.2 holds for the pair $(d, d+1)$, we have Assumption 2.1 holds for $d+1$, imply that $\Delta(d+1) = \theta$ and the result in Proposition 5.1 holds for $d+1$. Note that for this specific pair $(d, d+1)$, we do not require Assumption 2.1 to hold for d. Nevertheless, we also need a similar weak convergence result for $\sqrt{k}\big(\tilde{\Delta}_n(x,d) - \tilde{\Delta}_{0,n}(x,d)\big)$. Clearly, the asymptotic variance is not necessary in the same form because $D^{(d)}(u_n)$ condition is no longer guaranteed. However, the conditions in Theorem 2.2 ensure that the asymptotic variance exists and the tightness of the process continues to hold for d specified above. Following the same line of argument as in the proof of Proposition 5.1, we have $\big\{\sqrt{k}\big(\tilde{\Delta}_n(x,d) - \tilde{\Delta}_{0,n}(x,d)\big)\big\}_{x\in\mathscr{X}} \Rightarrow \{G(x)\}_{x\in\mathscr{X}}$, as $n\to\infty$, where G is a zero-mean continuous Gaussian process with covariance function:

$$\mathbb{E}\big(G(x)G(y)\big) = \lim_{n\to\infty}\frac{n}{k}\sum_{t=0}^{d-1}\mathbb{P}\Big(U_{2,d}^{\max} < 1-\frac{kx}{n} < U_1,\, U_{t+2,t+d}^{\max} < 1-\frac{ky}{n} < U_{t+1}\Big) + \Lambda_2(x,y) + \Lambda_2(y,x),$$

for $x \le y$. If Assumption A2 also holds for d, then $\Lambda_2(x,y) = 0$ and $G \overset{d}{=} W_d$, where W_d is defined in Proposition 5.1. It is clear that $\mathbb{E}\big(G^2(1)\big) = \Delta(d) + 2\Lambda_2(1,1)$. The remaining steps are similar to those in Theorem 2.1. Full details are deferred to the supplement. □

Proof. (**Proof of Proposition 2.1**) By the definition of $\hat{d}_L(k)$, $\{\hat{d}_L(k) \geq d_L + 1\} \subset \{\max_{d_L \leq s \leq d_U} \hat{\delta}_n(s) \geq 1/\sqrt{k}\}$. Note that the latter event has a probability tending to zero by Corollary 2.1. On the other hand, for any $j \leq d_L - 1$, $\{\hat{d}_L(k) = j\} \subset \{\max_{j \leq s \leq d_U} \hat{\delta}_n(s) < 1/\sqrt{k}\} \subset \{\hat{\delta}_n(d_L - 1) < 1/\sqrt{k}\}$, which also has a probability tending to zero. Therefore, $\mathbb{P}(\hat{d}_L(k) \neq d_L) \to 0$. $\square$

Estimation and inference for the persistence of extremely high temperatures: Online supplement

JUAN-JUAN CAI[†,‡], YICONG LIN[†,‡], JULIA SCHAUMBURG[†,‡] AND CHENHUI WANG[†]

[†] *Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands.*
E-mail: {j.cai, yc.lin, j.schaumburg, c.h.wang}@vu.nl

[‡] *Tinbergen Institute, Valkenierstraat 59, 1018 XE, Amsterdam, the Netherlands.*

Summary Section S1 provides detailed proofs for the propositions and lemmas used in the main results, as well as the full steps for Theorem 2.2. Section S2 then illustrates the computation of d_L for the simulation DGPs and presents additional results showing that the $D^{(d)}(u_n)$ condition is not fulfilled for an ARCH model. Section S3 provides the detailed steps of the bootstrap algorithm used in our empirical analysis. Section S4 reports the complete set of empirical results for the study of summer apparent temperatures in Europe.

S1. ADDITIONAL PROOFS OF THE MAIN RESULTS

PROOF. (**Proof of Theorem 5.1**) Recall the definition of $\Delta(\cdot)$ in (2.5). Then (A.1) follows from that for any $\tau > 0$,

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left(U_{2,s}^{\max} \le 1-\frac{\tau}{n} \,\middle|\, U_1 > 1-\frac{\tau}{n}\right) &= \lim_{n\to\infty} \frac{n}{\tau}\mathbb{P}\left(U_{2,s}^{\max} \le 1-\frac{\tau}{n}, U_1 > 1-\frac{\tau}{n}\right)\\
&= \lim_{n\to\infty} \frac{n}{\tau}\left\{\mathbb{P}\left(U_{2,s}^{\max} \le 1-\frac{\tau}{n}\right) - \mathbb{P}\left(U_{1,s}^{\max} \le 1-\frac{\tau}{n}\right)\right\}\\
&= \lim_{n\to\infty} \frac{n}{\tau}\left\{\mathbb{P}\left(U_{1,s}^{\max} > 1-\frac{\tau}{n}\right) - \mathbb{P}\left(U_{2,s}^{\max} > 1-\frac{\tau}{n}\right)\right\}\\
&= \ell_s - \ell_{s-1}.
\end{aligned}$$

Thus, under the assumption that ℓ_s exists, $\Delta(s)$ is well defined for any $s \le d_U$. Since $s \mapsto \Delta(s)$ is non-increasing, we have, for any $d_U \ge d$,

$$\Delta(d) \ge \Delta(d_U) \ge \lim_{n\to\infty} \mathbb{P}\left(U_{2,r_n}^{\max} \le 1-\frac{\tau}{n} \,\middle|\, U_1 > 1-\frac{\tau}{n}\right).$$

On the other hand, $D^{(d)}(u_n)$ condition is equivalent to the following equality:

$$\Delta(d) = \lim_{n\to\infty} \mathbb{P}\left(U_{2,r_n}^{\max} \le 1 - \frac{\tau}{n} \,\middle|\, U_1 > 1 - \frac{\tau}{n}\right). \tag{S.1}$$

Therefore, one has $\theta = \Delta(d) = \Delta(d+1) = \cdots = \Delta(d_U)$. Thus, d_L exists and $d_L = \min\left\{d \in \mathbb{Z}^+ : \Delta(d) = \theta\right\}$ as given in (2.6). By the definition of d_L and the monotonicity of Δ, we have for any $1 \le s < d_L$, $\Delta(s) > \theta$. □

Proof. (**Proof of Proposition 5.1**) When no ambiguity arises, we omit the subscript d in the notation of $W_d(\cdot)$ and write $\nu_n(\cdot, d)$ simply as $\nu_n(\cdot)$, where $d \in \mathbb{Z}^+$ is fixed. We establish the weak convergence in (A.5) by verifying the convergence of the finite-dimensional distributions together with uniform tightness.

Finite-dimensional distributional convergence Let $m \in \mathbb{Z}^+$. By the Cramér-Wold device, it suffices to show that for any $x_i \in \mathscr{X}$ and $a_i \in \mathbb{R}$, $i = 1, \ldots, m$,

$$\sum_{i=1}^{m} a_i \nu_n(x_i) \xrightarrow{d} \sum_{i=1}^{m} a_i W(x_i), \qquad n \to \infty. \tag{S.2}$$

We present the proof for the case $m = 2$. The argument for $m > 2$ is more involved but proceeds analogously. Let $1/2 \le x \le y \le 3/2$, and for convenience, $I_t = \mathbb{1}\left\{U_{t+1,t+d-1}^{\max} < 1 - kx/n < U_t\right\}$ and $J_t = \mathbb{1}\left\{U_{t+1,t+d-1}^{\max} < 1 - ky/n < U_t\right\}$ for $t = 1, \ldots, n-d+1$. By the definition in (A.4), we have

$$\begin{aligned} a_1\nu_n(x) + a_2\nu_n(y) &= \frac{1}{\sqrt{k}} \sum_{t=1}^{n-d+1} \Big(a_1 I_t + a_2 J_t - \mathbb{E}\big(a_1 I_t + a_2 J_t\big)\Big) + O\left(\frac{1}{\sqrt{k}}\right) \\ &=: \sum_{t=1}^{n-d+1} \xi_{t,n} + O\left(\frac{1}{\sqrt{k}}\right), \end{aligned} \tag{S.3}$$

where $\xi_{t,n} = \big(a_1 I_t + a_2 J_t - \mathbb{E}(a_1 I_t + a_2 J_t)\big)/\sqrt{k}$. We apply the central limit theorem (CLT) developed by Utev (1991) to prove that $\sum_{t=1}^{n-d+1} \xi_{t,n} \xrightarrow{d} a_1 W(x) + a_2 W(y)$. We begin by deriving the asymptotic variance. Note that

$$\mathbb{V}\mathrm{ar}\left(\sum_{t=1}^{n-d+1} \xi_{t,n}\right) = \frac{a_1^2}{k} \mathbb{V}\mathrm{ar}\left(\sum_{t=1}^{n-d+1} I_t\right) + \frac{a_2^2}{k} \mathbb{V}\mathrm{ar}\left(\sum_{t=1}^{n-d+1} J_t\right) + \frac{2a_1a_2}{k} \mathrm{Cov}\left(\sum_{t=1}^{n-d+1} I_t, \sum_{t=1}^{n-d+1} J_t\right).$$

Note that $\mathbb{E}(I_1) = \mathbb{P}\big(U_{2,d}^{\max} < 1 - kx/n < U_1\big) = O(k/n)$, and $\mathbb{V}\mathrm{ar}(I_1) = \mathbb{E}(I_1)\big(1 -$

$O(k/n)\big)$. Hence, by the strict stationarity and Lemmas S.1(i) and S.2, we have

$$\begin{aligned}\mathbb{V}\text{ar}\left(\sum_{t=1}^{n-d+1} I_t\right) &= (n-d+1)\,\mathbb{V}\text{ar}(I_1) + 2\sum_{i<j}\text{Cov}(I_i,I_j)\\ &= (n-d+1)\mathbb{E}(I_1)\left(1-O\left(\frac{k}{n}\right)\right) + 2\sum_{t=1}^{n-d}(n-d+1-t)\,\text{Cov}(I_1,I_{1+t})\\ &= (n-d+1)\mathbb{E}(I_1)\left(1-O\left(\frac{k}{n}\right)\right) + o(k).\end{aligned}$$

Therefore, $(a_1^2/k)\,\mathbb{V}\text{ar}\left(\sum_{t=1}^{n-d+1} I_t\right) = a_1^2\,(n/k)\,\mathbb{E}(I_1)+o(1)$. Similarly, one can also obtain $(a_2^2/k)\,\mathbb{V}\text{ar}\left(\sum_{t=1}^{n-d+1} J_t\right) = a_2^2\,(n/k)\,\mathbb{E}(J_1)+o(1)$. For the covariance term, we have, again by the strict stationarity and by Lemmas S.1(i) and S.2,

$$\begin{aligned}\frac{1}{k}\,\text{Cov}\left(\sum_{t=1}^{n-d+1} I_t, \sum_{t=1}^{n-d+1} J_t\right) &= \frac{1}{k}\sum_{t=-(n-d)}^{n-d}(n-d+1-|t|)\,\text{Cov}\big(I_1,J_{1+t}\big)\\ &= \frac{1}{k}\sum_{t=-(d-1)}^{d-1}(n-d+1-|t|)\,\text{Cov}\big(I_1,J_{1+t}\big)\\ &= \frac{1}{k}\sum_{t=0}^{d-1}(n-d+1-t)\mathbb{E}\big(I_1J_{1+t}\big) + o(1)\\ &= \frac{n}{k}\sum_{t=0}^{d-1}\mathbb{E}\big(I_1J_{1+t}\big) + o(1),\end{aligned}$$

where the second equality follows from $\text{Cov}\big(I_1,J_{1+t}\big) = 0$ for $|t|\geq d$. The third equality follows from $k^{-1}\sum_{t=1}^{d-1}(n-d+1-t)\;\text{Cov}\big(I_{t+1},J_1\big) = o(1)$, using the facts that $\mathbb{E}(I_1)\mathbb{E}(J_1) = O(k^2/n^2)$ and $\mathbb{E}(I_{t+1}J_1) = 0$ for $x\leq y$ and $t = 1,\ldots,d-1$. Putting together, we obtain

$$\begin{aligned}\mathbb{V}\text{ar}\left(\sum_{t=1}^{n-d+1}\xi_{t,n}\right) &= \lim_{n\to\infty}\left(a_1^2\,\frac{n}{k}\mathbb{E}(I_1) + a_2^2\,\frac{n}{k}\mathbb{E}(J_1) + 2a_1a_2\,\frac{n}{k}\sum_{t=0}^{d-1}\mathbb{E}\left(I_1J_{1+t}\right)\right)\\ &= a_1^2\,\mathbb{E}(W^2(x)) + a_2^2\,\mathbb{E}(W^2(y)) + 2a_1a_2\,\mathbb{E}\big(W(x)W(y)\big).\end{aligned}$$

Next, we verify Condition (2) in Utev (1991), which is essentially Lindeberg's condition. Let $\sigma_n^2 = \mathbb{V}\text{ar}\big(\sum_{t=1}^{n-d+1}\xi_{t,n}\big)$. By choosing $j_t = 1$ for all $t\in\mathbb{Z}^+$ in Utev (1991), we obtain, for any $\epsilon > 0$,

$$\begin{aligned}
\sigma_n^{-2} \sum_{t=1}^{n-d+1} \mathbb{E}\Big(\xi_{t,n}^2 \mathbb{1}\{|\xi_{t,n}| \geq \epsilon\sigma_n\}\Big) &\leq \sigma_n^{-3} n\, \epsilon^{-1} \mathbb{E}(|\xi_{1,n}|^3) \\
&= \sigma_n^{-3} n\, k^{-3/2} \epsilon^{-1} \mathbb{E}\Big(\big|a_1 I_1 + a_2 J_1 - \mathbb{E}(a_1 I_1 + a_2 J_1)\big|^3\Big) \\
&= \sigma_n^{-3} n\, k^{-3/2} \epsilon^{-1} \mathbb{E}\Big(\big|a_1 I_1 + a_2 J_1 - O(k/n)\big|^3\Big) \\
&= \sigma_n^{-3} n\, k^{-3/2} \epsilon^{-1} O\left(\frac{k}{n}\right) \to 0. \qquad \text{(S.4)}
\end{aligned}$$

Thus, the central limit theorem in Utev (1991) applies to the partial sums of the sequence $\{\xi_{t,n} : t = 1, \ldots, n-d+1\}$, yielding (S.2).

Tightness in $\mathcal{D}(\mathscr{X})$ We prove that for any $\lambda > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\substack{|x-y|<\delta_n \\ 1/2 \leq x < y \leq 3/2}} \big|\nu_n(x) - \nu_n(y)\big| \geq \lambda \right) = 0, \qquad \text{(S.5)}$$

where δ_n is a deterministic sequence such that $\lim_{n\to\infty} \delta_n = 0$. Then by Theorem 1 in Aldous (1978) and the explanation thereafter, $\nu_n(\cdot)$ is tight in $\mathcal{D}(\mathscr{X})$ and each weak limit has a.s. continuous sample paths. To prove (S.5), we decompose ν_n into two parts. For any $x \in \mathscr{X}$, it holds a.s. that $\mathbb{1}\big\{U_{t+1,t+d-1}^{\max} < 1 - kx/n < U_t\big\} = \mathbb{1}\big\{U_{t,t+d-1}^{\max} > 1 - kx/n\big\} - \mathbb{1}\big\{U_{t+1,t+d-1}^{\max} > 1 - kx/n\big\}$. By (A.4), we obtain

$$\nu_n(x) = \nu_{1,n}(x) - \nu_{2,n}(x), \qquad \text{(S.6)}$$

where $\nu_{1,n}(x) = k^{-1/2} \sum_{t=1}^{n-d+1} \Big(\mathbb{1}\big\{U_{t,t+d-1}^{\max} > 1 - kx/n\big\} - \mathbb{P}\big(U_{t,t+d-1}^{\max} > 1 - kx/n\big)\Big)$ and $\nu_{2,n}(x) = k^{-1/2} \sum_{t=1}^{n-d+1} \Big(\mathbb{1}\big\{U_{t+1,t+d-1}^{\max} > 1 - kx/n\big\} - \mathbb{P}\big(U_{t+1,t+d-1}^{\max} > 1 - kx/n\big)\Big)$. To prove (S.5), it suffices to prove the tightness condition for $\nu_{1,n}(x)$ and $\nu_{2,n}(x)$, respectively.

We provide the details for $\nu_{1,n}(x)$. Let $t_n = \lfloor (n-d+1)/(2r_n) \rfloor$. We split the sum into $2\,t_n$ blocks of length r_n and a remaining block of length less than $2r_n$. To simplify the notation, we denote $M_t = U_{t,t+d-1}^{\max}$ below and decompose $\nu_{1,n}(x)$ into $\nu_{1,n}(x) =$

$\mu_{1,n}(x)+\mu_{2,n}(x)+\mu_{3,n}(x)$, where

$$\mu_{1,n}(x)=\frac{1}{\sqrt{k}}\sum_{i=0}^{t_n-1}\sum_{j=1}^{r_n}\left(\mathbb{1}\left\{M_{2ir_n+j}>1-\frac{kx}{n}\right\}-\mathbb{P}\left(M_{2ir_n+j}>1-\frac{kx}{n}\right)\right),$$
$$\mu_{2,n}(x)=\frac{1}{\sqrt{k}}\sum_{i=0}^{t_n-1}\sum_{j=1}^{r_n}\left(\mathbb{1}\left\{M_{(2i+1)r_n+j}>1-\frac{kx}{n}\right\}-\mathbb{P}\left(M_{(2i+1)r_n+j}>1-\frac{kx}{n}\right)\right),$$
$$\mu_{3,n}(x)=\frac{1}{\sqrt{k}}\sum_{i=2t_nr_n+1}^{n-d+1}\left(\mathbb{1}\left\{M_i>1-\frac{kx}{n}\right\}-\mathbb{P}\left(M_i>1-\frac{kx}{n}\right)\right).$$

Since $\mu_{1,n}(x)$ and $\mu_{2,n}(x)$ share similar constructions, we first derive a generic result to establish their uniform tightness. Define $\tilde{\mu}_n(x)=k^{-1/2}\sum_{i=1}^{t_nr_n}\left(\mathbb{1}\{\tilde{M}_i>1-kx/n\}-\mathbb{P}(\tilde{M}_i>1-kx/n)\right)$, where

$$\{\tilde{M}_{(i-1)r_n+1},\ldots,\tilde{M}_{ir_n}\}\overset{d}{=}\{M_1,\ldots,M_{r_n}\},\qquad i=1,\ldots,t_n, \tag{S.7}$$

and $\{\tilde{M}_{(i-1)r_n+j},\, j=1,\ldots,r_n\}_{i=1}^{t_n}$ are t_n independent blocks. Thus, for each n, the sequence $\{\tilde{M}_i\}$ constitutes a special r_n-dependent array, which is not strictly stationary. We first apply a fluctuation inequality for m-dependent arrays given by Theorem 4.1 in Einmahl and Ruymgaart (2000) to prove the tightness of $\tilde{\mu}_n$. Then, the tightness of $\mu_{1,n}$ and $\mu_{2,n}$ follows from the bounded variation distance between $\tilde{\mu}_n$ and $\mu_{1,n}$ and between $\tilde{\mu}_n$ and $\mu_{2,n}$, respectively.

For each n, let $q=r_n^{1+\epsilon}$, where $\epsilon>0$ is some constant such that $q/\sqrt{k_n}\to 0$. Define $\mathcal{I}_i=\left[\frac{1}{2}+\frac{i}{q},\frac{1}{2}+\frac{i+1}{q}\right]\subset\mathscr{X}$ for $i=0,\ldots,q-1$. Choose $\delta_n=1/q$. For any $x,y\in\mathscr{X}$ and $|x-y|<\delta_n$, there exists an $i\in\{0,\ldots,q-1\}$ such that $\left|x-\frac{1}{2}-\frac{i}{q}\right|<q^{-1}$ and $\left|y-\frac{1}{2}-\frac{i}{q}\right|<q^{-1}$. Thus, for any $\lambda>0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\substack{|x-y|<\delta_n\\1/2\le x<y\le 3/2}}|\tilde{\mu}_n(x)-\tilde{\mu}_n(y)|\ge\lambda\right)\\
&\le\mathbb{P}\left(\max_{0\le i\le q-1}\sup_{\substack{|x-1/2-i/q|<1/q\\|y-1/2-i/q|<1/q}}\left(\left|\tilde{\mu}_n(x)-\tilde{\mu}_n\Big(\frac{1}{2}+\frac{i}{q}\Big)\right|+\left|\tilde{\mu}_n\Big(\frac{1}{2}+\frac{i}{q}\Big)-\tilde{\mu}_n(y)\right|\right)\ge\lambda\right)\\
&\le 2\mathbb{P}\left(\max_{0\le i\le q-1}\sup_{x\in\mathcal{I}_i}\left|\tilde{\mu}_n(x)-\tilde{\mu}_n\Big(\frac{1}{2}+\frac{i}{q}\Big)\right|\ge\frac{\lambda}{2}\right)+2\mathbb{P}\left(\max_{0\le i\le q-1}\sup_{x\in\mathcal{I}_i}\left|\tilde{\mu}_n(x)-\tilde{\mu}_n\Big(\frac{1}{2}+\frac{i+1}{q}\Big)\right|\ge\frac{\lambda}{2}\right)\\
&\le 2\sum_{i=0}^{q-1}\mathbb{P}\left(\sup_{x\in\mathcal{I}_i}\left|\tilde{\mu}_n(x)-\tilde{\mu}_n\Big(\frac{1}{2}+\frac{i}{q}\Big)\right|\ge\frac{\lambda}{2}\right)+2\sum_{i=0}^{q-1}\mathbb{P}\left(\sup_{x\in\mathcal{I}_i}\left|\tilde{\mu}_n(x)-\tilde{\mu}_n\Big(\frac{1}{2}+\frac{i+1}{q}\Big)\right|\ge\frac{\lambda}{2}\right)
\end{aligned}$$

$$\leq 4\sum_{i=0}^{q-1}\mathbb{P}\left(\sup_{x,y\in\mathcal{I}_i}|\tilde{\mu}_n(x)-\tilde{\mu}_n(y)|\geq\frac{\lambda}{2}\right)=:4\sum_{i=0}^{q-1}Q_i.$$

Next, we apply Einmahl and Ruymgaart (2000, Eq. (4.4)) to bound Q_i. As such, define

$$\Gamma_n(x)=\frac{1}{t_nr_n}\sum_{i=1}^{t_nr_n}\left(\mathbb{1}\left\{\tilde{M}_i>1-x\right\}-\mathbb{P}\big(\tilde{M}_i>1-x\big)\right),$$

which plays the role of Δ_n in Einmahl and Ruymgaart (2000). Then, adopting the notation of that paper with $\epsilon=1/2$ and $m=r_n$, by the distributional equivalence in (S.7), we obtain

$$\begin{aligned}
Q_i&=\mathbb{P}\left(\sup_{\frac{1}{2}+\frac{i}{q}\leq x<y\leq\frac{1}{2}+\frac{i+1}{q}}|\tilde{\mu}_n(x)-\tilde{\mu}_n(y)|\geq\lambda\right)\\
&=\mathbb{P}\left(\frac{r_nt_n}{\sqrt{k}}\sup_{\frac{1}{2}+\frac{i}{q}\leq x<y\leq\frac{1}{2}+\frac{i+1}{q}}\left|\Gamma_n\left(\frac{kx}{n}\right)-\Gamma_n\left(\frac{ky}{n}\right)\right|\geq\lambda\right)\\
&=\mathbb{P}\left(\sup_{\frac{k}{n}(\frac{1}{2}+\frac{i}{q})\leq a<b\leq\frac{k}{n}(\frac{1}{2}+\frac{i+1}{q})}\big|\Gamma_n(a)-\Gamma_n(b)\big|\geq\frac{\sqrt{k}\lambda}{r_nt_n}\right)\\
&\leq C\exp\left(\frac{-r_nt_n\frac{k\lambda^2}{r_n^2t_n^2}}{4r_np_i}\psi\left(\frac{\sqrt{r_nt_n}\frac{\sqrt{k}\lambda}{r_nt_n}}{\sqrt{r_nt_np_i}}\right)\right)=C\exp\left(-\frac{k\lambda^2}{4r_n^2t_np_i}\psi\left(\frac{\sqrt{k}\lambda}{r_nt_np_i}\right)\right),
\end{aligned}$$

where $C>0$ is some constant, $p_i=\mathbb{P}\left(1-\frac{k}{n}(\frac{1}{2}+\frac{i+1}{q})<U_{1,d}^{\max}\leq1-\frac{k}{n}(\frac{1}{2}+\frac{i}{q})\right)$ and ψ is a continuous and decreasing function such that $\psi(0)=1$. Observe that Assumption A4 implies that,

$$\sup_{x\in\mathscr{X}}\frac{n}{k}\left|\mathbb{P}\Big(U_{1,j}^{\max}>1-\frac{kx}{n}\Big)-x\ell_j\right|=O\left(\left(\frac{k}{n}\right)^{\rho}\right).\tag{S.8}$$

Thus, we have

$$\sup_{0\leq i\leq q-1}\left|\frac{n}{k}p_i-\frac{1}{q}\ell_d\right|=O\left(\left(\frac{k}{n}\right)^{\rho}\right).$$

If n is sufficiently large, we have $\frac{k}{2nq}\leq p_i\leq\frac{2dk}{nq}$, uniformly in i, due to the fact that $\ell_d\in[1,d]$. Then by the choice of q and that $n-r_n\leq2r_nt_n\leq n$, we obtain

$$\frac{k}{4r_n^2t_np_i}\geq\frac{k}{2r_nnp_i}\geq\frac{m}{4dr_n}=\frac{1}{4d}r_n^{\epsilon},$$

and

$$\psi\left(\frac{\sqrt{k}\lambda}{r_nt_np_i}\right)\geq\psi\left(\frac{3\sqrt{k}\lambda}{n\frac{k}{2nq}}\right)=\psi\left(\frac{6\lambda q}{\sqrt{k}}\right)\to\psi(0)=1.$$

Thus, as $n \to \infty$,

$$\mathbb{P}\left(\sup_{\substack{|x-y|<\delta_n \\ 1/2 \le x < y \le 3/2}} |\tilde{\mu}_n(x) - \tilde{\mu}_n(y)| \ge \lambda\right) \le 4q\exp(-c_2 r_n^{\epsilon}) = 4r_n^{1+\epsilon}\exp(-c_2 r_n^{\epsilon}) \to 0.$$

So the tightness of $\tilde{\mu}_n$ follows from the tightness criterion by Aldous (1978, Theorem 1).

Moreover, let $\Omega(X)$ denote the distribution of X. By Eberlein (1984, Lemma 2), we have, for the total variation-type norm $\|\cdot\|$ defined in Eberlein (1984),

$$\begin{aligned}
&\left\|\Omega\Big(\{\tilde{M}_{(i-1)r_n+1}, \dots, \tilde{M}_{ir_n}\}_{i=1}^{t_n}\Big) - \Omega\Big(\{M_{(i-1)r_n+1}, \dots, M_{ir_n}\}_{i=1}^{t_n}\Big)\right\| \\
&= \left\|\bigotimes_{i=1}^{t_n}\Omega\Big(M_{(i-1)r_n+1}, \dots, M_{ir_n}\Big) - \Omega\Big(\{M_{(i-1)r_n+1}, \dots, M_{ir_n}\}_{i=1}^{t_n}\Big)\right\| \le \phi_n(r_n)t_n \to 0,
\end{aligned}$$

by the absolutely regular assumption on the sequence, and Assumption A1. Thus, we obtain that for $i = 1, 2$,

$$\mathbb{P}\left(\sup_{\substack{|x-y|<\delta_n \\ 1/2 \le x < y \le 3/2}} \big|\mu_{i,n}(x) - \mu_{i,n}(y)\big| \ge \lambda\right) \to 0. \tag{S.9}$$

It remains to show that $\sup_{x\in\mathscr{X}} |\mu_{3,n}(x)| \xrightarrow{p} 0$. Note that by the definition of t_n, the number of summands in $\mu_{3,n}$ is bounded by $2r_n$.

$$\begin{aligned}
\mathbb{E}\left(\sup_{x\in\mathscr{X}} \big|\mu_{3,n}(x)\big|\right) &\le \mathbb{E}\left(\frac{1}{\sqrt{k}}\sum_{i=2t_n r_n+1}^{n-d+1}\left(\mathbb{1}\left\{M_i > 1 - \frac{3k}{2n}\right\} + \mathbb{P}\left(M_i > 1 - \frac{3k}{2n}\right)\right)\right) \\
&\le \frac{4r_n}{\sqrt{k}} \cdot \frac{3k}{2n} \to 0,
\end{aligned}$$

by the assumption that $r_n\sqrt{k}/n \to 0$ as $n \to \infty$. □

Proof. (**Proof of Theorem 2.2**) For some given $d \in \mathbb{Z}^+$, since Assumption 2.2 holds for the pair $(d, d+1)$, we have Assumption 2.1 holds for $d+1$, imply that $\Delta(d+1) = \theta$ and the result in Proposition 5.1 holds for $d+1$.

Note that for this specific pair $(d, d+1)$, we do not require Assumption 2.1 to hold for d. Nevertheless, we also need a similar weak convergence result for $\sqrt{k}\big(\tilde{\Delta}_n(x,d) - \tilde{\Delta}_{0,n}(x,d)\big)$. Clearly, the asymptotic variance is not necessary in the same form because $D^{(d)}(u_n)$ condition is no longer guaranteed. However, the conditions in Theorem 2.2 ensure that the asymptotic variance exists and the tightness of the process continues to hold for d specified above. Following the same line of argument as in the proof of

Proposition 5.1, we have

$$\left\{\sqrt{k}\big(\tilde{\Delta}_n(x,d)-\tilde{\Delta}_{0,n}(x,d)\big)\right\}_{x\in\mathscr{X}} \Rightarrow \big\{G(x)\big\}_{x\in\mathscr{X}}, \qquad n\to\infty, \tag{S.10}$$

where G is a zero-mean continuous Gaussian process with covariance function given by

$$\begin{aligned}\mathbb{E}\big(G(x)G(y)\big) = \lim_{n\to\infty}\frac{n}{k}\sum_{t=0}^{d-1}\mathbb{P}\Big(U_{2,d}^{\max} < 1-\frac{kx}{n} < U_1,\, U_{t+2,t+d}^{\max} < 1-\frac{ky}{n} < U_{t+1}\Big)\\ + \Lambda_2(x,y)+\Lambda_2(y,x),\end{aligned}$$

for $x\le y$. Note that if Assumption A2 also holds for d, then $\Lambda_2(x,y)=0$ and $G\overset{d}{=}W_d$, where W_d is defined in Proposition 5.1. It is clear that

$$\mathbb{E}\big(G^2(1)\big)=\Delta(d)+2\Lambda_2(1,1). \tag{S.11}$$

The theorem can now be proved in the same manner as Theorem 2.1. Recall that $\delta(\bullet):=\Delta(\bullet)-\Delta(\bullet+1)$ and $\hat{\delta}_n(\bullet)=\hat{\Delta}_n(\bullet)-\hat{\Delta}_n(\bullet+1)$. By the similar arguments for Parts (a)–(b) in Proof of Theorem 2.1, we have

$$\begin{aligned}\sqrt{k}\big(\hat{\delta}_n(d)-\delta(d)\big) &= \sqrt{k}\big(\tilde{\Delta}_n(e_n,d)-\tilde{\Delta}_n(e_n,d+1)-\delta(d)\big)\\ &= \sqrt{k}\big(\tilde{\Delta}_n(e_n,d)-\tilde{\Delta}_{0,n}(e_n,d)\big)-\sqrt{k}\big(\tilde{\Delta}_n(e_n,d+1)-\tilde{\Delta}_{0,n}(e_n,d+1)\big)\\ &\quad+\sqrt{k}\big(\tilde{\Delta}_{0,n}(e_n,d)-\tilde{\Delta}_{0,n}(e_n,d+1)-\delta(d)\big)\\ &= \nu_n(1,d)-\nu_n(1,d+1)-\delta(d)\nu_n(1,1)+o_{\mathbb{P}}(1).\end{aligned}$$

It suffices to show that

$$\nu_n(1,d)-\nu_n(1,d+1)-\delta(d)\nu_n(1,1)\xrightarrow{d}\mathcal{N}\big(0,\kappa^2(d)\big). \tag{S.12}$$

Define $I_t=\mathbb{1}\big\{U_{t+1,t+d-1}^{\max}<1-k/n<U_t\big\}$ and $J_t=\mathbb{1}\big\{U_{t+1,t+d}^{\max}<1-k/n<U_t\big\}$, and $K_t=\mathbb{1}\big\{U_t>1-k/n\big\}$. We shall show that

$$\frac{1}{\sqrt{k}}\sum_{i=1}^{n-d}\Big(\big(I_t-J_t-\delta(d)K_t\big)-\mathbb{E}\big(I_t-J_t-\delta(d)K_t\big)\Big)\xrightarrow{d}\mathcal{N}\big(0,\kappa^2(d)\big).$$

Note that

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{k}}\sum_{t=1}^{n-d}\big(I_t - J_t - \delta(d)K_t\big)\right) \\
&= \frac{1}{k}\,\mathbb{V}\mathrm{ar}\left(\sum_{t=1}^{n-d} I_t\right) + \frac{1}{k}\,\mathbb{V}\mathrm{ar}\left(\sum_{t=1}^{n-d} J_t\right) + \frac{\delta^2(d)}{k}\,\mathbb{V}\mathrm{ar}\left(\sum_{t=1}^{n-d} K_t\right) - \frac{2}{k}\,\mathrm{Cov}\left(\sum_{t=1}^{n-d} I_t, \sum_{t=1}^{n-d} J_t\right) \\
&\quad - \frac{2\delta(d)}{k}\,\mathrm{Cov}\left(\sum_{t=1}^{n-d} I_t, \sum_{t=1}^{n-d} K_t\right) + \frac{2\delta(d)}{k}\,\mathrm{Cov}\left(\sum_{t=1}^{n-d} J_t, \sum_{t=1}^{n-d} K_t\right) \\
&= \big(\Delta(d) + 2\Lambda_2(1,1)\big) + \theta + \delta^2(d)\big(1 + 2\Lambda_1(1,1)\big) + o(1) - \frac{2}{k}\,\mathrm{Cov}\left(\sum_{t=1}^{n-d} I_t, \sum_{t=1}^{n-d} J_t\right) \\
&\quad - \frac{2\delta(d)}{k}\,\mathrm{Cov}\left(\sum_{t=1}^{n-d} I_t, \sum_{t=1}^{n-d} K_t\right) + 2\delta(d)(\theta + \lambda_1) + o(1), \qquad \text{(S.13)}
\end{aligned}$$

where we use (S.11), $\mathbb{V}\mathrm{ar}(W_{d+1}(1)) = \theta$, and $(2/k)\,\mathrm{Cov}\left(\sum_{t=1}^{n-d} J_t, \sum_{t=1}^{n-d} K_t\right) = \theta + \lambda_1 + o(1)$, from the proof for Theorem 2.1. Next, we compute the two remaining covariance terms. By strict stationarity, Lemma S.2, and Assumption A2 (with d replaced by $d+1$), we have

$$\begin{aligned}
&\frac{1}{k}\,\mathrm{Cov}\left(\sum_{t=1}^{n-d} I_t, \sum_{t=1}^{n-d} J_t\right) \\
&= \frac{n}{k}\,\mathbb{P}\Big(U_{2,d+1}^{\max} < 1 - \frac{k}{n} < U_1\Big) + \frac{n}{k}\sum_{t=1}^{r_n}\mathbb{P}\Big(U_{2,d}^{\max} < 1 - \frac{k}{n} < U_1,\, U_{t+2,t+d+1}^{\max} < 1 - \frac{k}{n} < U_{t+1}\Big) \\
&\quad + \frac{n}{k}\sum_{t=1}^{r_n}\mathbb{P}\Big(U_{t+2,t+d}^{\max} < 1 - \frac{k}{n} < U_{t+1},\, U_{2,d+1}^{\max} < 1 - \frac{k}{n} < U_1\Big) + o(1) \\
&= \theta + \lambda_2 + o(1) + o(1). \qquad \text{(S.14)}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\frac{1}{k}\,\mathrm{Cov}\left(\sum_{t=1}^{n-d} I_t, \sum_{t=1}^{n-d} K_t\right) &= \frac{n}{k}\,\mathbb{P}\Big(U_{2,d}^{\max} < 1 - \frac{k}{n} < U_1\Big) + \frac{n}{k}\sum_{t=1}^{r_n}\mathbb{P}\Big(U_{2,d}^{\max} < 1 - \frac{k}{n} < U_1,\, U_{t+1} > 1 - \frac{k}{n}\Big) \\
&\quad + \frac{n}{k}\sum_{t=1}^{r_n}\mathbb{P}\Big(U_{t+2,t+d}^{\max} < 1 - \frac{k}{n} < U_{t+1},\, U_1 > 1 - \frac{k}{n}\Big) + o(1) \\
&= \Delta(d) + \lambda_3 + \tilde{\lambda}_1 + o(1). \qquad \text{(S.15)}
\end{aligned}$$

Combining (S.13)–(S.15) yields that

$$\lim_{n\to\infty} \mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{k}}\sum_{t=1}^{n-d}\big(I_t - J_t - \delta(d)K_t\big)\right) = \delta^2(d)\big(2\Lambda_1(1,1)-1\big) - 2\delta(d)\left(\tilde{\lambda}_1 - \lambda_1 + \lambda_3 - \frac{1}{2}\right) + 2\Lambda_2(1,1) - 2\lambda_2,$$

which is nothing but $\kappa^2(d)$. Finally, Condition (2) in Utev (1991) can be verified by the same argument used for (S.4). Hence, (S.12) follows. □

We present two auxiliary lemmas used to derive the covariance structures of W_d in Proposition 5.1, $\tilde{W}$ in Proposition 5.2, and G in the proof of Theorem 2.2.

Lemma S.1. *Let $x, y \in \mathscr{X}$. Define $I_t(x) := \mathbb{1}\{U^{\max}_{t+1,t+d-1} < 1 - kx/n < U_t\}$ for $i = 1,\dots,n-d+1$. Assume that $r_n k/n = o(1)$.*

(a) If Assumption A2 holds, then

$$\sum_{t=1}^{r_n}(n-t)\,\mathrm{Cov}(I_1(x), I_{1+t}(x)) = o(k), \qquad \sum_{t=d}^{r_n}(n-t)\,\mathrm{Cov}(I_1(x), I_{1+t}(y)) = o(k).$$

(b) If Assumption A3 holds, then

$$\lim_{n\to\infty}\frac{1}{k}\sum_{t=1}^{r_n}(n-t)\,\mathrm{Cov}\left(\mathbb{1}\{U_1 > 1-kx/n\}, \mathbb{1}\{U_{1+t} > 1-kx/n\}\right) = \Lambda_1(x,x),$$

$$\lim_{n\to\infty}\frac{1}{k}\sum_{t=1}^{r_n}(n-t)\,\mathrm{Cov}\left(\mathbb{1}\{U_1 > 1-kx/n\}, \mathbb{1}\{U_{1+t} > 1-ky/n\}\right) = \Lambda_1(x,y).$$

(c) If Eq. (2.13) holds, then

$$\lim_{n\to\infty}\frac{1}{k}\sum_{t=1}^{r_n}(n-t)\,\mathrm{Cov}(I_1(x), I_{1+t}(x)) = \Lambda_2(x,x),$$

$$\lim_{n\to\infty}\frac{1}{k}\sum_{t=d}^{r_n}(n-t)\,\mathrm{Cov}(I_1(x), I_{1+t}(y)) = \Lambda_2(x,y).$$

Proof. Observe that the results in Parts (a)–(b) are both special cases of Part (c). For Part (a), Assumption A2 implies that (2.13) holds and that $\Lambda_2(x,y) = 0$ for $x, y \in \mathscr{X}$. Part (b) is a special case of Part (c) with $d = 1$. When $d = 1$, the function Λ_2 in (2.13) coincides with Λ_1 in Assumption A3. We now present the proof of Part (c). Note that $\sup_{x\in\mathscr{X}} \mathbb{E}\big(I_1(x)\big) \le \mathbb{P}\big(U_1 > 1 - 3k/(2n)\big) = O(k/n)$. By construction, $\mathbb{E}\big(I_1(x) I_{1+t}(x)\big) =$

0 for $1 \le t \le d-1$. Thus, by (2.13),

$$\begin{aligned}\frac{1}{k}\sum_{t=1}^{r_n}(n-t)\operatorname{Cov}\big(I_1(x), I_{1+t}(x)\big) &= \frac{1}{k}\sum_{t=1}^{r_n}(n-t)\big[\mathbb{E}(I_1(x)I_{1+t}(x)) - \big(\mathbb{E}(I_1(x))\big)^2\big] \\ &= \frac{1}{k}\sum_{t=d}^{r_n}(n-t)\mathbb{E}(I_1(x)I_{1+t}(x)) - \frac{(\mathbb{E}(I_1(x)))^2}{k}\sum_{t=1}^{r_n}(n-t) \\ &\to \Lambda_2(x,x).\end{aligned}$$

The other term can be proved similarly. □

LEMMA S.2. *For $d_1, d_2 \in \mathbb{Z}^+$, let $A \in \sigma\big(\mathbb{1}\{U_j > 1-k/n\}, 1 \le j \le d_1\big)$ and $B_i \in \sigma\big(\mathbb{1}\{U_j > 1-k/n\}, i \le j \le i+d_2\big)$ for $i = 1, 2, \ldots, n-d_2$. If $(n/k)\,\mathbb{P}(A) = O(1)$ and $\sum_{i=r_n}^{n}(1-i/n)\phi_n(i) = o(1)$, then*

$$\frac{1}{k}\sum_{i=r_n}^{n-d_2-1}(n-i)\operatorname{Cov}(A, B_{i+1}) = o(1). \tag{S.16}$$

PROOF. By the definition of $\phi_n(\cdot)$ in (2.9), we have that $|\operatorname{Cov}(A, B_{i+1})| = |\mathbb{P}(A\cap B_{i+1}) - \mathbb{P}(A)\mathbb{P}(B_{i+1})| = \mathbb{P}(A)|\mathbb{P}(B_{i+1}|A) - \mathbb{P}(B_{i+1})| \le \mathbb{P}(A)\phi_n(i+1-d_1)$. Thus,

$$\frac{1}{k}\sum_{i=r_n}^{n-d_2-1}(n-i)|\operatorname{Cov}(A, B_{i+1})| \le \frac{n}{k}\,\mathbb{P}(A)\sum_{i=r_n}^{n-d_2-1}\left(1-\frac{i}{n}\right)\phi_n(i+1-d_1) = o(1).$$

This completes the proof. □

S2. ADDITIONAL VALIDATION FOR THE SIMULATIONS

We show how to compute θ and to validate $D^{(d)}(u_n)$ using Theorem 5.1 in the main manuscript for the example DGPs in Section 3, Table 1. We choose u_n such that $n\mathbb{P}(X > u_n) = 1$. The results for AR-C and ARCH model are provided in Propositions S.1 and S.2 below. For AR-N model, the asymptotic independence of multivariate normal random variables indicates that $\ell_s = s$, $s \ge 1$. Thus, $\theta = \Delta(s) = \ell_s - \ell_{s-1} = 1$ for any $s \ge 1$, and $d_L = 1$. For Moving Maximum DGP, $F(x) = \exp(-1/x)$, and for $s \ge 2$,

$$\ell_s = \lim_{n\to\infty} n\mathbb{P}(\max_{1\le t\le s} X_t > u_n) = \lim_{n\to\infty} n\left(1 - F^{\frac{s+m-1}{m}}(u_n)\right) = \frac{s+m-1}{m}.$$

We have $\Delta(s) = \ell_s - \ell_{s-1} = 1/m$, and thus $D^{(d)}(u_n)$ is satisfied for any $d \geq 2$. We obtain $d_L = 2$ and $\theta = 1/m$. Moreover, for the Max AR model, for $s \geq 2$,

$$\ell_s = \lim_{n\to\infty} n\mathbb{P}\Big(\max_{1\leq t\leq s} X_t > u_n\Big) = \lim_{n\to\infty} n\left(1 - \mathbb{P}\left(\epsilon_1 \leq (1-\varrho)u_n, \max_{2\leq t\leq s} \epsilon_t \leq u_n\right)\right)$$
$$= s - s\varrho + \varrho.$$

We have $\Delta(s) = \ell_s - \ell_{s-1} = 1 - \varrho$, and therefore, $d_L = 2$ and $\theta = 1 - \varrho$.

Proposition S.1. *For the AR(1) model with Cauchy margin specified in Table 1,*

(a) for $\varrho \geq 0$, $\ell_s = s - (s-1)\varrho$, for $s \geq 2$;

(b) for $\varrho < 0$, $\ell_2 = 2$ and $\ell_s = s - (s-1)|\varrho|^2$ for $s \geq 3$.

Proof. This result is easily derived by using the independence of $(X_1, \epsilon_2, \dots, \epsilon_s)$. Let v_n be such that $\lim_{n\to\infty} n\mathbb{P}(\epsilon_t > v_n) = 1$. Then $v_n = (1 - |\varrho|)u_n$. For $s \geq 2$, we have

$$\begin{aligned}
\ell_s &= \lim_{n\to\infty} n\,\mathbb{P}\big(X_1 > u_n \text{ or } \dots \text{ or } X_s > u_n\big)\\
&= \lim_{n\to\infty} n\,\mathbb{P}\big(X_1 > u_n \text{ or } \dots \text{ or } \varrho^{s-1}X_1 + \varrho^{s-2}\epsilon_2 + \dots + \epsilon_s > u_n\big)\\
&= \lim_{n\to\infty} n\,\mathbb{P}\Big(\frac{X_1}{u_n} > 1 \text{ or } \dots \text{ or } \varrho^{s-1}\frac{X_1}{u_n} + \varrho^{s-2}\frac{\epsilon_2}{u_n} + \dots + \frac{\epsilon_s}{u_n} > 1\Big)\\
&= \lim_{n\to\infty} n\,\mathbb{P}\Big(\frac{X_1}{u_n} > 1 \text{ or } \dots \text{ or } \varrho^{s-1}\frac{X_1}{u_n} + \varrho^{s-2}(1-|\varrho|)\frac{\epsilon_2}{v_n} + \dots + (1-|\varrho|)\frac{\epsilon_s}{v_n} > 1\Big)\\
&= \nu\big\{(t_1,\dots,t_s) : t_1 > 1 \text{ or } \dots \text{ or } \varrho^{s-1}t_1 + \varrho^{s-2}(1-|\varrho|)t_2 + \dots + (1-|\varrho|)t_s > 1\big\},
\end{aligned}$$

where ν denotes the exponent measure of $(X_1, \epsilon_2, \dots, \epsilon_s)$; see de Haan and Ferreira (2006, Section 6.1.3) for the definition of exponent measure. The last convergence follows from Theorem 6.1.11 in de Haan and Ferreira (2006) and the fact that the distribution of $(X_1, \epsilon_2, \dots, \epsilon_s)$ belongs to the max domain of attraction. Due to the exact independence between X_1 and the ϵ_t's, and hence their asymptotic independence, the exponent measure ν places mass only on the coordinate axes; that is, $\nu\{(t_1,\dots,t_s) : t_i > a_1 \text{ and } t_j > a_2\} = 0$, for any $i \neq j$ and positive a_1, a_2. Then, the result readily follows from the property that $\nu\{(t_1,\dots,t_s) :| t_i |> a_1\} = 1/a_1$. □

Proposition S.2. *An ARCH model specified in Table 1 does not satisfy the $D^{(d)}(u_n)$ condition for any finite d.*

Proof. Let $M_{i,j} = -\infty$ for $i > j$ and $M_{i,j} = \max_{i\leq t\leq j} X_t$ for $i \leq j$. We apply Proposition 6.2 of Ehlert et al. (2015) to show that for any finite d,

$$\lim_{n\to\infty} \mathbb{P}(M_{2,d} \le u_n < M_{d+1,r_n} \mid X_1 > u_n) > 0.$$

In this proof, all the cited equations are referred to the formulas in Ehlert et al. (2015). Note that ARCH(1,1)model is a special case of the model considered in that paper, which corresponds to $\delta_1 = \beta_1 = 0$ in the model given by relations (6.2) and (6.3) in that paper. Therefore, $\phi(x) = \alpha_1^{1/2}|x|$, for the ϕ appeared in the limit of (6.14) in that paper. Let W denote a random variable from Pareto distribution with parameter α and $(Z_i)_{i\ge 1}$ are i.i.d. standard normal random variables. Then by Proposition 6.2 of Ehlert et al. (2015),

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(M_{2,d} \le u_n < M_{d+1,r_n} \mid X_1 > u_n) \\
&\ge \lim_{n\to\infty} \mathbb{P}(M_{2,d} \le u_n < X_{d+1} \mid X_1 > u_n) \\
&= \mathbb{P}\left(\max_{2\le i\le d} \frac{W}{|Z_0|} Z_i \prod_{j=0}^{i-1} \phi(Z_j) \le 1 < \frac{W}{|Z_0|} Z_{d+1} \prod_{j=0}^{d} \phi(Z_j) \right) \\
&= \mathbb{P}\left(W \max_{2\le i\le d} \alpha_1^{i/2} Z_i \prod_{j=1}^{i-1} |Z_j| \le 1 < W \alpha_1^{(d+1)/2} Z_{d+1} \prod_{j=1}^{d} |Z_j| \right) \\
&\ge \mathbb{P}\left(W > \alpha_1^{-(d+1)/2}, \max_{2\le i\le d} Z_i \le -1, Z_{d+1} > 1 \right) = \alpha_1^{\alpha(d+1)/2} (\Phi(-1))^d > 0,
\end{aligned}$$

where $\alpha_1 \in (0,1)$ (which equals 0.7 in our simulation example), $\alpha > 0$, and Φ is a standard normal distribution function. □

S3. DETAILS OF THE BOOTSTRAP PROCEDURE AND ITS PERFORMANCE

We describe the block-type bootstrap scheme, originally proposed by Ferro and Segers (2003), which forms the building blocks of the bootstrap used in Section 4. It consists of two main steps.

Step A1 Construct cluster sets $\{\mathbb{C}_j\}_{j=1}^{J}$ and gap sets $\{\mathbb{G}_j\}_{j=1}^{J-1}$, where J is the number of clusters.

(i) Suppose that we observe k exceedance times $S_1 < \ldots < S_k$. Form clusters by assigning two consecutive exceedances S_i and S_{i+1} to the same cluster whenever $S_{i+1} - S_i < r$, for $i = 1, \ldots, k-1$, where $r > 0$ is some threshold. If $S_{i+1} - S_i \ge r$, then a new cluster starts.

(ii) Suppose the jth cluster contains the exceedance times $S_{a_j}, S_{a_j+1}, \ldots, S_{b_j}$, where

$a_j < b_j$ are the first and last exceedance positions. The corresponding cluster and gap are defined as

$$\mathbb{C}_j = \left\{X_{S_{a_j}}, X_{S_{a_j}+1}, \ldots, X_{S_{b_j}}\right\}, \qquad \mathbb{G}_j = \left\{X_{S_{b_j}+1}, \ldots, X_{S_{a_{j+1}}-1}\right\}.$$

Step A2 Resample cluster sets from $\{\mathbb{C}_j\}_{j=1}^{J}$ and gap sets from $\{\mathbb{G}_j\}_{j=1}^{J-1}$ with replacement. Arrange them consecutively until the length of the concatenated sequence reaches the sample size n, and then truncate the sequence to length n.

Throughout, two tuning parameters are required: the number of excedeence k and the length r that is used to define clusters. For bootstrapping the extremal index θ, Ferro and Segers (2003) recommend using $r = \lfloor 1/\hat{\theta} \rfloor$ while we propose using $r = \hat{d}_L$, which can be interpreted as the minimal block length to capture the extremal dependence. Our choice is motivated by Conditions $D(u_n)$ and $D^{(d)}(u_n)$. Condition $D(u_n)$ ensures that the two blocks separated by at least r_n observations are approximately independent. Meanwhile, Condition $D^{(d)}(u_n)$ implies

$$\mathbb{P}\left(X_{2,d}^{\max} \leq F^{-1}(u_n) \mid X_1 > F^{-1}(u_n)\right) - \mathbb{P}\left(X_{2,r_n}^{\max} \leq F^{-1}(u_n) \mid X_1 > F^{-1}(u_n)\right) \to 0,$$

or equivalently,

$$\mathbb{P}\left(S_2 - S_1 \geq d \mid X_{S_1} > F^{-1}(u_n)\right) - \mathbb{P}\left(S_2 - S_1 \geq r_n \mid X_{S_1} > F^{-1}(u_n)\right) \to 0,$$

which means that, under Condition $D^{(d)}(u_n)$, a separation of d is asymptotically equivalent to a separation of r_n. Thus, using d_L, i.e., the smallest d such that Condition $D^{(d)}(u_n)$ holds, is justified.

We compare the two choices by examining the empirical coverage rate and the empirical length of the resulting confidence intervals. For an AR–C model, as described in Section 3, with $\theta = 0.2$ ($d_L = 2$), the results are shown in Figure S.1. Our choice performs comparably in terms of both empirical coverage and interval length, while relying on a smaller r. As the sample size n increases, the empirical coverage improves and the interval length decreases, as one would expect.

S4. ADDITIONAL EMPIRICAL RESULTS

This section documents the following: additional descriptive figures; details on the data-driven selection of the local-dependence window; the construction of the marginal ex-

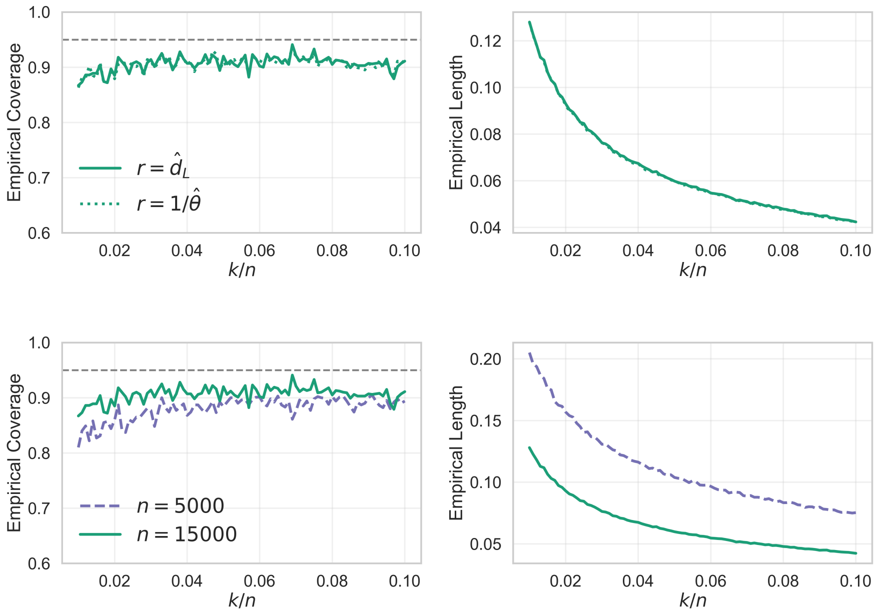

Figure S.1. Top panel: comparison of the two bootstrap parameter choices, $r = \hat{d}_L$ and $r = \lfloor 1/\hat{\theta} \rfloor$. Bottom Panel: comparison across sample sizes, using $r = \hat{d}_L$ in both cases. All results are based on 1000 Monte Carlo replications, each with 199 bootstrap samples.

ceedance probability together with additional comparative results for severity probability; and a detailed resampling scheme for uncertainty quantification.

Figure S.2 displays the time series plots of X_t (summer apparent temperatures) for Paris, Budapest, Munich, Rome, Barcelona, and Valencia over 1940–2025. All six cities exhibit a pronounced upward shift in the upper tail between periods, consistent with the patterns documented in the paper.

To determine the parameter d and assess Condition $D^{(d)}(u_n)$, we apply the selection rule in (2.19). Figure S.3 shows $\hat{\delta}_n(s), s = 1, \ldots, 4$ over $k/n \in [0.02, 0.08]$. The common pattern for Paris, Munich, Budapest, and Rome is clear: $\hat{\delta}_n(1)$ (solid green line) lies above $1/\sqrt{k}$ (dashed black line), while $\hat{\delta}_n(s)$ for $s = 2, 3, 4$ lie below $1/\sqrt{k}$, indicating that $\hat{d}_L = 2$. For Barcelona and Valencia, though we observe $\hat{\delta}_n(2)$ slightly above $1/\sqrt{k}$ for large k/n, it is still well separated from $\hat{\delta}_n(1)$. One may still conclude that $\hat{d}_L = 2$ for Barcelona and Valencia. Taken together, we conclude that $D^{(2)}(u_n)$ holds for all city–period pairs.

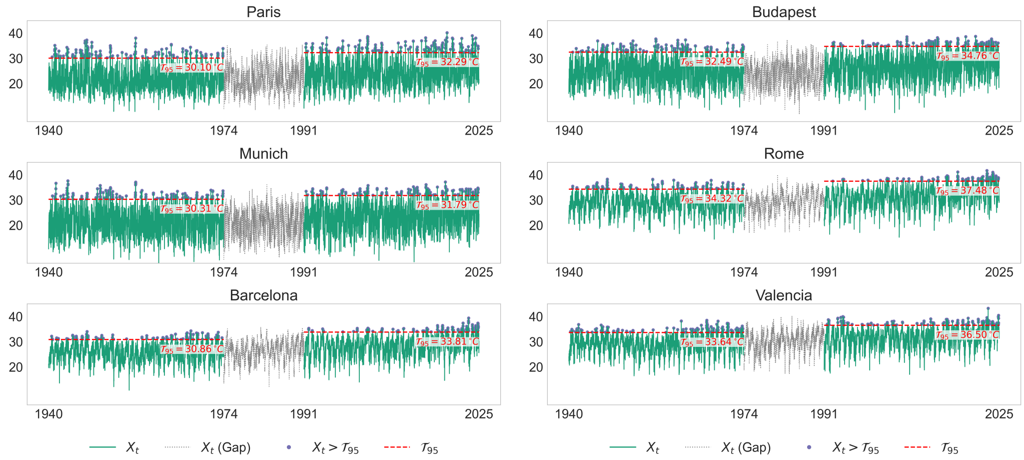

Figure S.2. Time series X_t during summer days from 1940 to 2025. The two subsamples covering 1940–1974 and 1991–2025 are shown as solid green lines, while observations in the intervening gap, denoted by X_t(Gap), are plotted as a gray dotted line. The red dashed lines labeled $\mathcal{T}_{95}$ represent the 95th percentiles within each subperiod, and the purple dots mark observations exceeding these respective thresholds.

Accordingly, we estimate the conditional probability in $S(\mathcal{T})$ by $1-\hat{\theta}$. For the marginal exceedance probability $P_{\mathcal{T}}$ in Section 4, it follows the standard approach in extreme value theory. Let $\hat{u} = X_{n-k,n}$, and $\hat{p}_{\hat{u}} = k/n$:

$$\hat{P}_{\mathcal{T}} = \begin{cases} n^{-1}\sum_{t=1}^{n} \mathbb{1}\{X_t > \mathcal{T}\}, & \mathcal{T} \le \hat{u}, \\ 0, & \mathcal{T} \ge \hat{x}^*, \\ \hat{p}_u \left(1 + \hat{\gamma}\, \frac{\mathcal{T}-\hat{u}}{\hat{\sigma}}\right)^{-1/\hat{\gamma}}, & \hat{u} < \mathcal{T} < \hat{x}^*, \end{cases}$$

where $\hat{\gamma}$, $\hat{\sigma}$, and $\hat{x}^*$ are obtained by maximum likelihood estimation under the generalized Pareto fit. This yields a coherent $\hat{P}_{\mathcal{T}}$ at high levels even when empirical exceedances are sparse.

Table S.1 compares our $\hat{S}(\mathcal{T})$ with the empirical joint exceedance estimator $\hat{S}_{\text{emp}}(\mathcal{T}) = (n-1)^{-1}\sum_{t=1}^{n-1} \mathbb{1}\{X_t > \mathcal{T}, X_{t+1} > \mathcal{T}\}$. Recall that $\mathcal{T}_p^{\text{fs}}$ denotes the $p\%$ percentile of the full sample from 1940 to 2025. We see that $\hat{S}_{\text{emp}}$ is zero for many cities for $\mathcal{T}_{99.9}^{\text{fs}}$, while $\hat{S}(\mathcal{T})$ remains informative.

Finally, we now provide the details of the resampling procedure used in Section 4, based on the bootstrap scheme described in Section S3. As discussed, we set $r = \hat{d}_L$. Recall that the original sample from 1940 to 2025 is split into three subsamples: Period 1 ($\mathbb{X}_{\text{P1}}$), a middle gap period ($\mathbb{X}_{\text{gap}}$), and Period 2 ($\mathbb{X}_{\text{P2}}$). The extremal indices $\hat{\theta}_{\text{P1}}$, $\hat{\theta}_{\text{P2}}$ are

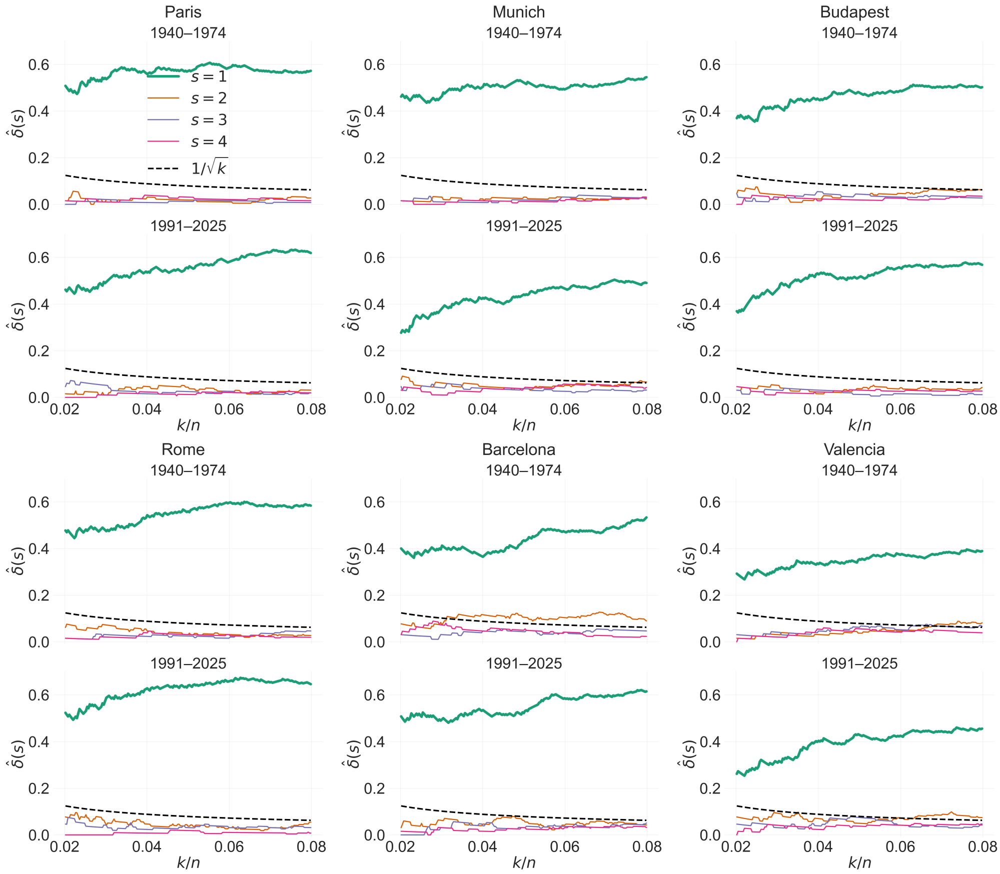

Figure S.3. $\hat{\delta}_n(s)$ for $s = 1, \ldots, 4$, as defined in (2.12), over a range of $k/n \in [0.02, 0.08]$ for Paris, Munich, Budapest (Top two panels), and Rome, Barcelona, Valencia (Bottom two panels).

estimated from $\mathbb{X}_{\mathrm{P1}}$ and $\mathbb{X}_{\mathrm{P2}}$, respectively. For estimating the heatwave severity probability $S(\mathcal{T})$, the threshold $\mathcal{T}$ is taken as a quantile of the full sample (1940–2025). The estimates $\hat{\theta}_{\mathrm{P1}}$, $\hat{\theta}_{\mathrm{P2}}$, as well as $\hat{\Theta} := \{\hat{\gamma}_{\mathrm{P1}}, \hat{\gamma}_{\mathrm{P2}}, \hat{u}_{\mathrm{P1}}, \hat{u}_{\mathrm{P2}}, \hat{\sigma}_{\mathrm{P1}}, \hat{\sigma}_{\mathrm{P2}}\}$ for the marginal probability $\hat{P}_{\mathcal{T}}$ will be used below in the bootstrap. For a given ratio k/n, in each bootstrap iteration, we proceed as follows:

Step B1 Construct the bootstrap sample $\mathbb{X}^*_{\mathrm{P1}}$ from $\mathbb{X}_{\mathrm{P1}}$ using the bootstrap scheme (Steps A1–A2) described in Section S3. Similarly, obtain the bootstrap samples $\mathbb{X}^*_{\mathrm{gap}}$ from $\mathbb{X}_{\mathrm{gap}}$ and $\mathbb{X}^*_{\mathrm{P2}}$ from $\mathbb{X}_{\mathrm{P2}}$, yielding three resampled segments.

Step B2 Re-estimate θ separately using $\mathbb{X}^*_{\mathrm{P1}}$ and $\mathbb{X}^*_{\mathrm{P2}}$, yielding bootstrap estimates θ^*_{P1} and θ^*_{P2}, respectively.

Table S.1. Estimates of the heatwave severity probability $S(\mathcal{T})$ across two periods P_1 (1940–1974) and P_2 (1991–2025) for $\mathcal{T} = \mathcal{T}_{97}^{\text{fs}}$ and $\mathcal{T}_{99.9}^{\text{fs}}$, where $\mathcal{T}_p^{\text{fs}}$ denotes the p% percentile of the full sample from 1940 to 2025. For display purposes, the $\hat{S}$ and $\hat{S}_{\text{emp}}$ are multiplied by 10^2. The "Parameters" columns reports the estimated extremal index $\hat{\theta}$ and the parameters from the GPD fit obtained by maximum likelihood, namely $\hat{\gamma}$, $\hat{u}$, and $\hat{\sigma}$, both using $k/n = 5\%$.

	$\mathcal{T}_{97}^{\text{fs}}$			$\mathcal{T}_{99.9}^{\text{fs}}$			Parameters			
City	$\mathcal{T}_{97}^{\text{fs}}$	$\hat{S}(P_1, P_2)$	$\hat{S}_{\text{emp}}(P_1, P_2)$	$\mathcal{T}_{99.9}^{\text{fs}}$	$\hat{S}(P_1, P_2)$	$\hat{S}_{\text{emp}}(P_1, P_2)$	$\hat{\theta}(P_1, P_2)$	$\hat{\gamma}(P_1, P_2)$	$\hat{u}(P_1, P_2)$	$\hat{\sigma}(P_1, P_2)$
London	28.37	0.99, 2.04	0.81, 2.02	33.75	0.046, 0.111	0.031, **0**	0.45, 0.53	-0.15, -0.17	25.94, 28.02	2.58, 2.42
Paris	32.52	0.96, 1.93	0.68, 1.77	38.10	0.006, 0.105	**0**, 0.062	0.48, 0.58	-0.29, -0.38	30.10, 32.29	2.79, 3.27
Munich	31.93	1.10, 2.06	0.96, 2.05	36.14	0.068, 0.039	0.093, **0**	0.48, 0.56	-0.12, -0.34	30.31, 31.79	1.96, 1.98
Budapest	34.51	0.68, 2.95	0.34, 3.14	37.69	0.010, 0.111	**0**, 0.031	0.52, 0.48	-0.30, -0.28	32.49, 34.76	1.92, 1.41
Milan	36.84	0.39, 3.70	0.22, 4.13	40.43	0.002, 0.206	**0**, 0.062	0.46, 0.40	-0.17, -0.21	34.45, 37.25	1.46, 1.55
Rome	37.08	0.18, 4.37	0.09, 4.47	40.09	0.001, 0.178	**0**, 0.031	0.44, 0.36	-0.21, -0.23	34.32, 37.48	1.33, 1.23
Barcelona	33.49	0.08, 3.35	**0**, 3.82	36.81	**0**, 0.146	**0**, 0.155	0.46, 0.48	-0.26, -0.12	30.86, 33.81	1.13, 1.23
Valencia	36.27	0.17, 2.49	0.12, 2.45	39.53	0.002, 0.967	**0**, **0**	0.65, 0.57	-0.12, -0.06	33.63, 36.50	1.30, 1.07
Athens	37.32	0.96, 2.45	0.96, 2.36	41.39	0.020, 0.687	0.031, **0**	0.43, 0.45	-0.10, -0.19	35.85, 37.13	1.42, 1.61

Step B3 Re-estimate marginal tail parameters, obtaining $\hat{\Theta}^*$

Step B4 Compute the bootstrap counterpart $\mathcal{T}^*$ of the full-sample threshold $\mathcal{T}$ by taking the corresponding percentile of the full bootstrap dataset $\{\mathbb{X}_{\text{P1}}^*, \mathbb{X}_{\text{gap}}^*, \mathbb{X}_{\text{P2}}^*\}$. Using θ_{P1}^*, θ_{P2}^*, and $\hat{\Theta}^*$ from the previous steps, obtain a bootstrap estimate $\hat{S}^*(\mathcal{T}^*)$.

For proper standardization, the confidence intervals for $S(\mathcal{T})$ are constructed by bootstrapping $\log S(\mathcal{T})$ and then transforming back to the original scale. We refer to Theorem 4.4.7 in de Haan and Ferreira (2006) for intuition. By repeating Steps B1–B4 B times, we obtain confidence intervals for θ for each city, as shown in Figure S.4, and confidence intervals for $S(\mathcal{T})$, as displayed in the bottom panel of Figure 4. For a wide range of choices of k/n, the estimates $\hat{\theta} \in (0.4, 0.8)$ support the presence of extremal serial dependence. The confidence intervals overlap across periods for all cities, providing no clear evidence of changes in extremal serial dependence. Note that the top panel of Figure 4 only displays the case of $k/n = 0.05$.

REFERENCES

Aldous, D. (1978). Stopping times and tightness. *Annals of Probability 6*(2), 335–340.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction.* Springer Verlag.

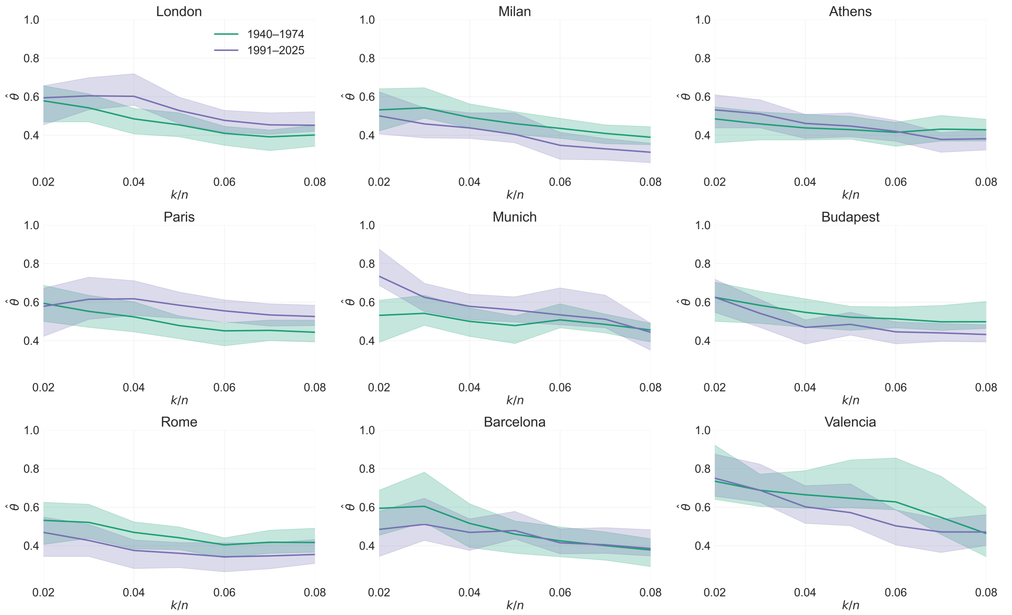

Figure S.4. Estimates of $\hat{\theta}$ for $k/n \in [2\%, 8\%]$, together with 90%-level bootstrap confidence intervals based on $B = 999$ bootstrap replicates.

Eberlein, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statistics & probability letters 2*(5), 291–293.

Ehlert, A., U.-R. Fiebig, A. Janßen, and Schlather (2015). Joint extremal behavior of hidden and observable time series with applications to GARCH processes. *Extremes 18*(1), 109–140.

Einmahl, J. and F. Ruymgaart (2000). Some results for empirical processes of locally dependent arrays. *Mathematical Methods of Statistics 9*(4), 399–414.

Ferro, C. A. T. and J. Segers (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 65*(2), 545–556.

Utev, S. (1991). On the central limit theorem for φ-mixing arrays of random variables. *Theory of Probability & Its Applications 35*(1), 131–139.