

ESTIMATION APPROACHES FOR A TIME-VARYING EXTREMAL INDEX

THESIS MSc ECONOMETRICS AND DATA SCIENCE

ARNE PLATTEAU

SUPERVISOR: JUAN JUAN CAI, SECOND READER: FRANCISCO BLASQUES

Vrije Universiteit Amsterdam, the Netherlands

In many time series, extreme values tend to be clustered, which can have large societal impact. One measure for the extent of extremal clustering is the extremal index θ . Most existing approaches for the estimation of θ assume a stable extremal index. However, in many applications, the extremal processes may change over time and assuming a stable θ may not yield accurate results. This paper develops two new approaches to estimate a time-varying extremal index: a non-parametric estimator, and a parametric filtering approach. We discuss a number of properties of these estimators, such as uniform invertibility and consistency. Although a parametric filtering approach may suffer from model misspecification, our simulation study shows that wrongly assuming a stable extremal index can lead to a worse performance. We apply the methods to estimate the expected duration of extreme temperature conditions over time in Milan, Italy. Both time-varying methods yield changes over time in the extremal index of extreme temperature conditions.

Keywords: extremal index, time-varying parameter model, clustering, extreme temperature duration

1 || INTRODUCTION

In many time series, the occurrence of extreme observations tends to be clustered, i.e. several extreme observations can happen in a short amount of time. This is of importance, as the occurrence of clustered extreme observations in a short time can have much larger consequences than extreme observations occurring in a standalone manner. For example, a period of very warm days can lead to large droughts, increases in mortality, and others, whereas several days of extreme precipitation can lead to floods, agricultural damage, infrastructural damage. Often, the risk or impact increases as the length of the period in which successive extreme observations take place increases.

Extreme events are values that cross a high threshold in the (empirical) cumulative distribution function and thus by definition seldom occur. Should the sequence $\{X_t\}$ be independently and identically distributed (iid), the extreme events should occur according to a homogeneous Poisson process (Davis et al., 2013). However, in many time series, these extreme events are often surrounded by other extreme events, and hence form a cluster of threshold exceedances (Davis et al., 2013). These clusters are independent, and the positions of the clusters can be described by a Poisson process (Moloney et al., 2019). As such, the occurrence of exceedances itself can be described by a compound Poisson process (Moloney et al., 2019).

One measure for the extremal dependency is the extremal index θ . The extremal index can be defined as follows:

DEFINITION (Extremal index (Hsing, 1993)). *Let $\{X_i\}$ be a strictly stationary sequence of continuous random variables with a continuous marginal distribution F . We define $\bar{F} = 1 - F$ and $x_0 = \sup_{x \in X}(x : F(x) < 1)$. We also define $M_{i,j} = \max_{i \leq t \leq j} X_t$ and $M_n = M_{1,n}$. The extremal index of $\{X_i\}$ is equal to θ if :*

$$\lim_{n \rightarrow \infty} \mathbb{P}(n\bar{F}(M_n) \geq x) = e^{-\theta x} \quad (1.1)$$

θ is a real number $\in (0, 1]$. Its inverse $\frac{1}{\theta}$ is the expected duration of successive extreme events, conditioned on the occurrence of an extreme observation (Moloney et al., 2019). Its inverse can thus be considered as the mean cluster size (Moloney et al., 2019). If $\theta = 1$, the extreme observations are independent, and no extremal clustering takes place, The closer θ gets to 0, the stronger the extremal dependency is.

Several estimators have been proposed for the extremal index. Cai (2019) investigates a non-parametric estimator for the extremal index which depends on a mixing condition. This estimator has the advantage of being asymptotically normal under mild conditions. Another estimator is proposed by Berghaus and Bücher (2018), which is based on (pseudo)-observations with an approximate exponential distribution. This estimator is also normal under certain conditions. Ferro and Segers (2003) developed an estimator based on interexceedance times. The block estimator defined in Smith and Weissman (1994) divides the sample in k_n blocks of length r_n , and considers the ratio of the number of extreme events and how many blocks contain an extreme value. The runs estimator defined in Smith and Weissman (1994) considers the runs of observations above or below the threshold defining the clusters.

Both the non-parametric estimator of Cai (2019) and the pseudo maximum-likelihood estimator of Berghaus and Bücher (2018) assume a stable tail process. However, this assumption could be invalid for longer time periods. For example, weather phenomena can be impacted by climate change, changing not only means and variances, but also extremal distributions. Several authors have therefore developed methods to accommodate the time-varying nature of extreme events. Hall and Tajvidi (2000) developed a non-parametric method to make time-varying extreme distributions and applied it to data on windstorm severity and on maximum temperatures. This enabled them to detect trends in extreme values. Another approach is taken by Huerta and Sansó (2007), who estimate a GEV-distribution with time-varying parameters based on a state space approach.

If the extreme value distribution may change over time, the extremal dependency and hence extremal index may do so as well. Some methods allow for a time-varying extremal index, such as the local-likelihood estimator of Süveges (2007). However, most methods, such as the non-parametric estimator of Cai (2019) or the pseudo-maximum-likelihood estimator of Berghaus and Bücher (2018), the interval estimator of Ferro and Segers (2003) and the block and runs estimator of Smith and Weissman (1994) all assume a constant extremal index.

This paper therefore addresses the following research question: Can time-varying extremal index estimation methods improve the characterisation of extremal clustering? To answer this

question, this main problem statement is divided into three subquestions.

The first subquestion addresses how a time-varying extremal index can be estimated. As described before, most existing estimators for the extremal index assume a stable extremal index. We specify a rolling horizon estimator, which is based on the non-parametric estimator introduced in Cai (2019), and parametric modelling approach, which focuses observation-driven filters for autoregressive processes. These filters can capture time-varying autoregressive behaviour and therefore don't need the assumption of a stable extremal index. Moreover, all parameters are estimated by maximum likelihood, so it is not necessary to finetune (hyper)parameters. Further, one is typically also interested in a number of properties of an estimator, such as consistency, and how to specify hyperparameters. We also address these issues.

The second subquestion addresses under which conditions a time-varying extremal index provides a better description of the extremal dependency in a process compared to a stable extremal index. After all, if a method assuming a stable extremal index yields better results than its time-varying counterparts, even for processes with a time-varying extremal index, it doesn't make sense to increase the complexity caused by using time-varying methods. Both time-varying extremal index methods developed in this paper have drawbacks: the non-parametric rolling horizon estimator needs to balance a sufficient long time horizon for accurate estimation with a time horizon which is not too long, as to capture the time-varying behaviour. The parametric filters carry a risk of model misspecification. Therefore, we set up a simulation study, and investigate how stable and time-varying estimators perform in terms of mean squared error (MSE) on processes with stable and time-varying extremal index, and on the other hand study the impact of model misspecification on performance in terms of MSE.

The third subquestion consists of analysing whether time-varying behaviour in the extremal index can be detected in real data. To this end, we apply the estimators in an empirical application, which estimates the extremal index for the detrended daily maximum temperatures in Milan, Italy. For this application, we use a dataset spanning 237 years. As many climatological processes are subject to complex temporal changes, this dataset presents an interesting case study about the potential time-varying behaviour of extremal index.

We find that time-varying methods for the estimation of the extremal index can lead to a more accurate characterisation of extremal dependency. By providing answers to the subquestions discussed earlier, we bring a number of contributions to the research on extremal dependency. First, we propose a time-varying version of the estimator of Cai (2019), and provide guidance on how to set the additional parameter. We develop efficient algorithms for the computation of both the stable non-parametric estimator and the rolling horizon estimator.

Second, we discuss a filter-based approach for extremal index estimation. We propose two observation-driven filters for the autoregressive coefficient of an AR(1) process with Student-t distributed errors. We discuss these filters in some detail, in particular regarding uniform invertibility and consistency. We present evidence from simulations about the performance of these time-varying approaches. The parametric filters has the lowest MSE of all methods in case of correct model specification. Its performance deteriorates when the model specification is not correct, although in different degrees depending on the extent of resemblance between the data generating process and the assumed process. Moreover, wrongly assuming a stable

extremal index can yield a higher MSE than a parametric filter with model misspecification. Of all estimators considered, the rolling horizon estimator is the most robust, as it doesn't need a parametric specification, but can at the same time allow for time-varying behaviour.

All time-varying approaches show that the extremal index of the difference between actual and expected daily maximum temperatures can vary over time. Even for long time horizons, time-varying behaviour can be detected, meaning that the expected length of days with a very high maximum temperature also varies over time.

The remainder of this paper is organised as follows. Chapter 2 discusses the estimators used. To investigate the performance of the respective estimators for different data generating processes, Chapter 3 discusses a simulation study, in which the extremal index is estimated by the different methods for sequences with a known extremal index. Chapter 4 presents an empirical application: the estimation of the extremal index of the difference between the realised daily maximum temperatures and expected daily maximum temperatures given the time of year and longer term trends, for the daily maximum temperatures measured at the Brera astronomical observatory in Milan. Chapter 5 concludes and discusses limitations and suggestions for further study.

2 || ESTIMATORS FOR θ

Several estimators for the extremal index have been proposed, which are based on different assumptions. Given the focus on changes in extremal clustering over time, two contrasting sets of estimators are considered: estimators assuming a constant extremal index over time, and estimators which allow for a time-varying extremal index. This Chapter describes the estimators used in this study, as well as their properties, and addresses a number of parameter choices. We start with the constant estimators for the extremal index, which are the non-parametric extremal index estimator introduced in Cai (2019), and the sliding block estimators described in Berghaus and Bücher (2018). We describe these estimators and propose an efficient implementation of the non-parametric estimator described in Cai (2019). In the second section of this, we develop two estimation approaches for a time-varying extremal index: a rolling horizon adaptation of the non-parametric estimator described in Cai (2019), and a parametric filtering approach. We also give some insight in choosing an appropriate horizon length and proposes an efficient implementation of this rolling horizon estimator. We discuss the uniform invertibility of the parametric filters, the estimation by maximum likelihood, and consistency. The Chapter ends with a discussion of the different estimators.

2.1 || STABLE EXTREMAL INDEX ESTIMATORS

Most estimators of the extremal index assume that the extremal index remains stable over time. Several estimators have been proposed, and this study considers two approaches: the non-parametric estimator proposed by Cai (2019), and the sliding block estimators proposed by Berghaus and Bücher (2018).

2.1.1 || Non-parametric estimator of Cai (2019)

Cai (2019) proposes a non-parametric estimator for the extremal index. The estimator is based on Corollary 1.3 of Chernick et al. (1991), which states the existence of the extremal index subject to the $D^{(d)}(u_n)$ and the $D(u_n)$ condition. To state these conditions, we define as described in Cai (2019): $u_n(\tau), \tau > 0$ such that

$$\lim_{n \rightarrow \infty} n\mathbb{P}(X_1 > u_n(\tau)) = \tau \quad (2.1)$$

As in Cai (2019), we write u_n instead of $u_n(\tau)$ when this doesn't cause confusion.

The $D(u_n)$ condition is described in Cai (2019) as well: for any integers $1 \leq i_1 < \dots < i_q < j_1 < \dots < j_{q'} \leq n$ for which $j_1 - i_q \geq l$:

$$|\mathbb{P}(\max_{1 \leq t \leq q} X_{i_t} \leq u_n, \max_{1 \leq t \leq q'} X_{j_t} \leq u_n) - \mathbb{P}(\max_{1 \leq t \leq q} \leq u_n)\mathbb{P}(\max_{1 \leq t \leq q'} X_{j_t} \leq u_n)| \leq \alpha_{n,l} \quad (2.2)$$

where $\lim_{n \rightarrow \infty} \alpha_{n,l_n} \rightarrow 0$ for some sequence $l_n = o(n)$ and $l_n \rightarrow \infty$ (Cai, 2019).

The $D^{(d)}(u_n)$ condition is: there exist a positive integer d , sequences of integers r_n and l_n such that $r_n \rightarrow \infty$ $n\alpha_{n,l_n}/r_n \rightarrow 0$, $l_n/r_n \rightarrow 0$, and

$$\lim_{n \rightarrow \infty} n\mathbb{P}(X_1 > u_n \geq M_{2,d}, M_{d+1,r_n} > u_n) \rightarrow 0 \quad (2.3)$$

where $M_{i,j} = -\infty$ if $i > j$ and $M_{i,j} = \max_{i \leq t \leq j} X_t$ for $i \leq j$ (Cai, 2019).

PROPOSITION 1 (Chernick et al., 1991). *Let $\{X_n\}$ be a stationary sequence of random variables such that for some $d \leq 1$ the conditions $D(u_n)$ and $D^{(d)}(u_n)$ hold for $u_n = u_n(\tau)$ for all $\tau > 0$. Then θ , the extremal index of $\{X_n\}$, exists if and only if*

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_{2,d} \leq u_n | X_1 > u_n) = \theta \quad (2.4)$$

for all $\tau > 0$.

We define the threshold $X_{n-k,n}$ where $\{X_{1,n} \leq X_{2,n}, \dots, X_{n,n}\}$ are the order statistics of the sample, k an intermediate sequence where $k \rightarrow \infty$ as $n \rightarrow \infty$, and F the empirical cumulative distribution function. The estimator considers for each time i whether simultaneously the maximum of observations $i+1, \dots, i+d-1 \leq X_{n-k,n}$ and $X_i > X_{n-k,n}$. This is done assuming d is known. If for each i for which $X_i > X_{n-k,n}$, the maximum of the next d observations $< X_{n-k,n}$, there is no extremal dependence and $\hat{\theta} = 1$. Reverse, the more often the maximum of the next d observations is larger than $X_{n-k,n}$, given $X_i > X_{n-k,n}$, the smaller $\hat{\theta}$.

We then write (Cai, 2019):

$$\theta \approx \frac{n}{k} \mathbb{P}(M_{2,d} \leq F^{-1}(1 - k/n) < X_1) \quad (2.5)$$

The estimator is given by:

$$\hat{\theta}_n(d) = \frac{1}{k} \sum_{i=1}^{n-d+1} \mathbb{1}\{M_{i+1,i+d-1} \leq X_{n-k,n} < X_i\} \quad (2.6)$$

where $\mathbb{1}$ is an indicator function.

In (2.6), d is assumed to be known. However, d is depending on the $D^{(d)}(u_n)$ condition. It is easy to see that $\hat{\theta}_n(d_i) \geq \hat{\theta}_n(d_j)$ if $d_i \leq d_j$. Whereas d is generally not unique, Cai (2019) proposes an estimator to find the smallest d for which the $D^{(d)}(u_n)$ condition is met. This estimator is given by:

$$\hat{d}^*(k) = \min\{h : \max_{h \leq i \leq d^u} (\hat{\theta}(i) - \hat{\theta}(i+1)) < \frac{1}{\sqrt{k}}\} \quad (2.7)$$

Where d^u is an imposed upper bound on d . This estimator selects the smallest d , for which the maximum difference between $\hat{\theta}(i)$ and $\hat{\theta}(i+1)$ with i larger or equal than d is smaller than $\frac{1}{\sqrt{k}}$.

The estimator can be implemented by iteratively going over each observation X_i in a dataset, and each time determine whether X_i is larger than $X_{n-k,n}$ and whether the maximum of the next $d-1$ observations is smaller than $X_{n-k,n}$ or not. In this constellation, the number of iterations equals the number of observations, which can be problematic for large samples. However, it is possible to come up with a more efficient implementation, as the indicator function can only become 1 if observation X_i is larger than threshold $X_{n-k,n}$. Therefore, it is sufficient to investigate whether the maximum of the next $d-1$ observations passes the $X_{n-k,n}$ threshold for those observations X_i that passed the threshold themselves. For these observations, the indicator function in (2.6) is one if the index of the next observation is maximally $d-1$ larger than the index of the current observation. Therefore, the following algorithm is proposed:

1. Select all elements $X_i > X_{n-k,n}$ and save their indices i .
2. Set d at 2.
3. Initialise $s = 0$
4. for $j = 1, \dots, k-1$; go through the indices i :
 - if $i(j) + d - 1 < i(j+1)$: $s = s + 1$
5. $\hat{\theta}_n(d) = \frac{s}{k}$
6. If $d < d^u$, set $d = d + 1$ and go back to step 4. Else, select \hat{d}^* , based on (2.7).
7. $\hat{\theta}_n = \hat{\theta}_n(t, \hat{d}^*)$

This algorithm assumes that d is known. When d is not known, this algorithm can be applied for several choices for d , until d^* is found. This algorithm only iterates over k elements instead of over n elements. Moreover, the asymptotic theory motivating the estimator requires that $k \rightarrow \infty$, it also requires that $\frac{n}{k} \rightarrow \infty$. It is therefore asymptotically infinitesimally more efficient.

2.1.2 || Sliding block estimators of Berghaus and Bücher (2018)

The estimators described by Berghaus and Bücher (2018) are sliding block estimators. Both compute a statistic for each block, \hat{Z}_{nt}^{sl} and \hat{Y}_{nt}^{sl} respectively, and then compute the inverse of

the average for all blocks. Each block has a length of b_n . The first block consists of observations $X_{1:b_n}$, the second $X_{2:b_n+1}$, and so on. The estimators are given by the following equations:

$$\hat{\theta}_n^{B,sl} = \left(\frac{1}{n - b_n + 1} \sum_{i=1}^{n-b_n+1} \hat{Z}_{nt}^{sl} \right)^{-1} \quad (2.8)$$

$$\hat{\theta}_n^{N,sl} = \left(\frac{1}{n - b_n + 1} \sum_{i=1}^{n-b_n+1} \hat{Y}_{nt}^{sl} \right)^{-1} \quad (2.9)$$

Where $\hat{Z}_{nt}^{sl} = b_n(1 - \hat{N}_{nt}^{sl})$ and $\hat{Y}_{nt}^{sl} = -b_n \log(\hat{N}_{nt}^{sl})$, and $\hat{N}_{nt}^{sl} = \hat{F}(M_{nt}^{sl})$, where \hat{F} is the empirical cdf of X_1, \dots, X_n and $M_{nt}^{sl} = M_{t:t+b_n-1} = \max\{X_t, \dots, X_{t+b_n-1}\}$.

2.2 || TIME-VARYING EXTREMAL INDEX ESTIMATORS

As pointed out in Süveges (2007), long time series are rarely stationary, and allowing the parameters to change slowly over time can yield more accurate results. This is also the case for the extremal index. As many processes exhibit time-varying behaviour, it is not unreasonable to assume that the extremal index is also subject to changes over time. Thus, when we assume a fixed model structure but with time-varying parameters, we still have a limiting Poisson process for the exceedances, but whereas a constant extremal index would imply a homogeneous Poisson process, the time-varying nature will now lead to an inhomogeneous Poisson process (Süveges, 2007). This leads to the definition of $\theta(t)$, which is the extremal index in function of time. In order to estimate $\theta(t)$, we proceed in a similar way as Süveges (2007), who assume that changes in extremal index are slow, and that there is a finite interval in which the inhomogeneous Poisson process can be approximated by a homogeneous Poisson process.

2.2.1 || Rolling horizon adaptation of Cai (2019)

We consider a rolling horizon estimator based on the estimator of Cai (2019). More specifically, the estimator defined in equation (2.7) is used, but instead of using all observations, only p observations are used. p is thus the horizon length of the estimator. Then, the subsample is changed by discarding the first observation and adding the $p + 1$ th observation. This process is continued until the end of the sample has been reached. Thus, the general form of the rolling horizon adaptation is thus given by the following equation:

$$\hat{\theta}_p(t) = \frac{1}{\tilde{k}} \sum_{i=t-\frac{p}{2}}^{t+\frac{p}{2}-d_t+1} \mathbb{1}\{M_{i+1,i+d_t-1} \leq \tilde{X}_{p-\tilde{k},t} < X_i\} \quad (2.10)$$

where $\tilde{X}_{p-\tilde{k},t} = X_{p-k,t-\frac{p}{2}:t+\frac{p}{2}}$, and assuming $\frac{p}{2}$ is an integer. $\tilde{X}_{p-\tilde{k},t}$ is equivalent to $X_{n-k,n}$ in the stable version, and represents the \tilde{k} 'th largest value in the horizon. The estimator is more or less centered around time t , but not completely, as $d_t \geq 1$. However, for many processes, d is quite small, and p quite large. Whereas the general form is similar to estimator defined in (2.6), there are a number of differences in the interpretation of this estimator and the estimator defined in (2.6). First, there is no single $\hat{\theta}_n$ anymore, instead $\hat{\theta}$ is a function of time. Further, not only $\hat{\theta}$ is time-varying, but so are d^* and $X_{n-k,n}$. In (2.10), the subscripts are changed to make this time varying relation transparent. Finally, in order to compare this rolling horizon

adaptation with the stable estimator, k is changed to \tilde{k} . The reason for this change is that k controls the quantile of the estimator. Considering the extremal index with $k = 50$ for a dataset with sample size 10,000 leads us to consider the observations in the 99.5% quantile, whereas $k = 50$ for a dataset of 1,000 leads to the analysis of observations in the 95% quantile. As $n/k \rightarrow \infty$, $p/\tilde{k} < n/k$, $\tilde{k} > kp/n$.

As is the case with (2.6), it is necessary to estimate d . This happens iteratively based on equation (2.7), with the only difference that instead of using the complete sample of n observations to estimate $\hat{d}^*(k)$, we use a subsample of size p , and we compute $\hat{d}^*(k)$ based on \tilde{k} instead of k .

The estimator can be implemented by applying the algorithm described in section 2.1.1 iteratively. However, a more efficient set-up is possible, which exploits the fact that by definition, the rolling horizon estimator is more likely to remain the same than to change when it is moved by one observation at the time. This is the case, because in (2.10) $p > k$, and ideally, $p \ll k$. As a consequence, when moving the time from t to $t + 1$, the likelihood of either the discarded observation or the new considered observation being larger than the threshold $X_{t+p-k,t+p}$ is quite small. In this case, $X_{t+p-k,t+p} = X_{t+1+p-k,t+1+p}$, and the set of observations larger than the threshold also remains the same. As a consequence, also $\hat{\theta}_n(t) = \hat{\theta}_n(t + 1)$. An algorithm to implement the estimator can then be constructed as follows:

1. Initialise at time $t = \frac{p}{2}$. Consider subsample X_1, \dots, X_p .
2. Select all elements in the subsample larger than the threshold $\tilde{X}_{p-\tilde{k},t}$ and save their indices i .
3. Set d at 2.
4. Initialise $s = 0$
5. for $j = 1, \dots, k - 1$; go through the indices i :
 - if $i(j) + d - 1 < i(j + 1)$: $s = s + 1$
6. $\hat{\theta}_p(t, d) = \frac{s}{\tilde{k}}$
7. If $d < d^u$, set $d = d + 1$ and go back to step 4. Else, select \hat{d}^* , based on (2.7).
8. $\hat{\theta}_n(t) = \hat{\theta}_n(t, \hat{d}^*)$
9. while $X_t < X_{t+p-k,t+p}$ and $X_{t+p+1} < X_{t+p-k,t+p}$:
 - set $t = t + 1$ and $\hat{\theta}_p(t + 1) = \frac{s}{\tilde{k}}$
 - end while
10. If $t < n - \frac{p}{2}$ go back to step 2 with $t = t + 1$. Else stop.

Given that the rolling horizon estimator at each point in time takes the form of the stable non-parametric estimator, it also inherits its asymptotic normality. However, one can expect larger confidence intervals, because the time horizon and \tilde{k} will be smaller than their counterparts in a stable setting.

Horizon length

In order to use rolling horizon estimator, it is necessary to define a horizon length p . When defining it, there is a trade-off between having a large p , which improves the accuracy of the estimator when the tail process is stable, and a small p , which is more reactive to changes in the tail process. In most practical applications, this is quite complex, as it is often not known if and how fast the tail process changes.

However, the degree of extremal clustering using different horizon lengths itself provides some insight in how large p needs to be for a rolling horizon estimator with horizon length p to be a valid estimator. As such, it is possible to construct a lower bound for p . To understand this, consider the application of the rolling horizon estimator with \tilde{k} . Define $l = \frac{\tilde{k}}{n}$. If the extremal index is θ , the expected duration of an extremal cluster is $\frac{1}{\theta}$. As such, when the expected duration of an extremal cluster is $\frac{1}{\theta}$, and the expected interexceedance time is $\frac{1}{l\theta} - \frac{1}{\theta}$. Adding the expected interexceedance time and expected cluster time, we arrive at a period of length $\frac{1}{l*\theta}$. For a given l , this function $\rightarrow \infty$ as $\theta \rightarrow 0$.

So, based on this result, we can consider a lower bound on the horizon length. After choosing a horizon length, one can compute $\hat{\theta}(t)$, and determine its minimum $\hat{\theta}_h^-$. The horizon length needs to be at least $\frac{1}{\hat{\theta}_h^- l}$. If the horizon length is smaller than this lower bound, the rolling horizon estimator will at moments consider a series in which no cluster is present, which will result in a spurious estimation.

Based on this reasoning, an algorithm which can be used to find the lower bound h is the following:

1. Set \tilde{k} and compute l .
2. Initialise horizon length at h .
3. Estimate $\hat{\theta}_h^- = \min_{t \in T}(\hat{\theta}_h(t))$
4. If $h > \frac{1}{\hat{\theta}_h^- l}$, stop. Else set $h_{i+1} = h_i + 1$ and go to step 3.

2.2.2 || Parametric Modelling Approach

For certain processes, the extremal index is known. This can be handy, as it can allow for the use of parametric approaches for the filtering of the extremal index. Moreover, some filters can allow a time-varying extremal index. This section starts with a description of a general class of models, and then discusses one model of particular interest: the time-varying AR(1) model with Student-t distributed errors, as well as potential filters for this model. The section ends with a description of the method to estimate parameters.

Generalised Autoregressive Score regression models

We consider the following set of models:

$$y_t = \phi_t X_t + \epsilon_t, \quad \epsilon_t \sim i i \tau(\lambda) \tag{2.11}$$

This model is a regression model with time varying parameter vector ϕ_t . X_t is a vector which can consist of exogenous regressors, a series of 1 (which amount to a time-varying intercept), lags of y_t , including y_{t-1} , but also further lags y_{t-2}, \dots, y_{t-p} , $p \in \mathbb{N}$, previous error terms $\epsilon_{t-1}, \dots, \epsilon_{t-q}$, $q \in \mathbb{N}$. One can filter out the parameter vector by the following filter:

$$\hat{\phi}_{t+1} = \omega + \alpha\hat{\phi}_t + \beta((X_i X_i')^{-1}(X_t y_t) - \hat{\phi}_t) \quad (2.12)$$

where $\phi_t, \hat{\phi}_t, X_t, \omega$ are $(n \times 1)$ vectors, and α and β are matrices of dimension $(n \times n)$ (Creal et al., 2008).

Depending on the elements involved in the regressor vector (for example previous lags), ϕ_t can be modelled as a transformation of another parameter. Further, restrictions can be needed, for example to ensure the uniform invertibility of the filter.

Time-varying AR(1) model

One process of particular interest for the data used in this study (see empirical sections) is the AR(1) model with identically and independently distributed errors which follow a Student-t distribution. More specifically:

$$y_{t+1} = \phi y_t + \epsilon_t, \quad \epsilon_t \sim ii\tau(\lambda) : \quad \theta = 1 - |\phi|^\lambda \quad (2.13)$$

with $y_1 \sim ii\tau(\lambda)$. $\tau(\lambda)$ refers to a Student-t distribution with λ degrees of freedom, and $|\phi| < 1$ (Mikosch, 2013). In this model, the larger the autocorrelation parameter ϕ is in absolute size, the stronger the extremal dependency. Also the degrees of freedom have an impact, and increases in the degrees of freedom lead to a smaller extremal dependency. When $\lambda \rightarrow \infty$, the Student-t distribution approximates the normal distribution. Hence, the extremal index for an AR(1) model with normally distributed errors is 1. When $\lambda = 1$, the Student-t distribution equals the Cauchy distribution, and for an AR(1) model with Cauchy errors, the extremal index is $1 - |\phi|$. In certain cases, the autoregressive parameter may also be varying over time. For example, imagine a data generating process described by the following equation:

$$y_{t+1} = \tanh(\rho_t) y_t + \epsilon_t, \quad \epsilon_t \sim ii\tau(\lambda) \quad (2.14)$$

with $y_1 \sim \tau(\lambda)$, $|\gamma| < 1$ and ρ_t a time-varying process. We assume $\lambda > 2$. The \tanh function is the hyperbolic tangent function: $\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$.

We consider the case in which ρ is changing slowly over time in a deterministic way. Then, it can locally be approximated by the data generating process described in (2.14), and can locally be considered approximately stationary and ergodic. The extremal index can then be approximated by $\theta(t) = 1 - |\rho(t)|^\lambda$.

For the data generating process defined in (2.14), an equivalent filter to the filter described in (2.12) would be:

$$\tanh(\hat{\rho}_{t+1}) = \omega + \alpha \tanh(\hat{\rho}_t) + \beta \left(\frac{(y_{t+1}) y_t}{y_t^2} - \tanh(\hat{\rho}_t) \right) \quad (2.15)$$

However, this filter will lead to problems, as the term $\frac{(y_{t+1})y_t}{y_t^2}$ could become large in absolute value, which may lead to expressions in the right hand side larger than 1 or smaller than -1, whereas the domain of the \tanh function is $(-1, 1)$.

Therefore, the following basic filter for ρ_t is proposed:

$$\hat{\rho}_{t+1} = \omega + \alpha \hat{\rho}_t + \beta \left(\frac{y_{t+1}y_t}{y_t^2} - \tanh(\hat{\rho}_t) \right) \quad (2.16)$$

with $|\alpha| + |\beta| < 1$. In this model, the time-varying autoregressive parameter $\tanh(\rho_t) \in (-1, 1)$ at all times. The filtering equation considers the previous filtered estimate, and an approximation of the error of the previous filtered estimate by comparing the filtered time-varying parameter with the OLS estimate based on the previous observation. However, the filter in (2.16) can be unstable, because when $\tanh(\hat{\rho}_t) \approx \tanh(\rho_t)$,

$$\frac{(y_{t+1})y_t}{y_t^2} - \tanh(\hat{\rho}_t) = \frac{(\tanh(\rho_t)y_t + \epsilon_t)y_t}{y_t^2} - \tanh(\hat{\rho}_t) \approx \frac{\epsilon_t}{y_t} \quad (2.17)$$

When y_t is close to zero and ϵ_t (relatively) large, this term may become very large, which can lead to an unstable filter. One way to deal with this is to instead of considering the OLS estimate based on the previous observation, using the previous κ observations. Given the fact that $\tanh(\rho_t)$ is varying over time, this is only approximately correct. However, the problem is alleviated by the nature of many practical applications, where the time-varying behaviour can be expected to change slowly, and considering the last κ observations will then lead to a smoother signal with relatively small approximation error. The κ -observations based filter then becomes:

$$\hat{\rho}^\kappa_{t+1} = \omega + \alpha \hat{\rho}^\kappa_t + \beta \left(\frac{\sum_{i=t-\kappa+1}^t y_{i+1}y_i}{\sum_{i=t-\kappa+1}^t y_i^2} - \tanh(\hat{\rho}^\kappa_t) \right) \quad (2.18)$$

In order to keep a Markovian structure, this filter can also be expressed in its companion form:

$$\hat{P}^\kappa_{t+1} = \Omega + A \hat{P}^\kappa_t + B((Y'_i Y_i)^{-1} Y'_i Y_{i+1} - \tanh(P^\kappa_t)) \quad (2.19)$$

where

$$\hat{P}^\kappa_t = \begin{pmatrix} \hat{\rho}^\kappa_t \\ \hat{\rho}^\kappa_{t-1} \\ \dots \\ \hat{\rho}^\kappa_{t-\kappa} \end{pmatrix}, \quad \Omega = \begin{pmatrix} \omega \\ 0 \\ \dots \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} \alpha & \mathbf{0} \\ \mathbf{0} & I_{\kappa-1} \end{pmatrix}, \quad Y_i = \begin{pmatrix} y_t & 0 & \dots & 0 \\ y_{t-1} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ y_{t-\kappa+1} & 0 & \dots & 0 \end{pmatrix} \quad (2.20)$$

where \hat{P}^κ_t and Ω are $(\kappa \times 1)$ vectors, and A and Y_i are $(\kappa \times \kappa)$ matrices. Another way to accommodate the problem is by considering a robust filter. A robust filter can be constructed as:

$$\hat{\rho}^*_{t+1} = \omega + \alpha \hat{\rho}^*_t + \beta \tanh \left(\frac{y_{t+1}y_t}{y_t^2} - \tanh(\hat{\rho}^*_t) \right) \quad (2.21)$$

As the image of the \tanh function is $(-1, 1)$, the robust filter limits the magnitude of the score part and also leads to a smoother signal.

The data generating process described in (2.14), initialised at $y_1 \in \mathbb{R}$ is converging e.a.s. to a stationary and ergodic sequence, and that $\mathbb{E}|y_t(y_1, \psi)|^2 < \infty$ provided that ρ_t innovation sequence in the data generating process of is strictly stationary and ergodic and that ρ_t is generated by a measurable function.

This can be proven by applying the uniform contraction theorem.

THEOREM 1 (Uniform contraction theorem (Blasques, 2019)). *A random sequence $y_t(\psi, y_1)_{t \in \mathbb{N}}$, initialized at $t = 1$ with value $y_1 \in Y \subseteq \mathbb{R}$ and generated by $y_{t+1} = \phi(y_t, \epsilon_t, \psi)$, $\forall t \in \mathbb{N}$ for some $\psi \in \Psi$, converges e.a.s to a unique strictly stationary and ergodic sequence as $t \rightarrow \infty$, and $\mathbb{E}|y_t(y_1, \psi)|^n < \infty$ if for some $n > 0$:*

1. *The innovation sequence $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a n_ϵ -variate stationary and ergodic sequence.*
2. *There exists a $y_1 \in Y$ such that $\mathbb{E}\log_+|\phi(y_1, \epsilon_1)|^n < \infty$.*
3. *$\sup_{y \in Y, \epsilon} |\partial\phi(y, \epsilon_t)/\partial y| < 1$.*

PROPOSITION 2. *The sequence defined by (2.14), converges e.a.s. to a strictly stationary and ergodic sequence with two bounded moments, provided that all innovation sequences in the updating equation of ρ_t are strictly stationary and ergodic.*

Proof. We apply the condition of the uniform contraction theorem to prove e.a.s. convergence to a strictly stationary sequence with two bounded moments.

To prove that the data generating process y_t described in (2.14), when initialised at some $y_1 \in \mathbb{R}$ is converging e.a.s. to a stationary and ergodic sequence, we start with the first condition of theorem 1. The innovation sequence for y_t consists of $(\tanh(\rho_t), \epsilon_t)$. As described before, we impose the condition that any innovation sequence in ρ_t is strictly stationary and ergodic, and that ρ_t is generated by a measurable function. As the \tanh function is also a measurable function, by Krengel's theorem, $\tanh(\rho_t)$ is also a strictly stationary and ergodic sequence (Krengel, 2011). As ϵ_t is also strictly stationary and ergodic, the first condition of the theorem is satisfied.

With regard to the second condition,

$$\mathbb{E}|\phi(y_1, \epsilon_1)|^2 = \mathbb{E}|\tanh(\rho_1)y_1 + \epsilon_1|^2 \quad (2.22)$$

$$\leq c\mathbb{E}|\tanh(\rho_1)y_1|^2 + c\mathbb{E}|\epsilon_1|^2 \quad (2.23)$$

$$\leq c|\tanh(\rho_1)y_1|^2 + c\mathbb{E}|\epsilon_1|^2 \quad (2.24)$$

$$\leq c|y_1|^2 + c\mathbb{E}|\epsilon_1|^2 < \infty \quad (2.25)$$

with $c > 1$. The second inequality follows from the c_n -inequality (Blasques, 2019). $\epsilon \sim ii\tau(\lambda)$ with $\lambda > 2$ and hence has at least two bounded moments. So, this condition is satisfied for all $|y_1| < \infty$.

Finally, the third condition is also satisfied, as

$$\sup_{y \in Y, \rho \in \mathbb{R}} |\partial\phi(y, \epsilon_t)/\partial y| \quad (2.26)$$

$$= \sup_{y \in Y, \rho \in \mathbb{R}} |\partial(\tanh(\rho_t)y_t + \epsilon_t)/\partial y| \quad (2.27)$$

$$= \sup_{\rho \in \mathbb{R}} |\tanh(\rho_t)| < 1 \quad (2.28)$$

Thus, the data generating process thus converges e.a.s. to a strictly stationary and ergodic sequence with two bounded moments. \square

Note that proposition 2 includes processes for the true sequence ρ_t in the same form of filters (2.18) and (2.21).

Uniform invertibility of the filters

DEFINITION (Uniform invertibility (Blasques, 2019)). *A filter is uniform invertible for $\psi \in \Psi$ if $\sup_{\psi \in \Psi} |\hat{\alpha}_t(\psi, \hat{\alpha}_1) - \alpha(\psi)| \rightarrow 0$ almost surely as $t \rightarrow \infty$ for every initialisation $\alpha_1 \in \mathbb{R}$.*

The uniform invertibility of the filters discussed in the previous section can be shown using Bougerol's theorem. However, in contrast to the application on the data generating process in the previous section, the innovations in the filtering equation now consist of the observations themselves.

THEOREM 2 (Bougerol's theorem (Blasques, 2019)). *A random sequence $y_t(\theta, y_1)_{t \in \mathbb{N}}$, initialized at $t = 1$ with value $y_1 \in Y \subseteq \mathbb{R}$ and generated by $y_{t+1} = \phi(y_t, \epsilon_t, \theta)$, $\forall t \in \mathbb{N}$ for some $\theta \in \Theta$, converges e.a.s to a unique strictly stationary and ergodic sequence as $t \rightarrow \infty$ if:*

1. *The innovation sequence $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a n_ϵ -variate stationary and ergodic sequence.*
2. *There exists a $y_1 \in Y$ such that $\mathbb{E} \log_+ |\phi(y_1, \epsilon_1)| < \infty$.*
3. *$\mathbb{E} \log \sup_{y \in Y} |\partial\phi(y, \epsilon_t)/\partial y| < 0$.*

In what follows, we argue the filter defined in (2.18) is uniformly invertible when $|\alpha| + |\beta| < 1$ and $\lambda > 2$ for the data generating process defined in (2.14). The proof is shown for $\kappa = 1$ as to have a univariate filter.

To ease notation, we drop the superscript κ from $\hat{\rho}_t^\kappa$ and write $\hat{\rho}_t$.

PROPOSITION 3. *The filter for the time-varying parameter in the process defined in (2.14), defined in (2.18) with $\kappa = 1$, is uniformly invertible when $|\alpha| + |\beta| < 1$ and $\lambda > 2$.*

Proof. To show the uniform invertibility of the filter, we apply the conditions of theorem 2. The first condition holds as the innovation sequence in (2.18) is $\beta \frac{y_t y_{t+1}}{y_t^2}$. As y_t is a stationary and ergodic sequences (shown in the previous section), and $\beta \frac{y_t y_{t+1}}{y_t^2}$ is a measurable function of this stationary and ergodic sequences, it is also stationary and ergodic by Krengel's theorem (Krengel, 2011).

With regard to the second condition $\mathbb{E} \log_+ |\phi(y_1, \epsilon_1)| < \infty$, one can write for a value $0 < k < \frac{1}{3}$

:

$$\mathbb{E} \log_+ |\omega + \alpha \hat{\rho}_1 + \beta((y_1)^{-2}(y_1 y_2) - \tanh(\hat{\rho}_1))| \quad (2.29)$$

$$= \mathbb{E} \log_+ |\omega + \alpha \hat{\rho}_1 + \beta((y_1)^{-2}(y_1(y_1 \tanh(\rho_1) + \epsilon_1) - \tanh(\hat{\rho}_1)))| \quad (2.30)$$

$$\leq \mathbb{E} |\omega + \alpha \hat{\rho}_1 + \beta((y_1)^{-2}(y_1(y_1 \tanh(\rho_1) + \epsilon_1) - \tanh(\hat{\rho}_1)))|^k \quad (2.31)$$

$$\leq c \mathbb{E} |\omega + \alpha \hat{\rho}_1 - \beta \tanh(\hat{\rho}_1)|^k + c \mathbb{E} |\beta(y_1 y_1)^{-1}(y_1 y_1 \tanh(\rho_1))|^k + c \mathbb{E} |\beta(y_1)^{-2}(y_1 \epsilon_1)|^k \quad (2.32)$$

$$\leq c \mathbb{E} |\omega + \alpha \hat{\rho}_1 - \beta \tanh(\hat{\rho}_1)|^k + c \mathbb{E} |\beta \tanh(\rho_1)|^k + c \mathbb{E} |\beta(y_1)^{-2}(y_1 \epsilon_1)|^k \quad (2.33)$$

$$\leq c \mathbb{E} |\omega + \alpha \hat{\rho}_1 - \beta \tanh(\hat{\rho}_1)|^k + c |\beta|^k + c \mathbb{E} |\beta(y_1)^{-2}(y_1 \epsilon_1)|^k \quad (2.34)$$

where c is a real number > 0 . The fourth equation is the result of the c_n -inequality (Blasques, 2019). In these equations, caution is necessary for the interpretation of y_1 , ρ_1 and ϵ_1 . Whereas the indexing of these terms correspond to the indexing of the filtered estimates, these sequences were initialised at the infinite past and not time $t = 1$. For every finite initialisation of $\hat{\rho}_1$, the first term is also finite, because the parameter vector $\psi \in \Psi$ and Ψ is a bounded subset of \mathbb{R} for every value of k . The second term is also finite for the same reason. Finally, the last term is also finite. By the Holder inequality (see for example (Blasques, 2019)), the term can be written as:

$$\mathbb{E} |\beta(y_1)^{-2}(y_1 \epsilon_1)|^k \leq |\beta|^k \mathbb{E} |(y_1)^{-2}(y_1 \epsilon_1)|^k \leq \beta^k \mathbb{E} |(y_1)^{-2}|^p \mathbb{E} |(y_1 \epsilon_1)|^q \quad (2.35)$$

where $k = \frac{pq}{p+q}$, and $p < 0.5$ and $q = 1$.

$\mathbb{E}|y_1 \epsilon_1| = \mathbb{E}|y_1| \mathbb{E}|\epsilon_1| = 0$, as y_t has two bounded moments (proven earlier), and ϵ_1 and y_1 are independent. The term $\mathbb{E}|(y_1)^{-2}|$ is complex. Even though $\mathbb{E}|y_t|^2$ is bounded, this does not imply that its inverse is. Moreover, when analysing $\mathbb{E}\frac{1}{y_t^2}$, one obtains:

$$y_i^2 = (\phi_{i-1} y_{i-1} + \epsilon_i)^2 \quad (2.36)$$

$$= (\phi_{i-1}(\phi_{i-2} y_{i-2} + \epsilon_{i-1}) + \epsilon_i)^2 \quad (2.37)$$

$$= (\phi_{i-1} \phi_{i-2} y_{i-2} + \phi_{i-1} \epsilon_{i-1} + \epsilon_i)^2 \quad (2.38)$$

$$= \left(\sum_{t=1}^{i-1} \prod_{j=t}^{i-1} \phi_j \epsilon_t + \epsilon_i \right)^2 \quad (2.39)$$

where $\phi_t = \tanh(\rho_t)$, and the ϵ -sequence was moved by 1 for notational ease. As ϕ_t is also a random variable, y_t has a distribution for which no closed form is available. This is also the case when ϕ_t is deterministic, as the linear combination of t-distributed variables has no general closed form solution (Ahsanullah et al., 2014). In some cases, a solution is available: in case $\phi_{i-1} = 0$, $y_i^{-2} = \epsilon_i^{-2} \sim F(n, 1)$, which has finite moments of order p with $p < 0.5$, which means that under those circumstances, the term is finite. Moreover, we can obtain the following result:

$$\mathbb{E}(y_i)^2 = \mathbb{E}\left(\sum_{t=1}^{i-1} \prod_{j=t}^{i-1} \phi_j \epsilon_t + \epsilon_i\right)^2 \quad (2.40)$$

$$= \mathbb{E}\left(\sum_{t=1}^i A_t^2 \epsilon_t^2 + 2 \sum_{t=1}^{i-1} \sum_{u=t+1}^i A_t A_u \epsilon_t \epsilon_u\right) \quad (2.41)$$

$$= \mathbb{E}\left(\sum_{t=1}^i A_t^2 \epsilon_t^2\right) + 2\mathbb{E}\left(\sum_{t=1}^{i-1} \sum_{u=t+1}^i A_t A_u \epsilon_t \epsilon_u\right) \quad (2.42)$$

$$= \mathbb{E}\left(\sum_{t=1}^i A_t^2 \epsilon_t^2\right) + 2 \sum_{t=1}^{i-1} \sum_{u=t+1}^i \mathbb{E}(A_t A_u) \mathbb{E}(\epsilon_t) \mathbb{E}(\epsilon_u) \quad (2.43)$$

$$= \mathbb{E}\left(\sum_{t=1}^i A_t^2 \epsilon_t^2\right) \geq \mathbb{E}(\epsilon_i^2) \quad (2.44)$$

where $A_u = \prod_{j=t}^u \phi_j$ and $A_i = 1$. The fourth equation follows from the fact that ϵ_u is independent from ϵ_t for $u \neq t$, and that ϵ_i is independent from A_j for all i, j . As such, given that for $k < 0.5$, $\mathbb{E}|(\epsilon_i^{-2})|^k < \infty$, and $\mathbb{E}(y_i^2) \geq \mathbb{E}(\epsilon_i^2)$ it is assumed that also $\mathbb{E}|(y_i^{-2})|^k < \infty$.

we can therefore consider

$$c\mathbb{E}|\omega + \alpha\hat{\rho}_1 - \beta\tanh(\hat{\rho}_1))|^k + c|\beta|^k + c\mathbb{E}|\beta(y_1)^{-2}(y_1\epsilon_1)|^k < \infty \quad (2.45)$$

Finally, with regard to the third condition, $\mathbb{E} \log \sup_{y \in Y} |\partial\phi(y, \epsilon_t)/\partial y|$ becomes:

$$\mathbb{E} \log \sup_{\hat{\rho}_t \in \mathbb{R}} \left| \frac{\partial(\omega + \alpha\hat{\rho}_t + \beta(\frac{y_t y_{t+1}}{y_t^2} - \tanh(\hat{\rho}_t)))}{\partial\hat{\rho}_t} \right| \quad (2.46)$$

$$= \mathbb{E} \log \sup_{\hat{\rho}_t \in \mathbb{R}} |\alpha - \beta(1 - \tanh(\hat{\rho}_t)^2)| \quad (2.47)$$

It is useful to note that $\mathbb{E} \log \sup_{\hat{\rho}_t \in \mathbb{R}} |\alpha - \beta(1 - \tanh(\hat{\rho}_t)^2)| < 0$ is implied by $\mathbb{E} \sup_{\hat{\rho}_t \in \mathbb{R}} |\alpha - \beta(1 - \tanh(\hat{\rho}_t)^2)| < 1$ (Blasques, 2019).

$$\mathbb{E} \sup_{\hat{\rho}_t \in \mathbb{R}} |\alpha - \beta(1 - \tanh(\hat{\rho}_t)^2)| = \sup_{\hat{\rho}_t \in \mathbb{R}} |\alpha - \beta(1 - \tanh(\hat{\rho}_t)^2)| \quad (2.48)$$

$$< |\alpha| + |\beta| |1 - \tanh(\hat{\rho}_t)^2| < |\alpha| + |\beta| \quad (2.49)$$

Hence, when $|\alpha| + |\beta| < 1$, the third condition is fulfilled.

□

The proof above discusses the filter defined in (2.18). However, it is also possible, and even easier, to prove the uniform invertibility for the filter defined in (2.21). Moreover, instead of Bougerol's theorem, the uniform contraction theorem can be used, which in addition to the e.a.s. convergence to a strictly stationary and ergodic sequence also adds that this strictly stationary and ergodic sequence has bounded moments of a certain order, in this case even of any order. We therefore analyse whether the conditions of the uniform contraction theorem are fulfilled. These conditions are listed in the previous section.

PROPOSITION 4. *The filter, defined in (2.21), is uniformly invertible for the process defined in (2.14), and converges e.a.s. to a limit filter with bounded moments of any order when $|\alpha| + |\beta| < 1$.*

Proof. To prove the uniform invertibility, we apply the conditions of the uniform contraction theorem.

The \tanh function is a measurable function, and as such, the innovation sequence, which is the same as in the proof for proposition 3, remains strictly stationary and ergodic by Krengel's theorem (Krengel, 2011) and the first condition is satisfied. For the second condition, considering $\tanh(\frac{y_i y_{i+1}}{y_i^2})$ instead of $\frac{y_i y_{i+1}}{y_i^2}$ actually makes life easier, as $\mathbb{E}|\tanh(z)|^k < \infty$ for all random variables z and for every $k > 0$ (Blasques, 2019). The condition is thus fulfilled without having to investigate the term $\mathbb{E}|1/(y_t^2)|^k$, and as all the elements outside the \tanh function are non-random, the term $\mathbb{E}|\omega + \alpha\hat{\rho}_1 + \beta\tanh((y_1)^{-2}(y_1 y_2) - \tanh(\hat{\rho}_1))|^n < \infty \forall n > 0$. Finally, with regard to the third condition of the uniform contraction theorem $\sup_{y \in Y, \epsilon \in E} |\partial\phi(y, \epsilon_t)/\partial y|$ can be written as:

$$\sup_{\hat{\rho}, \rho, y, \epsilon \in \mathbb{R}} \left| \frac{\partial(\omega + \alpha\hat{\rho}_t + \beta\tanh(\frac{y_t y_{t+1}}{y_t^2} - \tanh(\hat{\rho}_t)))}{\partial\hat{\rho}_t} \right| \quad (2.50)$$

$$= \sup_{\hat{\rho}, \rho, y, \epsilon \in \mathbb{R}} |\alpha + \beta(1 - \tanh(\frac{y_t y_{t+1}}{y_t^2} - \tanh(\hat{\rho}_t)))^2(-1)(1 - \tanh(\hat{\rho}_t)^2)| \quad (2.51)$$

$$\leq \sup_{\hat{\rho}, \rho, y, \epsilon \in \mathbb{R}} |\alpha - \beta| < |\alpha| + |\beta| \quad (2.52)$$

As such, the robust filter defined in (2.21) is also uniformly invertible for $|\alpha| + |\beta| < 1$ and $\lambda > 2$, and converges e.a.s. to a unique stationary and ergodic sequence with bounded moments of any order. \square

This is quite remarkable, especially because it is only possible to show that the data generating process itself only converges e.a.s to a limit sequence with bounded moments smaller than the degrees of freedom of the error term.

Parameter Estimation

The parameters of the model can be estimated by maximum likelihood. This is an advantage over the estimators based on Cai (2019) and Berghaus and Bücher (2018), as it is not necessary to set hyperparameters. For a given observation y_t , the observations y_1, \dots, y_{t-1} are exogenous. Moreover, given the Markovian structure of the model, conditioning on y_1, \dots, y_{t-1} boils down to conditioning on y_{t-1} . Then the likelihood function can be decomposed by using the prediction error decomposition:

$$\mathcal{L}(y, \psi) = f(y_1) \prod_{t=2}^n f(y_t | y_{t-1}, \psi) \quad (2.53)$$

where $f(y_t | y_{t-1}, \psi)$ is the probability density function (pdf) of y_t given previous observations y_{t-1} and parameter vector ψ .

In practice, it is easier to maximise the log likelihood function, which can be expressed as:

$$\log \mathcal{L}(y, \psi) = \log(f(y_1 | \psi)) + \sum_{t=2}^n \log(f(y_t | \psi, y_{t-1})) \quad (2.54)$$

For the likelihood of the filtering equation described by (2.18), the log likelihood function is:

$$\log \mathcal{L}(y, \psi) = \sum_{t=2}^n \log \left(\frac{\Gamma(\frac{\lambda+1}{2})}{\sqrt{\lambda \pi} \Gamma(\frac{\lambda}{2})} \left(1 + \frac{(y_t - \tanh(\hat{\rho}_{t-1}) y_{t-1})^2}{\lambda} \right)^{-\frac{\lambda+1}{2}} \right) \quad (2.55)$$

Where as before, $\hat{\rho}_{t+1} = \omega + \alpha \hat{\rho}_t + \beta \left(\frac{\sum_{i=t-\kappa+1}^t y_{i+1} y_i}{\sum_{i=t-\kappa+1}^t y_i^2} - \tanh(\hat{\rho}_t) \right)$ in case of the κ -based filter and $\hat{\rho}^*_{t+1} = \omega + \alpha \hat{\rho}^*_t + \beta \tanh \left(\frac{y_{t+1} y_t}{y_t^2} - \tanh(\hat{\rho}^*_t) \right)$ in case of the robust filter. Thus,

$$\hat{\psi} = \operatorname{argmax}_{\psi \in \Psi} \log \mathcal{L}(\psi) \quad (2.56)$$

Given the complex form of the log likelihood function, the optimisation is done numerically.

Based on the filtered time-varying autoregression coefficient, it is then possible to also determine a time-varying extremal index $\hat{\theta}_t$. This is equal to:

$$\hat{\theta}_t = 1 - |\tanh(\hat{\rho}_t)|^{\hat{\lambda}} \quad (2.57)$$

Consistency of the Estimators

DEFINITION (Consistency (Blasques, 2019)). *An estimator $\hat{\psi}_t$ is consistent for a parameter ψ_0 , if and only if $\hat{\psi}_t \xrightarrow{p} \psi_0$ (Blasques, 2019).*

Based on theorem 7.1 and theorem 7.14 of Blasques (2019), we can construct the following theorem:

THEOREM 3 (Consistency filter-based criterion (Blasques, 2019)). *The parameters of the filters estimated by maximum likelihood are consistent if:*

1. *Parameter space Ψ is compact;*
2. *The criterion function is continuous in both parameters and observations;*
3. *The criterion function of the limit filter converges uniformly over Ψ to the limit deterministic function Q_∞ as $t \rightarrow \infty$: $\sup_{\psi \in \Psi} |Q_t(x_t, \psi) - Q_\infty| \xrightarrow{p} 0$*
4. *The estimator is the identifiable unique maximizer of the criterion function: $\sup_{\psi \in S^c(\psi_0, \delta)} Q_\infty(\psi) < Q_\infty(\psi_0)$, where $S^c(\psi_0, \delta)$ is the complement of a set of points contained in a ball around ψ_0 in Ψ*
5. *The filter is uniformly invertible.*

PROPOSITION 5. *The parameters of the filters defined in (2.18) and (2.21) obtained by optimizing the criterion function defined in (2.55) over a compact set are consistent, provided the criterion function is well-behaved of order 2 in both ψ and ρ .*

Proof. We obtain the consistency by applying the conditions of theorem 3.

Regarding the first condition, choosing a compact Ψ is controlled by the user of the estimator, and this condition can be considered to be fulfilled when the user maximises the parameters over a compact set.

The second condition is also fulfilled. When looking at equation (2.55), one can see that for $\lambda > 0$, the function is continuous for every $\psi \in \Psi$ and $y_t \in \mathbb{R}$, for every value in $y_t \in \mathbb{R}$ and $\psi \in \Psi$ respectively.

A sequence is uniform convergent if it is pointwise convergent and stochastically equicontinuous (Blasques, 2019). A criterion function converges pointwise to a limit deterministic function $Q_\infty = \mathbb{E}(q(y_t, y_{t-1}, \psi))$ if $Q_t(y_t, \psi) \xrightarrow{p} Q_\infty(\psi)$ as $t \rightarrow \infty$ (Blasques, 2019). A sequence is stochastically equicontinuous if $\mathbb{E}\sup_{\psi \in \Psi} \|\partial Q_t(\psi)/\partial \psi\| < \infty$ (Blasques, 2019). Pointwise convergence can be shown easily for M-estimators, as it only requires that the sequence is stationary and ergodic, and has a bounded first moment (Blasques, 2019). (2.55) can be considered as $\sum_{t=2}^n q(y_t, y_{t-1}, \psi)$ and is thus an M-estimator. Further, q is a sum of a measurable function of SE elements, by Krengel's theorem, and hence $q(y_t, y_{t-1}, \psi)$ is also a SE sequence. Further,

$$\mathbb{E}|q(y_t, y_{t-1}, \psi)|^2 = \mathbb{E}|\log\left(\frac{\Gamma(\frac{\lambda+1}{2})}{\sqrt{\lambda\pi}\Gamma(\frac{\lambda}{2})}(1 + \frac{(y_t - \tanh(\hat{\rho}_{t-1})y_{t-1})^2}{\lambda})^{-\frac{\lambda+1}{2}}\right)|^2 \quad (2.58)$$

$$= \mathbb{E}|\log\left(\frac{\Gamma(\frac{\lambda+1}{2})}{\sqrt{\lambda\pi}\Gamma(\frac{\lambda}{2})}\right)^2 - (\lambda+1)\log(1 + \frac{(y_t - \tanh(\hat{\rho}_{t-1})y_{t-1})^2}{\lambda})| \quad (2.59)$$

$$= \mathbb{E}|A| + \mathbb{E}|(\lambda+1)\log(1 + \frac{(y_t - \tanh(\hat{\rho}_{t-1})y_{t-1})^2}{\lambda})| \quad (2.60)$$

$$\leq A + \mathbb{E}|(\lambda+1)(1 + \frac{(y_t - \tanh(\hat{\rho}_{t-1})y_{t-1})^2}{\lambda})| \quad (2.61)$$

$$\leq A + |\lambda+1| + |\frac{\lambda+1}{\lambda}|\mathbb{E}|y_t - \tanh(\hat{\rho}_{t-1})y_{t-1}|^2 \quad (2.62)$$

$$\leq B + c|\frac{\lambda+1}{\lambda}|\mathbb{E}|y_t|^2 + c|\frac{\lambda+1}{\lambda}|\mathbb{E}|\tanh(\hat{\rho}_{t-1})y_{t-1}|^2 \quad (2.63)$$

$$\leq B + c|\frac{\lambda+1}{\lambda}|\mathbb{E}|y_t|^2 + c|\frac{\lambda+1}{\lambda}|\sup_{\hat{\rho}_{t-1} \in \mathbb{R}}|\tanh(\hat{\rho}_{t-1})|^2\mathbb{E}|y_{t-1}| \quad (2.64)$$

$$\leq B + c|\frac{\lambda+1}{\lambda}|\mathbb{E}|y_t|^2 + c|\frac{\lambda+1}{\lambda}|\mathbb{E}|y_{t-1}| < \infty \quad (2.65)$$

Where $A = \log\left(\frac{\Gamma(\frac{\lambda+1}{2})}{\sqrt{\lambda\pi}\Gamma(\frac{\lambda}{2})}\right)^2$, and $B = A - (\frac{\lambda+1}{2})$ and c a number > 0 . The third last equation is the result of applying the c_n -inequality, and the second last equation of the Holder inequality, as $|\tanh(\hat{\rho}_{t-1})|^2 < 0$ almost surely for every value of $\hat{\rho}_{t-1} \in \mathbb{R}$. (Blasques, 2019). The result $< \infty$ follows from the fact that A is $< \infty$ as all $\psi < \infty \forall \psi \in \Psi$, y_t has two bounded moments (proven earlier) and $\mathbb{E}|\tanh(z)|^k < \infty$ for every random variable z and every k . As such, $\mathbb{E}|q(y_t, y_{t-1}, \psi)|^2 < \infty$ and hence its first moment is also bounded. Thus, the criterion function is pointwise converging.

Stochastic equicontinuity can be established if $q(y_t, \rho(\psi), \psi)$ is well behaved of order 2 in both ρ and ψ (Blasques, 2019). Proving this is non-trivial as the derivative of the log will lead to having an expression with random values in the denominator. This is outside the scope of this paper, and stochastic equicontinuity is just assumed, based on the assumption that the criterion function is well-behaved of order 2 in both ρ and ψ . Proving the well-behavedness of

the criterion function can be a suggestion for further research.

With regard to the fourth condition, the model is not observationally equivalent to other models, and is thus identifiable. The criterion function has a bounded moment, and when assuming correct model specification, one can use the information inequality theorem to conclude that the true parameter is the unique maximizer of the loglikelihood function (Blasques, 2019). As discussed before, Ψ is a compact set, and as based on Blasques (2019), the ψ_0 is then identifiable unique. In case of misspecification, the estimators are set consistent if all other criteria are fulfilled (Blasques, 2019). Set consistency is defined as $d(\psi_T, \Psi_0) \xrightarrow{P} 0$ as $T \rightarrow \infty$ with $\Psi_0 = \arg \max_{\psi \in \Psi} Q_\infty(\psi)$ (Blasques, 2019).

Finally, the uniform invertibility of the filter has been discussed in the previous section.

Thus, the estimator is consistent upon correct specification and when the criterion function is well behaved of order 2 in ρ and ψ , and is set consistent upon misspecification and if the criterion function is well behaved of order 2 in ρ and ψ . □

2.3 || COMPARISON OF THE ESTIMATORS

The estimators are based on different assumptions, and as such behave differently depending on whether these assumptions are met or not. The stable approaches of Cai (2019) and Berghaus and Bücher (2018) both are asymptotically normal. Cai (2019) reports the results of a number of simulations which find that the non-parametric estimator defined in (2.10) has better finite sample performance compared to Berghaus and Bücher (2018). As shown before, choosing the threshold parameter k can be difficult, and the estimator can be very far from the truth if an inappropriate value is chosen. Whereas the estimators of Berghaus and Bücher (2018) rely on a similar parameter b_n , and their performance can also suffer from inappropriate choices for b_n , the impact is not as severe as in case of the non-parametric estimator of Cai (2019).

The parametric filters have a number of advantages compared to the other approaches. First, by having an explicit assumption for the data generating process, it can deliver accurate estimations if the data generating process has a closed form solution for the extremal index, including for processes with a time-varying extremal index. Compared to the rolling horizon estimator, the parametric filters are not as reliant on a slow-changing nature of the extremal index, and thus remain to work well when the extremal index is fluctuating quite strongly. Further, its parameters are estimated by maximum likelihood, and it is hence not necessary to finetune the (hyper)parameters. The only exception is κ in case of the kappa-based filter. However, this parameter usually doesn't need to be large, and can be set at a point which renders a stable time-varying filter. Moreover, the filter is quite robust with regard to this parameter. Finally, the filters consider all n observations for the estimation of the parameters and hence the extremal index. The non-parametric and sliding block estimators only consider the observations above the threshold. The asymptotic normality of the non-parametric estimator of Cai (2019) is based on k , not on n . Similarly, the asymptotic normality of the non-parametric estimator of Berghaus and Bücher (2018) is based on the number of blocks. As such, upon correct specification, the parametric filters are more accurate. The main drawback of this approach is that the assumption of a certain parametric model entails the risk of model misspecification, in which case the

performance may deteriorate significantly. Further, the optimisation happens numerically, which can be a complex problem.

Comparing the non-parametric rolling horizon estimator with its stable version, one can easily see that by choosing the horizon to match the length of a given sample, one obtains the same estimator. As the stable version in general contains more data, and hence a higher n and k , its asymptotic properties are better in case the process has a stable extremal index. Moreover, choosing for a rolling horizon estimator also leads to the introduction of another parameter, which controls the horizon length. However, the rolling horizon estimator brings the advantage that the assumptions with regard to stability are considerably weakened. As such, the risk wrongly assuming a stable extremal index reduces as well. It doesn't reduce completely, as the extremal index should be approximately stable on the horizon length. So, the rolling horizon parameter regulates the speed of changes in the extremal index assumed.

3 || SIMULATION

It is hard to compare the performance of the different stable and time-varying estimators for the extremal index on a real dataset, as the extremal index is typically not known. Therefore, this chapter aims to give some insights in the conditions which affect the performance of the respective approaches by assessing how the estimators perform in terms of MSE on a number of models with known extremal index is. The models are chosen in a way to have differing underlying assumptions. The first section describes the models used in the simulation study, and is followed by a description of the overall simulation setup. Then, the results are presented. The chapter ends with a discussion of the results and their implications.

3.1 || SIMULATION MODELS

To compare the performance of the different estimators for extremal index θ , we compare it for several models for which θ is known. These take the form of an AR(1) model with Student t-distributed error terms (ARt(1)) and an GARCH(1,1) model, both with fixed θ , and an AR(1) model with Student t-distributed error terms and a MaxAR model with time-varying θ_t .

The inclusion of models which take the form of an AR(1) model, both with stable θ and time-varying θ_t allows to assess the performance of the parametric modelling approach when the model is (approximately) correctly specified. One can expect that this approach should perform relatively well in this case. The inclusion of GARCH(1,1) and MaxAR(1) generated data gives insight in the impact of model misspecification on the performance of the parametric approach. Further, by including both data generating processes with fixed and time-varying components, it is possible to assess the impact of on the one hand having an estimator which assumes stable tail dependence when the data generating process varies in time, and on the other hand of having an estimator which allows for time-varying behaviour when the data generating process remains fixed throughout time. Finally, the GARCH(1,1) model allows for assessing the impact of a series for which the $D^{(d)}(u_n)$ does not hold for any finite d .

PROPOSITION 6. *For any GARCH(1,1) model, defined by*

$$y_t = \sigma_t \epsilon_t, t \in \mathbb{Z} \quad (3.1)$$

$$\sigma_{t+1}^2 = \omega + \alpha \sigma_t^2 \epsilon_t^2 + \beta \sigma_t^2, t \in \mathbb{Z}; \quad (3.2)$$

the $D^{(d)}(u_n)$ does not hold for any finite d .

Proof. It is sufficient to look at the proof of proposition 6.3 in Cai (2019), which proves that the $D^{(d)}(u_n)$ does not hold for any finite d in case of an ARCH model. This proof is based on a proposition in Ehlert et al. (2015). To this end, Cai (2019) considers equations (6.2) and (6.3) of Ehlert et al. (2015), where it assumes $\delta_1 = \beta_1 = 0$. The only difference between this model and the GARCH(1,1) model is that in case of the GARCH(1,1) model, $\beta_1 \neq 0$. However, this doesn't violate the form of equations (6.2) and (6.3) in Ehlert et al. (2015). As such, the proof for proposition 6.3 in Cai (2019) is valid for the GARCH(1,1) model when the assumption $\beta = 0$ is dropped.

□

ARt(1) data with fixed autoregressive parameter

The data generating process is given by:

$$y_{t+1} = 0.80y_t + \epsilon_t, \quad \epsilon_t \sim ii\tau(3) \quad (3.3)$$

with $y_1 \sim ii\tau(3)$. For this model, θ is given by $1 - |\phi|^\lambda = 1 - |0.8|^3 = 0.488$ (Mikosch, 2013).

An example run is shown in figure I.

GARCH(1,1) data with fixed parameters

The data generating process is given by:

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 1) \quad (3.4)$$

$$\sigma_{t+1}^2 = 2 \times 10^{-5} + 0.25y_t^2 + 0.70\sigma_t^2 \quad (3.5)$$

In this model, ϵ_t is identically and independently distributed. For this model, $\theta = 0.447$ (Laurini and Tawn, 2012).

A sample is depicted in figure II.

ARt(1) data with time-varying autoregressive parameter

This data generating process is given by:

$$y_{t+1} = \tanh(\phi_t)y_t + \epsilon_t, \quad \epsilon_t \sim ii\tau(3) \quad (3.6)$$

$$\phi_t = 0.6 - \frac{1}{3}\cos\left(\frac{t}{1000}\right) \quad (3.7)$$

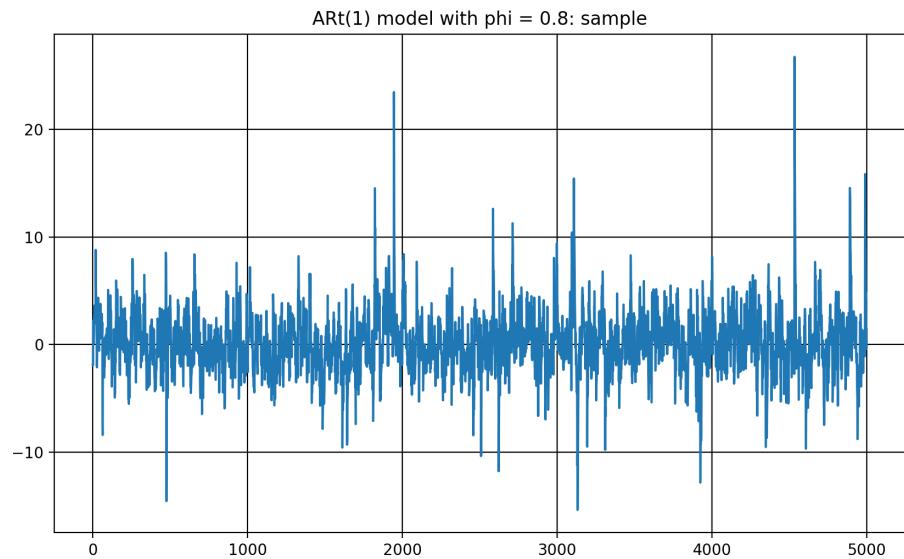


FIGURE I
Simulation sample generated by ARt(1) model with $\phi = 0.8$ and $\lambda = 3$

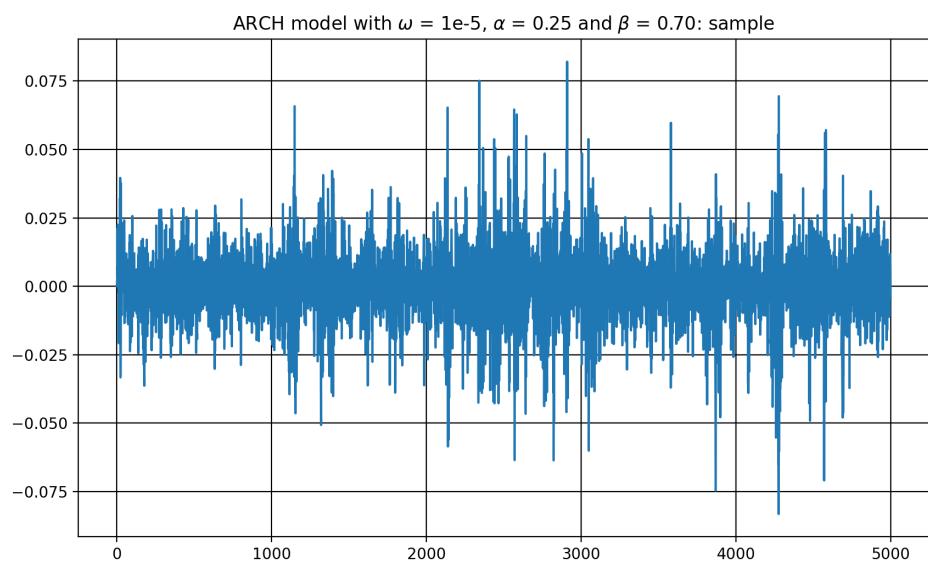


FIGURE II
Simulation sample generated by GARCH(1,1) model with $\omega = 1e - 5$, $\alpha = 0.25$, and $\beta = 0.70$

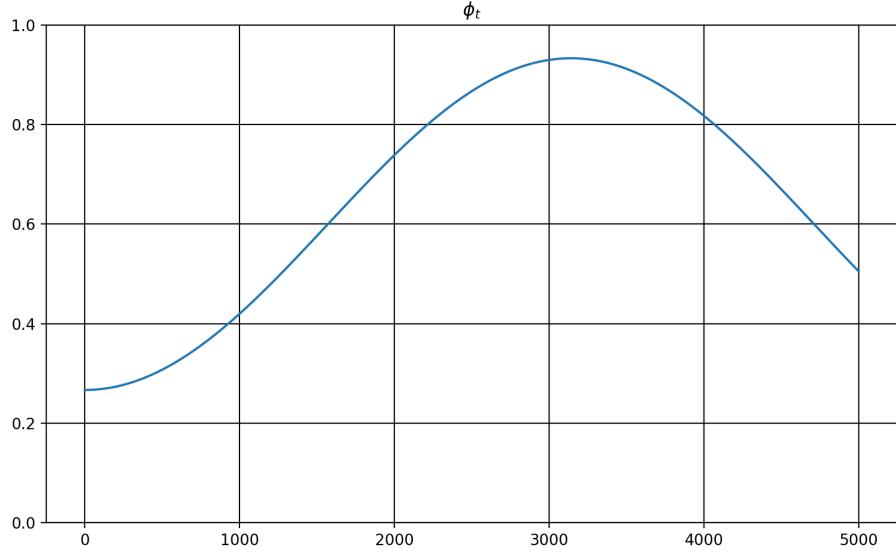


FIGURE III

Exogenous autoregressive parameter $\tanh(\phi_t)$, used in the time-varying ARt(1) and MaxAR model

with $y_1 \sim i\tau(3)$. For y_t , ϕ_t is a exogenous time-varying parameter which is changing slowly over time. It remains the same for every simulation run. $\tanh(\phi_t)$ is depicted graphically in figure III. For this model, $\theta(t)$ is given by $1 - |\tanh(\phi_t)|^\lambda = 1 - |\tanh(\phi_t)|^3$ (Mikosch, 2013). $\theta(t)$ is depicted in figure IV.

Finally, figure V depicts a single sample run of size 5000.

MaxAR data with time-varying parameter

This data generating process is given by:

$$y_{t+1} = \max(|\tanh(\phi_t)|y_t, \epsilon_{t+1}), \quad \epsilon_t \sim i.i.inverseWeibull(1, 1) \quad (3.8)$$

$$\phi_t = 0.6 - \frac{1}{3}\cos\left(\frac{t}{1000}\right) \quad (3.9)$$

with $y_1 = \frac{\epsilon_1}{1-\phi_1}$. For y_t , ϕ_t is a exogenous time-varying parameter which is changing slowly over time. It is depicted in figure III. The error terms ϵ_t are generated by an inverse Weibull distribution with parameters $\alpha = 1$ and $\beta = 1$. The cdf of the inverse Weibull distribution is $F(x) = \exp(-(\frac{x}{\alpha})^\beta)$ (De Gusmao et al., 2011). For this model, $\theta(t) = 1 - |\tanh(\phi_t)|$ (Cai, 2019).

The ground truth for $\theta(t)$ is depicted in figure VI, whereas figure VII shows a sample generated by the described MaxAR model.



FIGURE IV

Ground truth for θ in ARt(1) model with time-varying autoregressive parameter and $\lambda = 3$

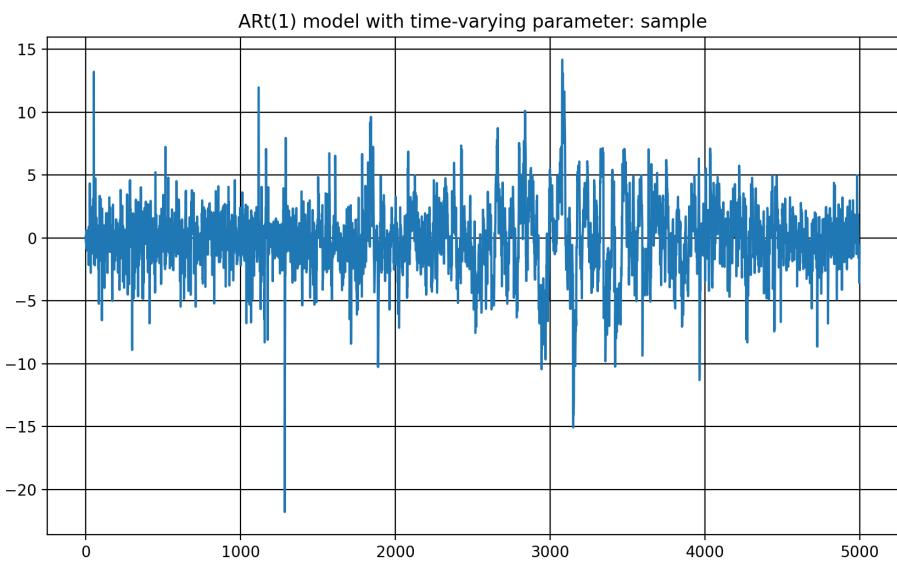


FIGURE V

Simulation sample generated by ARt(1) model with time-varying autoregressive parameter and
 $\lambda = 3$



FIGURE VI

Ground truth for θ in MaxAR model with time-varying autoregressive parameter

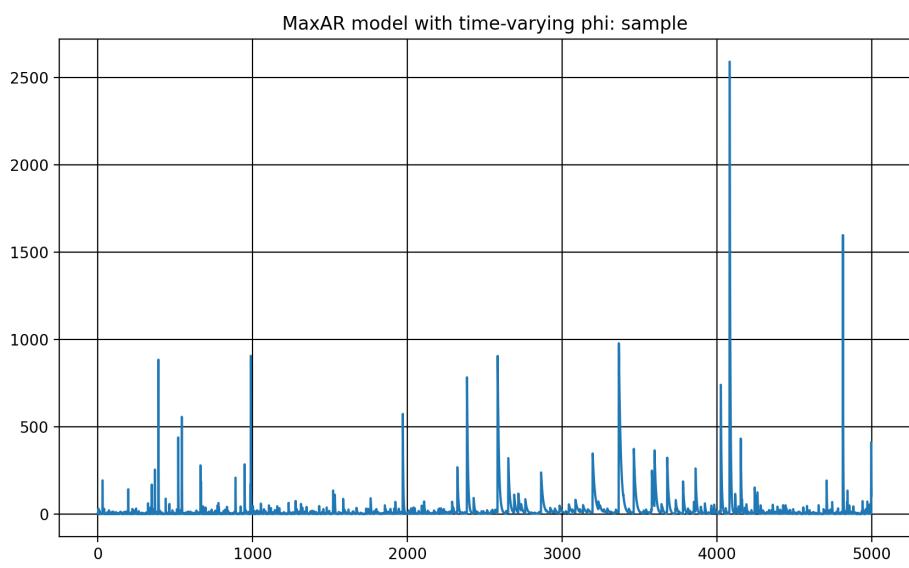


FIGURE VII

Simulation sample generated by MaxAR model with time-varying autoregressive parameter

3.2 || SIMULATION SETUP

For each of the data generating processes, 1000 samples with a sample size of 5000 are generated. Then, the extremal index θ is estimated by the following estimators:

- The stable non-parametric estimator $\hat{\theta}_n$ described in equation (2.6) with $d^u = 10$, k determined in a way to provide an as good fit with the ground truth as possible.
- The rolling horizon non-parametric estimator $\hat{\theta}_p(t)$ described in (2.10) with $d^u = 10$, horizon length p and k determined in a way to provide an as good fit with the ground truth as possible.
- The sliding block estimator $\hat{\theta}^{B,sl}$ described by (2.8) with parameter b_n determined in a way to provide an as good fit with the ground truth as possible.
- The sliding block estimator $\hat{\theta}^{N,sl}$ described in (2.9) with parameter b_n determined in a way to provide an as good fit with the ground truth as possible.
- The robust parametric AR(1) filter described by (2.21). The parameters ω , α , β and λ are estimated by maximum likelihood.
- The smoothed parametric AR(1) filter based on the last κ observations, described by (2.18), with $\kappa = 30$. The parameters ω , α , β and λ are estimated by maximum likelihood.

The finetuning of parameters poses some challenges. In particular, for the non-parametric estimators based on Cai (2019), and to a lesser extent the sliding block estimators described in Berghaus and Bücher (2018), the outcome of the estimator depends greatly on the parameters used. In order to make the comparison as fair as possible, the parameters are set in a way which is either depending on the data, or such that the estimators will yield results which are on average close to the ground truth. The disadvantage of this latter method is that this cannot be done when the ground truth is not known, and consequentially, the estimators may not exhibit the same performance when applied in empirical applications.

The length of the estimated $\hat{\theta}(t)$ depends on the estimator. The rolling horizon estimator $\hat{\theta}_p(t)$ described in (2.10) considers the period $[t - \frac{p}{2}, t + \frac{p}{2}]$ for estimation at time t . In a series of infinite length, this doesn't pose problems, but in finite samples, the estimator cannot be estimated for $t < \frac{p}{2}$ and $t > \frac{p}{2}$. As such, the estimator is only defined for period $t \in T = [t - \frac{p}{2}, t + \frac{p}{2}]$. A similar problem exists for the parametric AR(1) modelling approach, as all finite series have to be initialised. Therefore, the filtered estimate for the first observation is undefined. Moreover, the filter also needs to be initialised, and may be inaccurate for the first observations.

For each of these estimators, the mean square error is calculated for each run i as follows:

$$MSE(i) = \frac{1}{|T|} \sum_{t \in T} (\theta(t) - \hat{\theta}(t))^2 \quad (3.10)$$

where T is the domain for which $\hat{\theta}$ is defined and $|T|$ the number of elements in domain T . Based on these MSE's per run, an average MSE is then calculated:

	samples	sample size	$\hat{\theta}$ length	MSE	std MSE
Robust parametric ARt(1) filter	1000	5000	4999	0.0006	0.0006
Parametric ARt(1) filter, kappa: 30	1000	5000	4999	0.0005	0.0006
Stable non-parametric estimator, k: 50	1000	5000	5000	0.0050	0.0068
Rolling horizon non-parametric estimator, k: 24	1000	5000	3000	0.0105	0.0080
Sliding blocks Y-estimator, bn: 400	1000	5000	5000	0.0146	0.0309
Sliding blocks Z-estimator, bn: 400	1000	5000	5000	0.0147	0.0314

TABLE I

Simulation output for stable ARt(1) data: samples, sample size, length of $\hat{\theta}$ sequence estimated, average MSE and standard deviation of the MSE

$$\text{average } MSE = \frac{1}{n} \sum_{i=1}^n MSE(i) \quad (3.11)$$

3.3 || SIMULATION RESULTS

3.3.1 || ARt(1) data with fixed autoregressive parameter

The results of the simulation are described in table I. The parametric ARt(1) filter based on the last κ observations has the lowest average MSE, followed by the robust parametric ARt(1) filter. This is not surprising, as in this case, the model specification is correct. The non-parametric estimators follow, with the stable estimator benefiting from a larger sample size and hence better accuracy compared to the rolling horizon non-parametric estimator. The sliding block estimators have the highest average MSE for this data generating process. Moreover, the sliding block estimators have the highest standard deviation.

Figure VIII depicts the performance of the different estimators graphically. Although the parametric approaches lead to very low average MSEs, the estimated parameters tend to lead to filters which wrongly assume the existence of some time-varying behaviour, even though the filter specification allows for the estimation of a process without time-varying behaviour. This is potentially the case because the error term comes from a Student-t distribution with 3 degrees of freedom, leading to a lot of noise.

3.3.2 || GARCH(1,1) data with fixed parameters

The results for the simulation are summarised in table II. As can be seen, the sliding block estimators obtain the lowest MSE in this case.

Figure IX depicts the performance of the different estimators graphically. It is quite remarkable that the parametric ARt(1) filters almost always estimate θ to be very close to 0. This is because the lack of autoregressive component in the GARCH data leads the filter to consider the data to be generated by a (near) unit root process with very high noise. The misspecification leads to a high average MSE for the parametric ARt(1) filters in this case, followed by the rolling horizon estimator. When comparing several curves for $\hat{\theta}$ in function of k , it was noted that no

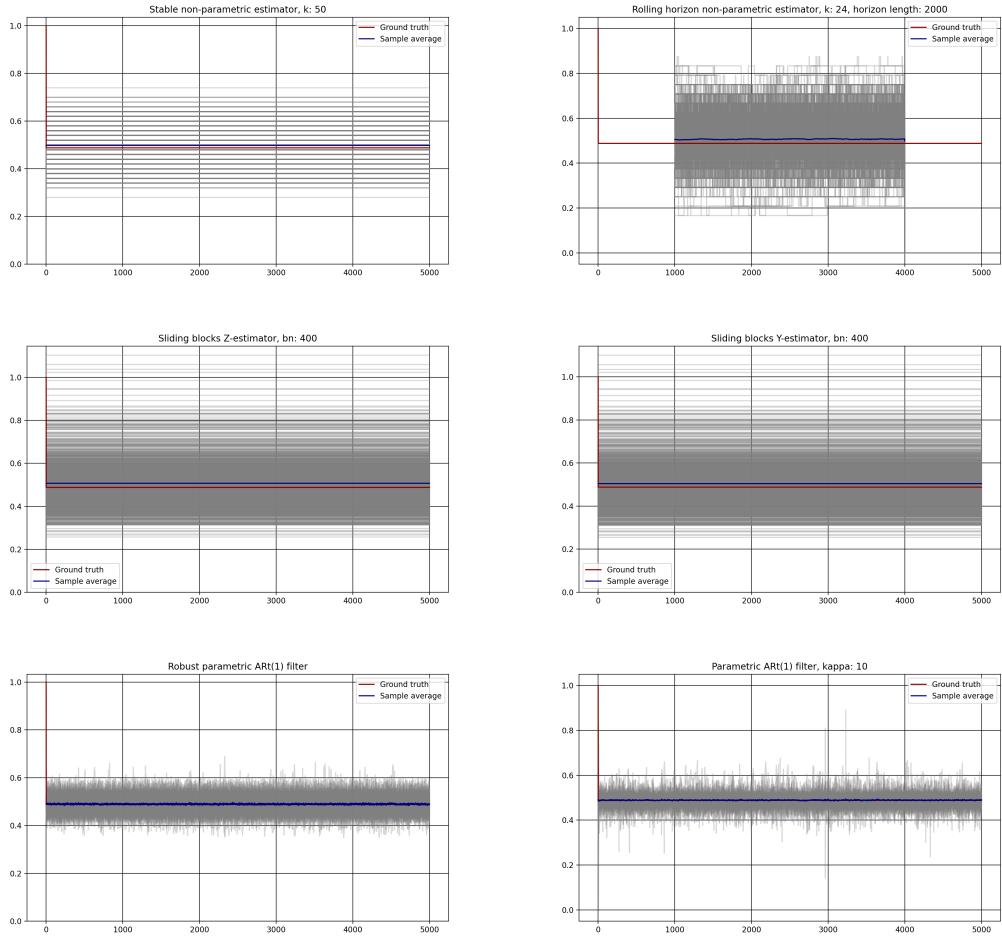


FIGURE VIII

Simulation output for ARt(1) data with $\phi = 0.8$ and $\lambda = 3$: ground truth, sample average and density for different estimators. The estimators are: stable non-parametric estimator with $k = 250$ (above left), rolling horizon non-parametric estimator with $k = 100$ and $p = 2000$ (above right), sliding block estimator $\hat{\theta}_n^{B,sl}$ with $b_n = 20$ (middle left), $\hat{\theta}_n^{N,sl}$ with $b_n = 20$ (middle right), robust parametric ARt(1) filter (under left), parametric ARt(1) filter based on $\kappa = 30$ last observations (under right)

	samples	sample size	$\hat{\theta}$ length	MSE	std MSE
Robust parametric ARt(1) filter	1000	5000	4999	0.1998	3.83369e-06
Parametric ARt(1) filter, kappa: 30	1000	5000	4999	0.1996	0.0071
Stable non-parametric estimator, k: 250	1000	5000	5000	0.0130	0.0136
Rolling horizon non-parametric estimator, k: 120	1000	5000	3000	0.0330	0.0183
Sliding blocks Y-estimator, bn: 80	1000	5000	5000	0.0053	0.0076
Sliding blocks Z-estimator, bn: 80	1000	5000	5000	0.0065	0.0089

TABLE II

Simulation output for GARCH(1,1) data: samples, sample size, length of $\hat{\theta}$ sequence estimated, average MSE and standard deviation of the MSE

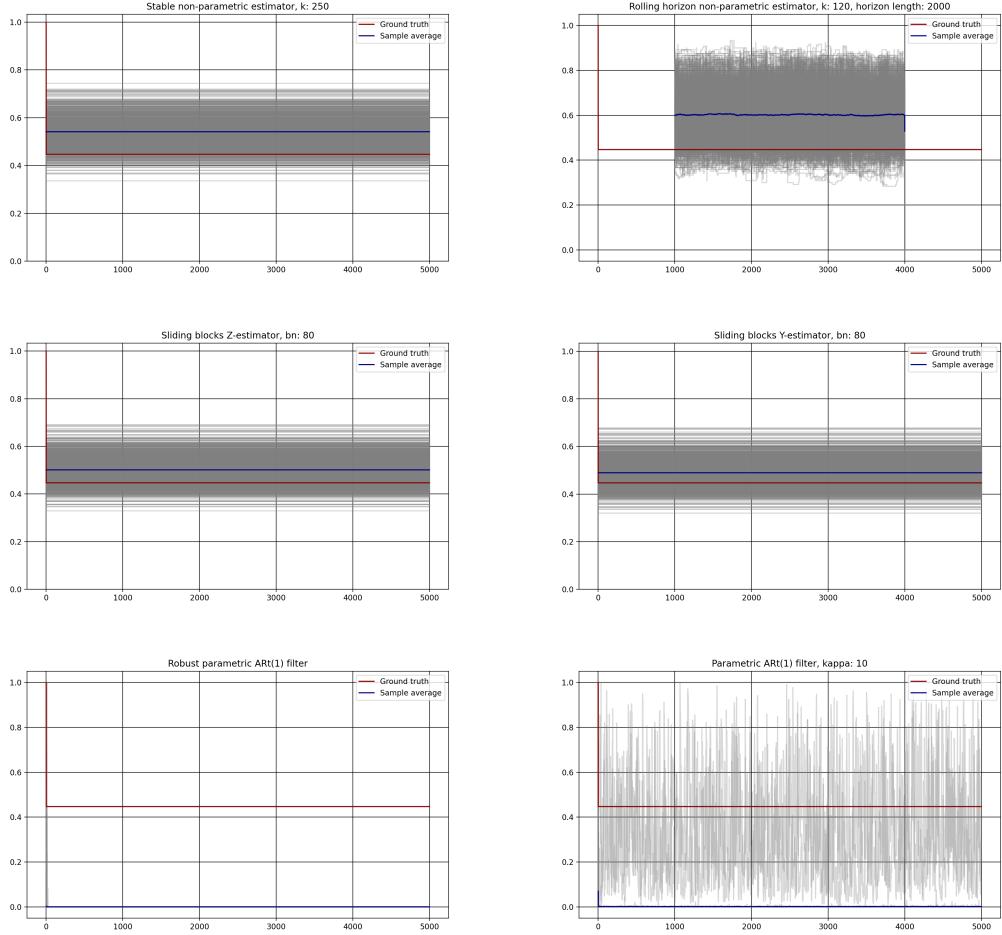


FIGURE IX

Simulation output for GARCH(1,1) data with $\omega = 1e - 5$, $\alpha = 0.25$, and $\beta = 0.70$: ground truth, sample average and density for different estimators. The estimators are: stable non-parametric estimator with $k = 250$ (above left), rolling horizon non-parametric estimator with $k = 100$ and $p = 2000$ (above right), sliding block estimator $\hat{\theta}_n^{B,sl}$ with $b_n = 20$ (middle left), $\hat{\theta}_n^{N,sl}$ with $b_n = 20$ (middle right), robust parametric ARt(1) filter (under left), parametric ARt(1) filter based on $\kappa = 30$ last observations (under right)

	samples	sample size	$\hat{\theta}$ length	MSE	std MSE
Robust parametric ARt(1) filter	1000	5000	4999	0.0126	0.0060
Parametric ARt(1) filter, kappa: 30	1000	5000	4999	0.0107	0.0031
Stable non-parametric estimator	1000	5000	5000	0.1296	0.0408
Rolling horizon non-parametric estimator, k = 25, p = 500	1000	5000	4500	0.0267	0.0099
Sliding blocks Z-estimator	1000	5000	5000	0.1712	0.0307
Sliding blocks Y-estimator	1000	5000	5000	0.1780	0.0317

TABLE III

Simulation output for time-varying ARt(1) data: samples, sample size, length of $\hat{\theta}$ sequence estimated, average MSE and standard deviation of the MSE

curve provided a good fit. This is potentially the case because $D^{(d)}(u_n)$ does not hold for any finite d in case of the GARCH(1,1) model.

3.3.3 || ARt(1) data with time-varying autoregressive parameter

The results for the simulation are summarised in table III. The average MSE is the lowest for the parametric ARt(1) filtering approaches. This is not very surprising, as the parametric model is approximately correctly specified in this case. The rolling horizon non-parametric estimator is the second best performing, as the benefit of having a time-varying estimator clearly outweighs the disadvantage of a reduced sample size. Of the estimators assuming a single extremal index for the complete time series, the stable non-parametric estimator has a lower MSE than the sliding blocks estimators, who perform in a relatively similar manner.

Figure X depicts the performance of the different estimators graphically. On this figure, each of the estimators' sample average $\hat{\theta}$ is depicted, as well as its density. The intensity of the gray scale informs about the number of estimations at a certain point. As can be seen, both the parametric ARt(1) approaches and the rolling horizon estimator follow the trend in the time-varying θ_t to some extent. As described earlier, this is not so surprising given the approximately correct model specification of this parametric approach. Further, although the rolling horizon estimator varies over time, it assumes a stable tail process for the horizon it is calculated on for every time t it is defined. As this is approximately correct in processes with slow time-varying behaviour, it delivers a performance which is relatively good.

3.3.4 || MaxAR data with time-varying parameter

The results for the simulation are summarised in table IV. The rolling horizon non-parametric estimator has the lowest average MSE. This implies that the time-varying nature of the data is an important aspect. Moreover, unlike in the case of the GARCH(1,1) data, the data generating process has some autoregressive form, and the general trend is picked up by the parametric approaches, although the time-varying κ -filter seems to have a bias and overestimate θ_t . This is probably because there is not always an autoregressive behaviour available (namely when the error term is larger than the autoregressive component). As a consequence, the parametric approach will estimate the autoregressive component to be smaller, which leads a larger $\hat{\theta}_t$. However, the parametric filter approaches still obtain a lower MSE than the estimators which

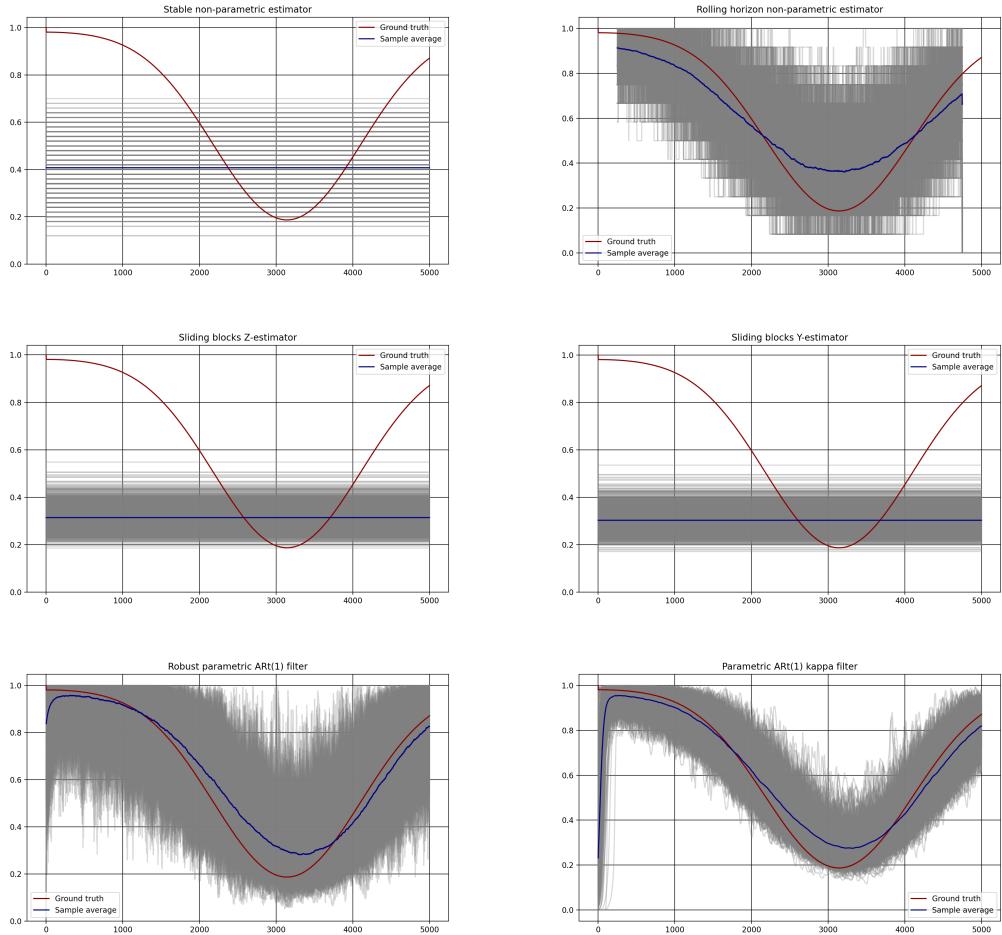


FIGURE X

Simulation output for time-varying ARt(1) data: ground truth, sample average and density for different estimators. The estimators are: stable non-parametric estimator with $k = 50$ (above left), rolling horizon non-parametric estimator with $k = 12$, $p = 500$ (above right), sliding block estimator $\hat{\theta}_n^{B,sl}$ with $b_n = 100$ (middle left), $\hat{\theta}_n^{N,sl}$ with $b_n = 100$ (middle right), robust parametric ARt(1) filter, kappa-based filter with $\kappa = 30$ (under)

	samples	sample size	$\hat{\theta}$ length	MSE	std MSE
Robust parametric ARt(1) filter	1000	5000	4999	0.0441	0.0131
Parametric ARt(1) filter, kappa: 10	1000	5000	4999	0.0468	0.0247
Stable non-parametric estimator, k: 250	1000	5000	5000	0.0984	0.0499
Rolling horizon non-parametric estimator, k = 25, p = 500	1000	5000	4500	0.0070	0.0026
Sliding blocks Y-estimator, bn: 20	1000	5000	5000	0.0551	0.0023
Sliding blocks Z-estimator, bn: 20	1000	5000	5000	0.0499	0.0010

TABLE IV

Simulation output for time-varying MaxAR data: samples, sample size, length of $\hat{\theta}$ sequence estimated, average MSE and standard deviation of the MSE

assume a stable extremal index.

3.4 || DISCUSSION OF SIMULATION RESULTS

Based on the simulations, a few interesting observations can be made. First, there is no estimator or filter which obtains minimal average MSE across all scenarios. Whereas the parametric filter approaches perform very well in case of (approximately) correct model specification. They also use n observations to estimate the parameters by maximum likelihood and hence the extremal index, whereas the estimation by non-parametric approaches is based on k observations. However, they struggle in absence of autoregressive behaviour in the data. Moreover, incorrect model specification can lead to biases. However, even in case of incorrect model specification, parametric filtering approaches often seem to get the overall long term time-varying behaviour right. For example, in the GARCH(1,1) data, no or very little time-varying behaviour took place, whereas the kappa-based filter correctly identified the time-varying behaviour of the MaxAR data, apart from a bias. On the other hand, the stable sliding block estimators and the stable non-parametric estimator struggle in case of processes with time-varying extremal dependency. This can be more severe than model misspecification, as the parametric filtering approaches were able to have a lower average MSE for the extremal index of the time-varying MaxAR model, despite suffering from model misspecification. The case of the rolling horizon is somewhat particular, as it is usually not the best performing estimator (with the exception of the time-varying MaxAR process). However, it is the most robust of all approaches, as it tends to get the general shape of the time-varying behaviour of θ right. Depending on the situation at hand, increasing or reducing the horizon length can further improve the accuracy of this approach.

Thus, all approaches assume some underlying model specification, although some have stricter assumptions than others: the ARt(1) filters assume a specific parametric model, the sliding block estimators are based on assumptions regarding the form of the tail process and the stability of the extremal index, and the non-parametric estimators assume a stable extremal index as well, and that some conditions are met (stable tail process and $D^{(d)}(u_n)$ condition). Whereas some of these conditions are broader than others, it is quite easy to come up with counterexamples for each of the approaches where the conditions are not met.

Further, as described before, this simulation aimed at giving a fair comparison between the approaches, and as such finetuning parameters were set in a way that the approaches would

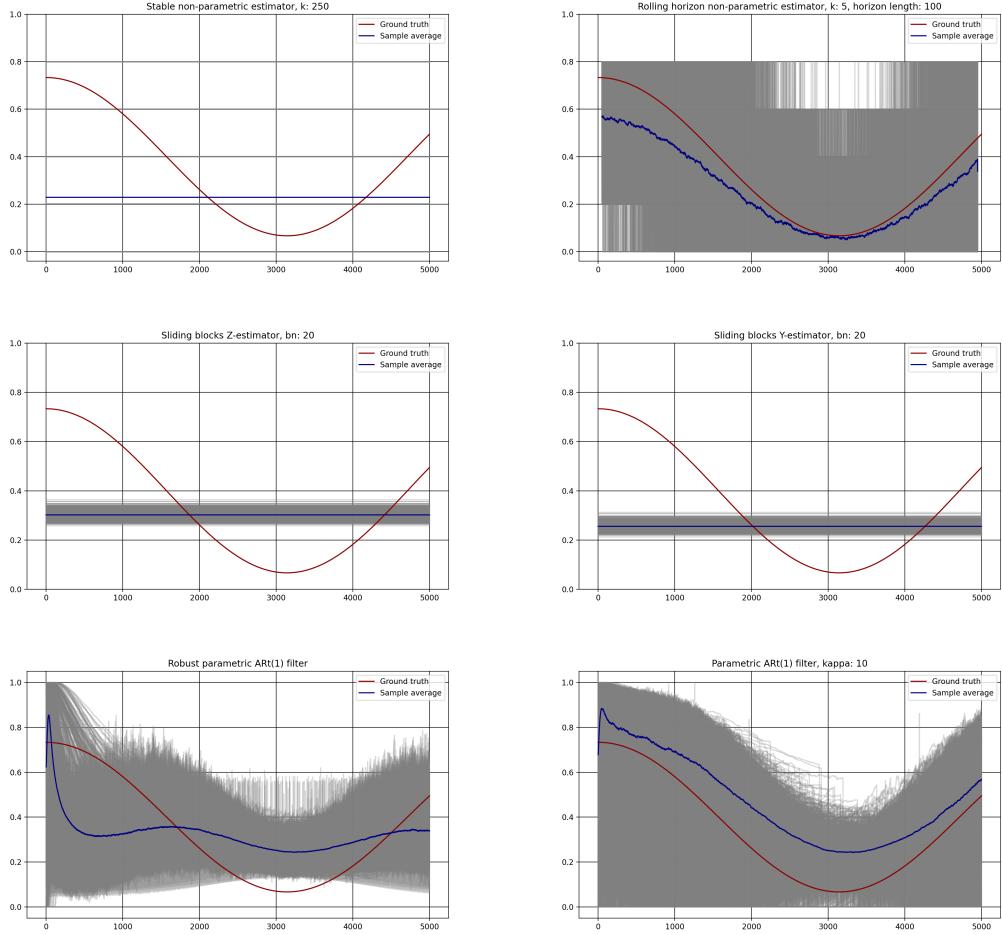


FIGURE XI

Simulation output for time-varying MaxAR data: ground truth, sample average and density for different estimators. The estimators are: stable non-parametric estimator with $k = 250$ (above left), rolling horizon non-parametric estimator with $k = 25$ and $p = 500$ (above right), sliding block estimator $\hat{\theta}_n^{B,sl}$ with $b_n = 100$ (middle left), $\hat{\theta}_n^{N,sl}$ with $b_n = 100$ (middle right), robust parametric ART(1) filter (under left), and parametric ART(1) filter based on $\kappa = 30$ last observations (under right)

on average be close to the ground truth. This is a luxury not available in case of empirical applications. In this regard, the parametric approach offers a clear advantage, as it is not necessary to finetune parameters. This is primarily the case for the non-parametric estimators, as the sliding block estimators are more robust to wrong finetuning parameters. However, even for the sliding block estimators the impact can nonetheless be large.

4 || EMPIRICAL APPLICATION: EXTREMAL CLUSTERING OF DAILY MAXIMUM TEMPERATURES RELATIVE TO EXPECTED MAXIMUM TEMPERATURE

This Chapter studies the extremal clustering in a practical application. We consider a temperature series where annual and long term trends have been removed. It thus concerns extremes in relative sense, i.e. the difference between actual and expected daily maximum temperature given seasonal and long term trends. The extremal index refers to the extremal clustering of days with very high daily maximum temperature, relative to the expected daily maximum temperatures. As we consider this difference, it is possible that actual daily maximum temperatures in winter can be considered more extreme than daily maximum temperatures in summer which are larger in absolute value, when they deviate more from the expected temperature. Several methods are available for estimating seasonalities and long term trends, including empirical mode decomposition and polynomial regression. However, when applying these techniques to the data, it was found that detrending by polynomial did not provide a good fit to the data. The results are reported in Appendix A. We therefore proceed with detrending the observations by the empirical mode algorithm.

Another approach to overcome the issue of seasonality is to consider a selection of months, assuming that in this limited time frame the data can be considered to have no seasonality. This is done in Appendix B, which describes the estimation of the extremal index for the summer months June, July and August.

The Chapter starts with a description of the raw data. The raw data contains seasonalities and long term trends, and it is necessary to remove those to come to a stationary sample. To estimate the seasonality and long term trends, we use the empirical mode decomposition algorithm. This algorithm iteratively sifts oscillations from the data. We test the stationarity of the detrended series by applying advanced Dickey-Fuller tests.

We then turn to the estimation of the extremal index by the methods described in Chapter 2. For non-parametric and sliding block estimators, it is necessary to choose a set of parameters. These include k for the stable and rolling horizon non-parametric estimators, the horizon length p for the rolling horizon non-parametric estimator, and the block length b_n of the sliding block estimators. These parameters are discussed in the following sections. Then, the partial auto-correlation function is analysed in order to inform on which components to use in a parametric modelling approach. We then present the results for the different estimators. We finish with a discussion of the estimation results.

Maximum temperature (°C)	
first	1763-01-01
last	2000-12-31
count	86928
mean	16.8254
std	9.45979
min	-11
25%	8.7
50%	17.1
75%	25
max	38.3

TABLE V

Summary statistics of daily maximum temperatures measured at Brera astronomical observatory, Milan, Italy

4.1 || RAW DATA

The data consists of daily maximum temperatures spanning the period 1763-2000, measured in the Brera astronomical observatory in Milan, Italy (Klein Tank et al., 2002). The data is summarised in table V. Figure XII depicts the observations graphically, as well as the average yearly maximum temperature. On this figure, several things can be observed. First, whereas the average maximum temperature is a bit volatile in year-on-year comparison, there seems to be an upward trend in the data. This trend seems to be getting stronger in the last 20 years of the sample. Second, most observations are in a band of around 15°C above and under the yearly average daily maximum temperature, but those observations that are not in this band often seem to be lying very closely to other observations also falling outside of this band. This may indicate that indeed extremal clustering is finding place.

Upon further inspection, the period 1763-1836 seems to be showing a different behaviour than the rest of the data. It turns out that this is partly caused by a reorganisation of the observatory which started in 1835 (Maugeri et al., 2002). We therefore process the data for two distinct periods: 1763-1837 and 1838-2000. The reason to include 1836 and 1837 in the first period is that the reorganisation started in 1835, but took several years. In considering 1836 and 1837 to be part of the first period, we follow the approach of (Maugeri et al., 2002).

4.2 || REMOVING SEASONALITY AND TRENDS

Analysing extremal clustering of daily maximum temperatures using this dataset has a drawback, namely that there is a strong (annual) seasonality in the data. Furthermore, the data may contain longer term trends, and the data may not be stationary. This causes problems for the stable estimators of the extremal index, as they assume the observations to be stationary.

Therefore, we remove the seasonalities and long term trends, in the assumption that this will lead to a stationary sample. The extremal behaviour here can then be interpreted as

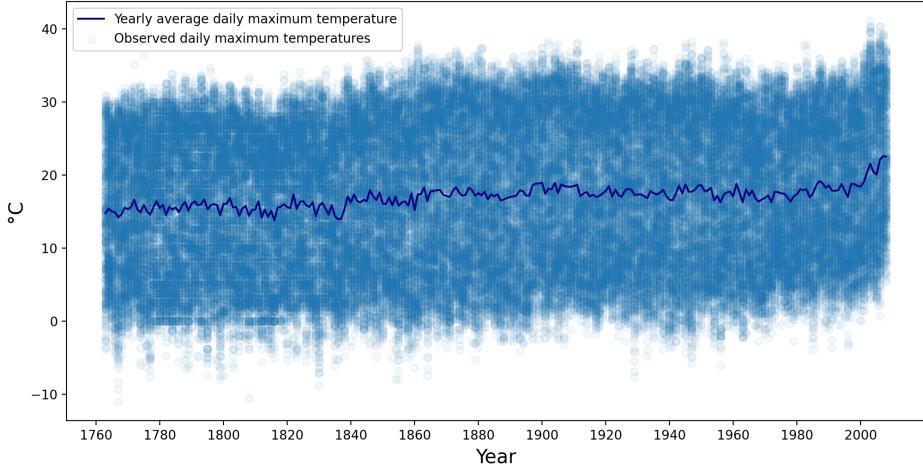


FIGURE XII

Daily maximum temperatures for the period 1763-2000, measured at Brera astronomical observatory, Milan, Italy: observations (scatter), yearly average (line).

deviations from the seasonal and trend patterns. This means that relatively warm days in spring or autumn can be considered more extreme than very hot days in the summer if they deviate more what is expected based on trends and seasonal effects. As such, the extremal index needs to be interpreted as a relative extremity, namely of the extremal dependence of daily maximum temperatures relative to the expected daily maximum temperatures given that time of year and given longer term trends.

We start with estimating the seasonalities and trends, and then remove these from the data. That way, the data can be considered to have the same mean.

We estimate the seasonality and long term trends with empirical mode decomposition. This is a method which decomposes a signal in a finite number of intrinsic mode functions (IMF), which represent different oscillations in the data (Huang et al., 1998). An IMF is a function which satisfies two conditions: the number of zero-crossings and the number of extrema is either equal or differs at most by 1, and the mean value of the cubic smoothing spline connecting all the maxima and the one connecting all the minima should be 0 at any point (Huang et al., 1998). The IMFs are sifted from the data. The sifting process starts with connecting all the maxima with a cubic smoothing spline, and connecting the minima as well by a different cubic smoothing spline. Then, their mean is calculated, and subtracted from the data, and the whole procedure is repeated on the residual. The sifting process continues until a stopping criterion is reached, and then the result is saved as an IMF. The procedure then continues iteratively from the remainder of the data, until no more IMF can be found. The original signal is then decomposed in a number of IMFs and a residual. The Empirical mode decomposition method can be applied to non-stationary and non-linear data. It fits the purpose of decomposing data describing natural phenomena well, as natural phenomena often are dependent on several processes, which may have a different frequency. For example, amongst others, air temperature measured at a specific

IMF	Period length (years)
1	0.018140
2	0.046490
3	0.094178
4	0.194785
5	0.996540
6	0.910305
7	1.917179
8	3.568032
9	6.265246

TABLE VI

IMF period (in years) for the empirical mode decomposition of the daily maximum temperatures in the period 1763-1837, measured at Brera, Milan, Italy

weather station follows a daily pattern, an annual pattern, and longer term trends. A key advantage of the EMD method is that it is not necessary to specify the oscillation frequencies in advance.

Again, given the differences in observations as a result from the reorganisation in 1835, the empirical mode decomposition is run on the data of period 1763-1837 and on the data of period 1838-2000 separately. The results can be seen in figure XIII for the period 1763-1837, and in figure XIV the period 1838-2000 respectively. Due to the smaller observation set, the results for period 1763-1837 contain not as many IMFs as the results for period 1838-2000. In both cases, the IMF with the largest range is IMF 5, which corresponds with the annual process. Tables VI and VII detail the periods of the IMFs. In both cases, the first 4 components have periods less than one year, the first components is one year, and the other components are longer, with exception of the sixth period. We consider the trends to be consisting of IMF 5 to 9 for the period 1763-1837 and IMF 5 to 14 for the period 1838-2000. The residuals of the detrending by empirical mode decomposition are therefore defined as the sum of the first four IMFs and the residuals of the empirical mode decomposition. The trends and residuals are shown in figure XV for the period 1763-2000. The residuals are centered around zero with seemingly constant variation. When analysing a shorter timeframe, it can be seen that the empirical mode decomposition trend follows the data quite well. The residuals also seem distributed homogeneously.

Based on a visual inspection of figures XV and XVI, no obvious linear or quadratic time trend can be detected. However, it is possible that the data contains a stochastic trend. This can be tested by running an augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979). This test is an extension of the simple Dickey-Fuller test. Under the null hypothesis it is assumed the time series is non-stationary and has a unit root:

$$H_0 : \lambda = 1 \text{ vs } H_1 : |\lambda| < 1 \quad (4.1)$$

where λ is the autoregressive coefficient in the following regression:

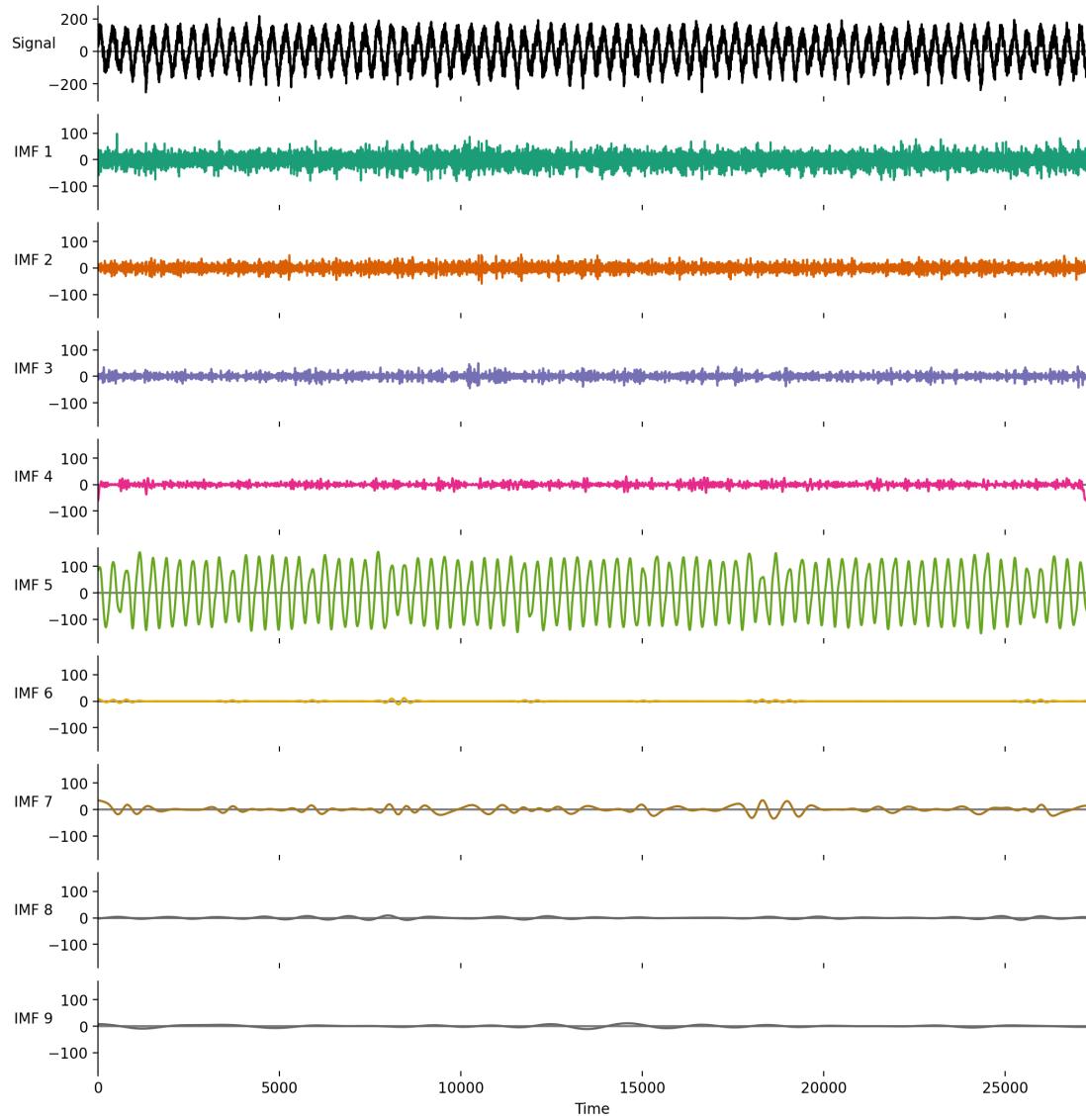


FIGURE XIII

Empirical mode decomposition: signal (first row) and IMFs (from the second row) for the daily maximum temperatures in the period 1763-1837, measured at Brera astronomical observatory, Milan, Italy

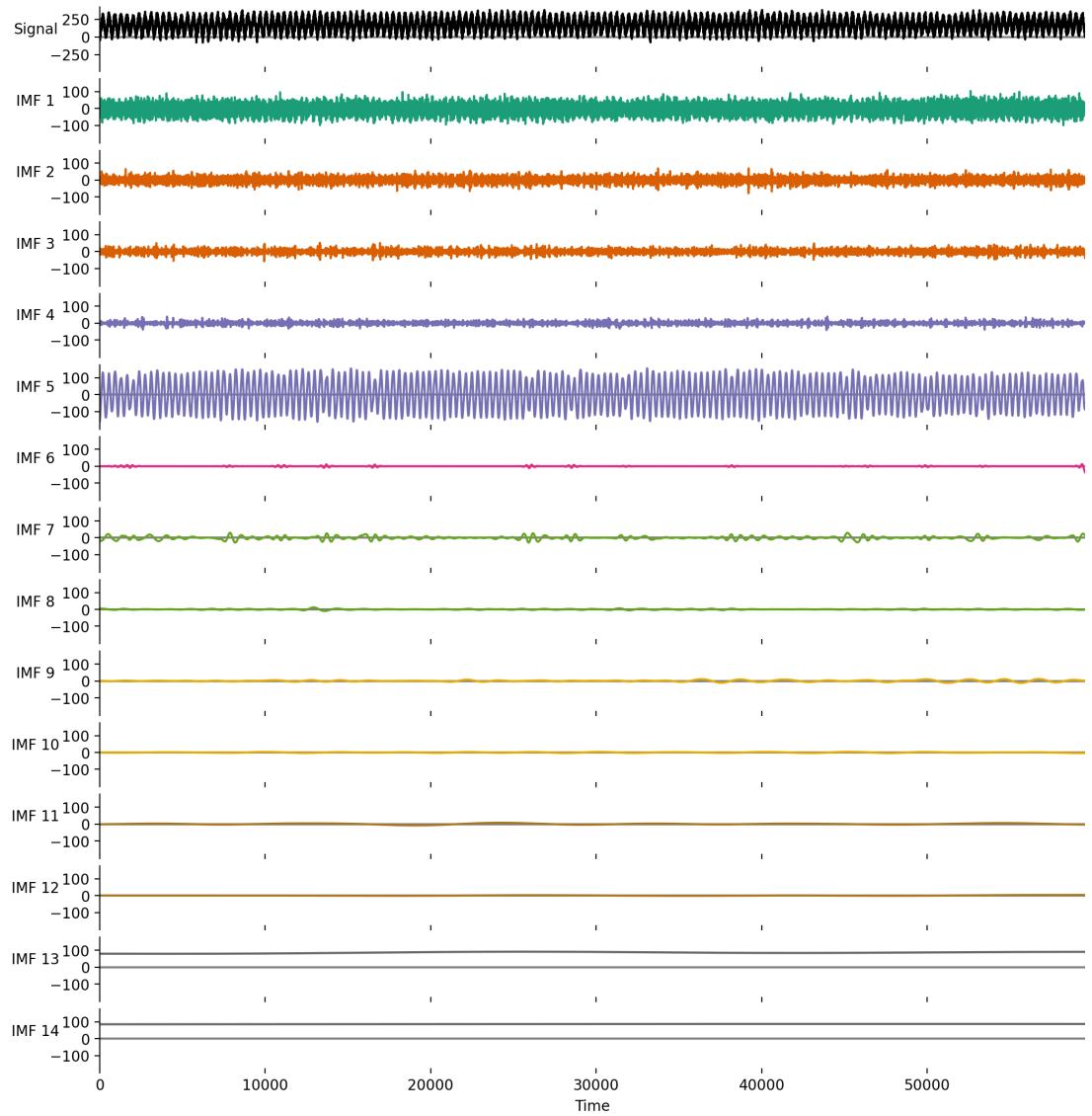


FIGURE XIV

Empirical mode decomposition: signal (first row) and IMFs (from the second row) for the daily maximum temperatures in the period 1838-2000, measured at Brera astronomical observatory, Milan, Italy

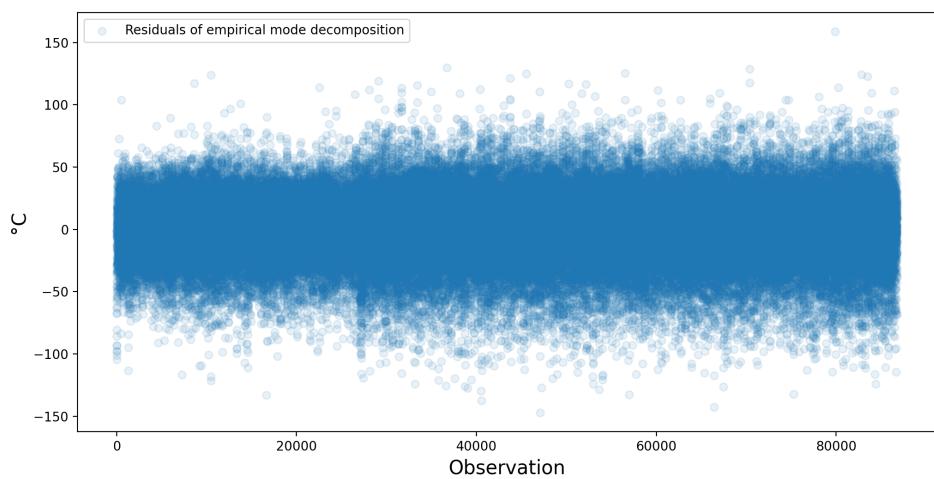
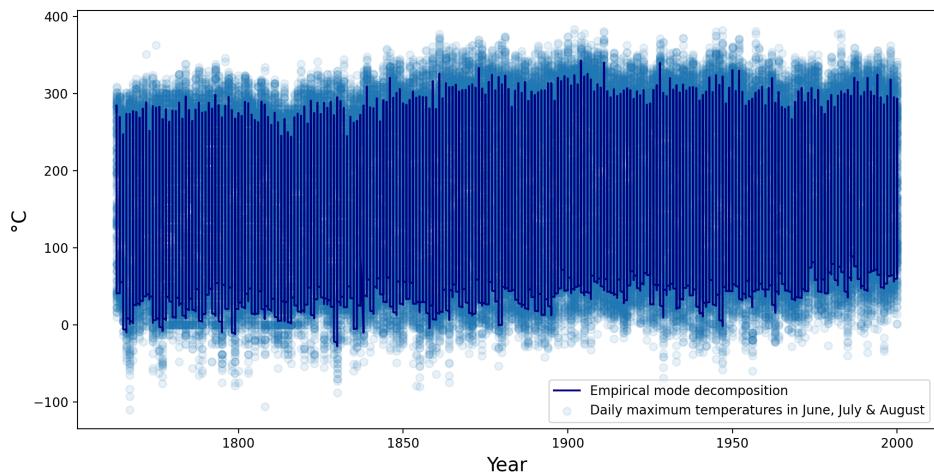


FIGURE XV

Daily maximum temperatures for 1763-2000, measured at Brera astronomical observatory, Milan, Italy, empirical mode decomposition trends with period of 1 year or more (above), values after detrending by IMFs with a period of one year or more (under)

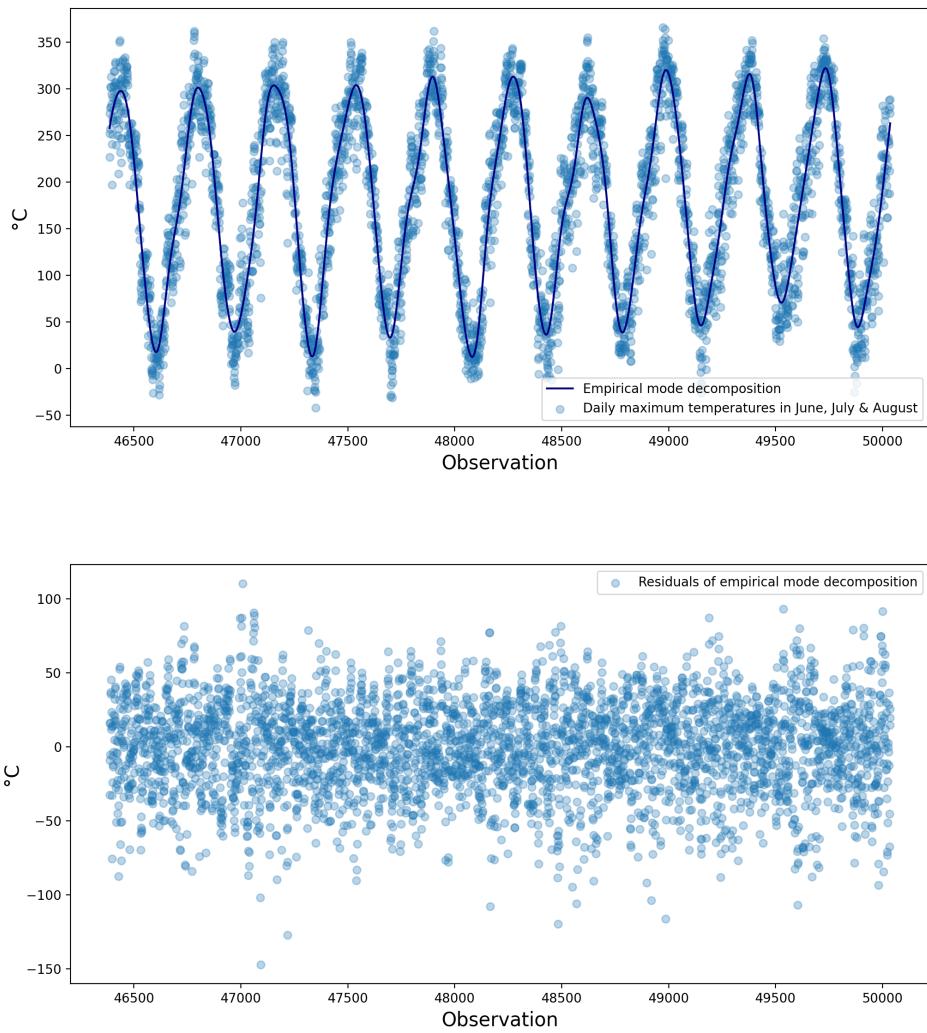


FIGURE XVI

Daily maximum temperatures for 1890-1899, measured at Brera astronomical observatory, Milan, Italy, empirical mode decomposition trends with period of 1 year or more (above), values after detrending by IMFs with a period of one year or more (under)

IMF	Period length (years)
1	0.017601
2	0.045016
3	0.092842
4	0.186926
5	1.000034
6	0.886548
7	2.065250
8	3.712266
9	5.819399
10	13.659144
11	27.286110
12	54.338715
13	unknown
14	unknown

TABLE VII

IMF period (in years) for the empirical mode decomposition of the daily maximum temperatures in the period 1838-2000, measured at Brera, Milan, Italy

$$y_t = \lambda y_{t-1} + \epsilon_t \Leftrightarrow \Delta y_t = (\lambda - 1)y_{t-1} + \epsilon_t \quad (4.2)$$

Equation (4.3) specifies a unit root without drift or a deterministic time trend. The simple Dickey-Fuller test assumes the absence of serial correlation in the error terms. In case this assumption is not met, the augmented Dickey-Fuller test can be used. This test is based on the following regression:

$$\Delta y_t = (\lambda - 1)y_{t-1} + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \epsilon_t \quad (4.3)$$

where k is chosen by the Akaike Information Criterion (AIC). The AIC is given by the following equation:

$$AIC = 2k - L(y, \psi) \quad (4.4)$$

where k is the number of parameters, and $L(y, \psi)$ the log likelihood given the data y and the parameter vector ψ (Akaike, 1974).

Again, equation (4.3) assumes a unit root without drift or deterministic time trend. These factors can be added as well, and will lead to a changed asymptotic distribution. The critical values for the test statistics are tabulated by Dickey and Fuller (1979). The results of the ADF test with constant are reported in table VIII. The unit root hypothesis is rejected and it can hence be concluded that the series do not contain a stochastic trend. As the series also look homoscedastic, we conclude that the series is stationary.

	test statistic	p-value	number of lags	observations used in regression
Deterministic components				
Constant	-55.2331	<0.00001	63	86714

TABLE VIII

ADF regression results for model with constant for the detrended daily maximum temperatures with empirical mode decomposition in the period 1763-2000 measured at Brera, Milan, Italy

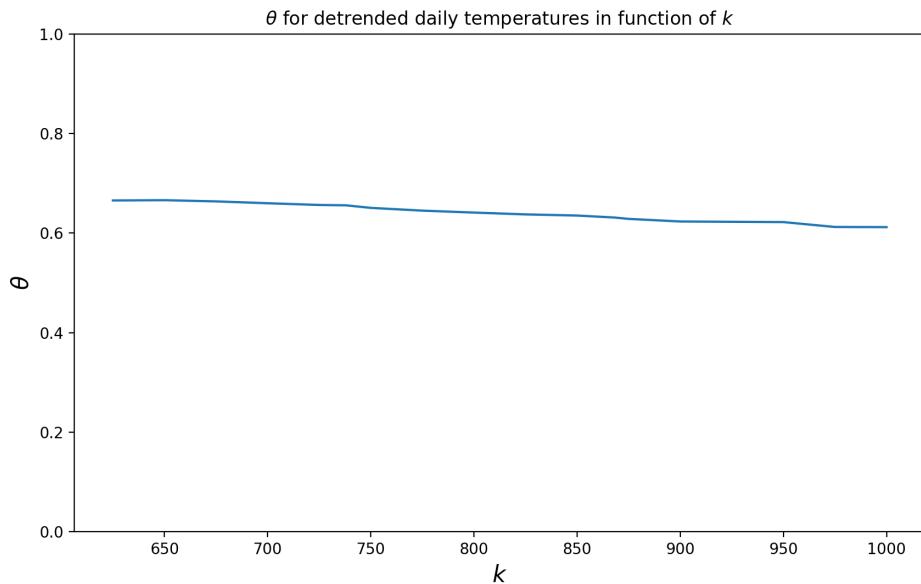


FIGURE XVII

$\hat{\theta}^{np}$ for the detrended daily maximum temperatures measured at Brera observatory, Milan, Italy in function of k for $k \in 625 - 1000$.

4.3 || STABLE NON-PARAMETRIC ESTIMATOR

We start with the stable estimator. This is depicted visually in figure XVII. As can be seen, when k is between 600 and 1000, it is quite stable. Table IX reports the outcomes of the stable non-parametric extremal index estimator for a number of values for k .

4.4 || ROLLING HORIZON ESTIMATOR

In case of the rolling horizon estimator, the horizon length and \tilde{k} need to be set. The value for \tilde{k} is determined to be in roughly the same order of magnitude $\frac{k*p}{n}$, but a little bit larger. Figure XIX shows how the rolling horizon estimator with horizon length of 30 years evolves for several values of \tilde{k} . The estimator is quite stable in when \tilde{k} is between 100 and 160.

Several horizon lengths are tried, where \tilde{k} equals $\frac{k*p*365}{n}$, where the horizon length p is expressed in years, and $k = 800$. These are shown in figure XVIII. As can be seen, the shortest horizon length (2 years) show very strong variation. However, when comparing the rolling

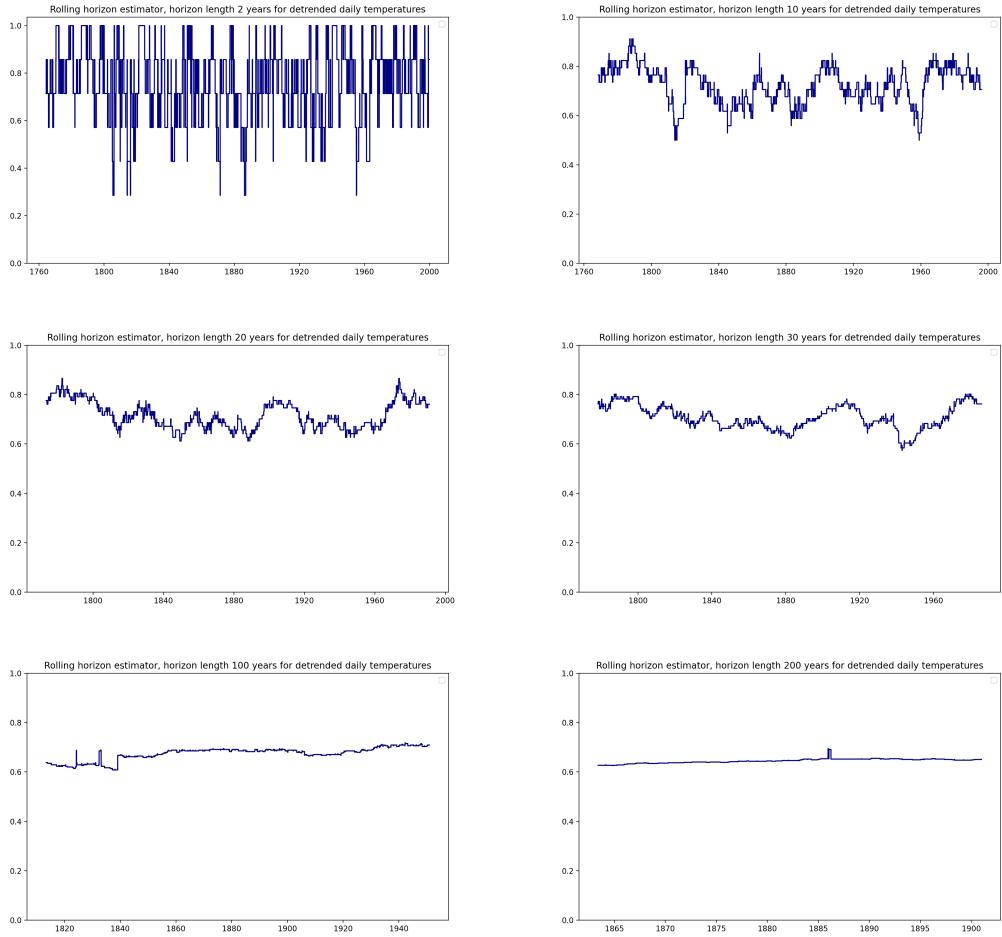


FIGURE XVIII

$\hat{\theta}_p(t)$ for the detrended daily maximum temperatures measured at Brera observatory, Milan, Italy for different horizon lengths: 2 years, 10 years, 20 years, 30 years, 100 years, 200 years (left-to-right, top-to-bottom)

horizon estimator with horizon length 10 years for the extremal index for the whole year and the extremal index for the summer months, the one for the whole year looks already smoother than the one which only took the summer months into consideration. The estimators with very long horizon lengths (100 and 200 years) display very little time-varying behaviour but seem to contain a slight upward trend. The estimators with horizon length 20 and 30 years show some time varying behaviour, but are relatively smooth functions. Table X reports the outcomes of the stable non-parametric extremal index estimator for a number of values for k , assuming a horizon length of 30 years. As is the case for the stable estimator, this range is smaller compared to the range for the rolling horizon estimator for the absolute daily maximum temperatures in June, July and August.

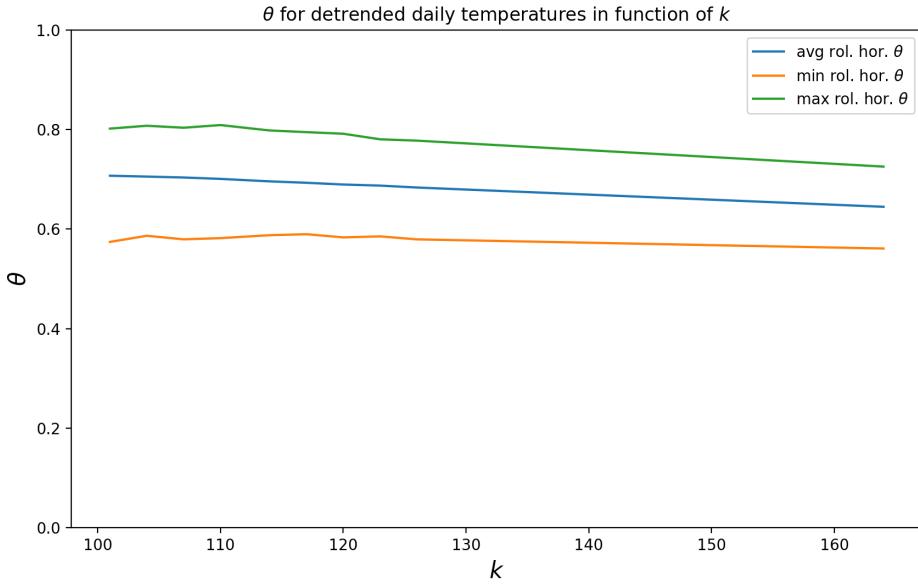


FIGURE XIX

$\hat{\theta}_p(t)$ for the detrended daily maximum temperatures measured at Brera observatory, Milan in function of k , with a horizon length of 30 years.

4.5 || SLIDING BLOCK ESTIMATORS

The sliding block estimators defined by Berghaus and Bücher (2018) show a similar trade-off between in parameter choice as the non-parametric estimator of Cai (2019). Choosing a block length which is too large or small leads to an estimator which can be far from the true value. Figure XX shows how the sliding block estimators change in function of b_n . As can be seen, the estimators slowly increase up until a block length b_n of about 2000, and then increases at a higher pace. Whereas the range of possible values is smaller compared to the non-parametric estimator, it is still quite large. To come to a choice for b_n , the same method is used as in Cai (2019), which sets $b_n = \frac{n}{k} = 108$. The estimators for a number of values for b_n obtained are reported in table XI.

4.6 || PARAMETRIC ARt(1) FILTERS

Before applying the parametric modelling approach, the autocorrelation is investigated. To this end, the partial autocorrelation function (PACF) is analysed. This function estimates the autocorrelation between two lags, while controlling for the lags in between. The PACF informs about to what extent a certain lag is informative. Figure XXI depicts the PACF graphically. As can be seen, the first lag has a relatively high correlation, but from the second lag on, the correlation after controlling for all other lags is quite close to zero. Given this characterisation, a parametric ARt(1) modelling approach can be appropriate.

The robust and kappa-based parametric ARt(1) filters are presented in figure XXII. The figures above show the filters for the whole period, and the figures under show the filters for the

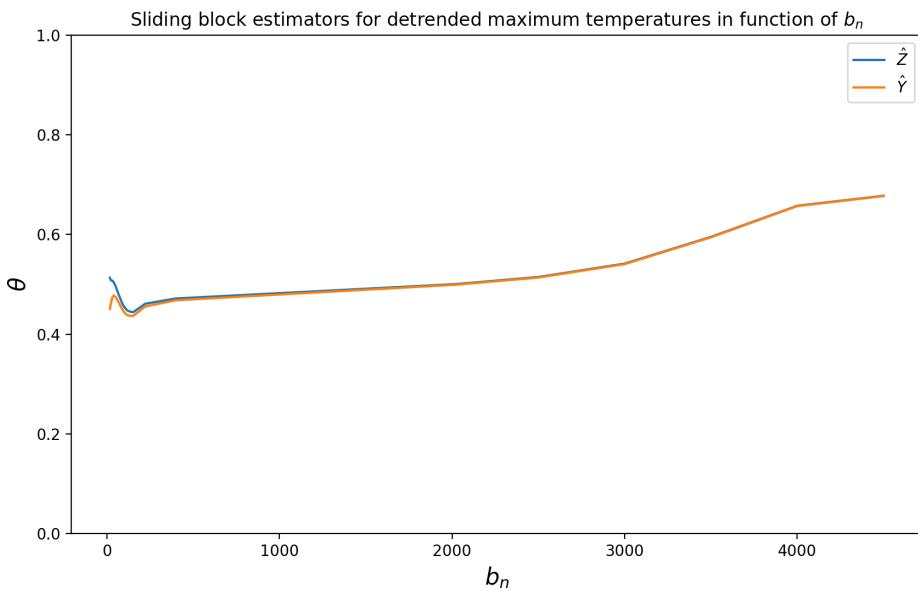


FIGURE XX

Sliding block estimators $\hat{\theta}_n^{B,sl}$ and $\hat{\theta}_n^{N,sl}$ for the detrended daily maximum temperatures measured at Brera observatory, Milan in function of b_n

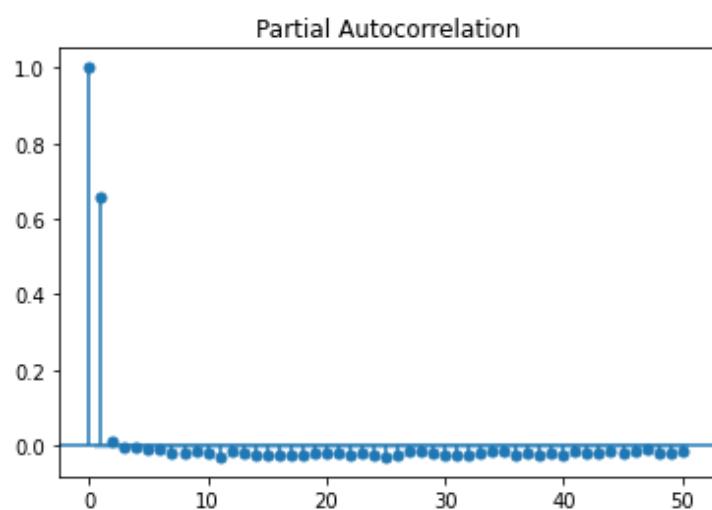


FIGURE XXI

Partial autocorrelation function for the detrended daily maximum temperatures measured at Brera observatory, Milan

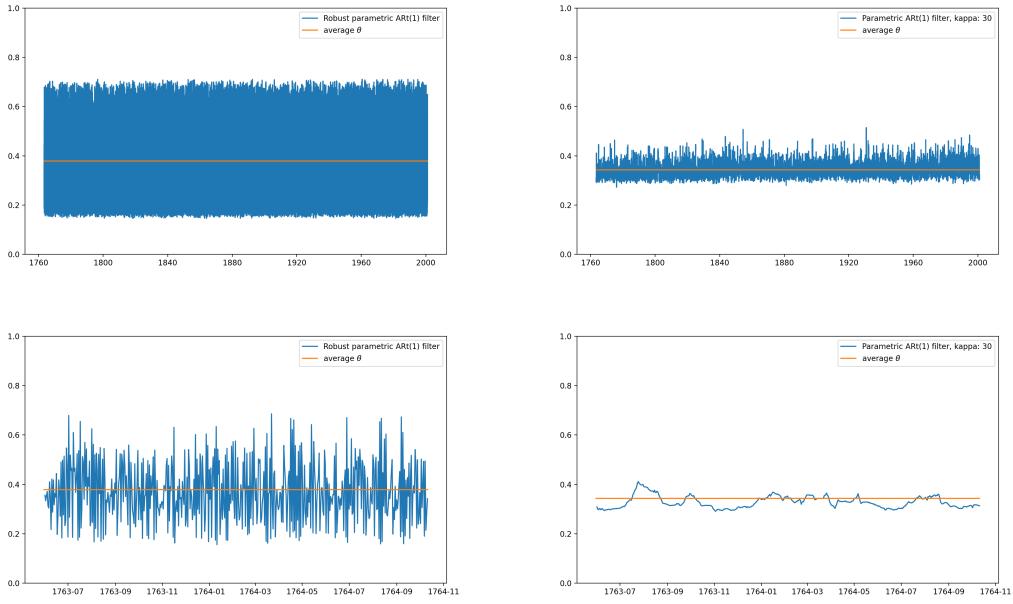


FIGURE XXII

Parametric approaches for the extremal index of the detrended daily maximum temperatures measured at Brera observatory, Milan: Robust ARt(1) filter (above left) and kappa-based filter with $\kappa = 30$ (above right) for the whole period, Robust ARt(1) filter (under left) and kappa-based filter with $\kappa = 30$ (under right) for the first two years.

first two years, to show how the filters vary on shorter time periods. The robust filter shows a lot of variation, also on short time periods. The kappa-based filter shows some variation but remains relatively smooth. A slight upward trend is noticeable in the filtered estimates. Table XII reports the average $\hat{\theta}$ and its standard deviation for the filtering approaches.

k	$\hat{\theta}$
625	0.67
800	0.64
1000	0.61

TABLE IX

$\hat{\theta}$ estimated with the stable non-parametric estimator of Cai (2019) for the detrended maximum temperatures measured at Brera observatory, Milan, Italy for different values of k .

k	$\hat{\theta}$	min	max
76	0.74	0.62	0.86
117	0.69	0.59	0.79
126	0.68	0.58	0.78

TABLE X

Average $\hat{\theta}$, minimum and maximum estimated with the rolling horizon non-parametric estimator Cai (2019) with a horizon length of 30 years for the detrended maximum temperatures measured at Brera observatory, Milan, Italy for different values of k .

b_n	Z-estimator	Y-estimator
96	0.46	0.45
108	0.45	0.44
222	0.46	0.46

TABLE XI

$\hat{\theta}$ estimated with the Z-estimator and Y-estimator of Berghaus and Bücher (2018) for the detrended maximum temperatures measured at Brera observatory, Milan, Italy for different values of b_n .

Filter	average $\hat{\theta}$	standard deviation
Robust parametric ARt(1) filter	0.38	0.13
Parametric ARt(1) filter, kappa: 30	0.34	0.027

TABLE XII

Average $\hat{\theta}$, and its standard deviation for the parametric ARt(1) filtering approaches for the detrended maximum temperatures measured at Brera observatory, Milan, Italy.

4.7 || DISCUSSION OF THE EMPIRICAL RESULTS

The respective approaches lead to quite different estimates of the extremal index. Based on the non-parametric estimators and the choice for k as described based on how the non-parametric estimators behave in function of k , one would obtain an extremal index around 0.66, whereas the sliding block estimators would estimate the extremal index to lie between 0.4 and 0.5. The parametric filter approaches lead to an average extremal index between 0.35-0.40. This divergence in results shows that fine-tuning the different estimators is quite difficult.

Further, both the filtering approaches and the non-parametric rolling horizon estimator show that θ can show quite some variation from the average θ over the whole period. This provides support for the assumption that θ may not be stable for long time periods. Moreover, both the filtering approaches and the rolling horizon estimator with a long time horizon show a slight upward trend, which would imply that the average duration of events of extreme daily maximum temperatures compared to expected maximum temperatures is decreasing. However, it is important to stress that this is after removal of seasonal and longer term trends. As such, the definition of what is considered as an extreme observation changes over time.

5 || CONCLUSION

This paper investigates how extremal dependency in time series can be better characterised by time-varying extremal index estimators. To this end, we propose two new estimation methods: a time-varying non-parametric estimator and a time-varying parametric filtering approach. For the parametric filtering approach, we discuss uniform invertibility and consistency and prove these characteristics depending on a number of assumptions. Further, we propose an efficient algorithm for the computation of the non-parametric estimator of Cai (2019).

We show the importance of allowing time-varying behaviour in a number of simulations. Overall, the time-varying non-parametric estimator is the most robust of the estimators assessed, as it is based on very light assumptions. Further, the parametric time-varying approach was the best performing of all estimators in case of correct model specification, and as such can be interesting in cases where the form of the data generating process can be established with some degree of confidence. Misspecification leads to decreased performance, which varies depending on the extent of misspecification. The simulations also showed that wrongly assuming a stable extremal index can be more severe than misspecification. Moreover, the simulations considered hyperparameters which optimised performance, in order to have a fair comparison. As the parametric approach doesn't contain hyperparameters, its performance in practice can be expected to be similar to the simulational setting. However, the other estimators will likely perform worse, as the optimal hyperparameter settings are typically unknown.

The empirical application applies the estimators on a real dataset, the daily maximum temperatures measured at the Brera observatory in Milan, Italy. The application shows that the rolling horizon estimator shows quite some variation when a medium time horizon is taken. Even for very long time periods (over 100 years), time-varying behaviour was detected. This suggests that the extremal index is not stable over time. The empirical application also highlights the complexity of parameter tuning, as results can be quite different for different parameter settings.

As such, we conclude that approaches which can estimate a time-varying extremal index can lead to a more accurate characterisation of extremal dependency processes, both in simulational and empirical settings.

There are a number of limitations to this study. First, the proof of consistency of the observation-based filters is not completely finished: the stochastic equicontinuity was not established. Showing this condition would finish the proof. Second, this paper did not address confidence intervals for the extremal index. Further, if the conditions mentioned in Blasques (2019) can be verified, asymptotic normality of the parameters of the filtering approach can be shown. Third, the extremal index was estimated by several methods, and the challenges of finetuning the hyperparameters have been highlighted. Even though the estimators are asymptotically normal, the study did not consider confidence intervals of the extremal index. Yet, as estimates of the extremal index can be quite different depending on the method on use, confidence intervals can be useful to further expand the understanding of time-varying behaviour of extremal clustering.

6 || ACKNOWLEDGEMENTS

We acknowledge the data providers in the ECA&D project. Klein Tank, A.M.G. and Coauthors, 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. of Climatol.*, 22, 1441–1453. Data and metadata available at <https://www.ecad.eu>. I further want to thank prof. dr. Juan Juan Cai for the helpful and insightful feedback I received over the last months, and Alena Bourd always for putting a smile on my face in the last months, even when I was struggling with very unpleasant inequalities.

REFERENCES

- AHSANULLAH, M., B. G. KIBRIA, AND M. SHAKIL (2014): “Normal distribution”, in: *Normal and Student t Distributions and Their Applications*, Springer, 7–50.
- AKAIKE, H. (1974): “A new look at the statistical model identification”, *IEEE transactions on automatic control*, 19(6), 716–723.
- 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.
- BLASQUES, F. (2019): “Advanced Econometric Methods: A Guide to Estimation and Inference for Nonlinear Dynamic Models”.
- CAI, J. J. (2019): “A nonparametric estimator of the extremal index”, *arXiv preprint arXiv:1911.06674*.
- CHERNICK, M. R., T. HSING, AND W. P. McCORMICK (1991): “Calculating the extremal index for a class of stationary sequences”, *Advances in applied probability*, 835–850.
- CREAL, D., S. J. KOOPMAN, AND A. LUCAS (2008): “A general framework for observation driven time-varying parameter models”, tech. rep., Tinbergen Institute Discussion paper.
- DAVIS, R. A., T. MIKOSCH, AND Y. ZHAO (2013): “Measures of serial extremal dependence and their estimation”, *Stochastic Processes and their Applications*, 123(7), 2575–2602.
- DE GUSMAO, F. R., E. M. ORTEGA, AND G. M. CORDEIRO (2011): “The generalized inverse Weibull distribution”, *Statistical Papers*, 52(3), 591–619.
- DICKEY, D. A. AND W. A. FULLER (1979): “Distribution of the estimators for autoregressive time series with a unit root”, *Journal of the American statistical association*, 74(366a), 427–431.
- EHLERT, A., U.-R. FIEBIG, A. JANSSEN, AND M. SCHLATHER (2015): “Joint extremal behavior of hidden and observable time series with applications to GARCH processes”, *Extremes*, 18(1), 109–140.
- FERRO, C. A. AND J. SEGERS (2003): “Inference for clusters of extreme values”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2), 545–556.
- HALL, P. AND N. TAJVIDI (2000): “Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data”, *Statistical Science*, 153–167.
- HSING, T. (1993): “Extremal index estimation for a weakly dependent stationary sequence”, *The Annals of Statistics*, 2043–2071.
- HUANG, N. E., Z. SHEN, S. R. LONG, M. C. WU, H. H. SHIH, Q. ZHENG, N.-C. YEN, C. C. TUNG, AND H. H. LIU (1998): “The empirical mode decomposition and the Hilbert spectrum

- for nonlinear and non-stationary time series analysis”, *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971), 903–995.
- HUERTA, G. AND B. SANSÓ (2007): “Time-varying models for extreme values”, *Environmental and Ecological Statistics*, 14(3), 285–299.
- KLEIN TANK, A., J. WIJNGAARD, G. KÖNNEN, R. BÖHM, G. DEMARÉE, A. GOCHEVA, M. MILETA, S. PASHIARDIS, L. HEJKRLIK, C. KERN-HANSEN, et al. (2002): “Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment”, *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 22(12), 1441–1453.
- KRENGEL, U. (2011): *Ergodic theorems*, vol. 6, Walter de Gruyter.
- LAURINI, F. AND J. A. TAWN (2012): “The extremal index for GARCH (1, 1) processes”, *Extremes*, 15(4), 511–529.
- MAUGERI, M., L. BUFFONI, B. DELMONTE, AND A. FASSINA (2002): “Daily Milan temperature and pressure series (1763–1998): completing and homogenising the data”, in: *Improved Understanding of Past Climatic Variability from Early Daily European Instrumental Sources*, Springer, 119–149.
- MIKOSCH, T. (2013): “Heavy tail phenomena and dependence of extremes”.
- MOLONEY, N. R., D. FARANDA, AND Y. SATO (2019): “An overview of the extremal index”, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(2), 022101.
- SMITH, R. L. AND I. WEISSMAN (1994): “Estimating the extremal index”, *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3), 515–528.
- SÜVEGES, M. (2007): “Likelihood estimation of the extremal index”, *Extremes*, 10(1-2), 41–55.

	test statistic	p-value	number of lags	observations used in regression
Deterministic components				
Constant	-41.1184	<0.00001	27	86900

TABLE A.I

ADF regression results for model with constant for the values detrended by the polynomial regression of the daily maximum temperatures in the period 1763-2000 measured at Brera, Milan, Italy

APPENDIX

A || DETRENDING THE DAILY MAXIMUM TEMPERATURES WITH POLYNOMIAL REGRESSION

Polynomial regression refers to the fitting of a polynomial of degree p of independent variables to a dependent variable. In this case, the independent variables are the day of the year, and the general time index variable to allow for longer term trends. The day of year is fitted to degree 4, and the general time index up to degree 2. Given the observational differences caused by the reorganisation of observatory starting in 1835, the period 1763-1837 and the period 1838-2000 are fitted separately. Figure A.I depicts the fitted polynomial curve graphically. Above, the separation into two periods for fitting the polynomial is clearly visible, and it can also be seen that the polynomial curve shows some longer term behaviour. The graph under shows the residuals of the polynomial regression. Whereas the temperatures are centered around 0, there may be some longer term trends which are not captured by the polynomial model. The long term trends seem not to be fitting well to the data. In order to investigate the shorter term behaviour, figure A.II shows the polynomial curve for the period 1890-1899. The polynomial curve follows the annual variation, but for this period seems to be underestimating the summer peaks and overestimating the lowest winter temperatures. This is also visible in the residuals, which seem to contain some seasonality which isn't captured by the model.

In order to check the stationarity of the residuals of the polynomial regression, A.I reports the results of an ADF test. The hypothesis of a unit root is rejected at the 1% significance level. The residuals of the polynomial regression can thus be considered to be stationary.

B || EMPIRICAL APPLICATION 2: EXTREMAL CLUSTERING OF ABSOLUTE DAILY MAXIMUM TEMPERATURES

This section describes a second empirical application: the extremal clustering of absolute daily maximum temperatures in the summer months (June, July, and August). In contrast to the application described in Chapter 4, this application considers the extremal index of the absolute daily maximum temperatures, whereas Chapter 4 described the extremal index of the difference between the actual and expected daily maximum temperatures given the time of year and long term trends. This second application can thus be considered as a measure for the extremal dependency of very warm days, and hence for the length of heat waves.

In this application, it is not possible to compute the extremal index by means of the sliding block estimators defined by equations (2.8) and (2.9), nor by the (robust) parametric filtering

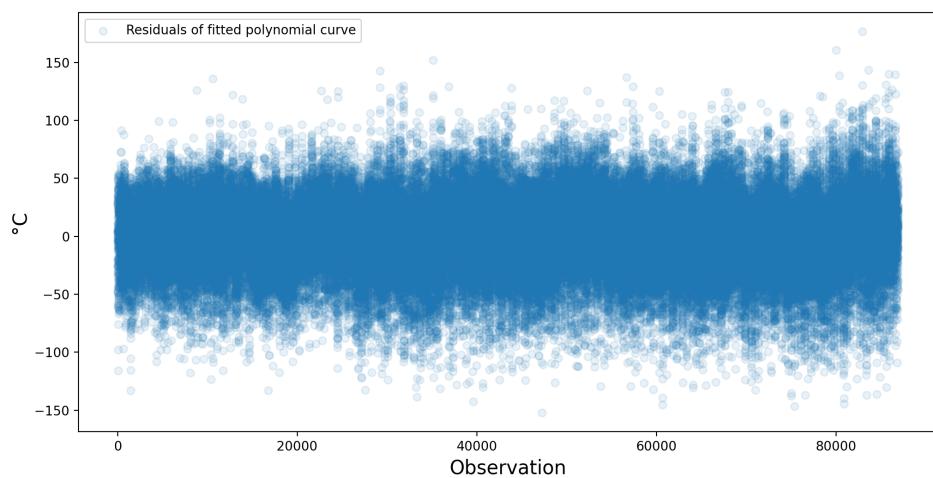
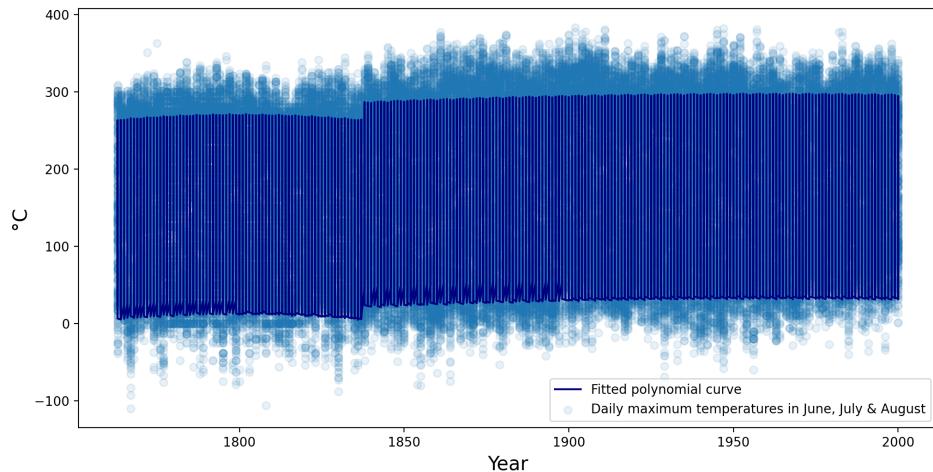


FIGURE A.I

Daily maximum temperatures measured at Brera astronomical observatory, Milan, Italy, and fitted polynomial curve measured for the period 1763-2000 (above), detrended values (under)

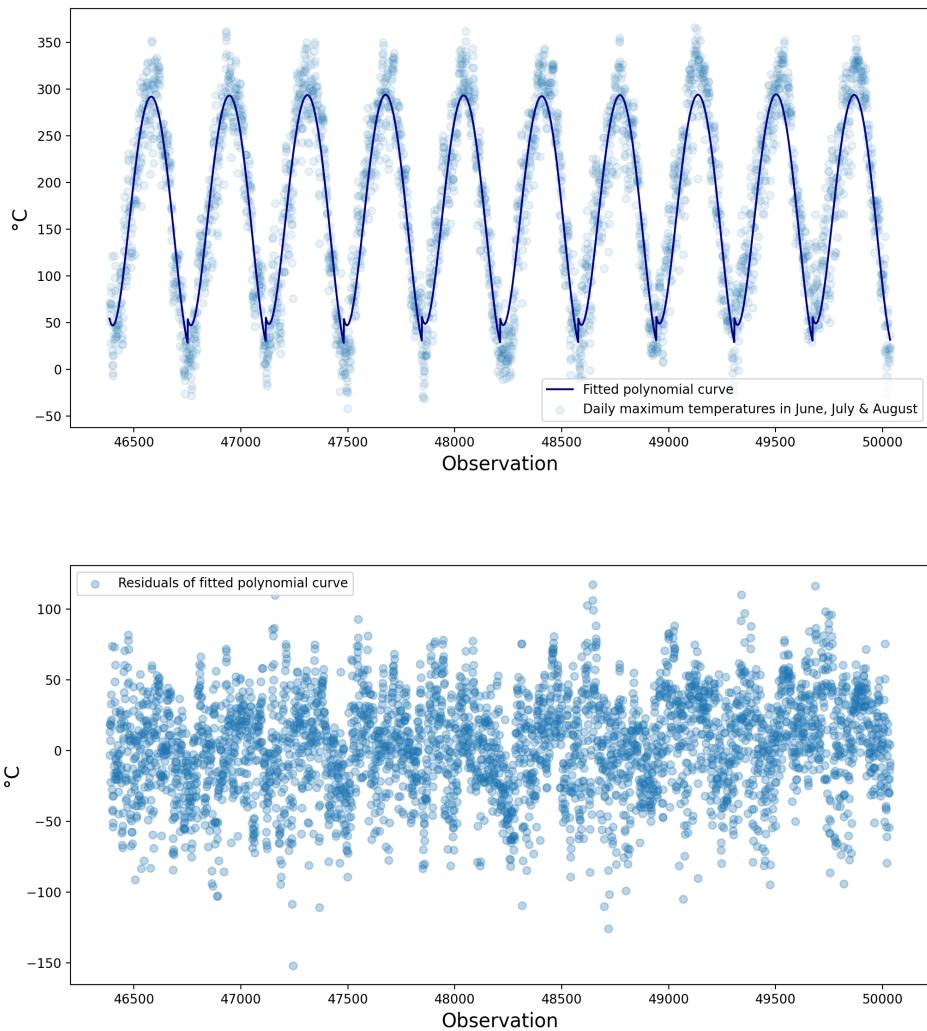


FIGURE A.II

Daily maximum temperatures measured at Brera astronomical observatory, Milan, Italy, and fitted polynomial curve measured for the period 1890-1899 (above), detrended values (under)

approach because of the discontinuous nature of this data set. We can however adapt equations (2.6) and (2.10) to deal with discontinuities. The stable estimator becomes:

$$\hat{\theta}_n(d) = \frac{1}{k} \sum_{i=0}^{N-1} \sum_{j=1}^{L-d+1} \mathbb{1}\{M_{iL+j+1,iL+j+d-1} \leq X_{n-k,n} < X_{iL+j}\} \quad (\text{B.1})$$

Note that it is possible that the number of observations deviates in the first and/or last year of the sample. If this is the case, the estimator should be modified by changing the summation in j for those years.

In a similar spirit, as in the stable non-parametric case, (2.10) can be adapted to accommodate for this aspect as follows:

$$\hat{\theta}_p(t) = \frac{1}{\tilde{k}} \sum_{i=t}^{Q-1} \sum_{j=u}^{L-d_t+1} \mathbb{1}\{M_{iL+j+1,iL+j+d_t-1} \leq X_{t+p-\tilde{k},t+p} < X_{iL+j}\} \quad (\text{B.2})$$

where Q is the number of years in the subsample, and the other parameters defined as above. As is the case for the stable non-parametric estimator, the first and last year of the subsample may not contain all observations of the year. Again, when this happens, the estimator should be modified by changing the summation in j for those years such that only the elements of the subsample are taken into consideration.

Note that the algorithms defined in 2.1.1 and 2.2.1 implicitly use these estimators, when the original indexing is kept.

This remainder of this application is structured as follows: We start with a description of the data, and discuss the stationarity of the sample. Then, the stable non)parametric estimator is applied, followed by the rolling horizon estimator. This application ends with a discussion of the results.

Selection of summer months

Only the months June, July and August are considered, based on the assumption that observations in these months are drawn from the same distribution. Table B.I summarises the descriptive statistics for the observations in June, July and August, whereas figure B.I depicts these observations and the yearly average of daily maximum temperatures. It seems that the extreme values behave differently over time, with moments in which peaks consist of a single observation, and others where peaks of several observations can be seen.

When observing figure B.I, the effect of the organisational change in the observatory in 1835 is even more pronounced than is the case for the yearly maximum temperature data. Further, it seems that there are long term trends in the data, with some periods with an upward trend, and other periods with a downward trend. Hence, even though the data can be considered as not containing seasonality, the trending behaviour needs to be assessed.

There are no obvious linear or quadratic trends, but the data could contain a stochastic trend. However, the ADF test cannot be applied straightforwardly to the data as there is an annual gap in the data between the last observation of August and the first observation of July the next year. To overcome this issue, the problem is split into two subproblems, which are analysed separately. First, in order to assess whether the data within one year is stationary, the ADF test

Maximum temperature in June, July and August (°C)	
first	1763-06-01
last	2000-08-31
count	21896
mean	27.8781
std	3.58529
min	12.2
25%	25.6
50%	28.1
75%	30.4
max	38.3

TABLE B.I

Summary statistics of daily maximum temperatures in June, July and August, measured at Brera astronomical observatory, Milan, Italy

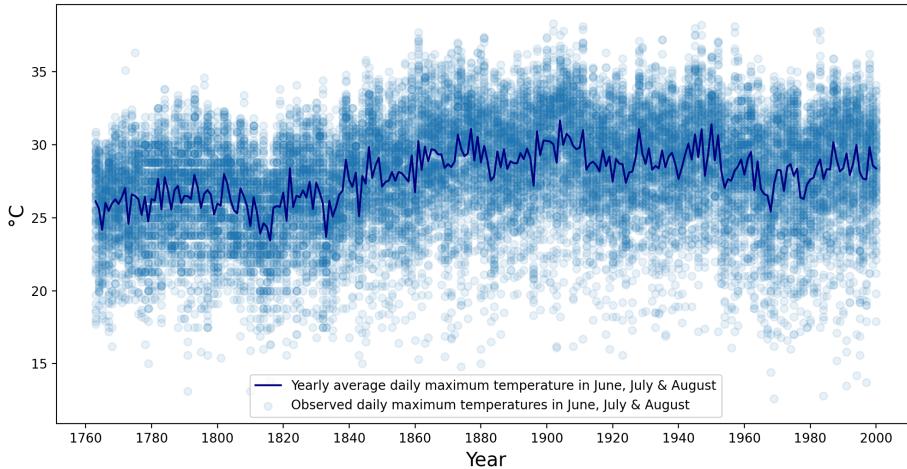


FIGURE B.I

Daily maximum temperatures measured in June, July and August for the period 1763-2000, measured at Brera astronomical observatory, Milan, Italy: observations (scatter), average for June, July and August for a given year (line).

	test statistic	p-value	number of lags	observations used in regression
count	238	238	238	238
mean	-3.63	0.05	0.99	90.01
std	0.94	0.10	2.05	2.05
min	-6.13	0	0	79
25%	-4.28	0	0	90
50%	-3.62	0.01	0	91
75%	-3.02	0.03	1	91
max	-1.29	0.63	12	91

TABLE B.II

Summary of the yearly ADF tests for the daily maximum temperatures in June, July and August in 1763-2000 measured at the Brera astronomical observatory, Milan, Italy. For each year, a separate ADF test was performed.

is run for each year separately. For this ADF test, we only specify a constant in the regression to reduce the number of parameters in the model. Second, in order to assess whether the data contains longer term stochastic trends, the average of the daily maximum temperature in the months of June, July and August is calculated for each year, and an ADF test is performed on this yearly average. For this ADF test, we consider both a model with constant only, with constant and linear trend, and with constant, linear and quadratic trend.

Table B.II shows the summary statistics of the ADF tests performed on each year. The interpretation of the test statistics is complex, as the critical values depend on the number of lags, which is not the same for each year. However, the p-values can be compared, and as can be seen, the average p-value is 5%. However, this average value is affected by a number of outliers, and the median is 3%. The hypothesis of having a unit root can be rejected at the 5 % significance level for 193 of the 238 years. Figure B.II depicts the p-values in a histogram. The unit root hypothesis is rejected for most of the years at the 5 % significance level.

The results of the ADF test for the yearly summer averages for different deterministic specifications are reported in table B.III. As can be seen, the unit root hypothesis cannot be rejected at the 5% significance level for the specification with constant. However, this ADF test does not take the aforementioned differences in observation due to the reorganisation of the Brera Observatory in 1835 into consideration. Therefore, table B.IV shows the results of the ADF test when applied on the period 1763-1837 and the period 1838-2000 separately. When considering this split, the ADF test rejects the hypothesis of a unit root for both periods, and for all deterministic component specifications at the 1% significance level.

In table B.IV, three possible trend models were specified. We proceed by selecting the model with the lowest AIC for each respective time period. As can be seen in table B.IV, for the period going from 1763 until 1837, a model with constant only obtains the lowest AIC. For the period 1838-2000, a model with constant and linear trend has the lowest AIC. The models used for detrending are reported in table B.V for the period 1763-1837 and in table B.VI for the period

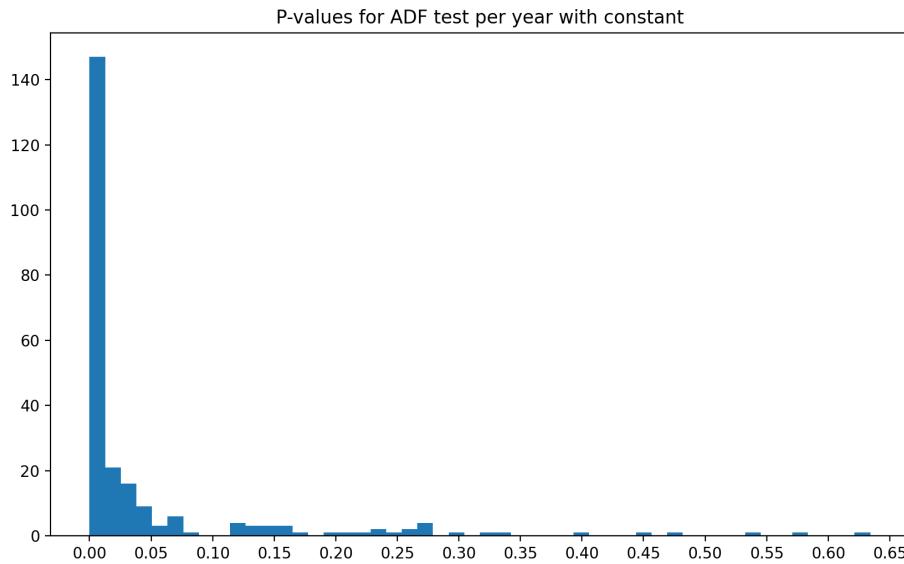


FIGURE B.II

Histogram of the p-values of yearly ADF tests for the daily maximum temperatures in June, July and August in 1763-2000 measured at the Brera astronomical observatory, Milan, Italy.

For each year, a separate ADF test was performed.

	test statistic	p-value	number of lags	observations used in regression	AIC
Deterministic components					
Constant only	-2.7955	0.0589	3	234	1677.4668
Constant and linear trend	-3.7623	0.0186	2	235	1676.8031
Constant, linear trend, quadratic trend	-4.5553	0.0054	2	235	1670.7028

TABLE B.III

Results of ADF tests for different configurations (constant only, constant and linear trend, constant, linear and quadratic trend) for the yearly average daily maximum temperatures in June, July and August, measured at the Brera astronomical observatory, Milan, Italy. In contrast to the ADF tests for testing for the stationary within a given year, this ADF considers the yearly average of the summer months.

Deterministic components	period	test statistic	p-value	number of lags	observations used in regression	AIC
Constant only	1763-1837	-4.3778	0.0003	1	73	465.7500
Constant and linear trend	1763-1837	-4.3670	0.0025	1	73	467.1081
Constant, linear trend, quadratic trend	1763-1837	-4.3775	0.0098	1	73	469.0227
Constant only	1838-2000	-4.8076	0.0001	1	161	1101.2171
Constant and linear trend	1838-2000	-4.8539	0.0004	1	161	1099.4715
Constant, linear trend, quadratic trend	1838-2000	-5.4201	0.0002	1	161	1100.2505

TABLE B.IV

Results of ADF tests for the yearly average daily maximum temperatures in June, July and August, for the period 1763-1837, and the period 1838-2000 measured at the Brera astronomical observatory, Milan, Italy. In contrast to the ADF tests for testing for the stationary within a given year, this ADF considers the yearly average of the summer months.

1838-2000 respectively. As these models have different coefficients, we subtract the constant and time trend, to have a comparable series. However, caution is needed for the interpretation, as the regression equation is:

$$\Delta y_i = c + \lambda y_{i-1} + \beta \Delta y_{i-1} + \gamma i + \epsilon_i \quad (\text{B.3})$$

where y_i is the annual average daily maximum temperature of year i .

It is necessary to compute the constant and linear trend. These are not equal to c and γ respectively, as this doesn't take the autoregressive process into consideration. To obtain these elements, equation (B.3) can be written as:

$$y_i = c + (\lambda + \beta + 1)y_{i-1} - \beta y_{i-2} + \gamma i + \epsilon_i \quad (\text{B.4})$$

This equation can be put in companion form:

$$\begin{pmatrix} y_i \\ y_{i-1} \end{pmatrix} = \begin{pmatrix} c \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda + \beta + 1 & -\beta \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{i-1} \\ y_{i-2} \end{pmatrix} + \begin{pmatrix} \gamma \\ 0 \end{pmatrix} i + \begin{pmatrix} \epsilon_i \\ 0 \end{pmatrix} = c^* + \phi^* y_{i-1}^* + \gamma^* i + \epsilon^* \quad (\text{B.5})$$

Based on this companion form, the trend δ and the constant ω can be computed as:

$$\delta = (I - \phi^*)^{-1} \gamma^* \quad (\text{B.6})$$

$$\omega = (I - \phi^*)^{-1} (c^* - \phi^* \delta) \quad (\text{B.7})$$

with I the (2×2) identity matrix. As the trends and constant were calculated for the annual average daily maximum temperature, the detrending also happens on yearly basis.

The data with linear trends removed is presented in figure B.III. Compared to the original data in figure B.I, the data looks more stationary, but may still contain some longer term trend behaviour not captured by the simple deterministic components specified in the Dickey-Fuller regression. However, given that these trends seem to be mean-reverting, it does not affect the overall stationarity of the process. The variance seems to be relatively constant in time as well. So, even when no specific test about the stability of the tail process is done, given that the first two moments look stationary, one can assume that also the tail process is also stable.

Dep. Variable:	Δy_i	R-squared:	0.455			
Model:	OLS	Adj. R-squared:	0.439			
Method:	Least Squares	F-statistic:	29.18			
Date:	Tue, 18 May 2021	Prob (F-statistic):	6.06e-10			
Time:	16:45:32	Log-Likelihood:	-269.41			
No. Observations:	73	AIC:	544.8			
Df Residuals:	70	BIC:	551.7			
Df Model:	2					
	coef	std err	t	P> t	[0.025	0.975]
c	174.1676	39.778	4.378	0.000	94.832	253.503
λ	-0.6675	0.152	-4.378	0.000	-0.972	-0.363
β	-0.2197	0.117	-1.881	0.064	-0.453	0.013
Omnibus:	1.882	Durbin-Watson:	2.003			
Prob(Omnibus):	0.390	Jarque-Bera (JB):	1.428			
Skew:	-0.339	Prob(JB):	0.490			
Kurtosis:	3.103	Cond. No.	8.96e+03			

TABLE B.V

ADF regression results for model with constant for the annual average daily maximum temperatures in the period 1763-1837 measured at Brera, Milan, Italy

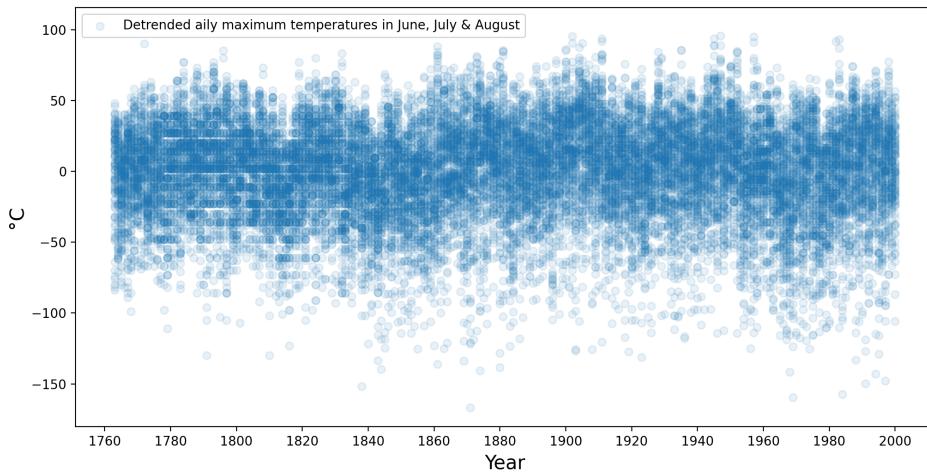


FIGURE B.III

Daily maximum temperatures in June, July and August after removing constant and linear trend for the Brera astronomical observatory, Milan, Italy

Dep. Variable:	Δy_i	R-squared:	0.417			
Model:	OLS	Adj. R-squared:	0.406			
Method:	Least Squares	F-statistic:	37.49			
Date:	Tue, 18 May 2021	Prob (F-statistic):	2.52e-18			
Time:	14:22:25	Log-Likelihood:	-598.81			
No. Observations:	161	AIC:	1206.			
Df Residuals:	157	BIC:	1218.			
Df Model:	3					
	coef	std err	t	P> t 	[0.025	0.975]
c	117.8667	24.283	4.854	0.000	69.903	165.830
λ	-0.4068	0.084	-4.854	0.000	-0.572	-0.241
β	-0.3695	0.073	-5.049	0.000	-0.514	-0.225
γ	-0.0127	0.017	-0.736	0.463	-0.047	0.021
Omnibus:	1.558	Durbin-Watson:	2.073			
Prob(Omnibus):	0.459	Jarque-Bera (JB):	1.179			
Skew:	0.079	Prob(JB):	0.555			
Kurtosis:	3.388	Cond. No.	9.11e+03			

TABLE B.VI

ADF regression results for model with constant and linear trend for the annual average daily maximum temperatures in the period 1838-2000 measured at Brera, Milan, Italy

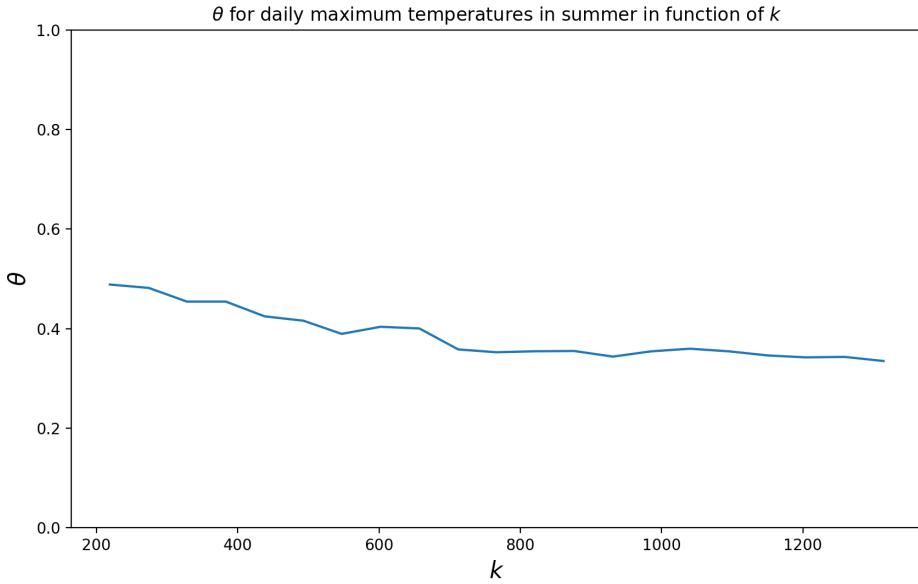


FIGURE B.IV

$\hat{\theta}^n$ in function of $k \in (200 - 1300)$ for the daily maximum temperatures in June, July and August measured at Brera observatory, Milan, Italy.

Stable non-parametric estimator

For the stable non-parametric estimator, it is necessary to choose k . As can be seen in B.IV, when $\theta \approx 0.4$, the estimator is quite stable when k is between 400 and 1200, it is quite stable. Table B.VII reports the outcomes of the stable non-parametric extremal index estimator for a number of values for k .

Rolling horizon estimator

The value for \tilde{k} for the rolling horizon estimator can be determined in a similar way as k in case for the stable non-parametric estimator. In a first step, a value for \tilde{k} is chosen so that $\frac{\tilde{k}}{p}$ is a little bit larger than $\frac{k}{n}$. Figure B.VI shows how the rolling horizon estimator with horizon length of 30 years evolves for several values of k . The estimator is quite stable in when k is between 60 and 160.

When considering the rolling horizon estimator, it is necessary to also choose a horizon length. This poses a trade-off, because on the one hand, one aims to have a horizon length which is as long as possible, given that the extremal index is an asymptotic process, and a short horizon length, which acknowledges the possibility that the extremal index may change over time. To this end, several horizon lengths have been tried out, where \tilde{k} equals $\frac{k * \text{horizonlength} * 92}{n}$, where the horizon length is expressed in years, and $k = 657$ (based on the analysis of k for the stable estimator). These are shown in figure B.V. As can be seen, the short horizon lengths (2 years and 10 years) show very strong variation. On the other hand, the very long horizon lengths (100 and 200 years) display some time-changing behaviour, but this time-varying behaviour is quite mild

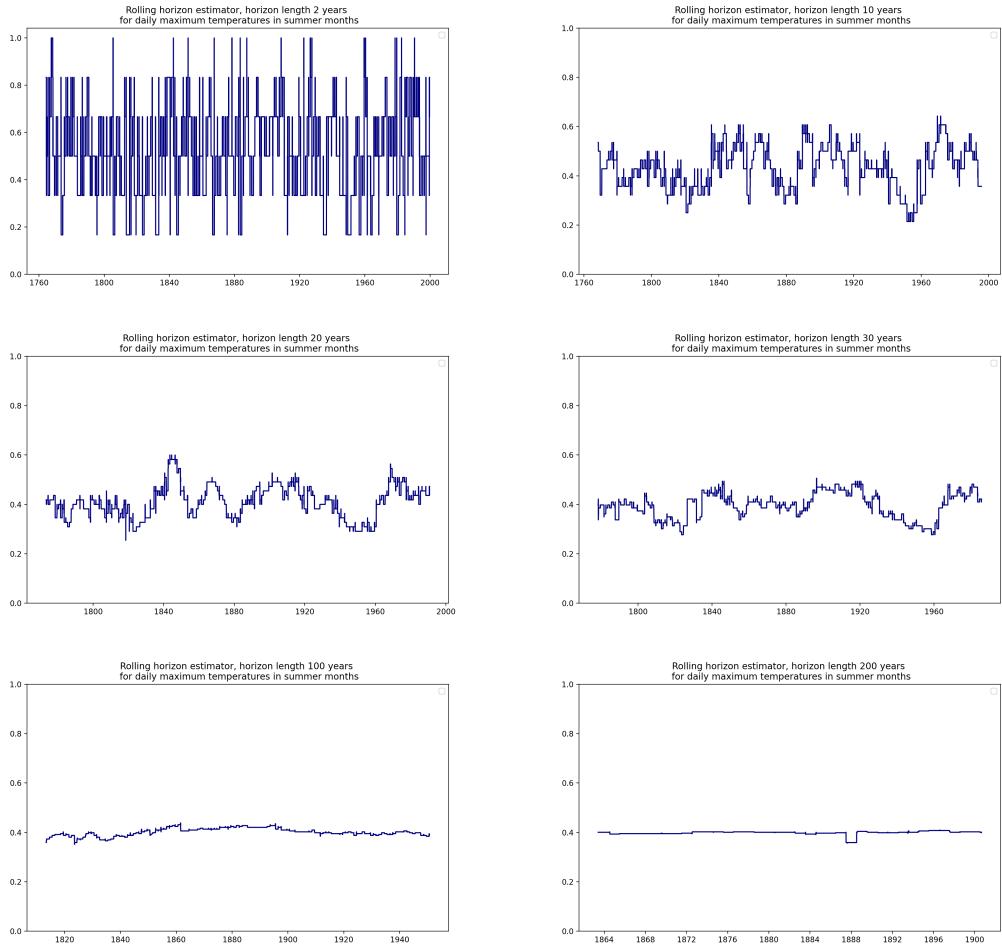


FIGURE B.V

$\hat{\theta}_p(t)$ for the daily maximum temperatures in June, July and August measured at Brera observatory, Milan, Italy for different horizon lengths: 2 years, 10 years, 20 years, 30 years, 100 years, 200 years (left-to-right, top-to-bottom)

and therefore does not have much impact on the extremal index. When considering a horizon length of 20 or 30 years, a relatively smooth curve is obtained with meaningful time-varying behaviour.

Figure B.VI shows the behaviour of the rolling horizon estimator $\hat{\theta}_p(t)$ for different values for k . The estimator is quite stable when k is between 60 and 160, and lies on average around 0.4. Table B.VIII reports the outcomes of the stable non-parametric extremal index estimator for a number of values for k , assuming a horizon length of 30 years.

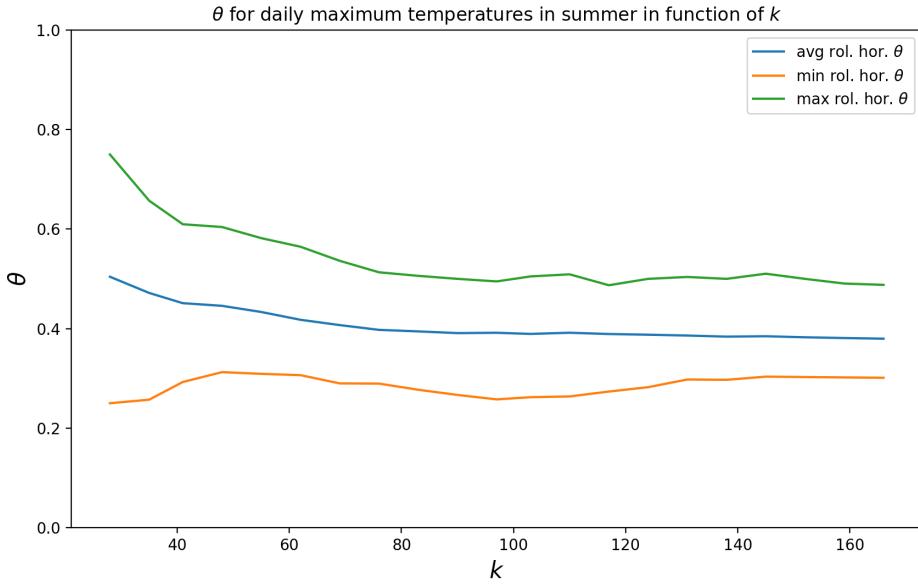


FIGURE B.VI

$\hat{\theta}_p(t)$ with a horizon length of 30 years in function of $k \in (28 - 166)$, for the daily maximum temperatures in June, July and August measured at Brera observatory, Milan, Italy:
maximum, minimum and average value in function of t

k	$\hat{\theta}$
547	0.39
766	0.35
1040	0.36

TABLE B.VII

$\hat{\theta}$ estimated with the stable non-parametric estimator of Cai (2019) for the daily maximum temperatures in June, July and August measured at Brera observatory, Milan, Italy for different values of k .

k	$\hat{\theta}$	min	max
76	0.40	0.29	0.51
117	0.39	0.27	0.49
159	0.38	0.30	0.49

TABLE B.VIII

Average $\hat{\theta}$, minimum and maximum estimated with the rolling horizon non-parametric estimator Cai (2019) with a horizon length of 30 years for the daily maximum temperatures in June, July and August measured at Brera observatory, Milan, Italy for different values of k .

Discussion of the empirical results

Figure B.VII compares the rolling horizon estimator with a horizon of 30 years and $k = 83$ to the stable estimator with $k = 657$. When analysing the time-varying behaviour for the estimator with a horizon length of 30 years, one can observe that the extremal index is decreasing slightly over the first decades, before increasing quite strongly. The timing of this increase coincides more or less with the reorganisation of the Brera observatory in 1835. The extremal index then

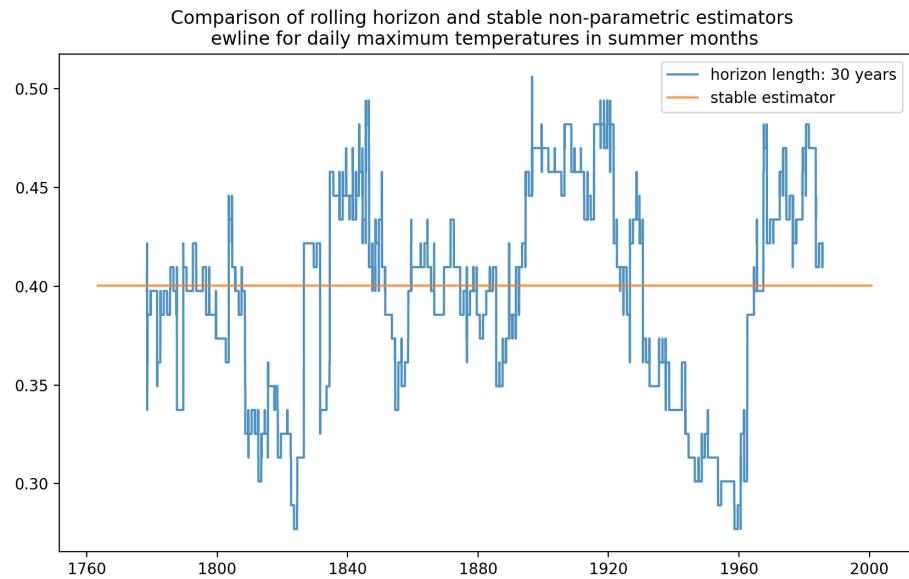


FIGURE B.VII

Stable non-parametric estimator and rolling horizon non-parametric estimator with a horizon length of 30 years for the daily maximum temperatures in June, July and August measured at Brera observatory, Milan, Italy.

fluctuates, before decreasing for a number of decades starting around 1920. As from 1960, the extremal index increases again. Whereas the average $\hat{\theta}$ for the rolling horizon estimator is quite similar to the stable $\hat{\theta}$, there is quite some variation.