REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

PERFORMANCE ASSESSMENT OF SANDWICH AND BLOCK BOOTSTRAP ESTIMATORS FOR TEMPORALLY DEPENDENT BIVARIATE EXTREMES

Authors: Alexander Engberg ✉
– Department of Statistics, Uppsala University, Sweden
alexander.engberg@statistics.uu.se

Received: Month 0000   Revised: Month 0000   Accepted: Month 0000

Abstract:

- Ignoring temporal dependence when modelling sequences of extreme observations yields underestimated standard errors which can lead to inaccurate risk assessment of extreme phenomena such as floods and economic crises. One remedy is to inflate standard errors with a sandwich or block bootstrap estimator. In this study, the performance of four such standard error estimators is investigated, through simulation, when modelling extremes from bivariate sequences. The results show that under strong temporal dependence, all considered estimators seriously underestimate standard errors, while under moderate to weak dependence both the sandwich and the bootstrap estimators can mitigate this underestimation.

Keywords:

- *Block bootstrap; Sandwich estimator; Logistic model; Hüsler-Reiss model; Bivariate extremes; Temporal dependence.*

AMS Subject Classification:

- 62G32, 62M10.

✉ Corresponding author

1. INTRODUCTION

Extreme phenomena such as floods, droughts and economic crises cause devastating damage to our societies and tend to be driven by concatenated rare events in complex systems such as rivers or financial networks. Flooding in the lower part of a river system is often the result of simultaneous high flows further up in the system [39, 17]. Severe droughts are the result of joint extremes of several meteorological variables [41]. The risk of economic crises is highly dependent on the connections between different financial institutions [40]. In all these examples, the multivariate dependence between individual rare events determines the risk of an extreme phenomenon, and multivariate extreme-value theory provides a suitable framework for modelling such dependence. Standard extreme-value modelling approaches, however, rely on the assumption that data consists of independent and identically distributed (i.i.d.) random vectors, while in applications, temporal dependence in data is common. Failure to account for temporal dependence leads to underestimating the uncertainty of estimated quantities, such as the strength of multivariate dependence, and, in consequence, inaccurate risk assessment. Against this background, the present study is concerned with methods to account for the effect that temporal dependence in data has on the estimated parameter uncertainty when modelling bivariate dependence between rare events.

An often more realistic assumption than that of i.i.d. observations is that data constitutes a stationary sequence of random vectors. Under stationarity, temporal dependence can be explicitly modelled in conjunction with the multivariate model, or accounted for after model estimation by adjusting standard errors with block bootstrap estimators [19, 25, 29] or sandwich covariance matrix estimators [13, 38, 34]. Explicit modelling might yield the most accurate results, but models can be application specific or intractable, and if the model fit is poor, results may be misleading. To adjust the standard errors with a block bootstrap or sandwich estimator is more robust and often simpler. However, block bootstraps are computationally expensive and sensitive to the choice of block length, and while sandwich estimators are computationally cheaper, they might be subject to numerical difficulties.

The goal of this study is to, through simulation, investigate and compare the performance of four standard error estimators when modelling the bivariate dependence between extreme values from bivariate stationary sequences. Focus is on the logistic extreme-value model due to its relative simplicity, which makes it a good model of reference. The results are, however, generalized both to higher dimensions and to additional models. The considered estimators are, first, the sandwich estimators of [13] and [34], where the former is combined with the Newey-West estimator [26] to account for temporal dependence. Second, the moving block- and stationary bootstrap estimators of [19] and [25], and [29] respectively. Focus is on how the uncertainty of the estimated dependence strength is affected by temporal dependence in data, and to what extent the considered methods can mitigate underestimation.

To the author's knowledge, no comparative performance assessment of sandwich and block bootstrap estimators has been conducted in the context of modelling extremes from bivariate (or multivariate) stationary sequences. The performance of some estimators have, however, been examined in specific settings. [11] show that the sandwich estimator of [34] yields notably less biased standard errors of parameter estimates when modelling univariate threshold exceedances from first order Markov chains; [27] uses the sandwich estimator of [38] and the stationary bootstrap [29] to adjust standard errors of different extremal index estimators for univariate block-maxima showing that the bootstrap generally provides a good bias reduction; and [15] show that the stationary bootstrap adequately captures temporal dependence in rare events when modelling extremes from spatial processes with a censored likelihood.

In Section 2, a brief overview of bivariate extreme-value theory is given. In Section 3, the block bootstrap and sandwich estimators are presented in more detail, and in Section 4, the performance of the estimators is assessed through simulation. The outcomes of the study are discussed in Section 5.

2. BIVARIATE EXTREMES

In this section, an overview of the bivariate extreme-value theory used in this study is given. For a more comprehensive account of multivariate extreme-value theory, see [7], [3, Ch. 8], and [2, Ch. 8-9].

2.1. Block-maxima

Let $\{(X_i, Y_i)\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. bivariate random vectors with joint cumulative distribution function $F(x, y)$, and denote the vector of componentwise maxima $\mathbf{M}_n = (M_{x,n}, M_{y,n}) = (\max_{i=1,\dots,n}(X_i), \max_{i=1,\dots,n}(Y_i))$. It should be noted that the maximum might not occur for the same index i in both X_i and Y_i which means that the joint maxima $\mathbf{M}_n$ might not correspond to an actual observation. Furthermore, assume that there exist sequences $(a_{x,n}, a_{y,n}) > 0$ and $(b_{x,n}, b_{y,n})$, and a non-degenerate distribution $G(x, y)$, such that the distribution of normalized maxima converges as

$$\lim_{n\to\infty} \Pr\left(\frac{M_{x,n} - b_{x,n}}{a_{x,n}} \le x, \frac{M_{y,n} - b_{y,n}}{a_{y,n}} \le y\right) = G(x, y), \tag{2.1}$$

where $G(x, y)$ has non-degenerate marginals $G_X(x)$ and $G_Y(y)$. Then, $G(x, y)$ is a bivariate extreme-value distribution, and the marginals belong to the generalized extreme-value (GEV) family of distributions, which has the distribution function

$$G_Z(z) = \begin{cases} \exp\left[-\left(1+\xi\left(\frac{z-\mu}{\tau}\right)\right)_+^{-1/\xi}\right], & \xi \neq 0, \\ \exp\left[-\exp\left(-\frac{z-\mu}{\tau}\right)\right], & \xi = 0, \end{cases} \tag{2.2}$$

defined on $\{z : 1+\xi(z-\mu)/\tau > 0\}$, and where $c_+ = \max(c, 0)$, and $\mu \in \mathbb{R}$, $\tau \in \mathbb{R}_+$ and $\xi \in \mathbb{R}$ are location, scale and shape parameters which can be different for the respective marginals. A customary assumption is that the univariate marginals of $\{(X_i, Y_i)\}_{i\in\mathbb{N}}$ follow the unit Fréchet distribution (corresponding to GEV($\mu = \tau = \xi = 1$)). This assumption is not restrictive because the marginals can be transformed with the probability integral transform to approximately follow the unit Fréchet distribution once the parameters of the GEV distribution have been estimated. With unit Fréchet marginals the joint distribution function $G(x, y)$ can be expressed as

$$G(x, y) = \exp[-V(x, y)],\ x > 0, y > 0, \tag{2.3}$$

[9]. The function V can be written as

$$V(x, y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right)\ \mathrm{d}H(w), \tag{2.4}$$

where H is a distribution function that determines the bivariate dependence structure and satisfies the mean constraint

$$\int_0^1 w\ \mathrm{d}H(w) = \int_0^1 (1-w)\ \mathrm{d}H(w) = 1/2, \tag{2.5}$$

[3, 24].

In practice, vectors of component-wise maxima $\mathbf{M}_n$ are obtained by splitting the data sequence into large disjoint blocks and extracting the maxima from each block, such as annual maximum daily rainfall. For sufficiently large blocks the block-maxima can be viewed as approximate realizations from a GEV distribution. Maximum likelihood estimation is often used to estimate the model parameters [30], and the marginal and dependence parameters can be estimated either separately or simultaneously. Simultaneous estimation has the benefit that information can be shared across marginals, although at a higher computational cost. The procedure of modelling maxima from large disjoint blocks of data is referred to as the *block-maxima* modelling approach.

2.2. Threshold exceedances

In many cases, rare events occur in clusters, such as consecutive days of heavy rainfall or high temperatures, which the analysis of block-maxima overlooks. A more efficient use of data is achieved with the *peaks over thresholds*

(POT) approach, in which exceedances over some high thresholds, u_x and u_y, are studied. More specifically, for a random variable X with distribution function $F_X(x)$, define the right endpoint of the distribution as $x_F = \sup\{x : F_X(x) < 1\}$. Then, if the distribution of normalized block-maxima from $F_X(x)$ converge to a GEV distribution, the limiting distribution of suitably normalized conditional excesses $((X-u)/a(u) \mid X > u)$, $a(u) > 0$, as $u \nearrow x_F$, is a generalized Pareto (GP) distribution with distribution function

$$H(z) = \begin{cases} 1 - \left(1 + \xi \frac{z}{\tilde{\tau}}\right)_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{z}{\tilde{\tau}}\right), & \xi = 0, \end{cases} \tag{2.6}$$

defined on $\{z : z > 0 \text{ and } (1 + \xi z/\tilde{\tau}) > 0\}$, with parameters $\tilde{\tau} \in \mathbb{R}_+$ and $\xi \in \mathbb{R}$, and where $\tilde{\tau} = \tau + \xi u$ [1, 28]. Thus, for high thresholds, a GP likelihood provides an appropriate model for threshold excesses.

In practice, for bivariate sequences $\{(X_i, Y_i)\}_{i \in \mathbb{N}}$, GP likelihoods are fitted to each marginal, and the marginal distributions are transformed to unit Fréchet scale. To model the bivariate dependence structure it can be shown that with unit Fréchet marginals, if the thresholds are sufficiently high and ($X_i \geq u_x$, $Y_i \geq u_y$), the limiting joint distribution of conditional excesses is approximately the multivariate extreme-value distribution in (2.3) [2, p. 276]. This fact was used by [23] as the basis for a censored likelihood for threshold exceedances in which contributions of observations are censored from below at the thresholds. To specify the censored likelihood in the bivariate case, let λ_x and λ_y be some small probabilities and set $r_x = -1/\ln(1-\lambda_x)$ and $r_y = -1/\ln(1-\lambda_y)$ which corresponds to marginal thresholds transformed to unit Fréchet scale. Then the bivariate censored likelihood can be expressed as

$$L(\theta; x, y) = \begin{cases} \frac{\partial^2}{\partial x \partial y} \exp[-V(x,y)] & \text{if } x > r_x, y > r_y, \\ \frac{\partial}{\partial x} \exp[-V(x, r_y)] & \text{if } x > r_x, y \leq r_y, \\ \frac{\partial}{\partial y} \exp[-V(r_x, y)] & \text{if } x \leq r_x, y > r_y, \\ \exp[-V(r_x, r_y)] & \text{if } x \leq r_x, y \leq r_y. \end{cases} \tag{2.7}$$

2.3. Models

To form tractable models from the distribution in (2.3), a common approach is to approximate the family of distributions with some parametric sub-family. A frequently used alternative is the logistic extreme-value model [14] which has the exponent function

$$V(x,y) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}, \; x > 0, y > 0. \tag{2.8}$$

The parameter $0 < \alpha \leq 1$ governs the dependence with $\alpha = 1$ corresponding to independence and the limiting case $\alpha \to 0$ to complete dependence. The logistic model is relatively easy to work with and is therefore chosen as a model of reference in this study. It is also related to more flexible models, such as the asymmetric logistic model presented below.

To assess the generality of the results, the asymmetric logistic [37] and the Hüsler-Reiss [16] models are also considered. Assuming unit Fréchet marginals, the asymmetric logistic model has the exponent function

$$V(x,y) = \frac{1-\psi_1}{x} + \frac{1-\psi_2}{y} + \left\{ \left(\frac{\psi_1}{x}\right)^{1/\beta} + \left(\frac{\psi_2}{y}\right)^{1/\beta} \right\}^{\beta}, \; x>0, y>0. \tag{2.9}$$

Here, dependence is governed by the parameter $0 < \beta \leq 1$, and the asymmetry parameters $0 \leq \psi_1, \psi_2 \leq 1$. As such, this model has two additional parameters compared to the (symmetric) logistic model and, including the marginals, it has a total of 9 parameters to be estimated for block-maxima, and 7 for POT. Independence can be achieved when either $\beta = 1$ or $\psi_1 = 0$ or $\psi_2 = 0$, or when $\psi_1 = \psi_2 = \beta = 1$, while complete dependence is obtained in the limit when $\psi_1 = \psi_2 = 1$ and $\beta \to 0$. The exponent function of the Hüsler-Reiss model has the form

$$V(x,y) = \frac{1}{x}\Phi\left[\frac{1}{r} + \frac{r}{2}\ln\left(\frac{y}{x}\right)\right] + \frac{1}{y}\Phi\left[\frac{1}{r} + \frac{r}{2}\ln\left(\frac{x}{y}\right)\right], \; x>0, y>0, \tag{2.10}$$

where $\Phi(\cdot)$ is the standard normal distribution function. The dependence is governed by $r > 0$, where independence is obtained in the limit as $r \to 0$ and complete dependence corresponds to $r \to \infty$.

3. EXTREMES OF DEPENDENT SEQUENCES

In this section, the modelling of extreme values from temporally dependent sequences is discussed, and the considered sandwich and block bootstrap estimators are described.

3.1. Asymptotic dependence

When analysing extreme values from bivariate (or multivariate) random vectors, the primary interest is often to describe the dependence in the tail, referred to as asymptotic dependence. As standard association metrics often perform poorly in the tail, asymptotic dependence can better be characterized by the tail dependence coefficient χ [4]. For two random variables X and Y with continuous distributions F_X and F_Y, the tail dependence coefficient is defined as

$$\chi = \lim_{u \to 1} P(F_X(X) > u \mid F_Y(Y) > u). \tag{3.1}$$

The variables X and Y are considered asymptotically dependent if $\chi > 0$ and asymptotically independent if $\chi = 0$.

3.2. Modelling of extremes from dependent sequences

Since the block-maxima and POT approaches in their simplest form rely on the assumption of i.i.d. observations, applying them to temporally dependent stationary sequences results in model misspecification. A common practice is therefore to remove, or at least diminish, the dependence before model estimation. For block-maxima, provided that long-range dependence in the sequence is weak at extreme levels, maxima are often close to independent and the limiting distribution still belongs to the GEV family [22]. Thus, applying the block-maxima approach to maxima from stationary sequences yields valid parameter point estimates, although different than if data had been independent [3, p. 96]. However, if some temporal dependence remains, standard errors associated with the parameter estimates are underestimated.

In the POT approach, exceedances from stationary sequences tend to form clusters of dependent extremes above the thresholds. In consequence, the resulting sequence of exceedances constitutes a series of concatenated clusters, with temporal dependence retained within the clusters. Perhaps the most common remedy is declustering, which entails identifying clusters of extremes and only using the maximum from each cluster in the analysis [8]. However, [11] suggest that declustering might induce serious bias in estimates, and showed that fitting GP likelihoods to all exceedances from first order Markov chains yields close to unbiased parameter estimates but underestimated standard errors. A better approach than declustering might therefore be to fit the model to all exceedances and to account for temporal dependence afterwards.

The underestimated standard errors can be inflated with sandwich or block bootstrap estimators to yield more correct uncertainty measurements. In Sections 3.3 and 3.4, the four standard error estimators considered in this study are presented in more detail.

3.3. Sandwich estimators

Under model misspecification, and given regularity conditions, the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$, of the vector of model parameters, $\boldsymbol{\theta}$, converges as

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \tilde{\mathrm{I}}(\boldsymbol{\theta})^{-1}) \text{ as } n \to \infty, \tag{3.2}$$

with $\tilde{\mathrm{I}}(\boldsymbol{\theta})^{-1} = \mathrm{H}(\boldsymbol{\theta})^{-1}\mathrm{J}(\boldsymbol{\theta})\mathrm{H}(\boldsymbol{\theta})^{-1}$. Here, $\tilde{\mathrm{I}}(\boldsymbol{\theta})^{-1}$ is the sandwich covariance matrix, $\mathrm{H}(\boldsymbol{\theta}) = -\mathbb{E}[\nabla^2 \ell(\boldsymbol{\theta}; x, y)]$ is the Fisher information matrix and $\mathrm{J}(\boldsymbol{\theta}) = \mathbb{V}[\nabla \ell(\boldsymbol{\theta}; x, y)]$ is the variance of the score vector, with $\ell(\boldsymbol{\theta}; x, y)$ being the model log-likelihood, and ∇ and ∇^2 the first and second order gradients [5, p. 147].

The negative Hessian matrix

$$-\mathsf{H}_{\hat{\boldsymbol{\theta}}} = -\frac{\partial^2 \ell(\boldsymbol{\theta}; x, y)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{3.3}$$

provides an estimate of $\mathrm{H}(\boldsymbol{\theta})$, and to estimate $\mathrm{J}(\boldsymbol{\theta})$, while also taking temporal dependence into account, the covariance matrix of [26] is used which has the form

$$\mathsf{J}_{\hat{\boldsymbol{\theta}}} = \mathsf{J}_{\hat{\boldsymbol{\theta}},0} + \sum_{j=1}^{m} w_{j,m}(\mathsf{J}_{\hat{\boldsymbol{\theta}},j} + \mathsf{J}_{\hat{\boldsymbol{\theta}},j}^{\mathsf{T}}), \tag{3.4}$$

with

$$\mathsf{J}_{\hat{\boldsymbol{\theta}},j} = \sum_{i=1}^{n-j} \nabla \ell(\hat{\boldsymbol{\theta}}; x_i, y_i) \nabla \ell(\hat{\boldsymbol{\theta}}; x_{i+j}, y_{i+j})^{\mathsf{T}} \text{ and } \ w_{j,m} = 1 - \left(\frac{j}{m+1}\right). \tag{3.5}$$

Here $\mathsf{J}_{\hat{\boldsymbol{\theta}},j}$ is the heavy tailed sample autocovariance [12] at lag j, and $w_{j,m}$ are weights that decline as j increases. The first term on the right hand side of (3.4) is the non-adjusted variance of the score vector, while the second term accounts for temporal dependence up to lag m. An optimal choice of m is determined by the dependence strength and the sequence length. In this study, however, m is chosen based on the actual asymptotic dependence in the data, by studying plots of $\hat{\chi}(u)$ for a range of lags. The "standard" sandwich covariance matrix estimator is then

$$\hat{\hat{\mathrm{I}}}(\boldsymbol{\theta})^{-1} = (-\mathsf{H}_{\hat{\boldsymbol{\theta}}})^{-1} \mathsf{J}_{\hat{\boldsymbol{\theta}}} (-\mathsf{H}_{\hat{\boldsymbol{\theta}}})^{-1}. \tag{3.6}$$

If data consists of blocks of observations, such as annual summer temperatures, with dependence within, but near independence between the blocks, [34] proposed a blocked sandwich estimator in which the sequence of n observations is split into K approximately independent blocks. The log-likelihood is divided accordingly as

$$\ell(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y}) = \sum_{k=1}^{K} h_k(\boldsymbol{\theta}; \mathbf{x}_k, \mathbf{y}_k), \tag{3.7}$$

where $h_k(\boldsymbol{\theta}; \mathbf{x}_k, \mathbf{y}_k)$, $k = 1, \ldots, K$, is the contribution to the log-likelihood from the kth block. The score vector can be expressed as

$$\nabla \ell(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y}) = \sum_{k=1}^{K} \nabla h_k(\boldsymbol{\theta}; \mathbf{x}_k, \mathbf{y}_k), \tag{3.8}$$

and score covariance estimator

$$\mathsf{J}_{\hat{\boldsymbol{\theta}},Block} = \sum_{k=1}^{K} \nabla h_k(\hat{\boldsymbol{\theta}}; \mathbf{x}_k, \mathbf{y}_k) \nabla h_k(\hat{\boldsymbol{\theta}}; \mathbf{x}_k, \mathbf{y}_k)^{\mathsf{T}}. \tag{3.9}$$

The blocked sandwich estimator is formed by replacing $\mathsf{J}_{\hat{\boldsymbol{\theta}}}$ with $\mathsf{J}_{\hat{\boldsymbol{\theta}},Block}$ in (3.6). This estimator was developed for spatial data with dependence in space but near independence in time, but [11] showed that when modelling first order Markov chains with GP likelihoods, the blocked sandwich estimator provided a notable improvement compared to non-adjusted standard errors, even when the blocks were dependent.

3.4. Block bootstrap estimators

The bootstrap of [10] provides a method to approximate the sampling distribution of a statistic through repeated random sampling of individual observations with replacement. In its original form, however, the method relies on the assumption that data are i.i.d. and fails to reproduce dependence in temporally dependent sequences [33]. Block bootstrap methods, on the other hand, sample blocks of length $l \in \{1, 2, \ldots, n-1\}$ of consecutive observations which retain some temporal dependence in data and, provided that

$$l \to \infty \quad \text{and} \quad n^{-1}l \to 0 \quad \text{as} \quad n \to \infty, \tag{3.10}$$

they produce asymptotically valid approximations of the underlying sequence for weakly dependent sequences. In this study, the moving block bootstrap (MBB) of [19] and [25], and the stationary bootstrap (SB) of [29] are considered, and descriptions of the methods are provided below. For a more comprehensive presentation of block bootstrap see e.g. [21] and [18].

To describe the MBB method, first, let $X_1, \ldots, X_n$ denote observations from a stationary sequence. To generate an MBB sample, $X_1^*, \ldots, X_n^*$, start by determining a block length l and define the set of overlapping blocks $\{\mathrm{B}_1, \ldots, \mathrm{B}_{n-l+1}\}$, where $\mathrm{B}_i = (X_i, \ldots, X_{i+l-1})$. Then draw $b = \min(k \geq 1 : kl \geq n)$ blocks with replacement from $\{\mathrm{B}_1, \ldots, \mathrm{B}_{n-l+1}\}$, and concatenate the blocks. The first n observations form the MBB sample.

The SB method is similar to the MBB, but the block lengths are random, drawn from a Geometric(p) distribution with $p = 1/l \in (0, 1]$ such that $l = 1/p$ is the mean, rather than the actual, block length. To form an SB sample, first draw a block length L_1 from the Geometric(p) distribution. Second, draw a block, B_{L_1}, of L_1 consecutive observations randomly from the original sequence. Repeat the procedure until $L_1 + \cdots + L_b \geq n$ and concatenate the $B_{L_1}, \ldots, B_{L_b}$ blocks. The first n observations in the sequence form the SB sample. As the name suggests, samples generated by the SB are stationary, in contrast to those generated by the MBB. However, in terms of bias and variance, [20] showed that MBB and SB variance estimators share asymptotic bias, but that the asymptotic variance of the SB estimator is considerably larger than that of the MBB.

To compute a block bootstrap standard error associated with an estimate of the dependence parameter θ of one of the considered extreme-value models, first generate a bootstrap sample $\{(X_i^*, Y_i^*)\}_{i=\mathbb{N}}$ with the MBB or SB. Second, fit GEV or GP likelihoods to the marginals, and a bivariate extreme-value likelihood to the joint sample to obtain a bootstrap estimate $\hat{\theta}^*$. Repeat the procedure N times and estimate the standard error as

$$\hat{\sigma}_{\hat{\theta}} = \left(\frac{1}{N-1} \sum_{j=1}^{N} (\hat{\theta}_j^* - \bar{\theta}^*)^2 \right)^{1/2}, \tag{3.11}$$

where $\bar{\theta}^*$ is the average of the N bootstrap estimates $\hat{\theta}_j^*$.

The bias and variance of block bootstrap estimators are largely affected by the choice of block length; longer blocks yield a better approximation of the temporal dependence and thus lower bias, but also fewer blocks to sample from and thereby higher variance [6, p. 397]. An investigation of block length estimators is outside the scope of this study. However, it is investigated through simulation how the choice of block length affects the standard error estimates.

4. SIMULATION STUDY

In this section, the implementation of, and the results from, the simulation study are presented. All numerical experiments are performed in `R` [31].

4.1. Data generating process

The considered data generating process is a bivariate process with first order temporal dependence imposed by a logistic extreme-value copula with parameter $0 < \varphi \leq 1$, and bivariate dependence imposed by one of the extreme-value models in Section 2.3. To achieve this, first, a bivariate sequence $\{(X_i, Y_i)\}_{i \in \mathbb{N}}$ of independent vectors is generated with algorithms from [36] which are implemented in the `R` package `evd` [35]. Second, temporal dependence is imposed on each marginal separately by inverse transform sampling of x_{t+1} from the conditional distribution of x_{t+1} given x_t and φ, i.e.

$$G_{X_{t+1}|X_t=x_t}(x_{t+1} \mid \varphi) = \Pr(X_{t+1} \leq x_{t+1} \mid X_t = x_t, \varphi), \tag{4.1}$$

where G is the logistic extreme-value distribution formed by (2.8). As such, the joint distribution function of two values (x_t, x_{t+1}) from the marginal $\{X_i\}$ is

$$F_{X_t,X_{t+1}}(x_t, x_{t+1}; \varphi) = \exp\left[-\left(x_t^{-1/\varphi} + x_{t+1}^{-1/\varphi}\right)^{\varphi}\right], \tag{4.2}$$

and correspondingly for $\{Y_i\}$. For simplicity, the same temporal dependence strength is used for both marginals. The marginal distributions of $\{(X_i, Y_i)\}$ are then transformed with the probability integral transform to GEV$(\mu = 0, \tau = 1, \xi = 0.1)$ for block-maxima, and GP$(\tau = 1, \xi = 0.1)$ for POT. To investigate if the heaviness of the marginal tails have an effect on the standard error estimator's performance, parts of the simulation were replicated with marginal distributions GEV$(\mu = 0, \tau = 1, \xi = -0.1)$ for block-maxima, and GP$(\tau = 1, \xi = -0.1)$ for POT. Changing the value of ξ from 0.1 to -0.1, however, had a negligible effect on the results and is therefore not further discussed.

To assess the strength of temporal dependence in the generated data, the tail dependence coefficient χ in (3.1) is estimated for $\{X_i\}$ and $\{Y_i\}$ respectively.

For the POT approach, χ is estimated only from the exceedances as those are the only actually informative observations. The empirical estimator

$$\hat{\chi}(u) = \frac{\sum_{i=k+1}^{n} I(Z_i > \hat{F}_Z^{-1}(u)) I(Z_{i-k} > \hat{F}_Z^{-1}(u))}{\sum_{i=k+1}^{n} I(Z_i > \hat{F}_Z^{-1}(u))}, \tag{4.3}$$

is used, where I is the indicator function and $\hat{F}^{-1}$ is the inverse empirical distribution function [3, p. 165]. Estimates are computed with $u = 0.95$ at lags $k = 1, \ldots, 5$, and in Figure 1 the averages over 1000 estimates from the logistic model for different temporal dependence strengths are presented. Corresponding estimates from the asymmetric logistic and Hüsler-Reiss models look similar. When temporal dependence is weak ($\varphi = 0.9$), the estimated asymptotic dependence is close to 0 already at lag 2, while under strong temporal dependence ($\varphi = 0.3$) there is still notable dependence in the tail at lag 5.

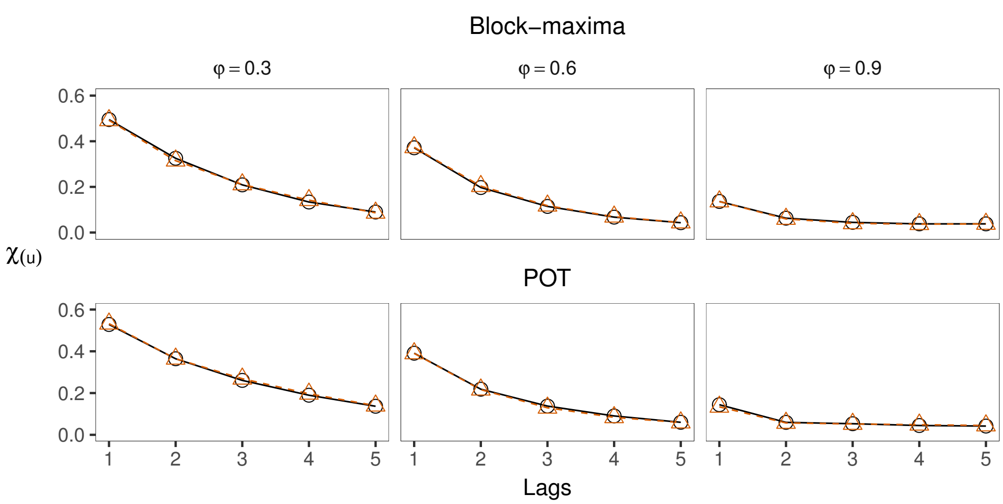

Figure 1: Average estimates of χ under different levels of temporal dependence for the marginals $\{X_i\}$ (solid, circles) and $\{Y_i\}$ (dashed, triangles) at lags 1 to 5, computed from 1000 simulated data sets from the logistic model.

4.2. Estimation

To compute the parameter estimates and associated standard errors, a bivariate extreme-value model with GEV (block-maxima) or GP (POT) marginals is fitted with maximum likelihood estimation to each generated data sequence. The model fitting is done in one step, estimating marginal and dependence parameters simultaneously. Once the model is fitted, standard errors are computed with each method described in Sections 3.3 and 3.4. Furthermore, 95% confidence intervals are computed by using either the normal distribution approximation of the maximum likelihood estimator for the sandwich estimators or basic bootstrap confidence intervals for the block bootstrap estimators [6, p. 194]. For comparison, confidence intervals are also computed with the non-adjusted "naive" standard errors

$$\hat{\sigma}_{Naive} = -\frac{\partial^2 \ell(\theta; x, y)}{\partial \theta^2}\bigg|_{\theta=\hat{\theta}}, \tag{4.4}$$

where $\ell(\theta; x, y)$ is the log-likelihood of the considered model with bivariate dependence parameter θ. As a benchmark, the sample standard deviation of the maximum likelihood estimator (MLE) is used

$$\sigma_R = \left(\frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\theta}_r - \bar{\theta}\right)^2\right)^{1/2}. \tag{4.5}$$

Here $\hat{\theta}_r$ is the parameter estimate from the rth data sequence and $\bar{\theta}$ is the average over R estimates.

4.3. Main simulation: logistic model

This section presents the main simulation study performed with the logistic extreme-value model with bivariate dependence parameter $0 < \alpha \leq 1$. Complementary results from the asymmetric logistic and the Hüsler-Reiss models are presented in Section 4.4.

To investigate how the standard error estimators perform under different dependence strengths, data sets are generated from the bivariate logistic model with dependence parameter $\alpha \in \{0.3, 0.7\}$ (strong, moderate) and temporal dependence parameter $\varphi \in \{0.1, 0.2, \ldots, 0.9\}$ (strong to weak). In each scenario $R = 1000$ independent sequences of length $n = 100$ are generated for block-maxima, and $n = 2000$ for POT. In the POT approach, the thresholds are set to the 0.95 quantiles such that there are 100 exceedances in each marginal. The block length for the blocked sandwich estimator is specified such that the data sequences are divided into 10 equally sized blocks, i.e. of size 10 for block-maxima and 200 for POT. Different block lengths were considered but with small effects on the results and, as such, the chosen lengths are deemed to be adequate.

To assess the effect that the choice of bootstrap block length has on estimation, root mean squared error (RMSE) is computed for the bootstrap estimators for each block length $l \in \{2, 3, \ldots, 12\}$ for block-maxima, and $l \in \{2, 4, \ldots, 24\}$ for POT, with 300 bootstrap replicates. This is done for strong and weak temporal dependence, and the difference in block lengths between block-maxima and POT is due to the different lengths of the underlying sequences (100 and 2000).

The results are presented in Figure 2 together with associated 95% bootstrap percentile confidence intervals. Under weak dependence, the shortest possible block length ($l_{MBB} = l_{SB} = 2$) yields the lowest RMSE for both estimators. This is unsurprising as dependence is weak and close to 0 at lag 2 (see Figure 1). Under strong dependence RMSE decreases with block length and has a minimum at $l_{MBB} = 10$ and $l_{SB} = 9$ for block-maxima, and $l_{MBB} = 24$ and $l_{SB} = 18$ for POT. Since the lowest RMSE for the moving block bootstrap was achieved with $l_{MBB} = 24$, longer block lengths were investigated but with no notable improvements of the results.

In Section 4.3.1, block bootstrap results are computed with the block lengths that yield the lowest RMSE, henceforth referred to as the "minimum RMSE block lengths".

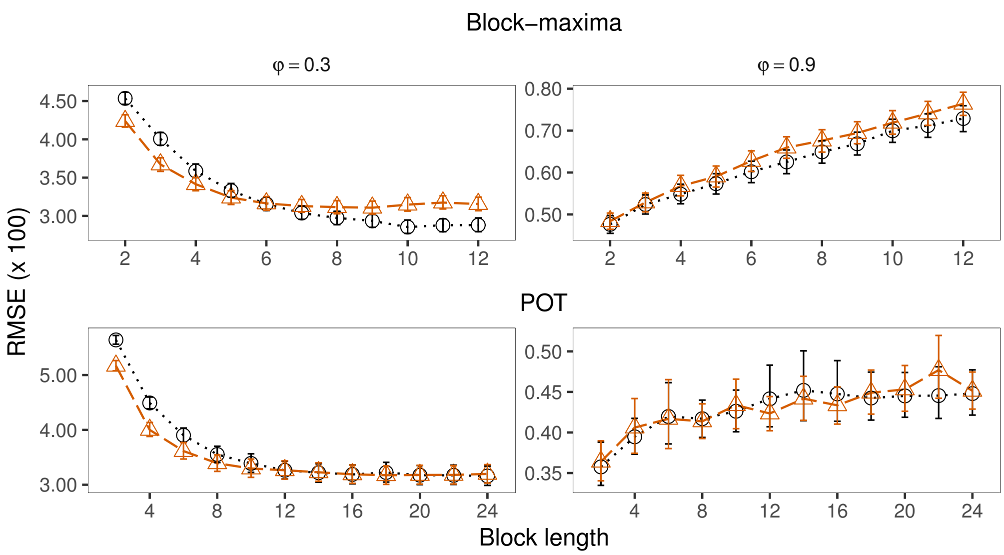

Figure 2: RMSE with 95% bootstrap percentile confidence intervals under strong (left panel) and weak (right panel) dependence for the moving block bootstrap (dotted, circles) and the stationary bootstrap (long-dash, triangles) estimators, computed from 1000 simulated data sets with 300 bootstrap replicates.

4.3.1. Results for the bivariate logistic model

In this section, results from the main simulation with the bivariate logistic model are presented. Only the results from the scenarios with strong bivariate dependence ($\alpha = 0.3$) are shown, as these summarise the general outcomes of the study. In the following, "dependence" refers to temporal dependence unless stated otherwise.

Figure 3 presents coverage probabilities of 95% confidence intervals for the bivariate logistic dependence parameter α, computed from data sets with dependence $\varphi \in \{0.1, 0.2, \ldots, 0.9\}$. Here, the effect of ignoring temporal dependence is clearly seen as the naive estimator has poor coverage probabilities under strong dependence. Noteworthy is that the sandwich and block bootstrap estimators also have coverage probabilities far from 0.95 under strong dependence; although they still provide a clear improvement compared to the naive estimator. As expected, when dependence weakens all estimators approach the nominal coverage probability. The standard sandwich estimator slightly outperforms the other estimators for the POT approach, but overall the differences are quite small.

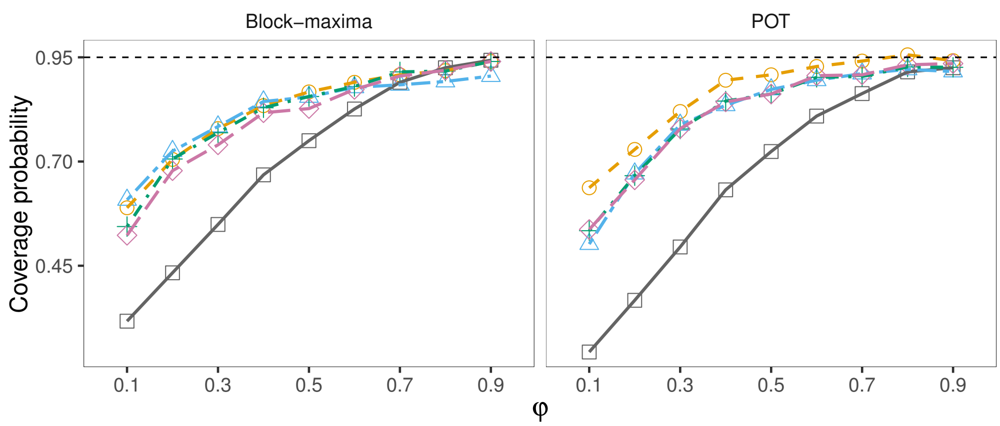

Figure 3: Coverage probabilities of 95% confidence intervals from the bivariate logistic model under strong ($\alpha = 0.3$) bivariate dependence for the naive (solid, squares), standard sandwich (dashed, circles), blocked sandwich (two-dash, triangles), MBB (dash-dotted, plus) and SB (long-dash, diamonds) estimators, computed from 1000 simulated data sets.

In Figure 4, box-plots of $\hat{\sigma}_r/\sigma_R$ are illustrated, where $\hat{\sigma}_r$ is the estimate from one replicate and σ_R is the benchmark in (4.5). This gives a view of the sampling distributions of the estimators in relation to the benchmark σ_R. Under strong dependence, the estimates are overall negatively biased. The block bootstraps perform slightly better than the standard sandwich estimator, while the blocked sandwich estimates are almost as biased as the naive ones for POT. Under weak dependence, the distributions of all estimates are more or less centred around the benchmark with similar variability.

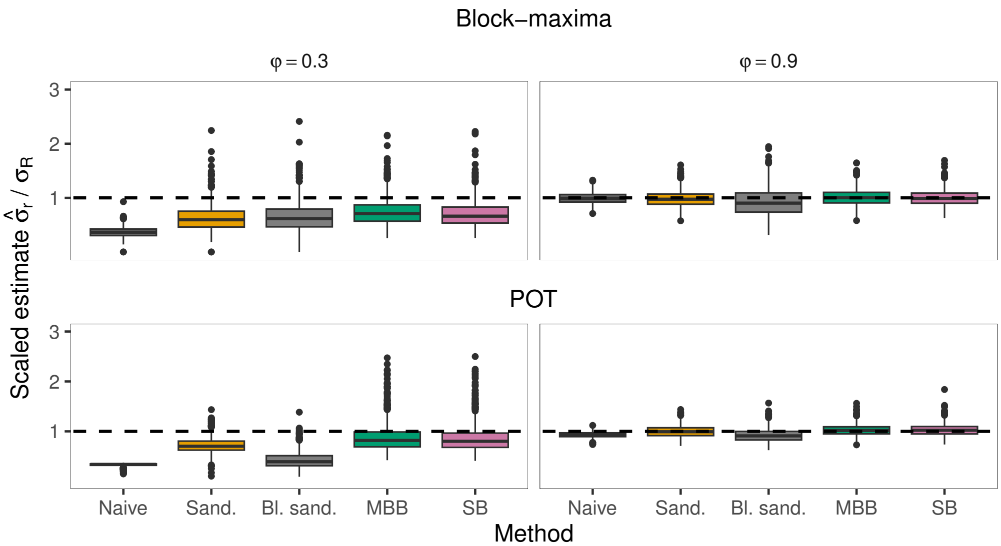

Figure 4: Distributions of $\hat{\sigma}_r/\sigma_R$ for strong (left panel) and weak (right panel) temporal dependence, under strong ($\alpha = 0.3$) bivariate dependence computed from 1000 simulated data sets. Two large estimates of $\sigma_{Sand.}$ were excluded for visibility purposes.

To summarise, when dependence in data is moderate to weak all considered standard error estimators perform similarly and quite well. Under strong dependence, however, all estimators are negatively biased, and confidence intervals computed from the estimates have coverage probabilities far from the nominal coverage probability. It is known that both sandwich and block bootstrap estimators might perform poorly when dependence in data is strong, which is confirmed by the results of this study. Furthermore, the standard sandwich estimator might provide inaccurate results if the data sequence is too short. This is supported by results from complementary numerical experiments (not shown) where substantially increasing the size of data sets clearly improved the coverage probabilities. Hence, altogether the results suggest that if the number of observations is large, the standard sandwich estimator can provide acceptable results under strong dependence, while the considered block bootstrap estimators should only be used when temporal dependence in data is moderate to weak.

4.3.2. Results for the logistic model in higher dimensions

In order to assess how the results from Section 4.3.1 generalize to higher dimensions, numerical experiments are conducted with the logistic model in dimensions $d = 5$ and $d = 10$, in the scenarios with temporal dependence

$\varphi \in \{0.3, 0.6, 0.9\}$ and multivariate dependence $\alpha = 0.3$. The data generating process is the same as described in Section 4.1 although extended to higher dimensions. Furthermore, the marginal distributions are directly transformed to unit Fréchet to avoid the estimation of a large number of marginal parameters. Thus, only the multivariate logistic dependence parameter α is estimated. The multivariate logistic log-likelihood is fitted efficiently using a representation from [32], and only the block-maxima approach is considered as the results from the bivariate logistic model were similar for the block-maxima and the POT approaches.

Figure 5 shows coverage probabilities of 95% confidence intervals for the multivariate logistic dependence parameter α computed from 1000 independent data sets. The relation between the estimators is similar to what was observed in the bivariate case. The coverage probabilities are, however, lower and decrease with increasing dimension, due to increasing bias. Thus, the results suggest that even greater care must be taken when modelling extremes from temporally dependent sequences in higher dimensions.

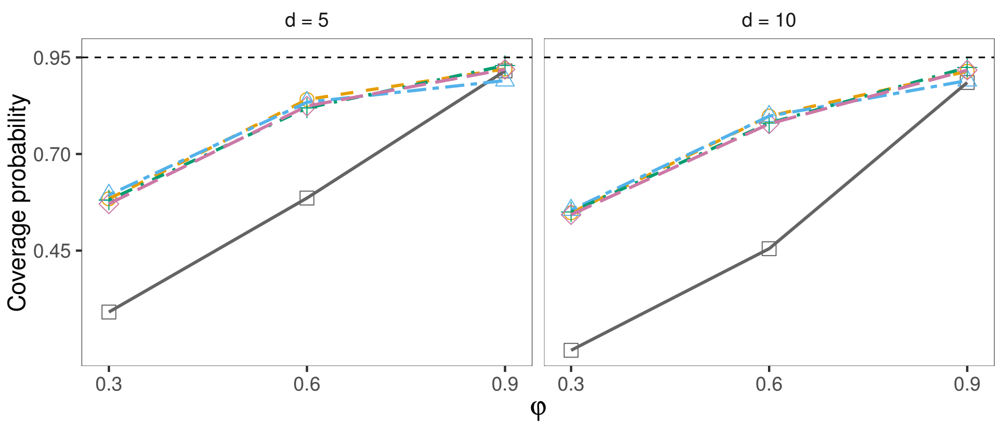

Figure 5: Coverage probabilities of 95% confidence intervals from the multivariate logistic model in dimension $d = 5$ (left) and $d = 10$ (right) under strong ($\alpha = 0.3$) multivariate dependence for the naive (solid, squares), standard sandwich (dashed, circles), blocked sandwich (two-dash, triangles), MBB (dash-dotted, plus) and SB (long-dash, diamonds) estimators, computed from 1000 simulated data sets.

4.4. Results for the bivariate asymmetric logistic and Hüsler-Reiss models

To evaluate if the results from the bivariate logistic model in Section 4.3.1 translates to other models, numerical experiments are conducted with the asymmetric logistic and Hüsler-Reiss models. Data is generated as previously, described in Section 4.1, but the bivariate dependence is imposed with either the asymmetric logistic model in (2.9), or the Hüsler-Reiss model in (2.10). Again, only the block-maxima approach is considered in the scenarios with temporal dependence $\varphi \in \{0.3, 0.6, 0.9\}$.

With the asymmetric logistic model, dependence is governed by the parameter $0 < \beta \leq 1$ and the asymmetry parameters $0 \leq \psi_1, \psi_2 \leq 1$. Numerical experiments are performed with $\beta = 0.3$, $\psi_1 = 0.5$ and $\psi_1 = 0.8$ which corresponds to quite strong bivariate dependence. The Hüsler-Reiss model has a single dependence parameter $r > 0$ which, in the experiments, is set to $r = 1.5$ which corresponds to moderate bivariate dependence. The results are presented in Figure 6, which shows coverage probabilities of 95% confidence intervals for the dependence parameter β of the asymmetric logistic model, and r of the Hüsler-Reiss model, computed from 1000 independent data sets. Corresponding results from the bivariate logistic model with $\alpha = 0.3$ are also shown for comparison.

The relation between the estimator's performance is similar for all models, although the block bootstraps perform somewhat worse than the sandwich estimators for the asymmetric logistic model under strong dependence. It is noteworthy that the coverage probabilities of all estimators for the asymmetric logistic model are far from the nominal coverage probability, even when temporal dependence is weak. A possible explanation is the added complexity from the two additional parameters in the model, which has a total of 9 parameters to be estimated compared to 7 for the bivariate logistic and Hüsler-Reiss models. The estimator's performance for the Hüsler-Reiss model is similar to that of the logistic model and, hence, the conclusions from Section 4.3.1 seem to also hold for the Hüsler-Reiss model.

5. DISCUSSION

The goal of this study was to investigate and compare the performance of four sandwich and block bootstrap standard error estimators when modelling the bivariate dependence between extremes from bivariate stationary sequences. The performance of the estimators was assessed through coverage probabilities of 95% confidence intervals, and by comparing the estimators sampling distributions to

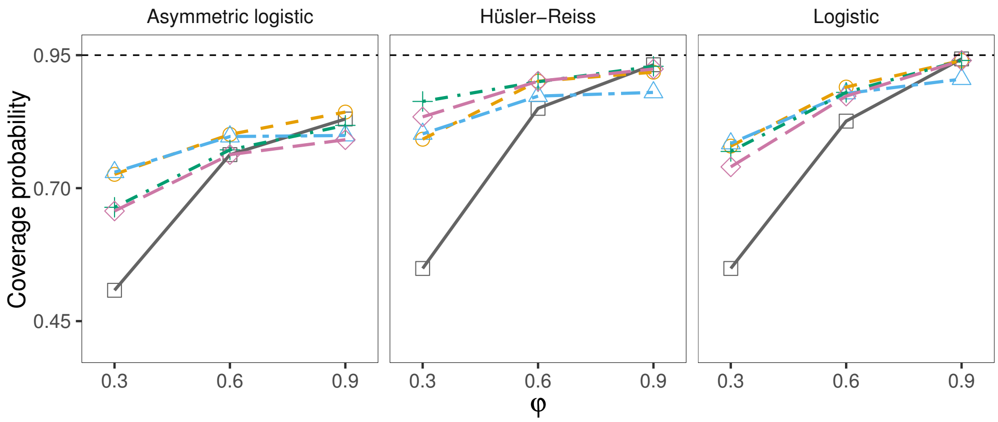

Figure 6: Coverage probabilities of 95% confidence intervals under different levels of temporal dependence for the naive (solid, squares), standard sandwich (dashed, circles), blocked sandwich (two-dash, triangles), MBB (dash-dotted, plus) and SB (long-dash, diamonds) estimators. Values are computed from 1000 simulated data sets from the asymmetric logistic model (left) with parameter $\beta = 0.3$, the Hüsler-Reiss model (middle) with parameter $r = 1.5$, and logistic model (right) with parameter $\alpha = 0.3$.

a benchmark. Focus was on the bivariate logistic extreme-value model, but the results were generalized both to higher dimensions and to additional models.

In the cases with the logistic and Hüsler-Reiss models, when temporal dependence in data is moderate to weak, all considered estimators perform quite well and provide viable alternatives to account for underestimation. With the asymmetric logistic model, however, all estimators performed poorly under both strong and weak dependence, and extra care should be taken if modelling extremes from temporally dependent sequences with this model. Under strong temporal dependence, all estimators have coverage probabilities far from the nominal coverage probability and are notably biased due to an inability to fully capture the strong dependence in data. Furthermore, for the standard sandwich estimator, the considered data sequences seem too short, at least in the block-maxima case, for the estimator to perform well.

Altogether, the results suggest that when encountered with smaller data sets with strong temporal dependence, none of the considered estimators are preferable. Instead, it may be more successful either to model the dependence explicitly or, if feasible, to decrease the dependence by e.g. declustering. There also exist bootstrapping techniques that might provide more accurate standard errors, such as residual bootstraps where a parametric model is first fitted to data and the residuals resampled, or bootstrap methods tailored for Markov processes. An investigation of additional bootstrapping methods is, however, left for future research.

ACKNOWLEDGMENTS

I would like to thank Jennifer Wadsworth for valuable ideas and discussions during this project. I would also like to thank the anonymous referees for valuable suggestions that improved the paper.

REFERENCES

[1] Balkema, A.A. and de Haan, L. (1974). Residual life time at great age, *The Annals of Probability*, **2**, https://doi.org/10.1214/aop/1176996548.

[2] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, UK.

[3] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

[4] Coles, S.; Heffernan, J. and Tawn, J. (1999). Dependence measures for extreme value analyses, *Extremes*, **2**, 339–365, https://doi.org/10.1023/A:1009963131610.

[5] Davison, A.C. (2003). *Statistical Models*, Cambridge University Press, New York.

[6] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*, Cambridge University Press, New York.

[7] Davison, A.C. and Huser, R. (2015). Statistics of extremes, *Annual Review of Statistics and Its Application*, **2**, 1, 203–235, https://doi.org/10.1146/annurev-statistics-010814-020133.

[8] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society: Series B (Methodological)*, **52**, 3, 393–442, https://doi.org/10.1111/j.2517-6161.1990.tb01796.x.

[9] de Haan, L. and Resnick, S.I. (1977). Limit theory for multivariate sample extremes, *Zeitschrift für Wahrscheinlichkeitstheorie Und Verwandte Gebiete*, **40**, 317–337, https://doi.org/10.1007/BF00533086.

[10] Efron, B. (1979). Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, **7**, 1, 1–26, https://doi.org/10.1214/aos/1176344552.

[11] Fawcett, L. and Walshaw, D. (2007). Improved estimation for temporally clustered extremes, *Environmetrics*, **18**, 173–188, https://doi.org/10.1002/env.810.

[12] Feigin, P.D. and Resnick, S.I. (1999). Pitfalls of fitting autoregressive models for heavy-tailed time series, *Extremes*, **1**, 391–422, https://doi.org/10.1023/A:1009946321478.

[13] Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation, *The Annals of Mathematical Statistics*, **31**, 4, 1208–1211.

[14] Gumbel, E.J. (1961). Bivariate logistic distributions, *Journal of the American Statistical Association*, **56**, 294, 335–349, https://doi.org/10.2307/2282259.

[15] Huser, R. and Wadsworth, J.L. (2019). Modeling spatial processes with unknown extremal dependence class, *Journal of the American Statistical Association*, **114**, 525, 434–444, https://doi.org/10.1080/01621459.2017.1411813.

[16] Hüsler, J. and Reiss, R.D. (1989). Maxima of normal random vectors: Between independence and complete dependence, *Statistics & Probability Letters*, **7**, 4, 283-286, https://doi.org/10.1016/0167-7152(89)90106-5.

[17] Keef, C.; Svensson, C. and Tawn, J.A. (2009). Spatial dependence in extreme river flows and precipitation for Great Britain, *Journal of Hydrology*, **378**, 3-4, 240–252, https://doi.org/10.1016/j.jhydrol.2009.09.026.

[18] Kreiss, J.P. and Lahiri, S.N. (2012). *Bootstrap methods for time series*. In "Handbook of Statistics" (T.S. Rao, S.S. Rao and C.R. Rao, Eds.), Volume 30, Elsevier, 3–26.

[19] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations, *The Annals of Statistics*, **17**, 3, 1217–1241, https://doi.org/10.1214/aos/1176347265.

[20] Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods, *The Annals of Statistics*, **27**, 1, 386–404, https://doi.org/10.1214/aos/1018031117.

[21] Lahiri, S.N. (2003). *Resampling Methods for Dependent Data*, Springer, New York.

[22] Leadbetter, M.R. (1974). On extreme values in stationary sequences, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **28**, 289–303.

[23] Ledford, A.W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values, *Biometrika*, **83**, 1, 169–187, https://doi.org/10.1093/biomet/83.1.169.

[24] Ledford, A.W. and Tawn, J.A. (1997). Modelling dependence within joint tail regions, *Journal of the Royal Statistical Society: Series B (Methodological)*, **59**, 2, 475–499, https://doi.org/10.1111/1467-9868.00080.

[25] Liu, R.Y. and Singh, K. (1992). *Moving blocks jackknife and bootstrap capture weak dependence*. In "Exploring the Limits of the Bootstrap" (R. LePage and L. Billard, Eds.), Wiley, 225–248.

[26] Newey, W.K. and West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 3, 703–708.

[27] Northrop, P.J. (2015). An efficient semiparametric maxima estimator of the extremal index, *Extremes*, **18**, 585–603, https://doi.org/10.1007/s10687-015-0221-5.

[28] Pickands, J. (1975). Statistical inference using extreme order statistics, *The Annals of Statistics*, **3**, 1, 119–131, https://doi.org/10.1214/aos/1176343003.

[29] Politis, D.N. and Romano, J.P. (1994). The stationary bootstrap, *Journal of the American Statistical Association*, **89**, 428, 1303–1313, https://doi.org/10.1080/01621459.1994.10476870.

[30] Prescott, P. and Walden, A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution, *Biometrika*, **67**, 3, 723–724, `https://doi.org/10.2307/2335152`.

[31] R Core Team (2023). R: A language and environment for statistical computing, *R Foundation For Statistical Computing*, Vienna, Austria, `https://www.R-project.org/`.

[32] Shi, D. (1995). Multivariate extreme value distribution and its Fisher information matrix, *Acta Mathematicae Applicatae Sinica*, **11**, 4, 421–428, `https://doi.org/10.1007/BF02007180`.

[33] Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap, *The Annals of Statistics*, **9**, 6, 1187–1195, `https://doi.org/10.1214/aos/1176345636`.

[34] Smith, R.L. (1990). Regional estimation from spatially dependent data, *Preprint*.

[35] Stephenson, A. (2002). Evd: Extreme value distributions, *The R Journal*, **2**, 2, 31-32.

[36] Stephenson, A. (2003). Simulating multivariate extreme value distributions of logistic type, *Extremes*, **6**, 49–59, `https://doi.org/10.1023/A:1026277229992`.

[37] Tawn, J.A. (1988). Bivariate extreme value theory: Models and estimation, *Biometrika*, **75**, 3, 397-415, `https://doi.org/10.1093/biomet/75.3.397`.

[38] White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica*, **50**, 1, 1–25, `https://doi.org/10.2307/1912526`.

[39] Zheng, F.; Westra, S. and Sisson, S.A. (2013). Quantifying the dependence between extreme rainfall and storm surge in the coastal zone, *Journal of Hydrology*, **505**, 15, 172–187, `https://doi.org/10.1016/j.jhydrol.2013.09.054`.

[40] Zhou, C. (2010). Dependence structure of risk factors and diversification effects, *Insurance: Mathematics and Economics*, **46**, 3, 531–540, `https://doi.org/10.1016/j.insmatheco.2010.01.010`.

[41] Zscheischler, J. and Seneviratne, S.I. (2017). Dependence of drivers affects risks associated with compound events, *Science Advances*, **3**, 6, e1700263, `https://doi.org/10.1126/sciadv.1700263`.