U-Statistics of Extremes

Torben Staud, Axel Bücher

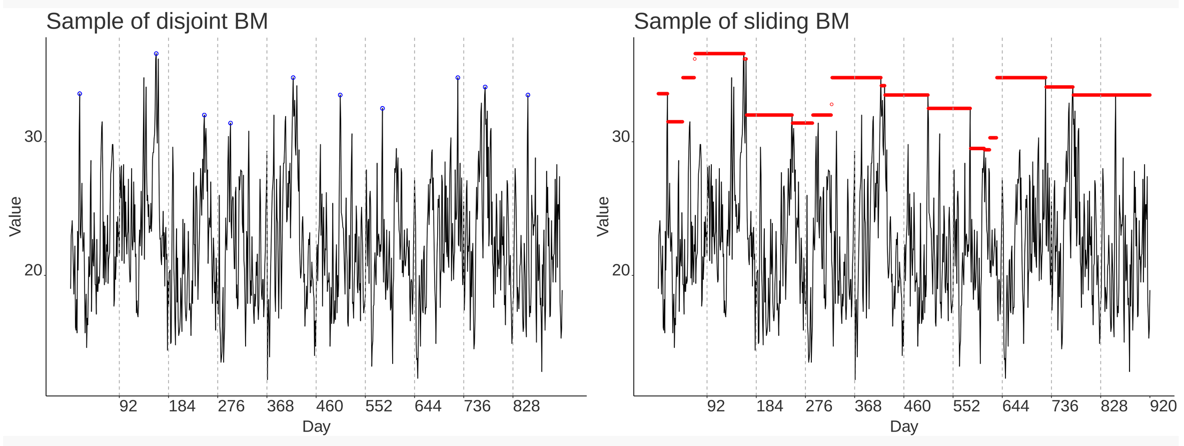
Heinrich-Heine-University Düsseldorf

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Motivation

Block maxima method:

- A classical method in extremes.
- Recently, new insights were gained by regarding the block size parameter as a sequence converging to infinity [3].
- Certain estimators based on the sliding block maxima method were found to provide lower estimation variances [1,2].



U-statistics:

- Constitute a broad class of statistics that are known to satisfy certain optimality conditions in classical situations.
- Some common estimators like the PWM-estimator for block maxima may be identified as a U-Statistic.

Aims

- Derive asymptotic theory for **U-Statistics applied to (multivariate) block maxima**.
- Compare disjoint and sliding blocks estimators theoretically and by means of Monte Carlo simulation studies.

Model

- **Observations:** Excerpt of a strictly stationary d-variate time series $\mathbf{X}_1, \dots, \mathbf{X}_n$.
- **Block maxima:** $\mathbf{M}_{r,i}$ denotes the componentwise block maxima of block length $r = r_n \to \infty$ starting at time i.
- Multivariate domain of attraction condition:
$$\left(\frac{M_{r,1}^{(j)} - b_r^{(j)}}{a_r^{(j)}} \right)_{j=1,\dots,d} \xrightarrow{d} \mathbf{Z},$$
as $r \to \infty$, for certain $(\mathbf{b}_r)_r, (\mathbf{a}_r)_r$.
- **Block maxima samples:** Obtain the disjoint and sliding sample $\mathcal{M}^{(\mathrm{db})} = (\mathbf{M}_{r,1+ri})_{i=1,\dots,m}$, $\mathcal{M}^{(\mathrm{sb})} = (\mathbf{M}_{r,i})_{i=1,\dots,n-r+1}$.

Objects of interest

- **Kernel function** $h\colon \mathbb{R}^{2d} \to \mathbb{R}$ with enough regularity including the existence of functions f, ℓ such that, for all $\mathbf{x}, \mathbf{y}, \mathbf{b} \in \mathbb{R}^d, \mathbf{a} \in (0,\infty)^d$,
$$h\left(\frac{\mathbf{x}-\mathbf{b}}{\mathbf{a}}, \frac{\mathbf{y}-\mathbf{b}}{\mathbf{a}} \right) = \frac{h(\mathbf{x},\mathbf{y})}{f(\mathbf{a},\mathbf{b})} + \ell(\mathbf{a},\mathbf{b}).$$
- **U-statistic induced by** h:
$$U_{n,r}^{(\mathrm{mb})} := \binom{n_{\mathrm{mb}}}{2}^{-1} \sum_{1 \le i < j \le n_{\mathrm{mb}}} h(\mathbf{M}_{r,i}, \mathbf{M}_{r,j}),$$
for $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$ and n_{mb} the respective block maxima sample size.
- **Target parameter:** $U_{n,r}^{(\mathrm{mb})}$ is an estimator for
$$\theta_r := \int h(\mathbf{x},\mathbf{y}) \, \mathrm{d}\mathbb{P}^{\otimes 2}_{\mathbf{M}_{r,1}}(\mathbf{x},\mathbf{y}),$$
but under dependence generally not unbiased.

Theorem (Bücher, S., 2023)

Under certain regularity conditions it holds, that

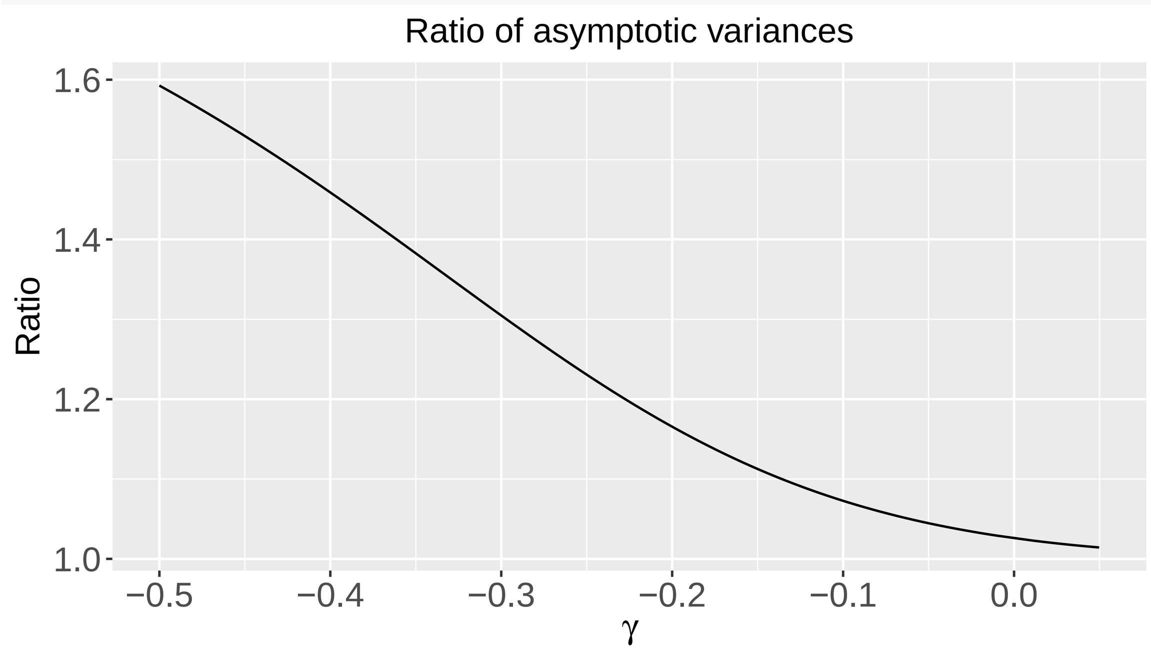
$$\frac{\sqrt{m}}{f(\mathbf{a}_r, \mathbf{b}_r)} (U_{n,r}^{(\mathrm{mb})} - \theta_r) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{(\mathrm{mb})}),$$

where $\sigma^2_{(\mathrm{sb})} \le \sigma^2_{(\mathrm{db})}$.

- Disjoint variance $\sigma^2_{(\mathrm{db})}$ is described by a term from asymptotic U-statistic theory and the EVD $\mathbf{Z}$.
- The sliding variance has a component quantifying asymptotic overlap $\mathbf{Z}_\xi$ and a U-statistic component, where $\mathbf{Z}_\xi$ is a $2d$-variate EVD and its distribution can be described via $\mathbf{P}_{\mathbf{Z}}$. This generalizes a well known univariate overlap copula from [1].

Examples

- **Variance estimation**: the empirical variance is a U-statistic with kernel $h_{\mathrm{Var}}(x,y) = (x-y)^2/2$



- **Estimation of Kendall's** τ: for bivariate block maxima we have, using the concordance kernel $h_\tau(\mathbf{x},\mathbf{y}) = \mathbf{1}\big((x^{(1)} - y^{(1)})(x^{(2)} - y^{(2)}) > 0\big)$,
$$\sqrt{m}\big(\hat{\tau}_{n,r}^{(\mathrm{mb})} - \tau_r\big) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{(\mathrm{mb})}).$$
- **Other estimators:** in the literature similar effects were observed, e.g., for the PWM estimator [2].

Simulation studies

Variance estimation:

- The sliding blocks estimator performs significantly better than its disjoint blocks version.
- Both for serial independence and strong serial dependence, the estimators show qualitatively similar behaviour.

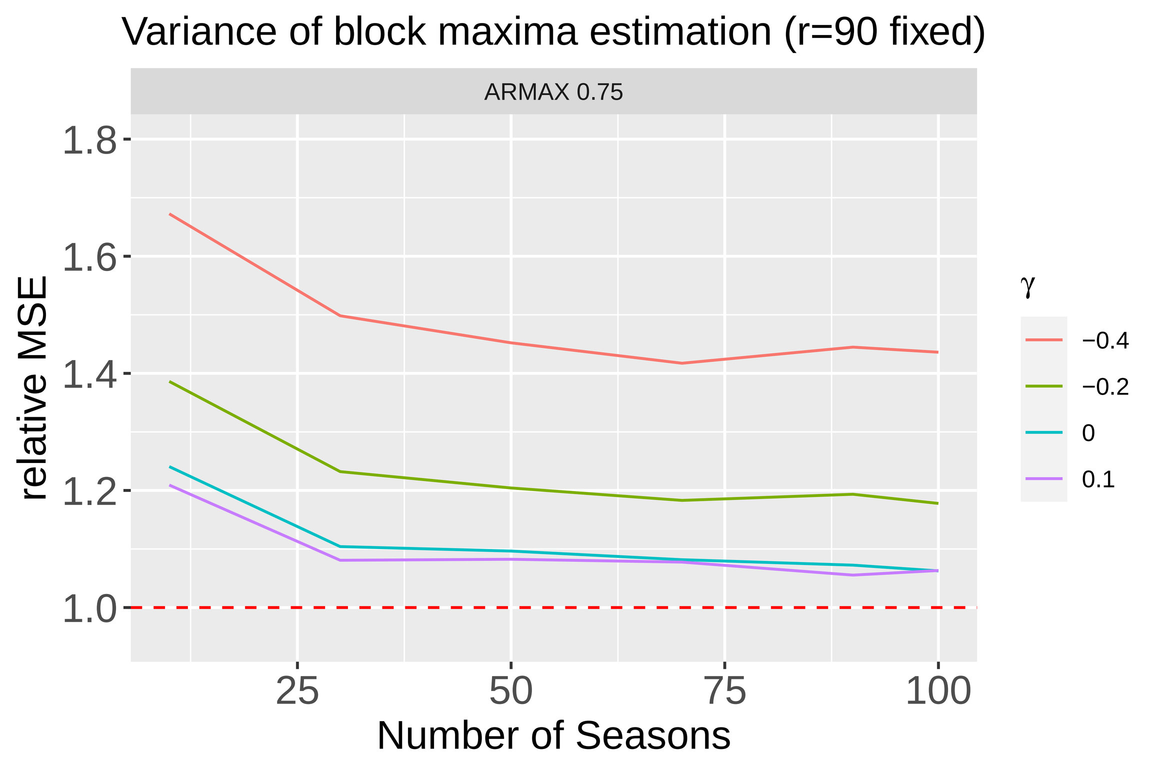


Figure: Relative MSEs for the estimation of σ_r^2 in an (transformed) ARMAX(1) model with parameter 0.75 plotted against the number of seasons m for the disjoint and sliding blocks estimator $\hat{\sigma}^2_{n,r,(\mathrm{db})}, \hat{\sigma}^2_{n,r,(\mathrm{sb})}$ and different shape parameters γ, respectively.

Estimation of Kendall's τ:

- Also better performance when using sliding blocks.
- The stronger the dependence between the margins, the smaller the improvement.

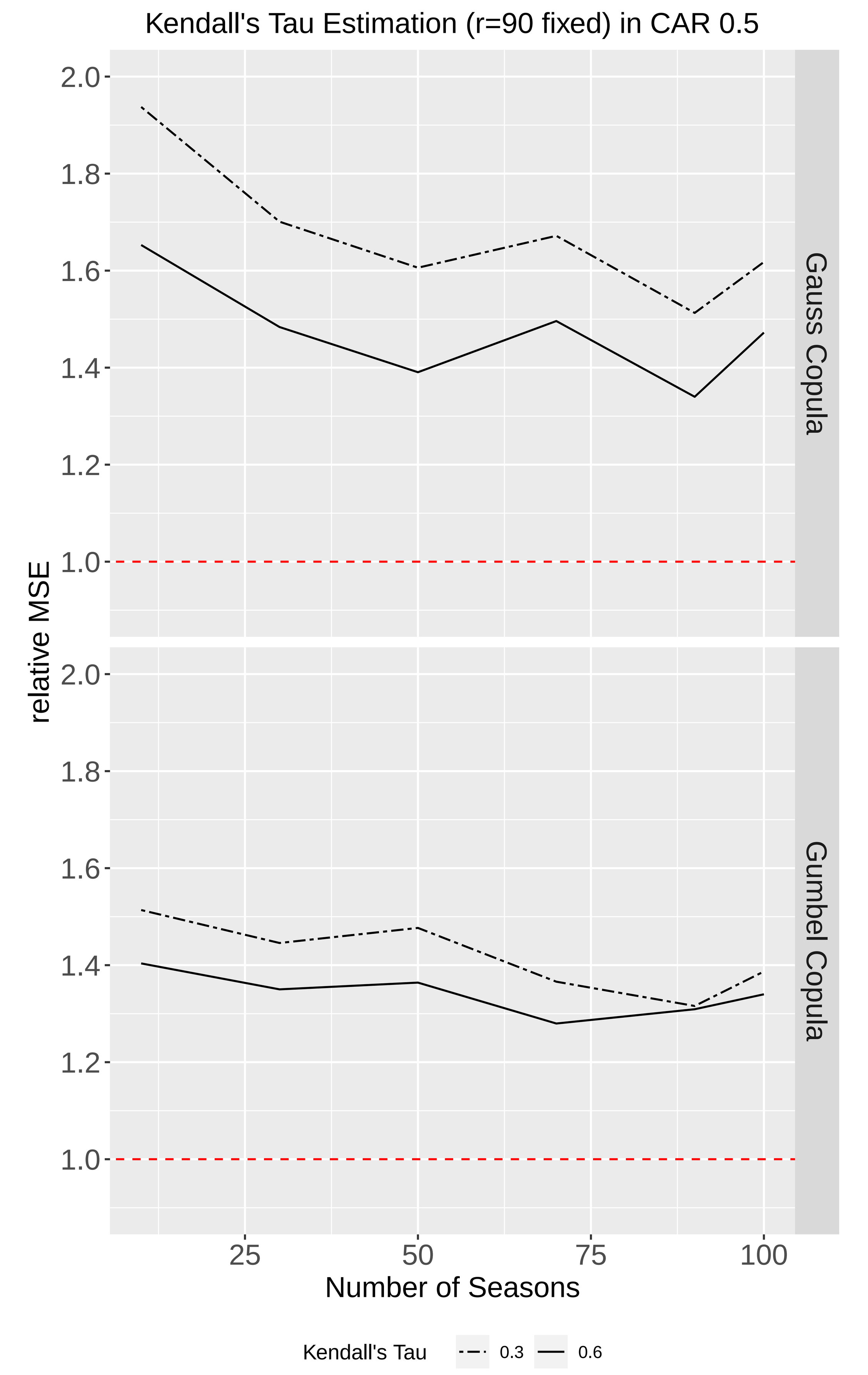


Figure: Relative MSEs of disjoint and sliding blocks estimators plotted against the number of seasons m.

Extensions

- The results hold under the assumption of β-mixing with rate, but generalizations regarding α-mixing are possible.
- The results may also be extended to the framework of *piecewise stationary* time series which was recently embedded into extreme value statistics and potentially offers more realistic seasonal modelling [2].

Conclusion

For the broad class of U-statistics of block maxima, statistics based on sliding blocks exhibit better asymptotic properties than their disjoint counterparts. The superiority is visible in finite sample situtations and especially for small sample sizes.

References

[1] Bücher, A. and Segers, J. (2018). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Stat.*, 12, No. 1, 1098–1125

[2] Bücher, A. und Zanger, L. (2021). On the Disjoint and Sliding Block Maxima method for piecewise stationary time series. *Ann. Stat.*, 51 No. 2, 573 - 598

[3] Ferreira, Ana and de Haan, Laurens (215). On the block maxima method in extreme value theory: PWM estimators. *Ann. Stat.*, 43, 276–298.

[4] Gumbel, E. J. (1958). Statistics of extremes. *New: Columbia University Press.* XX, 375 p.