ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Econometrics & Management Science - General Econometrics

Extreme quantile estimation under serial dependence

Ying Wu

451968

Supervisor: Chen Zhou

Second assessor: Mikhail Zhelonkin

Date of final version: November 22, 2018

Abstract

This paper evaluates the finite-sample performances of six extreme quantile estimators in the heavy-tailed series under serial dependence. Through Monte Carlo simulations, we show that the performances of the estimators are related to the degree of the serial dependence and the linearity/nonlinearity of the serial dependence. The maximum likelihood estimator based on the sliding block maxima is optimal to handle the linear serial dependence in data. The probability-weighted moment estimators are likely to be distorted by strong linear serial dependence. When the serial dependence is nonlinear, the excess kurtosis would affect the quantile estimation. The Weissman estimator outperforms when data has nonlinear serial dependence and a low excess kurtosis. The probability-weighted moment estimators based on the disjoint blocks is preferable when the data has a relatively high excess kurtosis. Additionally, this paper investigates an approach to improve the maximum likelihood estimators based on the block maxima in the GARCH models.

Key words: peaks-over-threshold; block maxima; maximum likelihood estimation; probability-weighted moment; heavy tails; Monte Carlo simulation

1 Introduction

One of the popular traditional market risk measures is the volatility. The main drawback of volatility, however, is that it ignores the direction of the investment's movement, i.e. gain or loss. By contrast, Value-at-Risk (VaR) sheds light on risk management by focusing on portfolio's losses only. It is defined as the high quantile of the negative log-returns, and it measures the potential bad scenario for a given low probability over a certain time period that a investor wants to be aware of. Mathematically, given a confidence level $\alpha \in (0, 1)$, the probability that a loss L exceeds its VaR is no higher than $1 - \alpha$:

$$\text{VaR}_\alpha := \inf\{l; \Pr(L > l) \leq 1 - \alpha\}.$$

There are some existing methods to estimate VaR, for instance, historical simulation, Monte Carlo simulation method, delta-normal method, variance-covariance method, etc. One of the limitations of these methods, except for historical simulation, is that they all make parametric assumptions on the loss distribution (Linsmeier & Pearson, 2000). For example, Monte Carlo simulation method requires a pre-determined distribution (e.g. normal mixture model) to generate a large number of samples. Delta-normal method and variance-covariance method assume a Gaussian distribution on the loss. However, since financial time series are usually not normally distributed, the estimation based on parametric assumptions is problematic if the assumption fails to capture the underlying distribution of the data. Although historical simulation releases the normality restriction, the high quantile is estimated by an ordered statistic at the cost of inaccuracy, especially if the sample size is low. In order to make inference about the tail behavior without specifying a global parametric form for the distribution function, one may use the extreme value theory (EVT) which makes rather mild structural assumptions on the tail of the distribution of loss.

There are two prevailing approaches to implement EVT for extreme quantile estimation, the peaks-over-threshold (POT) and the block maxima (BM). More specifically, the POT extracts the ordered statistics above a high threshold (referred as exceedances) and the exceedances approximately follow a scaled generalized Pareto distribution (GPD). Differently, the BM splits the sample into blocks and collects the block maxima. And the distribution of the scaled block maxima converges to the generalized extreme-value (GEV) distribution.

The extreme quantile estimator in EVT framework is constructed through extrapolation, requiring estimations for the parameters of the approximated distribution (GPD or GEV). Several parameter estimation methods have been employed, such as the maximum likelihood estimation (MLE), the probability-weighted moments (PWM) estimation and the method of moment. Although there are multiple parameter estimators available, one should be careful about the choices by considering the tail behaviour of the series, which can be measured by the extreme value index. For example, the moment estimator in Hosking and Wallis (1987) needs that the exceedances or block maxima have a finite variance, meanwhile, some estimators are proposed for the heavy-tailed behaviour and others can be adopted in a more general case. Moreover, the asymptotic normalities of the parameter estimators are usually established under a more restrictive condition on the range of the extreme value index as briefly discussed in Section 2. In gen-

eral, the asymptotic normalities of the parameter estimators are proved in the identical and independent distributed (i.i.d.) sample. The asymptotic property of extreme quantile estimator follows consequently. However, the real time series of log-returns usually exhibits serial dependence such as volatility clustering. Hence independence is not a realistic assumption for application to financial data. The extreme quantile estimators remain consistent under weak serial dependence[1], while the asymptotic variance usually has a complex structure. Drees (2003) proved the asymptotic normality of a class of the POT extreme quantile estimators for stationary β-mixing time series. However, to the best of my knowledge, the closed-form asymptotic variance of the BM extreme quantile estimators under serial dependence is not derived yet. This is because within the BM framework, the extreme quantile estimation requires the estimation of an extremal index (Leadbetter, 1983), which quantifies the serial dependence of extremes.

The performance of the POT and the BM extreme quantile estimators under serial dependence are therefore of interest. On the one hand, the finite-sample bias is non-negligible though the consistency can be obtained theoretically. On the other hand, the asymptotic variances in both the POT and the BM methods blow up due to serial dependence. Heuristically, it is straightforward to follow the POT estimators, but the estimating uncertainty is higher due to the serial dependence of the exceedances. For the BM method, the additional estimation of extremal index is expected to induce extra uncertainty, though the block maxima are still considered to be i.i.d.. The asymptotic variances of parameter estimators of the GEV remain unchanged. A sliding blocks approach may gain efficiency compared to the disjoint BM thanks to this approximate i.i.d. structure.

As mentioned above, the performances of the estimators may relate to their conditions on the extreme value index. Since there is no a universal rule to decide the optimal estimator based on that, it is the motivation of this paper to compare the MLE estimator with the PWM estimator. Overall, there are six extreme quantile estimators considered here, i.e. the POT-MLE/PWM estimator, the disjoint BM-MLE/PWM estimator, and the sliding BM-MLE/PWM estimator.

Given a theoretical comparison of the POT and the BM estimators under serial dependence is currently unavailable, we address the following research question by simulations:

In the existence of serial dependence, which extreme quantile estimator outperforms in a finite sample?

We evaluate six extreme quantile estimators under linear and nonlinear serial dependence. And we take different degrees of the serial dependence into consideration. Furthermore, we allow the excess kurtosis varying across the data generating processes where the serial dependence is nonlinear. The superiority of the extreme quantile estimator is determined by the minimum squared error. The simulation results show that the sliding BM-MLE estimator is preferable in most scenarios when the serial dependence is linear. Under nonlinear serial dependence, the POT-MLE estimator (referred as the Weissman estimator in the rest of the paper) outperforms when the excess kurtosis is low, and the disjoint BM-PWM estimator is preferred when the excess kurtosis is high. Besides the performance examination, we investigate a

[1]The conditions on serial dependence are different for the POT and the BM methods, see Section 2.2 for details.

procedure to improve the performances of the disjoint and sliding BM-MLE estimators under nonlinear serial dependence. The simulation results present that the procedure we propose reduces the bias and variance of the quantile estimator with proper block sizes.

The paper is organized as follows. Section 2 contains the literature reviews focusing on the extreme quantile estimation methods and serial dependence. Section 3 provides the POT and the BM estimators, moreover, demonstrates the evaluation criteria and the data generating processes. Section 4 presents the finite-sample performances which are evaluated via Monte Carlo simulation. Section 5 depicts the procedure which improves the performances of the BM-MLE estimators under nonlinear serial dependence. The corresponding simulation results are shown in Section 5 as well. The final section discusses the implications and limitations of this paper and concludes.

2 Literature review

2.1 Extreme value theory

Classical Extreme Value Theory shows that if the maximum

$$M_n = \max(X_1, \ldots, X_n)$$

of n i.i.d. random variables (with cumulative distribution function F) has a non-degenerate limiting distribution G as $n \to \infty$, then G must be one of the GEV distribution function. That is, for some normalizing constants $a_n > 0$ and b_n,

$$\lim_{n\to\infty} \Pr(\frac{M_n - b_n}{a_n}) = \lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x), \tag{1}$$

where G_γ is the GEV distribution function with extreme value index γ, i.e.

$$G_\gamma(x) = \exp(-(1+\gamma x)^{-\frac{1}{\gamma}}), \quad 1+\gamma x > 0, \quad \gamma \in \mathbb{R}. \tag{2}$$

It is also called that a continuous distribution F is in the domain of attraction of a GEV distribution denoted by $F \in D(G_\gamma)$. Obviously, the extreme value index is essential to capture the tail behaviour of a distribution. Pickands (1975) proved that a continuous distribution function F has a generalized Pareto upper tail is equivalent to $F \in D(G_\gamma)$, and showed that the shape parameters of two distributions are identical.[2] Specifically, the building block of the POT is that given a high threshold t, the excesses $X - t$ are asymptotically generalized Pareto distributed. Denote the conditional distribution function of $X - t$ given $X > t$ as

$$F_t(x) := \Pr(X - t \le x | X > t) = \frac{F(t+x) - F(t)}{1 - F(t)},$$

[2]The condition of the equivalence is presented in Theorem 3.1.

with $1 - F(t) > 0$, $t < x^*$ and $x > 0$ where $x^* := \sup\{x : F(x) < 1\} \leq \infty$ is the upper endpoint of F. Then there exists a normalizing function $\sigma(t) > 0$, such that

$$\lim_{t \to x^*} F_t(x\sigma(t)) = H_\gamma(x) := \begin{cases} 1 - (1+\gamma x)^{-\frac{1}{\gamma}}, & \gamma \neq 0, \\ 1 - \exp(-x), & \gamma = 0, \end{cases}$$

for all $1 + \gamma x > 0$ and $x > 0$, where H_γ is the GPD function and the Pareto distribution is obtained when $\gamma > 0$.

Within the POT framework, estimation methods such as the MLE and the PWM estimation are well-explored. Hill (1975) proposed a semiparametric maximum likelihood approach to infer the tail behavior of a Zipf type distribution, i.e. the case $\gamma > 0$. Suppose that an observed sequence $(X_n)_{n \in \mathbb{Z}}$ has a cumulative distribution function F. Then consider the ordered statistics $X_{1:n} \leq X_{2:n}, \ldots, \leq X_{n:n}$ and a high threshold $X_{n-k_n:n}$. Since the exceedances are asymptotically Pareto distributed, the exponential distribution with mean γ provides an approximation to the distribution of the logarithm-transformed excess ratio $\log(\frac{X_{n-i+1:n}}{X_{n-k_n:n}})$, $i = 1, \ldots, k_n$, that is,

$$\Pr[\log(\frac{X_{n-i+1:n}}{X_{n-k_n:n}}) < x] \approx 1 - \exp(-\frac{1}{\gamma}x), \qquad x > 0.$$

The weak consistency of the Hill estimator is achieved for any sequence $k_n \to \infty, \frac{k_n}{n} \to 0$ as $n \to \infty$ (Mason, 1982) and strong consistency is proved for any sequence $\frac{k_n}{\log\log n} \to 0, \frac{k_n}{n} \to 0$ as $n \to \infty$ (Deheuvels, Haeusler, & Mason, 1988). Moreover, Haeusler and Teugels (1985) proved that under certain extra conditions, the Hill estimator is asymptotically normally distributed with convergence rate $\sqrt{k_n}$, and its asymptotic variance is γ^2. Given that the Hill estimator is only appropriate for $\gamma > 0$, Dekkers et al. (1989) proposed a moment estimator which handles the general case $\gamma \in \mathbb{R}$ and provided its asymptotic normality. Furthermore, Drees et al. (2004) showed a MLE estimator that can be applied for $\gamma > -\frac{1}{2}$ by constructing the GPD likelihood functions based on the empirical excesses $Y_i := X_{n-i+1:n} - X_{n-k_n:n}$ for $i = 1, \ldots, k_n$. Zhou (2009) showed the existence and consistency of the solution of likelihood equations in Drees et al. (2004) using the first order condition only. Additionally, the second order condition implies the asymptotic theory of the MLE estimator for $\gamma > -\frac{1}{2}$ (Drees et al., 2004). Furthermore, Zhou (2010) proved the asymptotic normality for $-1 < \gamma \leq -\frac{1}{2}$. As an alternative of the MLE estimator, Hosking and Wallis (1987) motivated the PWM estimators for the GPD parameters. The existence of unbiased estimators for probability-weighted moments are given when $\gamma < 1$, and the asymptotic normality is obtained when $\gamma < \frac{1}{2}$. Via simulations in the i.i.d. case, they suggested that the PWM parameter estimators would be preferable when $\gamma > \frac{1}{5}$ due to a smaller bias.

Another prevailing approach to estimate the extreme value index is the BM method. Different from the POT, the BM firstly divides the i.i.d. sample into k_n blocks with constant block size m_n. Then from the domain of attraction condition (1), the block maxima are asymptotically i.i.d. GEV distributed with extreme value index γ. By fitting the block maxima into the GEV distribution in (2), one obtains the MLE estimators for the GEV parameters. The existence of consistent MLE estimators is proved

by Dombry (2015) under the first order extreme value condition with $\gamma > -1$ and for any sequence $m := m(n)$ such that $\frac{m(n)}{\log n} \to \infty$ as $n \to \infty$. Moreover, Dombry and Ferreira (2017) established the asymptotic normality of the MLE estimators under both the first order condition (with $\gamma > -\frac{1}{2}$) and the second order condition. Particularly, when γ is positive, the distribution function G_γ is the Fréchet distribution. Bücher and Segers (2018b) proved the unique existence of the solution of the likelihood function based on the Fréchet distribution. The consistent PWM parameter estimators for the BM approach are established by Hosking et al. (1985), given $\gamma < 1$. The asymptotic normality of the PWM estimators are proved for $\gamma < \frac{1}{2}$. Ferreira and de Haan (2015) established the asymptotic normality for the disjoint BM-PWM extreme quantile estimator in the i.i.d. case and carried out a theoretical comparison with the POT-PWM estimator in terms of the extreme value index estimation and quantile estimation. The disjoint BM-PWM estimator is suggested to be more efficient. Furthermore, Dombry and Ferreira (2017) theoretically compared the MLE/PWM estimators under the POT/BM (the disjoint BM) methods in the i.i.d. case. In light of extreme value index estimation, it is indicated that the BM-MLE estimator is the most efficient, while the POT-MLE estimator has the smallest asymptotic bias and the minimal optimal asymptotic mean squared error.

2.2 SERIAL DEPENDENCE

As mentioned in the introduction, the assumption of an i.i.d. underlying sequence may be restrictive and unrealistic in practice. We weaken the i.i.d. assumption to a strictly stationary sequence, that is, for any $h \in \mathbb{Z}$,

$$(X_{i_1}, \ldots, X_{i_n}) \stackrel{d}{=} (X_{i_1+h}, \ldots, X_{i_n+h}).$$

For the POT approach, Drees (2003) showed that if the serial dependence of the underlying sequence is weak, i.e. the underlying sequence is β-mixing, then the POT estimators are still consistent. Recall that the β-coefficients are defined as

$$\beta(\ell) := \sup_{m \in \mathbb{N}, A_i \in \mathcal{A}_1^m, B_j \in \mathcal{B}_{m+\ell+1}^\infty} \frac{1}{2} \sum_i^I \sum_j^J |\Pr(A_i \cap B_j) - \Pr(A_i)\Pr(B_j)|,$$

where $\mathcal{A}_1^m := \sigma(X_1, \ldots, X_m)$ is the σ-algebra generated by $(X_1, \ldots, X_m)$ and $\mathcal{B}_{m+\ell+1}^\infty := \sigma(X_{m+\ell+1}, X_{m+\ell+2}, \ldots)$ is the σ-algebra generated by $(X_{m+\ell+1}, X_{m+\ell+2}, \ldots)$. Then the sequence is called β-mixing (or absolute regular) if

$$\lim_{\ell \to \infty} \beta(\ell) = 0. \tag{3}$$

By dividing $\Pr(A_i)$ on the both sides of (3), we obtain

$$\lim_{\ell \to \infty} \sup_{m \in \mathbb{N}, A_i \in \mathcal{A}_1^m, B_j \in \mathcal{B}_{m+\ell+1}^\infty} \frac{1}{2} \sum_i^I \sum_j^J |\Pr(B_j|A_i) - \Pr(B_j)| = 0.$$

Note that A_i is a set containing past events and B_j is a set containing future events. Therefore, if a strictly stationary sequence is β-mixing, the dependence between the past and future events vanishes as the time

interval increases. Furthermore, the condition (C1) in Drees (2003) ensures that the dependence vanishes sufficiently fast such that it is considered as weak. Many time series models satisfy these conditions. For instance, the autoregressive moving average (ARMA), the autoregressive conditional heteroskedasticity (ARCH) and the generalized ARCH (GARCH) time series are geometrically β-mixing. Furthermore, the asymptotic normality of the extreme quantile estimator is established under such serial dependence conditions.

Within the BM framework, Leadbetter (1983) proved that if the underlying strictly stationary sequence satisfies some mixing condition, then the normalized block maxima extracted from the stationary sequence with extremal index θ ($\theta \in [0,1]$) are asymptotically i.i.d. GEV distributed with shape parameter γ. Consequently, the asymptotic normality of the MLE estimator based on the block maxima extracted from a stationary sequence is established for the two-parameter Fréchet distribution by Bücher and Segers (2018b). The serial dependence does not affect the consistency and efficiency of the MLE estimator under certain conditions. Moreover, the shape parameter is the same as in the i.i.d. case, only the scale and location parameters are affected by the extremal index (McNeil, 1998). Therefore, the estimation of the extremal index is required for the extrapolation for extreme quantile estimation. Northrop (2015) constructed a semiparametric maxima estimator $\hat{\theta}^N$ for the extremal index and showed that it is more efficient than parametric counterparts via simulation. Since the asymptotic distribution of the Northrop estimator $\hat{\theta}^N$ is difficult to derive, Berghaus and Bücher (2018) proposed an asymptotic equivalent variant $\hat{\theta}^B$ of $\hat{\theta}^N$ and improved the bias reduction scheme. The consistency and asymptotic normalities of the MLE estimator $\hat{\theta}^B$ based on both disjoint blocks and sliding blocks are proved. It is also verified that $\hat{\theta}^B$ based on sliding blocks can be substantially more efficient than based on disjoint blocks. The reduction in asymptotic variance is independent of the value of shape parameter.

The sliding blocks approach can also be applied to the PWM estimators and the MLE estimator to gain efficiency for the BM approach. The consistency and other asymptotic property of the sliding BM-PWM estimator require further research. It is reasonable to expect that the sliding PWM estimators may be distorted, causing a larger asymptotic bias compared to the disjoint PWM estimators or the sliding PWM estimators in the i.i.d. case. This is because that on the one hand, the sliding block maxima are heavily correlated and not asymptotically independent, not even for an i.i.d. underlying sequence. On the other hand, the situation is worse under serial dependence. Hence, the degree of serial dependence in sliding block maxima is "doubled" in a way, which introduces a further approximation between F and G_γ. Nevertheless, Bücher and Segers (2018a) proved the asymptotic normality of the sliding BM-MLE estimator for the Fréchet distribution under serial dependence. Actually it should be referred as maximum quasi-likelihood estimator, since the log-likelihood is constructed by taking the sliding block maxima as asymptotic independent. It is shown that the asymptotic variance of the sliding BM-MLE estimator is substantially smaller than the disjoint BM-MLE estimator, while the asymptotic bias is the same.

The main contribution of this paper is that it focuses on the estimation of the extreme quantile, rather than the single extreme value index estimation, under serial dependence. In applications, it is often the extreme quantile that is of interest. And there are still some gaps between the asymptotic properties of the extreme quantile estimators and the asymptotic normalities of the parameter estimators, especially for

the BM method where the extremal index involves into the quantile estimation under serial dependence. Therefore, this paper presents the finite-sample performance comparisons of six widely-used extreme quantile estimators in order to provide some insights.

3 Methodology

In this section, we review the EVT at first, then introduce the POT approach and the BM approach in the i.i.d. case. The extrapolation follows by a discussion of the estimations under serial dependence. Consequently, the sliding BM method is employed to gain efficiency. Lastly, the six competing extreme quantile estimators are summarized and the evaluation criteria are listed.

3.1 Extreme value theory

The necessary and sufficient condition for $F \in D(G_\gamma)$ with $\gamma \in \mathbb{R}$ can be presented in various ways, and one of them is the following criterion.

Theorem 3.1 *Let F be a common but unknown continuous distribution function. Then $F \in D(G_\gamma)$, if and only if for some $\gamma \in \mathbb{R}$, the following condition holds*

$$\lim_{t \to x^*} \inf_{0<a<\infty} \sup_{0 \le x < \infty} |[1 - F_t(x)] - \exp\{-\int_0^{\frac{x}{a}} [(1+\gamma y)_+]^{-1} dy\}| = 0, \tag{4}$$

where for any y, $y_+ = \max(0, y)$.[3]

When and only when the condition (4) holds, it follows that

$$\lim_{t \to x^*} \sup_{0 \le x < \infty} |[1 - F_t(x\sigma(t))] - \exp\{-\int_0^{x} [(1+\gamma y)_+]^{-1} dy\}| = 0,$$

which means that if t is sufficiently large, the conditional distribution of excesses $X - t$ given $X > t$ is very nearly of the GPD function form

$$\lim_{t \to x^*} \sup_{0 < x < \infty} |F_t(x\sigma(t)) - H_\gamma(x)| = 0, \qquad 1 + \gamma x > 0. \tag{5}$$

In the case $\gamma > 0$, G_γ is the Fréhcet distribution and (1) becomes

$$\lim_{n \to \infty} \Pr(\frac{M_n}{a_n} \le x) = \lim_{n \to \infty} F^n(a_n x) = \exp(-x^{-\frac{1}{\gamma}}), \tag{6}$$

for $x > 0$, and some scale constants $a_n > 0$. Further, $F \in D(G_\gamma)$ with $\gamma > 0$ is equivalent to

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\frac{1}{\gamma}}, \tag{7}$$

[3] It is showed that the extremal distribution functions have the form $G_\gamma(x) \equiv \exp\{-\int_0^{\frac{x-b}{a}} [(1+\gamma y)_+]^{-1} dy\}$, where a, b and γ are the scale, location and shape parameter respectively, with $0 < a < \infty$, $-\infty < b$, $\gamma < \infty$ (Pickands, 1975).

for all $x > 0$. From (7), the excess ratios X/t are asymptotically i.i.d. Pareto distributed with shape parameter $\frac{1}{\gamma}$, that is,

$$\lim_{t\to\infty} \Pr(\frac{X}{t} > x | X > t) = x^{-\frac{1}{\gamma}}, \qquad x > 1.$$

And it follows that the log excess ratio $\log(\frac{X}{t})$ is asymptotically i.i.d. exponentially distributed with mean γ, i.e.

$$\Pr[\log(\frac{X}{t}) < x] = 1 - \exp(-\frac{1}{\gamma}x), \qquad x > 0. \tag{8}$$

3.2 Estimation in the identical and independent distributed case

3.2.1 Estimation base on the POT approach

Let $(X_n)_{n\in\mathbb{Z}}$ be an i.i.d. sequence with cumulative distribution function F, and $X_{1:n} \leq X_{2:n}, \ldots, \leq X_{n:n}$ be the ordered statistics. Naturally, $X_{n-k_n:n}$ can be considered as a high threshold as if k_n is an intermediate sequence of integers

$$k_n \to \infty, \qquad \frac{k_n}{n} \to 0, \ \ as\ n \to \infty. \tag{9}$$

Now consider the PWM estimator based on the POT approach. The PWM's of a continuous random variable X with distribution function F are the quantities

$$M_{p,r,s} = E[X^p (F(X))^r (1 - F(x))^s].$$

For the GPD, it is convenient to work with the quantities

$$\alpha_s = M_{1,0,s} = E[X(1 - F(x))^s],$$

which exist for $\gamma < 1$. Consequently, the scale parameter $\sigma(t)$ and the shape parameter γ for the GPD are respectively computed by

$$\sigma(t) = \frac{2\alpha_0\alpha_1}{\alpha_0 - 2\alpha_1}, \qquad \gamma = 2 - \frac{\alpha_0}{\alpha_0 - 2\alpha_1}.$$

From (5), the PWM estimators $\hat{\sigma}(t)$ and $\hat{\gamma}$ are obtained when replacing α_0 and α_1 above by their empirical estimators (Ferreira & de Haan, 2015)

$$\hat{\alpha_0} = \frac{1}{k_n}\sum_{i=1}^{k_n}(X_{n-i+1:n} - X_{n-k_n:n}) \tag{10}$$

and

$$\hat{\alpha_1} = \frac{1}{k_n}\sum_{i=1}^{k_n}\frac{i-1}{k_n}(X_{n-i+1:n} - X_{n-k_n:n}). \tag{11}$$

Therefore, the scale estimator and the extreme value index estimator are respectively given as

$$\hat{\sigma}(t) = \frac{2\hat{\alpha}_0\hat{\alpha}_1}{\hat{\alpha}_0 - 2\hat{\alpha}_1}, \qquad \hat{\gamma} = 2 - \frac{\hat{\alpha}_0}{\hat{\alpha}_0 - 2\hat{\alpha}_1}.$$

Differently, the Hill estimator, a MLE estimator designed for the Pareto distribution (i.e. the GPD with $\gamma > 0$), is obtained by solving the following log-likelihood equation based on (8)

$$(\frac{1}{\gamma})^{n-1} \exp(-\frac{1}{\gamma}\sum_{i=1}^{n} \ln \frac{X_{n-i+1:n}}{X_{n-k_n:n}})(n - \frac{1}{\gamma}\sum_{i=1}^{n} \ln \frac{X_{n-i+1:n}}{X_{n-k_n:n}}) = 0,$$

that is,

$$\hat{\gamma}^H = \frac{1}{k_n}\sum_{i=1}^{k_n} \ln \frac{X_{n-i+1:n}}{X_{n-k_n:n}}.$$

3.2.2 Estimation based on the BM approach

Split the i.i.d. sequence $(X_n)_{n\in\mathbb{Z}}$ into k_n disjoint blocks with block size m, where m satisfying the following condition

$$\frac{m(n)}{\log n} \to \infty, \qquad n \to \infty. \tag{12}$$

And the ith disjoint block maximum is defined as

$$M_{i,m}^d = \max(X_{(i-1)m+1}, \ldots, X_{im}), \quad i = 1, \ldots, k_n.$$

The PWM estimators of the GEV distribution for $\gamma \neq 0$ are given by

$$\beta_r = \frac{1}{r+1}[b_m - \frac{a_m}{\gamma}(1 - (r+1)^\gamma \Gamma(1-\gamma))], \qquad \gamma < 1,$$

where $a_m > 0$ and b_m are the scale and the location parameter of the GEV distribution, respectively. An empirical estimator of β_r is based on the ordered block maxima $M_{1:k_n}^d \leq M_{2:k_n}^d \leq \ldots \leq M_{k_n:k_n}^d$,

$$\hat{\beta}_r = \frac{1}{k_n}\sum_{i=1}^{k_n} \frac{(i-1)(i-2)\ldots(i-r)}{(k_n-1)(k_n-2)\ldots(k_n-r)} M_{i:k_n}^d,$$

and $\hat{\beta}_0 = k_n^{-1}\sum_{i=1}^{k_n} M_{i:k_n}^d$. Ferreira and de Haan (2015) provided an estimator of $\hat{\gamma}$ as

$$\hat{\gamma} = \frac{1}{\ln 2} \ln(\frac{4\hat{\beta}_3 - \hat{\beta}_0}{2\hat{\beta}_1 - \hat{\beta}_0} - 1),$$

which is the solution of $(4\beta_3 - \beta_0)(2\beta_1 - \beta_0)^{-1} = (1 - 4^{\hat{\gamma}})(1 - 2^{\hat{\gamma}})^{-1}$. Given $\hat{\gamma}$, the scale parameter can be computed as

$$\hat{a}_m = \frac{(2\hat{\beta}_1 - \hat{\beta}_0)\hat{\gamma}}{\Gamma(1-\hat{\gamma})(2^{\hat{\gamma}-1})}.$$

Consequently, the location estimator, which regarded as the estimator of the intermediate quantile in the

extrapolation discussed in Section 3.3, is given by

$$\hat{b}_m = \hat{\beta}_0 + \hat{a}_m \frac{1 - \Gamma(1-\hat{\gamma})}{\hat{\gamma}}.$$

As an alternative, the MLE for the case $\gamma > 0$ is defined as fitting the block maxima to the Fréchet log-likelihood function:

$$L(\omega|\mathbf{x}) = \sum_{i=1}^{k_n} \ell_\omega(x_i), \qquad \omega = (\gamma, a_m) \in (0,\infty)^2 = \Omega, \tag{13}$$

where $x_i = M_{i,m}^d \bigvee c$, $c > 0$ is the left-truncated block maximum, and where

$$\ell_\omega(x) = \log(\frac{1}{\gamma a_m}) - (\frac{x}{a_m})^{-\frac{1}{\gamma}} - (\frac{1}{\gamma} + 1)\log(\frac{x}{a_m})$$

is the individual contribution to the Fréchet log-likelihood. The existence and uniqueness of the MLE estimator are provided that if the scalars $x_1, \ldots, x_{k_n} \in (0,\infty)$ are not all identical, then there exists a unique maximizer parameter vector

$$\hat{\omega}(\mathbf{x}) = (\hat{\gamma}(\mathbf{x}), \hat{a}_m(\mathbf{x})) = \underset{\omega\in\Omega}{argmax} L(\omega|\mathbf{x}).$$

Specifically, $\hat{\gamma}(\mathbf{x})$ is the unique solution of

$$\Psi_{k_n}(\gamma|\mathbf{x}) = \gamma + \frac{\frac{1}{k_n}\sum_{i=1}^{k_n} x_i^{-\frac{1}{\gamma}} \log(x_i)}{\frac{1}{k_n}\sum_{i=1}^{k_n} x_i^{-\frac{1}{\gamma}}} - \frac{1}{k_n}\sum_{i=1}^{k_n}\log(x_i) = 0. \tag{14}$$

From (14), the MLE estimator of γ is also scale invariant : $\hat{\gamma}(c\mathbf{x}) = \hat{\gamma}(\mathbf{x})$.

3.3 Extrapolation

Another characterization of the necessary and sufficient condition for $F \in D(G_\gamma)$ with $\gamma \in \mathbb{R}$ is the so-called first order condition (15).

Theorem 3.2 *Let $U = (\frac{1}{1-F})^{\leftarrow}$ be the left continuous inverse function of $\frac{1}{1-F}$. Then $F \in D(G_\gamma)$ if and only if there exists a function $a(t) > 0$ such that*

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{15}$$

for all $x > 0$.

By taking $tx = \frac{1}{p_n}$ and $t = \frac{n}{k_n}$ with k_n satisfying the condition (9), (15) implies that for the extreme quantile $x_{p_n} := F^{-1}(1-p_n)$ with $np_n = O(1)$, the extrapolation is as

$$\frac{np_n}{k_n} = (\gamma \frac{x_{p_n} - x_{\frac{k_n}{n}}}{a(\frac{n}{k_n})} + 1)^{-\frac{1}{\gamma}}, \tag{16}$$

and the extreme quantile estimator is the following

$$\hat{x}_{p_n} = \hat{U}(\frac{n}{k_n}) + \hat{a}(\frac{n}{k_n})\frac{(\frac{k_n}{np_n})^{\hat{\gamma}} - 1}{\hat{\gamma}}, \tag{17}$$

where $\hat{U}(\frac{n}{k_n})$, $\hat{a}(\frac{n}{k_n})$ and $\hat{\gamma}$ are proper estimators of $U(\frac{n}{k_n})$, $a(\frac{n}{k_n})$ and γ, respectively. In the POT approach, $U(\frac{n}{k_n})$ is estimated by the empirical intermediate quanitle $X_{n-k_n:n}$, meanwhile the estimates of $a(\frac{n}{k_n})$ and γ are provided by the estimators $\hat{\sigma}(\frac{n}{k_n})$ and $\hat{\gamma}$ in methods like the PWM method and the MLE. Under the condition (9), the estimators $\hat{\sigma}(\frac{n}{k_n})$ and $\hat{\gamma}$ respectively converge to $a(\frac{n}{k_n})$ and γ in probability as $n \to \infty$.[4]

This extrapolation approach is used in the general case $\gamma \in \mathbb{R}$. Regarding the estimators of $a(\frac{n}{k_n})$ and γ, the PWM is built-in to be applied in the case $\gamma < 1$, thus one should always use the extrapolation in (16) for the PWM estimator, so do the MLE estimator proposed by Drees et al. (2004), the moment estimator constructed by Dekkers et al. (1989) and the moment estimator in Hosking and Wallis (1987).

If F exhibits heavy tails, i.e. $\gamma > 0$, a possible choice for the auxiliary function $a(t)$ is $a(t) = \gamma U(t)$ (Dombry, 2015). Then the extrapolation turns to be

$$\frac{np_n}{k_n} \approx (\frac{x_{p_n}}{x_{\frac{k_n}{n}}})^{-\frac{1}{\gamma}},$$

which implies a quantile estimator as

$$\hat{x}_{p_n} = \hat{U}(\frac{n}{k_n})(\frac{k_n}{np_n})^{\hat{\gamma}}. \tag{18}$$

The quantile estimator in (18) can be regarded as a special case of (17), where there is a linear relationship between the location parameter and the scale parameter. This expression can be employed by the Hill estimator, and the corresponding quantile estimate is referred as the Weissman estimator (Weissman, 1978).

In the BM framework, we consider the extrapolation in the following way. Let m satisfy the condition (12), and (1) can be written as

$$\lim_{m\to\infty} \frac{1}{-m \log F(a_m x + b_m)} = (1 + \gamma x)^{\frac{1}{\gamma}}, \tag{19}$$

for all $1 + \gamma x > 0$, and $\gamma \in \mathbb{R}$. Define $V = (\frac{1}{-\log F})^{\leftarrow}$ as the left continuous inverse function of $\frac{1}{-\log F}$, then by taking

$$\lim_{m\to\infty} \frac{1}{-m \log F(a_m x + b_m)} = y, \qquad (1 + \gamma x)^{\frac{1}{\gamma}} = y,$$

one obtains

$$x = \lim_{m\to\infty} \frac{V(my) - b_m}{a_m}, \qquad x = \frac{y^{\gamma} - 1}{\gamma}.$$

[4]The condition (9) is obviously not the only condition of the convergence. For different estimators, the range of γ matters as well. For instance, the first order condition (15) supposes to hold for $\gamma > -1$ and $\gamma \neq 0$ for the MLE estimator proposed by Drees et al. (2004) (Zhou, 2009), and for $\gamma < 1$ for the PWM estimator in Hosking and Wallis (1987).

Consequently, $F \in D(G_\gamma)$ with $\gamma \in \mathbb{R}$ is equivalent to

$$\lim_{m\to\infty} \frac{V(mx) - b_m}{a_m} = \frac{x^\gamma - 1}{\gamma}.$$

Recall that $F(x_{p_n}) = 1-p_n$, then define $q_n := -\log(1-p_n) = -\log(F(x_{p_n}))$, and therefore, $x_{p_n} = V(\frac{1}{q_n})$. The quantile estimator is as

$$\hat{x}_{p_n} = \hat{b}_m + \hat{a}_m \frac{(mq_n)^{-\hat{\gamma}} - 1}{\hat{\gamma}}. \tag{20}$$

Note that for any distribution function F, $-\log F(x) \sim (1 - F(x))$ as $F(x) \to 1$, and thus $mq_n = -m\log(1 - p_n) \sim 1 - (1 - p_n)^m$ as $1 - p_n \to 1$. Further, taking the first order Taylor expansion of $1 - (1 - p_n)^m$ at $p_n = 0$, the quantile estimator becomes

$$\hat{x}_{p_n} = \hat{b}_m + \hat{a}_m \frac{(mp_n)^{-\hat{\gamma}} - 1}{\hat{\gamma}}. \tag{21}$$

(21) is sufficiently close to (20) and is adopted in several papers such as Ferreira and de Haan (2015). And it is attractive since it simplifies the theoretical comparison of the POT method and the BM method in extreme quantile estimation. With $m = \frac{n}{k_n}$, (21) shares the same structure as (17) except that the distribution parameters are estimated in the BM framework.

If $\gamma > 0$, the GEV distribution G_γ is the Fréchet distribution in (6). As a consequence, (19) becomes

$$\lim_{m\to\infty} \frac{1}{-m\log F(a_m x)} = x^{1/\gamma},$$

for all $x > 0$, and thus,

$$\lim_{m\to\infty} \frac{V(mx)}{a_m} = x^\gamma.$$

Similarly, one carries out a quantile estimator as

$$\hat{x}_{p_n} = \hat{a}_m (mq_n)^{-\hat{\gamma}}. \tag{22}$$

In (20) to (22), $\hat{a}_m$, $\hat{b}_m$ and $\hat{\gamma}$ are respectively the suitable estimators for a_m, b_m and γ that can be provided by the PWM method and the MLE.

Overall, in the POT approach, the location parameter is estimated by the empirical intermediate quantile, and if $\gamma > 0$, one can estimate the scale parameter via multiplying the empirical intermediate quantile by the estimated extreme value index. However, in the BM framework, the location parameter is ignored if $\gamma > 0$. Moreover, the estimate of the scale parameter equals to the estimate of the intermediate quantile.

3.4 Estimation under serial dependence

Denote a strictly stationary sequence as $(\tilde{X}_n)_{n=1}^\infty$ and the corresponding maximum of the sequence as $\tilde{M}_n$.

Under serial dependence, Drees (2003) established the asymptotic normality of a class of the POT

extreme quantile estimators, including the Hill estimator and the PWM estimator applied in this paper. Under the mild structural conditions on $(\tilde{X}_n)_{n=1}^{\infty}$, the estimation procedure is in line with the i.i.d. case. The estimator can be applied for time series that are geometrically β-mixing, such as ARMA models with balanced heavy tails and (G)ARCH models.

However, the quantile estimation in the BM approach is modified. The following theorems hold when the serial dependence exhibits.

Theorem 3.3 *Let $(\tilde{X}_n)_{n=1}^{\infty}$ be a strictly stationary sequence with marginal distribution function F, and a sequence of constants u_n, satisfies the distributional mixing condition $D(u_n)$ in Leadbetter (1983). Then*

$$\Pr(\tilde{M}_n \leq u_n) \to e^{-\theta\tau} \quad \textit{if and only if} \quad n[1 - F(u_n)] \to \tau, \tag{23}$$

where $\tilde{M}_n$ is the maximum of the stationary sequence. The series $(\tilde{X}_n)_{n=1}^{\infty}$ is said to have the extremal index θ, $\theta \in [0, 1]$. If the sequence (u_n) also satisfies the anti-cluster condition $D^{'}(u_n)$ in Leadbetter (1983), it implies that the extremal index θ equals to unity.

Theorem 3.4 *Suppose that the stationary sequence $(\tilde{X}_n)_{n=1}^{\infty}$ has extremal index θ, $\theta \in (0, 1]$. Denote its associate i.i.d. sequence as $(X_n)_{n=1}^{\infty}$ with the corresponding maximum M_n. Then*

$$\lim_{n\to\infty} \Pr(\frac{M_n - b_n}{a_n} \leq x) = G_\gamma(x)$$

holds for a non-degenerate $G_\gamma(x)$ if and only if

$$\lim_{n\to\infty} \Pr(\frac{\tilde{M}_n - b_n}{a_n} \leq x) = G_\gamma^\theta(x) \tag{24}$$

with $G_\gamma^\theta(x)$ also non-degenerate.

The two theorems show that if the long-range dependence in the stationary time series is weak, and there is no tendency to form clusters of large values, then the normalized block maxima of stationary series and associate i.i.d. series have the same type of limiting distribution. Moreover, the normalization constants are the same for the two series. However, the anti-cluster condition $D^{'}(u_n)$ may not be tenable for a financial series, for instance in a stationary (G)ARCH process, the clusters of volatility lead to clusters of extreme values (McNeil, 1998). Then the clustering of extreme values leads to the modifications of the location and the scale parameters in the i.i.d. case, such that the limiting distributions of block maxima of both the stationary and the associate i.i.d. series are of the same type. It can be verified that condition (24) is equivalent to the following condition

$$\lim_{n\to\infty} \Pr(\frac{\tilde{M}_n - b_n^*}{a_n^*} \leq x) = G_\gamma(x)$$

with normalization constants a_n^* and b_n^* given by

$$a_n^* = a_n\theta^\gamma, \quad b_n^* = \frac{a_n}{\gamma}(\theta^\gamma - 1) + b_n.$$

Thus, in order to use the extreme quantile estimator in (21) and (22), besides the necessary estimation for GEV parameters, an additional extremal index θ needs to be estimated.

Let $u_n = F^{\leftarrow}(1-\frac{x}{n})$, then $n\bar{F}(u_n) = x$ as $n \to \infty$ where $\bar{F} = 1 - F$. From (23), it can be shown that $Z_n = n(1-N_n)$ with $N_n = F(\tilde{M}_n)$ is asymptotically exponential distributed with mean $\frac{1}{\theta}$ as follows: for any $x > 0$,

$$\begin{aligned} \lim_{n\to\infty} \Pr(n(1-N_n) \geq x) &= \lim_{n\to\infty} \Pr(\tilde{M}_n \leq u_n) \\ &= \lim_{n\to\infty} \Pr(n\bar{F}(\tilde{M}_n) \geq n\bar{F}(u_n)) \\ &= \exp(-\theta x). \end{aligned} \tag{25}$$

Berghaus and Bücher (2018) provided a MLE $\hat{\theta}$ for the extremal index based on a sample of block maxima. More precisely, consider the disjoint block maxima extracted from $(\tilde{X}_n)_{n\in\mathbb{Z}}$, that is

$$\tilde{M}^d_{i,m} = \max(\tilde{X}_{(i-1)m+1}, \ldots, \tilde{X}_{im}), \quad i = 1, \ldots, k_n.$$

Let $N^d_{i,m} = F(\tilde{M}^d_{i,m})$ and $Z^d_{i,m} = m(1 - N^d_{i,m})$. If m is sufficiently large, then by (25), the limiting distribution of the unobserved random variables $Z^d_{1,m}, \ldots, Z^d_{k_n,m}$ is the exponential distribution with mean $\frac{1}{\theta}$. As a consequence, the pseudo-likelihood function is given by taking the block maxima as asymptotically independent (Northrop, 2015),

$$L_{exp}(\theta; Z^d_{i,m}) = \theta^{\gamma} exp(-\theta \sum_{i=1}^{k_n} Z^d_{i,m}). \tag{26}$$

Since the distribution function F is unknown, one can use the empirical cumulative distribution function $\hat{F}_n(x) = n^{-1}\sum_{s=1}^{n} \mathbb{1}(X_s \leq x)$ where $\mathbb{1}(\cdot)$ is an indicator function as an estimate of F. The MLE estimator for θ is then

$$\hat{\theta}_d = (\frac{1}{k_n}\sum_{i=1}^{k_n} \hat{Z}^d_{i,m})^{-1}, \tag{27}$$

where $\hat{Z}^d_{i,m} = m(1 - \hat{N}^d_{i,m})$ and $\hat{N}^d_{i,m} = \hat{F}_n(\hat{M}^d_{i,m})$. Furthermore, a bias correction is necessary since an asymptotic bias term may appear. In this paper, we adopt the bias reduction scheme suggested by Berghaus and Bücher (2018). Define $\hat{T}^d_m = \frac{1}{k_n}\sum_{i=1}^{k_n} \hat{Z}^d_{i,m}$. Since $\sqrt{k_n}(\hat{T}_m - \theta^{-1}) \rightsquigarrow N(0, \sigma^2)$ [5], through the Taylor expansion of $\hat{T}^{-1}_m - \theta$ at $\hat{T}^{-1}_m = \theta$, a bias-reduced estimator is as following

$$\hat{\theta}_{bc} = \hat{\theta} - k_n^{-1}\hat{\theta} - k_n^{-1}\hat{\theta}^3\hat{\sigma}^2, \tag{28}$$

where the first bias-component is due to the use of the empirical cumulative distribution function and $\hat{\sigma}^2$ is the variance estimator. Define

$$\hat{B}^d_{i,m} = \hat{Z}^d_{i,m} + \sum_{s\in I_i} \frac{1}{k_n}(\sum_{i=1}^{k_n} \mathbb{1}(\hat{F}_n(\tilde{X}_s)) > 1 - \frac{\hat{Z}^d_{i,m}}{m}) - 2\hat{T}^d_m,$$

[5] Here, $(\hat{T}_m, \sigma^2) \in \{(\hat{T}^d_m, \sigma^2_d), (\hat{T}^{sl}_m, \sigma^2_{sl})\}$.

where $I_i = \{(i-1)k_n + 1, \ldots, ik_n\}$ is the ith block of indices. Then denote $\hat{\sigma}^2$ based on the disjoint block maxima as $\hat{\sigma}_d^2$, it is estimated as

$$\hat{\sigma}_d^2 = \frac{1}{k_n} \sum_{i=1}^{k_n} (\hat{B}_{i,m}^d)^2.$$

By correcting the dominating bias-components in $\hat{\theta}$, the estimations of the location and scale parameters are more accurate, so does the extreme quantile estimation.

3.5 Sliding block maxima

As mentioned before, the serial dependence in the time series enlarges the asymptotic variance of extreme quantile estimators. In order to gain efficiency, a sliding block approach is considered. Now the series $(\tilde{X}_n)_{n\in\mathbb{Z}}$ is divided into $k_{sl} = n - m + 1$ sliding blocks with block length m. Then, define the sliding block maximum as following

$$\tilde{M}_{i,m}^{sl} = max(\tilde{X}_i, \ldots, \tilde{X}_{i+m-1}), \quad i = 1, \ldots, n - m + 1.$$

The sample of sliding block maxima is stationary but not asymptotically independent. Nevertheless, the limiting distribution of a single sliding block maximum is still Fréchet. Bücher and Segers (2018a) considered a maximum quasi-likelihood function by taking the sliding block maxima as independent. Then the sliding BM-MLE estimator is given by maximizing (13) where $x_i = \tilde{M}_{i,m}^{sl} \bigvee c$ with c an arbitrary positive truncation constant. Similarly, via replacing disjoint block maxima by sliding block maxima, one obtains the sliding BM-PWM estimators. While the asymptotic biases of the disjoint and the sliding BM-MLE estimators are the same, the efficiency gain of using sliding blocks is substantial. The asymptotic variances of the sliding BM-MLE estimator of shape and scale parameters are $\frac{0.4946}{\gamma^2}$ and $0.9578\gamma^2$, respectively, whereas those of the disjoint BM-MLE estimator are $\frac{0.6080}{\gamma^2}$ and $1.1087\gamma^2$, respectively (Bücher & Segers, 2018a). Note the efficiency improvement is independent of the values of γ and θ.

In addition, the sliding BM estimator for the extremal index is modified based on (27):

$$\hat{\theta}_{sl} = \Big(\frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \hat{Z}_{i,m}^{sl}\Big)^{-1}, \quad \hat{Z}_{i,m}^{sl} = m(1 - \hat{F}_n(\tilde{M}_{i,m}^{sl})).$$

The use of sliding block maxima induces a further approximation in the pseudo-likelihood function (26) since the blocks are dependent. Similar to the other sliding block estimators, Berghaus and Bücher (2018) showed that both $\hat{\theta}_{sl}$ and $\hat{\theta}_d$ are consistent and converge at the same rate to a normal distribution. Furthermore, $\hat{\theta}_{sl}$ is proved to be more efficient than $\hat{\theta}_d$ and the variance reduction does not affected by the value of γ. Precisely,

$$\hat{\sigma}_{sl}^2 = \hat{\sigma}_d^2 - \hat{\theta}_{sl}^{-2}(3 - 4\ln 2). \tag{29}$$

3.6 Summary

To clarify the six extreme quantile estimators that are evaluated in this paper, we wrap them up as follows.

(i) the POT-MLE estimator (the Weissman estimator)

$$\hat{x}_{p_n} = \tilde{X}_{n-k_n:n}(\frac{k_n}{np_n})^{\hat{\gamma}},$$

where γ is estimated by the Hill estimator

$$\hat{\gamma} = \frac{1}{k_n}\sum_{i=1}^{k_n} \ln \frac{\tilde{X}_{n-i+1:n}}{\tilde{X}_{n-k_n:n}}.$$

(ii) the POT-PWM estimator

$$\hat{x}_{p_n} = \tilde{X}_{n-k_n:n} + \hat{\sigma}(\frac{n}{k_n})\frac{(\frac{k_n}{np_n})^{\hat{\gamma}} - 1}{\hat{\gamma}},$$

where

$$\hat{\sigma}(\frac{n}{k_n}) = \frac{2\hat{\alpha}_0\hat{\alpha}_1}{\hat{\alpha}_0 - 2\hat{\alpha}_1}, \quad \hat{\gamma} = 2 - \frac{\hat{\alpha}_0}{\hat{\alpha}_0 - 2\hat{\alpha}_1}$$

with $\hat{\alpha}_0$ and $\hat{\alpha}_1$ respectively estimated by (10) and (11) based on the ordered statistics $\tilde{X}_{n-k_n:n}, \ldots, \tilde{X}_{n:n}$.

(iii) the disjoint BM-MLE estimator

$$\hat{x}_{p_n} = \hat{a}_m(mq_n)^{-\hat{\gamma}_d},$$

where

$$\hat{a}_m = \hat{a}^*_m\hat{\theta}_d^{-\hat{\gamma}_d}, \quad \hat{\theta}_d = (\frac{1}{k_n}\sum_{i=1}^{k_n}\hat{Z}^d_{i,m})^{-1},$$

and $(\hat{\gamma}_d, \hat{a}^*_m)$ are the MLE estimators by maximizing the log-likelihood function(13) based on the left-truncated disjoint block maxima $\tilde{M}^d_{i,m}$ for $i = 1, \ldots, k_n$.

(iv) the sliding BM-MLE estimator

$$\hat{x}_{p_n} = \hat{a}_m(mq_n)^{-\hat{\gamma}_{sl}},$$

where

$$\hat{a}_m = \hat{a}^*_m\hat{\theta}_{sl}^{-\hat{\gamma}_{sl}}, \quad \hat{\theta}_{sl} = (\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\hat{Z}^{sl}_{i,m})^{-1}$$

and $(\hat{\gamma}_{sl}, \hat{a}^*_m)$ are the MLE estimators by maximizing the log-likelihood function(13) based on the left-truncated sliding block maxima $\tilde{M}^{sl}_{i,m}$ for $i = 1, \ldots, n-m+1$.

(v) the disjoint BM-PWM estimator

$$\hat{x}_{p_n} = \hat{b}_m + \hat{a}_m\frac{(mq_n)^{-\hat{\gamma}_d} - 1}{\hat{\gamma}_d}.$$

The modifications of the scale and location estimators are

$$\hat{a}_m = \hat{a}^*_m\hat{\theta}_d^{-\hat{\gamma}_d}, \quad \hat{b}_m = \hat{b}^*_m - \hat{a}_m\hat{\gamma}_d^{-1}(\hat{\theta}_d^{\hat{\gamma}_d} - 1).$$

And the parameter estimators for the GEV distribution are

$$\hat{\gamma}_d = \frac{1}{\ln 2}\ln(\frac{4b_3 - 2b_1}{2b_1 - b_0}), \quad \hat{a}_m^* = \frac{(2b_1 - b_0)\hat{\gamma}_d}{(2^{\hat{\gamma}_d} - 1)\Gamma(1 - \hat{\gamma}_d)}, \quad \hat{b}_m^* = b_0 + \hat{a}_m^* \frac{1 - \Gamma(1 - \hat{\gamma}_d)}{\hat{\gamma}_d},$$

where b_0,b_1 and b_3 are computed by $\tilde{M}_{i,m}^d$ for $i = 1, \ldots, k_n$.

(vi) the sliding BM-PWM estimator

$$\hat{x}_{p_n} = \hat{b}_m + \hat{a}_m \frac{(mq_n)^{-\hat{\gamma}_{sl}} - 1}{\hat{\gamma}_{sl}}.$$

The modifications of $\hat{a}_m$ and $\hat{b}_m$ are of the same form as disjoint blocks. And all parameters are estimated based on $\tilde{M}_{i,m}^{sl}$ for $i = 1, \ldots, n - m + 1$.

The extremal index estimators $\hat{\theta}_d$ and $\hat{\theta}_{sl}$ are bias-reduced as in (28), and for the sliding BM, the variance estimator is $\hat{\sigma}_{sl}^2$ in (29).

3.7 Evaluation criteria

The evaluation of the methods is done in three aspects: the accuracy, the efficiency, and the bias-variance tradeoff. Furthermore, the three measurements are scaled by true extreme quantile in order to be comparable across different models. The squared bias is computed by

$$bias^2 = (\frac{\hat{x}_{p_n}}{x_{p_n}} - 1)^2.$$

The efficiency of an estimator is measured by the its sample variance

$$variance = \frac{1}{S}\sum_{i=1}^{S}(\frac{\hat{x}_{i,p_n}}{x_{p_n}} - \frac{\hat{x}_{p_n}}{x_{p_n}})^2.$$

And the mean squared error (MSE) captures the bias-variance tradeoff, which is defined as

$$MSE = bias^2 + variance = \frac{1}{S}\sum_{i=1}^{S}(\frac{\hat{x}_{i,p_n}}{x_{p_n}} - 1)^2.$$

3.8 Data generating processes

In this section we evaluate the performance of the above six extreme quantile estimators under serial dependence. Since many asset returns have the stylized facts such as excess kurtosis, we focus on the heavy-tailed case i.e. $\gamma > 0$ here. Consider three time series models for $(X_n)_{n\in\mathbb{Z}}$: independent and identical distributed random variables (r.v.s), the heavy-tailed ARMA(1,1) time series and the symmetric GARCH(1,1) time series. In the first two models, three choices are considered for the distribution r.v.s in the first model and the innovations in the second model: absolute value of a student t distribution with degree of freedom 3, Pareto distribution and Fréchet distribution with shape parameter $\frac{1}{3}$. The shape parameter is chosen to have a finite variance, since an infinite variance is unrealistic in practice.

For the ARMA(1,1) model

$$X_i = \lambda_1 X_{i-1} + \epsilon_i + \phi_1 \epsilon_{i-1},$$

the parameter vector (λ_1, ϕ_1) is set to be each of the following

(i) $\lambda_1 = 0.95, \phi_1 = 0.9$,

(ii) $\lambda_1 = 0.9, \phi_1 = 0.3$,

(iii) $\lambda_1 = 0.9, \phi_1 = -0.6$,

(iv) $\lambda_1 = 0.3, \phi_1 = 0.9$,

which partially based on the settings in Drees (2003). The degree of dependence is declining from setting (i) to (iii), and it is dominated by the autoregressive parameter λ_1. In setting (iv), the dependence has short memory but it is locally strong. Note for $\lambda_1 = -\phi_1$, the i.i.d. R.V.s are observed.

For the GARCH(1,1) model, the innovation are standard Gaussian distributed, that is

$$\begin{aligned} X_i &= \sigma_i \epsilon_i, \quad \epsilon_i \sim N(0,1), \\ \sigma_i^2 &= \nu + \rho_1 X_{i-1}^2 + \psi_1 \sigma_{i-1}^2 \\ &= \nu + (\rho_1 \epsilon_{i-1}^2 + \psi_1)\sigma_{i-1}^2, \end{aligned} \tag{30}$$

with

(v) $\nu = 0.5, \rho_1 = 0.07, \psi_1 = 0.55$,

(vi) $\nu = 0.5, \rho_1 = 0.08, \psi_1 = 0.91$,

(vii) $\nu = 0.5, \rho_1 = 0.55, \psi_1 = 0.07$,

(viii) $\nu = 0.5, \rho_1 = 0.6, \psi_1 = 0.25$.

The tuning parameters ρ_1 and ψ_1 are chosen to satisfy $\rho_1 + \psi_1 < 1$ such that the time series is covariance stationary.[6] And the parameters are non-negative to ensure the positivity of σ_i^2.

Unlike in the linear models where the clustering of extreme values is due to the auto-correlations in observations, in the GARCH models, the clustering of extremes is caused by volatility clustering. Since ρ_1 and ψ_1 simultaneously decide the degree of volatility clustering and the coefficient of excess kurtosis measured relative to the Gaussian distribution, we choose the values of ρ_1 and ψ_1 such that both aspects are concerned. The PWM estimation would probably fail to capture the heavy-tail behaviour if the excess kurtosis is positively small. Hence, we consider the excess kurtosis to be varying in models. It increases from Model (v) to Model (vii). By contrast, Model (viii) has an infinite excess kurtosis. Furthermore, we measure the degree of volatility clustering by the the second moment of the random parameter $(\rho_1 \epsilon_{i-1}^2 + \psi_1)$ in (30), that is,

$$E(\rho_1 \epsilon_{i-1}^2 + \psi_1)^2 = (\rho_1 + \psi_1)^2 + 2\rho_1^2.$$

[6] This is a sufficient but not necessary condition for strict stationarity.

Then, a preliminary ascending order of the degree of volatility is Model (v) < Model (vii) < Model (viii) < Model (vi).

The quantiles x_{p_n} are estimated for $p_n = 0.001$ and $p_n = 0.0005$. Since the true quantile is unknown for all models, they are computed by pre-simulation. We simulate $S = 1000$ time series of length $n = 5 \times 10^6$ and estimate x_{p_n} by the median of empirical $(1 - p_n)$ quantiles. Next, we conduct $S = 1000$ simulation runs for each of the above-mentioned models with the fixed sample sizes $n = 1000$ and $n = 2000$. Here we choose the sample sizes in a plausible way. For many assets we are capable to collect 1000 or 2000 observations. Also the sample sizes are sufficient high allowing the block size to be large enough. When sample size $n = 1000$, it is about four-year trading days of New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotations (NASDAQ). Via Monte Carlo simulations, we obtain the empirical estimates of the scaled squared bias, variance and MSE of the six candidate estimators. For $n = 1000$, the disjoint block size m is calculated by the integer part of n/k_0 with k_0 ranging from 10 to 150 with a step 2. The POT estimators are based on the respective values $k_n = \lfloor \frac{n}{m} \rfloor$. And for $n = 2000$, k_0 ranges from 20 to 200 with a step 5.

4 Results

The scaled MSE, squared bias and variance are depicted as the functions of the effective sample size k_n. k_n refers to the number of upper ordered statistics in the POT methods, and the number of disjoint blocks in the BM methods. For the sliding BM estimators, the actual effective sample size is $n - m + 1$. Here we plot the sliding BM estimators together with other competing estimators, such that it presents the MSE curves when the block size decreases. The finite-sample performances of six estimators are evaluated by the MSE. Since the MSE equals to the sum of the squared bias and the variance, we decompose the MSE to provide more insights of the estimators performances.

4.1 Identical and independent random variables

4.1.1 MSE comparison

Figure 1 shows the finite-sample performances of six extreme quantile estimators for the i.i.d. samples. The first row indicates that the sliding BM-MLE estimator outperforms the competing estimators for all three distributions. For the i.i.d. Fréchet sample, the BM-MLE estimators (the disjoint and the sliding BM-MLE estimators) have a monotonically declining MSE curve since the BM-MLE estimators are based on the Fréchet likelihood function, and the MSE curve of the Weissman estimator is slightly U-shaped. For the i.i.d. Pareto sample, the Weissman estimator has a decreasing MSE curve, while the MSE curves of the BM-MLEs are U-shaped. Furthermore, the U shapes of the MSE curves of the BM-MLE estimators and the Weissman estimator are more obvious when the sample follows the i.i.d. student t distribution. Although the MSEs of the BM-MLE estimators climb faster than the MSE of the Weissman estimator as k_n increases, the lowest MSEs at the optimal k_n are lower. The POT-PWM estimator performs almost identically for the three i.i.d. samples, as well as the BM-PWM estimators (the disjoint and sliding BM-PWM estimators). When k_n increases, the MSE curves of the POT-PWM estimator and the BM-PWM

estimators tend to coincide.

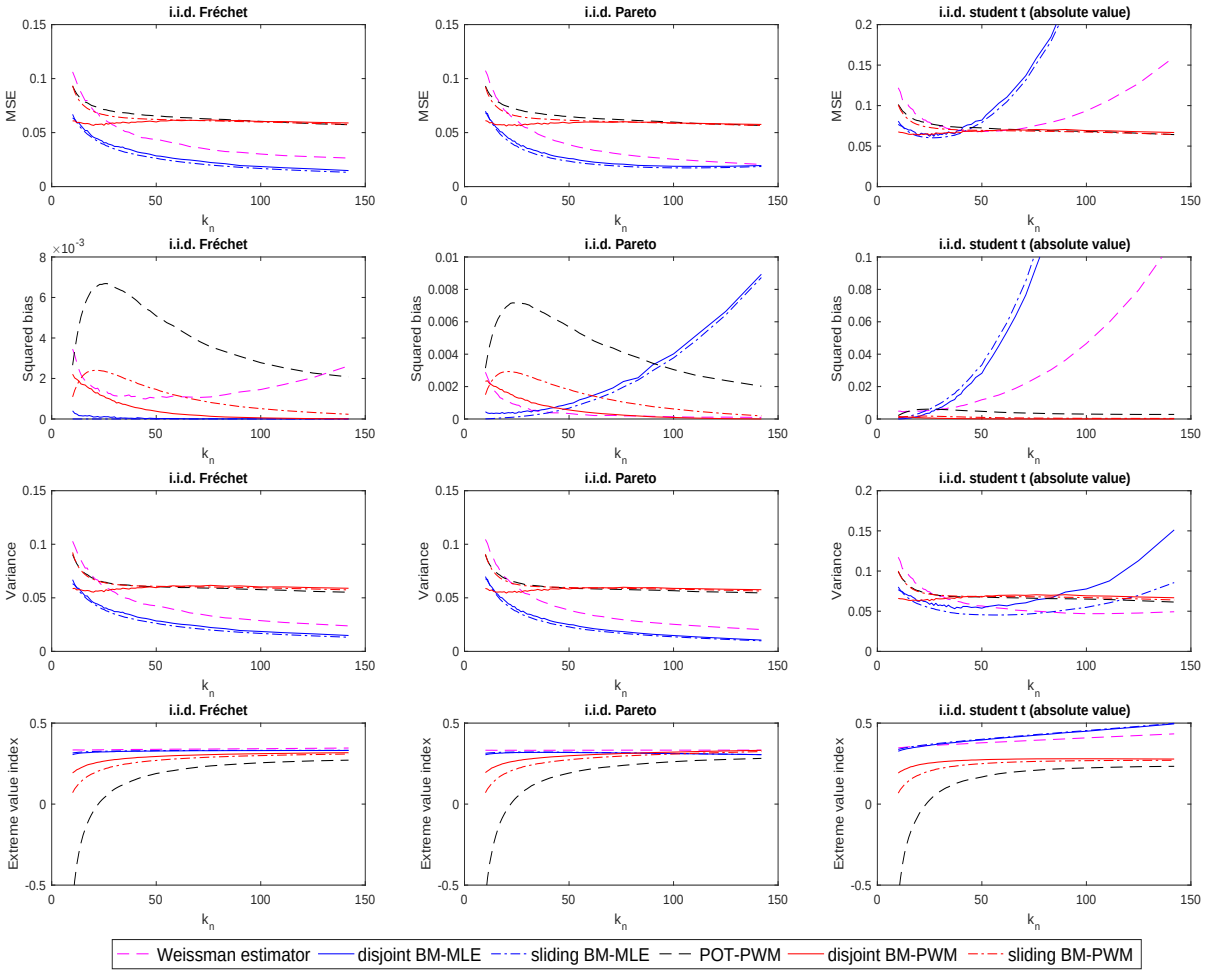

Figure 1: Performances (the scaled MSEs, scaled squared biases and scaled variances) and the extreme value index estimates in the i.i.d. samples as a function of the effective sample size, respectively. $n = 1000$, $p_n = 0.001$

4.1.2 MSE DECOMPOSITION

The second and the third rows of Figure 1 show the decomposition of MSE into the squared bias and the variance. The sliding BM-MLE processes the lowest variance for the i.i.d. Fréchet sample. The bias of the Weissman estimator is increasing in k_n, while the variance is decreasing. The disjoint BM-PWM estimator has a decreasing bias curve since its γ estimates are closer to the true γ, i.e. $\frac{1}{3}$; see the accompany figures on the γ estimates in the forth row. It has a higher variance than the POT-PWM estimator, although the variance differences among the three PWM estimators are minor. Since the MSE is dominated by the variance instead of the bias for the i.i.d. Fréchet case, the MSE of the disjoint BM-PWM estimator is higher than of the Weissman estimator and the BM-MLE estimators due to the high variance. Moreover, the POT-PWM estimator fails to capture the heavy-tailed feature in the series when k_n is low. There is an uptrend in the bias curves of the POT-PWM estimator and the sliding BM-PWM estimator. Both peaks appear when the γ estimates are adjacent to zero. One possible explanation is that the extrapolation for the POT-PWM estimator and the BM-PWM estimators requires the $\hat{\gamma}$ to be

non-zero, thus the POT-PWM estimator and the BM-PWM estimators have a high bias when the $\hat{\gamma}$ is zero or close to zero.

For the i.i.d. Pareto sample, the BM-MLE estimators have the increasing bias curves and decreasing variance curves in k_n. Therefore, the MSE decomposition suggests that the U shape is induced by the bias-variance tradeoff. Since the variance is dominant in the MSE, the U shape is not clearly presented. It also indicates that the sliding BM-MLE estimator outperforms the other estimator due to its low variance.

For the i.i.d. student t sample, the Weissman estimator and the BM-MLE estimators have higher biases than the POT-PWM estimator and the BM-PWM estimators with high values of k_n. Though the Weissman estimator has a lower variance with a larger effective sample size, the variance reduction is higher than in the other two samples. The variance reduction is negligible when the BM-MLE estimators overestimate γ. The variance increases faster in k_n if the overestimation is more serious. The overestimation of γ indicates a heavier tail than it actually is. Therefore, the estimation variances of the BM-MLE estimators are higher. Consequently, the variance reduction becomes negligible in this case. The biases and the variances of the POT-PWM estimator and the BM-PWM estimators are almost identical in three samples, thus we skip detailed discussion.

4.1.3 Robustness check

Figure 2, 3 and 4 present the simulation results in the i.i.d. case for $n = 1000$ $p_n = 0.0005$, $n = 2000$ $p_n = 0.001$ and $n = 2000$ $p_n = 0.0005$, respectively. Here we compare the results in Figure 1 with the results under different sample sizes n and probability levels p_n. The MSE, squared bias and variance reduce when estimating a less extreme quantile or using a sample with a larger sample size n. Theoretically, p_n should not affect the performance of the quantile estimators as long as $p_n = o(n)$. Nevertheless, we observe that higher p_n corresponds to a lower empirical bias and variance. Moreover, the sample with a large n would make the asymptotic theory on the estimator work better, consequently improve the finite-sample performances of the quantile estimators. In Figure 3, the POT-PWM estimator is able to estimate the heavy tails for low k_n as long as the sample size is sufficiently large. Also, the BM-PWM estimators do not yield γ estimates close to zero for low k_n. In the rest of this section, we provide the simulation results of the ARMA models and the GARCH models with $n = 1000$ and $p_n = 0.001$, the results under other combinations of n and p_n are in Appendix B.

4.2 ARMA models

4.2.1 MSE comparison

We consider the performances of the quantile estimators under linear serial dependence. Figure 5 depicts the MSEs of six estimators in the ARMA models. Firstly, the first column shows the results with Fréchet innovations. The sliding BM-MLE estimator is always the optimal quantile estimator, regardless of whether the serial dependence is strong, weak, or locally strong. The advantage of low variance of the sliding BM-MLE estimator over the disjoint BM-MLE estimator is insignificant. The Weissman

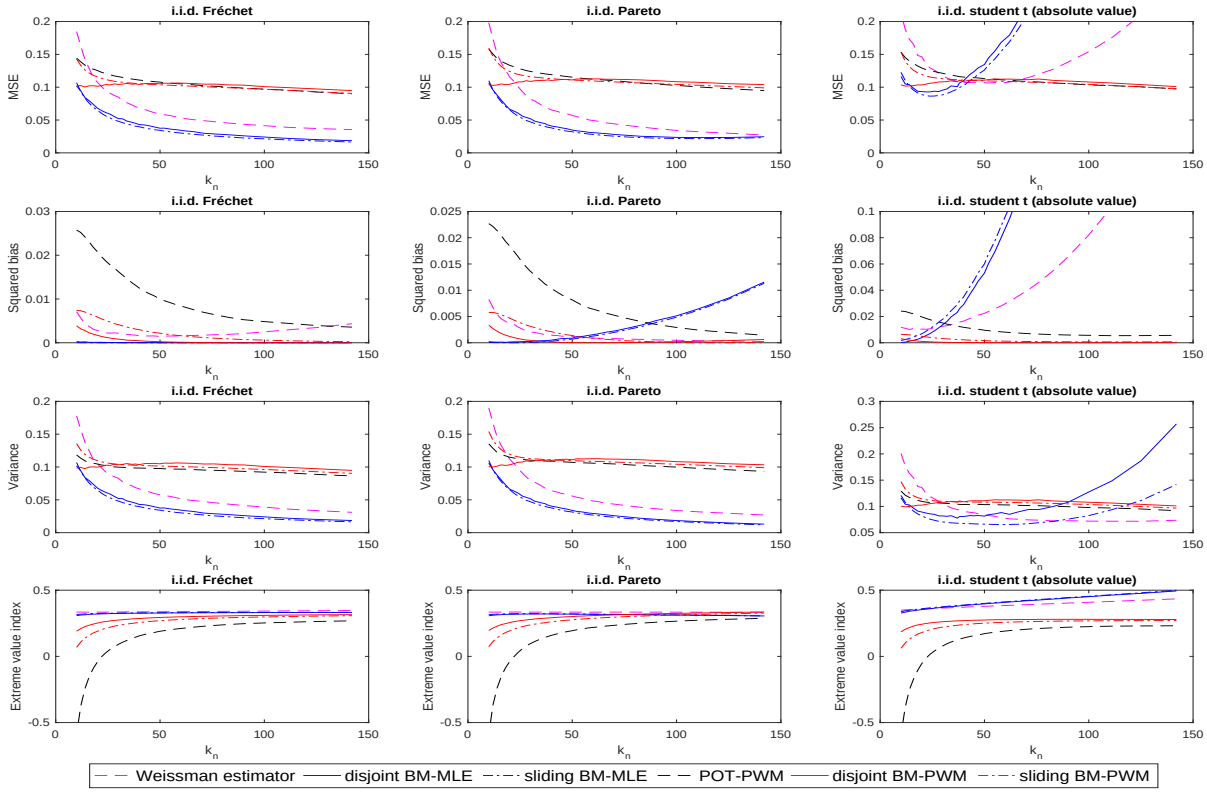

Figure 2: Performances (the scaled MSEs, scaled squared biases and scaled variances) and the extreme value index estimates in the i.i.d. samples as a function of the effective sample size, respectively. $n = 1000$, $p_n = 0.0005$

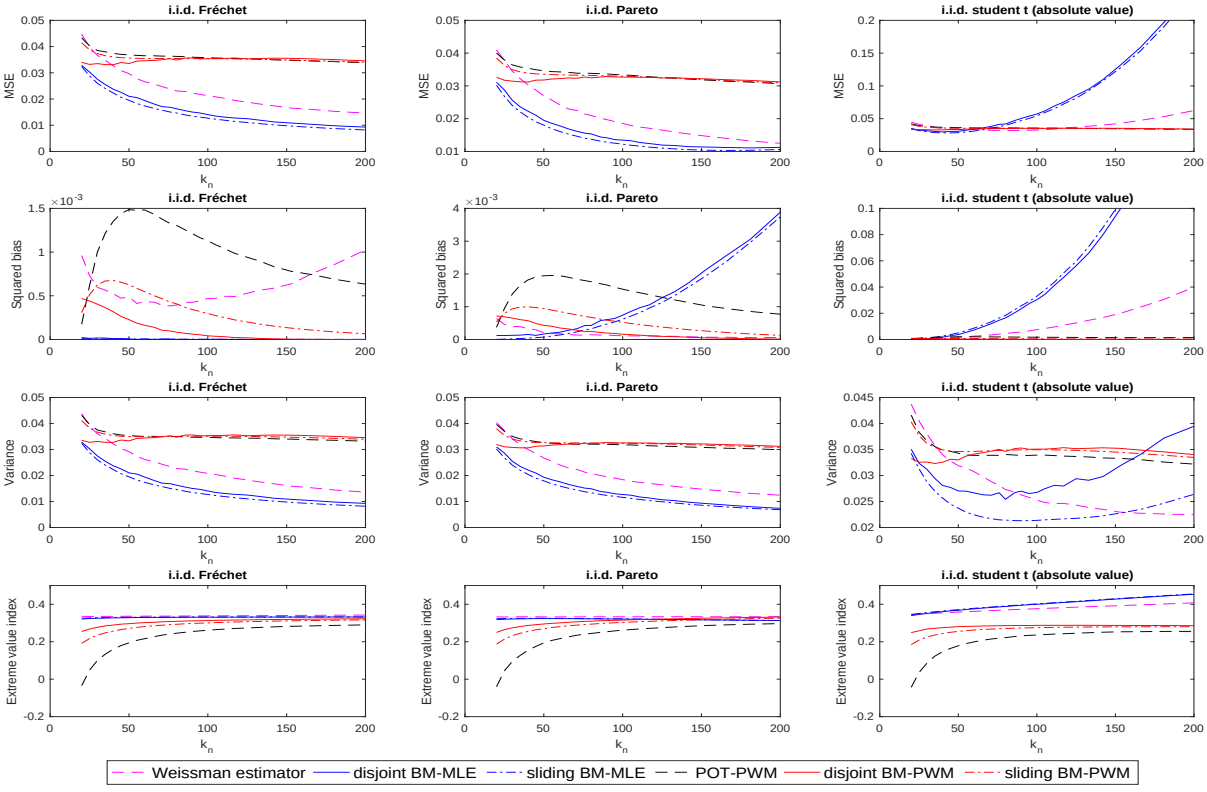

Figure 3: Performances (the scaled MSEs, scaled squared biases and scaled variances) and the extreme value index estimates in the i.i.d. samples as a function of the effective sample size, respectively. $n = 2000$, $p_n = 0.001$

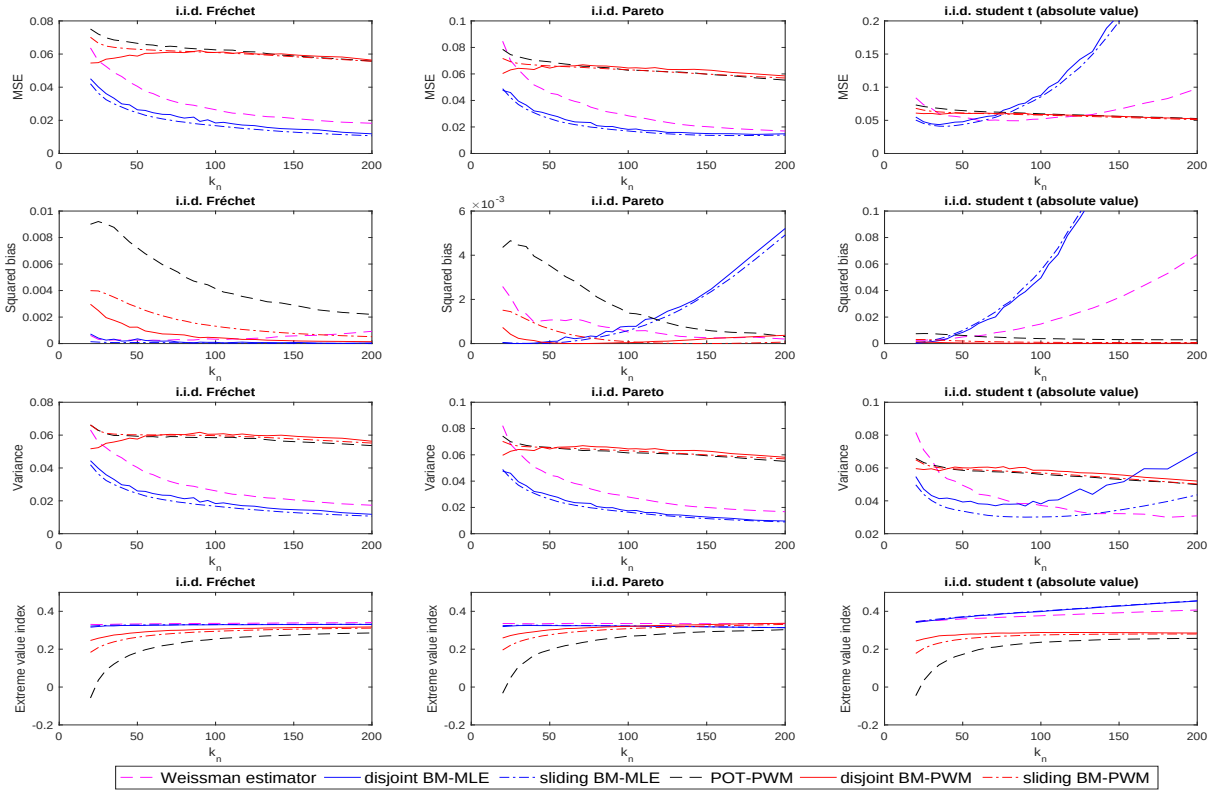

Figure 4: Performances (the scaled MSEs, scaled squared biases and scaled variances) and the extreme value index estimates in the i.i.d. samples as a function of the effective sample size, respectively. $n = 2000$, $p_n = 0.0005$

estimator has a hump-shaped MSE curve under (relatively) strong serial dependence as in the ARMA (i) and ARMA (ii) models. The hump shape is not observed when the serial dependence is weaker as in the ARMA (iii) and ARMA (iv) models. We shall discuss the reason behind in the MSE decomposition below. The MSE curve of the sliding BM-PWM estimator has a similar downward trend as the disjoint BM-PWM estimator, but smoother. The BM-PWM estimators are not preferable compared to the Weissman estimator and the BM-MLE estimators in the four ARMA models. Note that the POT-PWM yields an increasing MSE curve in k_n under strong serial dependence in the ARMA (i) model. When the serial dependence is weaker from the ARMA (i) to the ARMA (iii) model, the MSE curve is more horizontal. The MSE decreases with k_n increases in the ARMA (iv) model. For this model, the serial dependence which is locally strong but not persistent.

Secondly, the second column presents the results with Pareto innovations. The MSE curves of the six quantile estimators are similar as these in the first column. The sliding BM-MLE estimator outperforms the other candidate estimators under different degrees of serial dependence. Note that in the ARMA (iii) and (iv) models, the U-shaped MSE curves of the BM-MLE estimators are more pronounced, which suggests that the BM-MLE estimators have higher biases with Pareto innovations.

Lastly, consider the results with student t innovations in the third column. The POT-PWM estimator and the BM-PWM estimators have almost identical performances across the three different innovation distributions. By contrast, the BM-MLE estimators provide the minimal MSEs only when the serial dependence is rather strong. Once the serial dependence is weaker as in the ARMA (ii) models, the

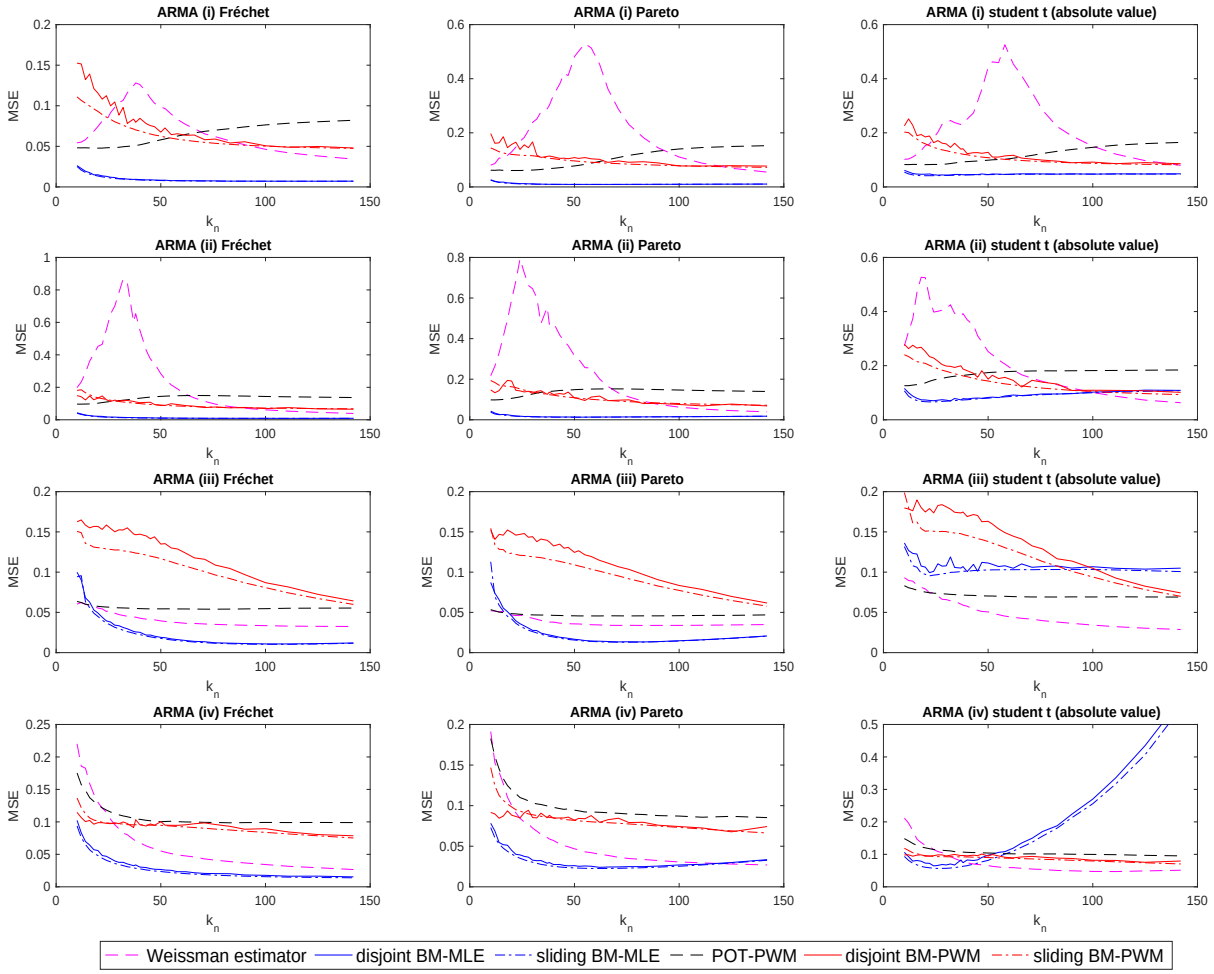

Figure 5: Scaled MSEs in the ARMA models as a function of the effective sample size. $n = 1000$, $p_n = 0.001$

Weissman estimator has the best performance given k_n is large. When the serial dependence reduces further in the ARMA (iii) model, the Weissman estimator outperforms the BM-MLE estimators. For this model, the BM-MLE estimators have higher MSEs than that for the models with Fréchet or Pareto innovations. In the ARMA (iv) model where the serial dependence has short memory, the MSEs of the BM-MLE estimators rise rapidly after a small decrease. We conclude that the BM-MLE estimators require the block size to be sufficiently large. Again, the Weissman estimator is recommended in this case and it reaches the minimal MSE with relatively low k_n.

4.2.2 MSE DECOMPOSITION

Figure 6 and Figure 7 demonstrate the variance and bias of each quantile estimator, respectively. Figure 8 reflects the effect of serial dependence on the γ estimation. The true γ equals to the extreme value index of the innovations, which is $\frac{1}{3}$. The first column of Figure 6, 7 and 8 show the results for the ARMA models with Fréchet innovations. The sliding BM-MLE is the most efficient method under all coefficient settings, since one of the advantages of the BM approach is its low variance by taking the block maxima as independent. Since the variances of the BM-MLE estimators dominate their MSEs, the bias-variance tradeoff is not well-observed in the MSE curves. The BM-MLE estimators are relatively accurate due to

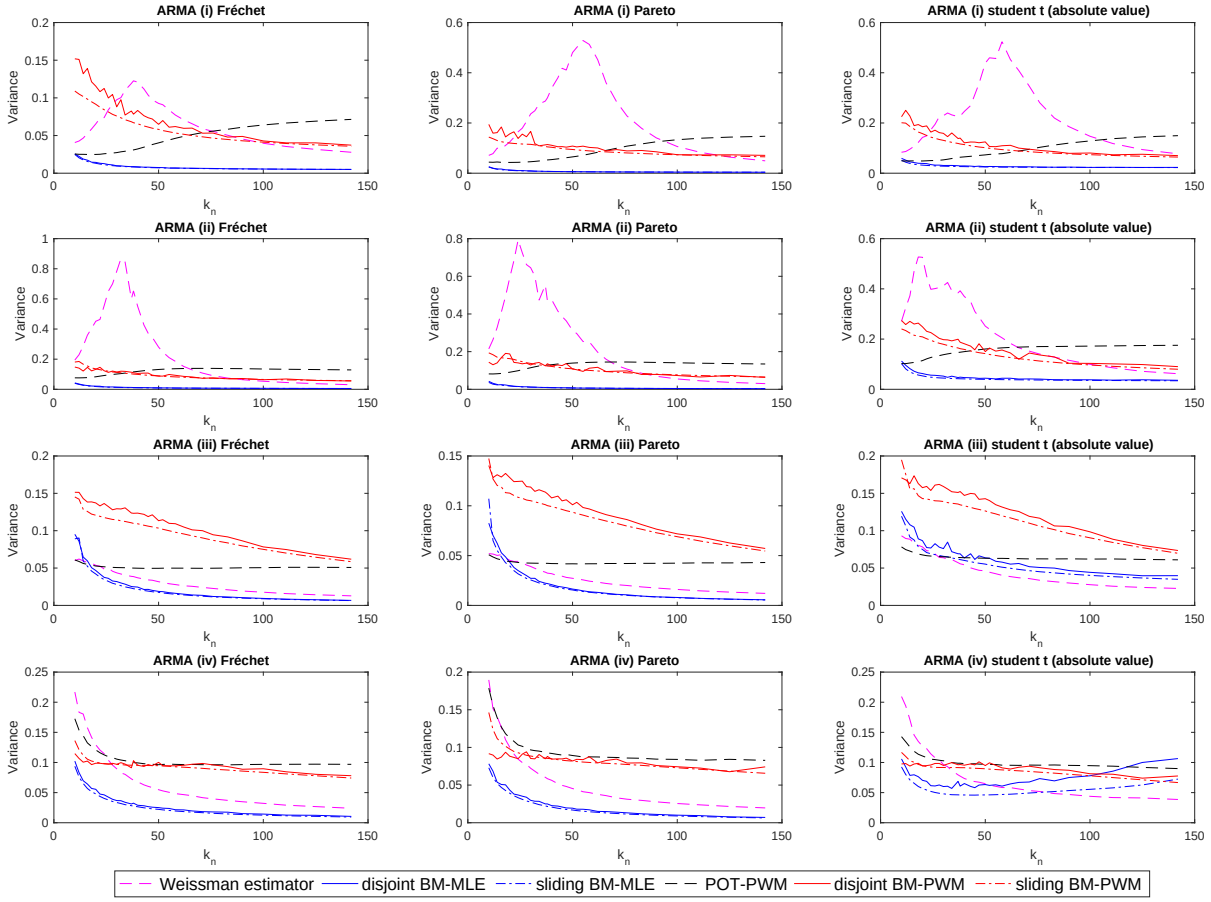

Figure 6: Scaled variances in the ARMA models as a function of the effective sample size. $n = 1000$, $p_n = 0.001$

low biases, although they generally have increasing biases in k_n. When the serial dependence is strong, the BM-MLE estimators underestimate γ even if the block size is sufficiently large. The underestimation is less pronounced when the serial dependence is weaker as in ARMA (i), (ii) and (iii) models. Compared to the first three ARMA models, the BM-MLE estimators are more accurate for the γ estimation due to lower biases, if the serial dependence is locally strong and not persistent as in the ARMA (iv) model.

Different from the other estimators, the variance curve of the Weissman estimator is hump-shaped which leads to the hump-shaped MSE curve, when the serial dependence is relatively persistent in the ARMA (i) and ARMA (ii) models. One possibility is that, when k_n is low and the sample is strongly dependent, the variance reduction is not sufficient to eliminate the serial dependence in the effective sample. With k_n increases, the variance reduction is greater. Therefore, there is no hump shape in the ARMA (iii) and ARMA (iv) models where the serial dependence is rather weak. Furthermore, the Weissman estimator is less accurate than the BM-MLE estimators in the first three ARMA models due to a higher bias, unless the threshold is considerably high. It is in line with the γ estimation results. The accuracy of the Weissman estimator is positively related to the Hill estimator, and the Hill estimator has a higher bias than the BM-MLE estimators except in the ARMA (iv) model.

Note that the POT-PWM estimator fails to capture the heavy-tailed feature if the serial dependence is persistent. The POT-PWM estimator provides negative γ estimates for all k_n values in the ARMA (i) and ARMA (ii) models. Hence, one observes an increasing variance curve and a declining bias curve

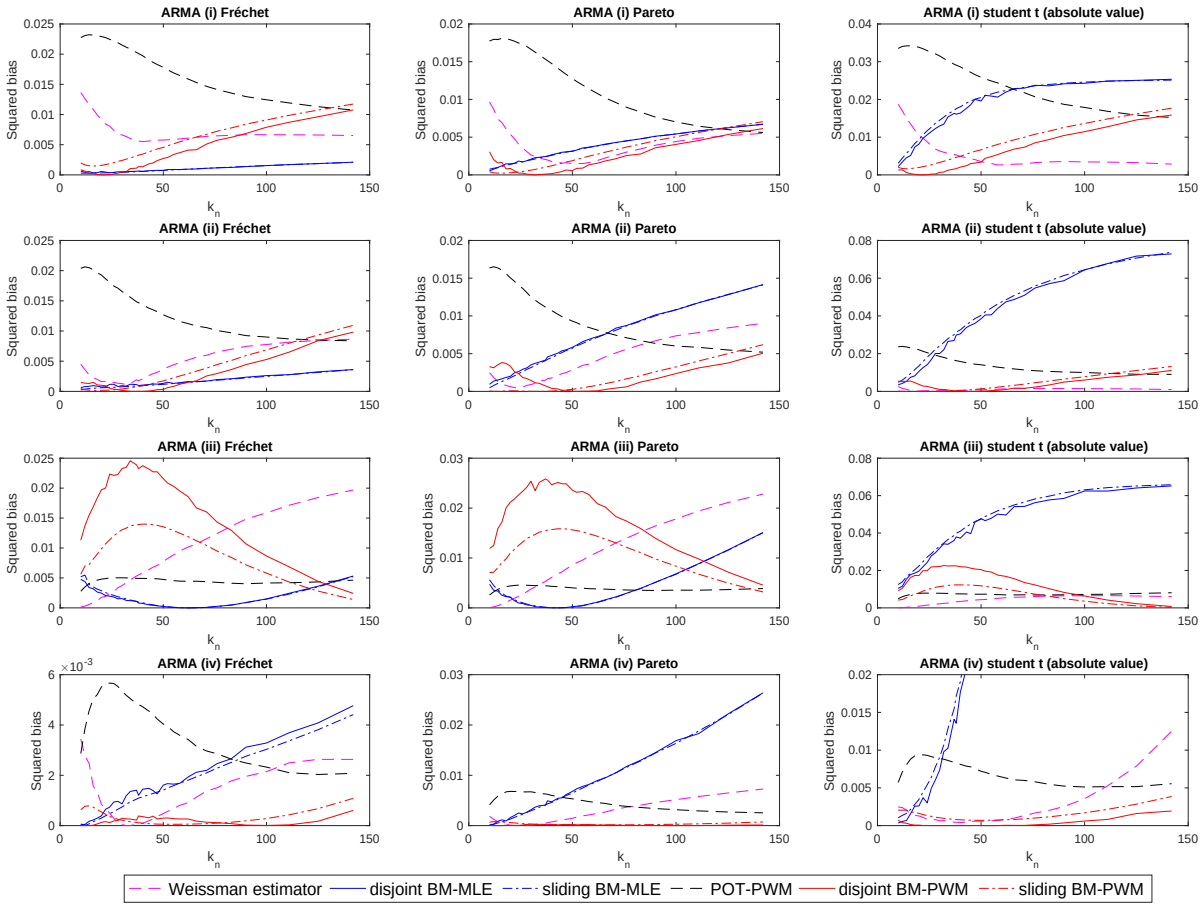

Figure 7: Scaled squared biases in the ARMA models as a function of the effective sample size. $n = 1000$, $p_n = 0.001$

due to the wrong γ estimates. The POT-PWM estimator is able to estimate the heavy tails if the serial dependence is weak or only locally strong, however, it requires a sufficiently large effective sample size. The hump shape in the bias curve of the POT-PWM estimator appears in the ARMA (iii) and ARMA (iv) models due to the same reason as explained in the i.i.d. cases. The serial dependence distorts the γ estimation of the BM-PWM as well but less serious than what is does for the POT-PWM. The sliding BM-PWM estimator fails to capture the heavy-tailed feature when the serial dependence is persistent as in the ARMA (i) model. The disjoint BM-PWM estimator yields positive γ estimates when k_n is low, however, the γ estimates are close to zero. The BM-PWM estimators have better performance in the γ estimation when the serial dependence is weaker in the ARMA (ii) model. The bias and variance curves of the BM-PWM estimators are similar in the ARMA (i) and ARMA (ii) models. Overall, if the γ estimates of the BM-PWM method are not sufficiently positive, the bias of the extreme quantile estimator is considerable and likely to be as the same magnitude as the bias under the negative γ estimates. The BM-PWM estimators can estimate the heavy tails under weak serial dependence and with relatively large effective sample size. Although the hump shape in the bias curves appears in the ARMA (iii) model, the BM-PWM estimators have lower biases than the POT-PWM estimator if k_n is high. In the ARMA (iv) model where the serial dependence is only locally strong, the BM-PWM estimators have lower biases than the other estimators. The POT-PWM estimator and the BM-PWM estimators are not preferable in

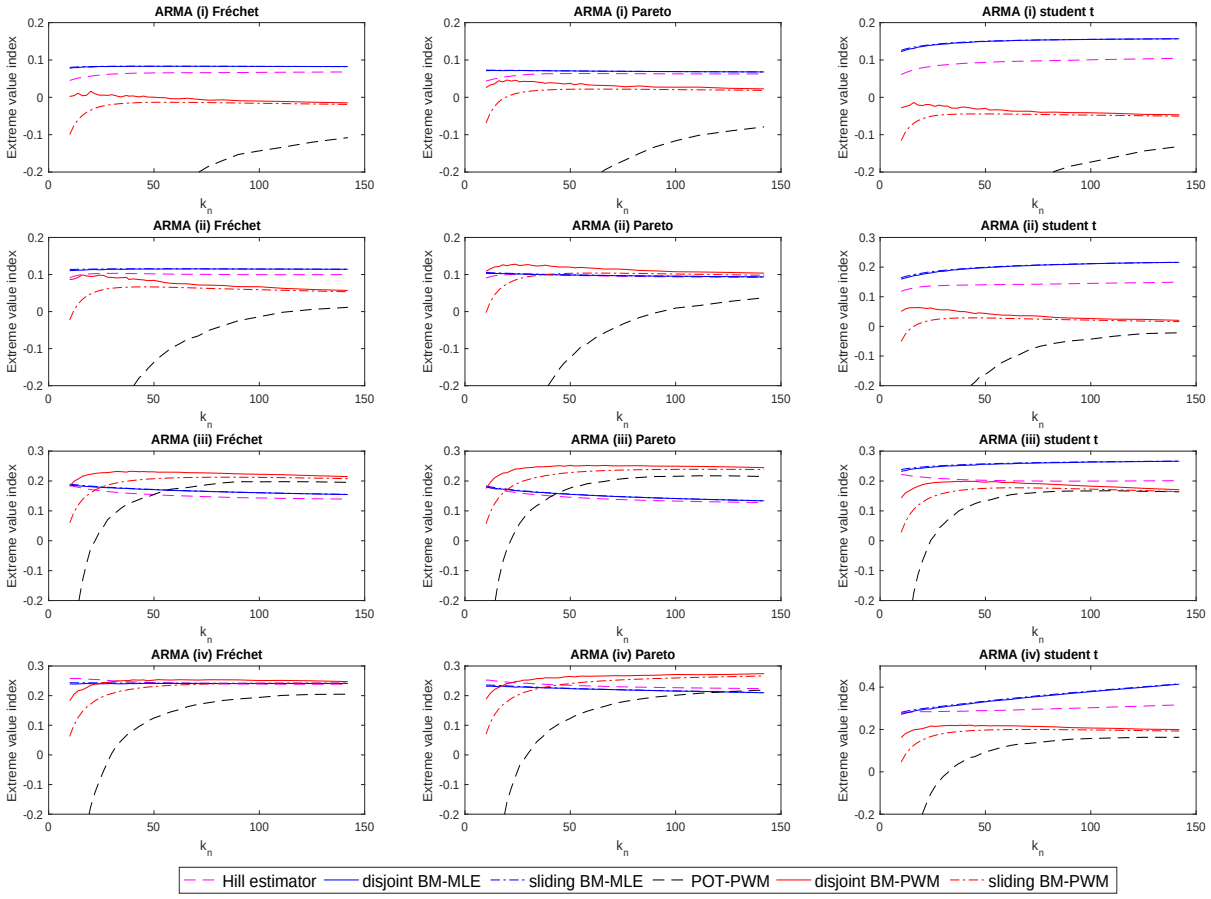

Figure 8: Extreme value index estimates in the ARMA models as a function of the effective sample size. $n = 1000$, $p_n = 0.001$

the ARMA (iii) and ARMA (iv) model, due to the high variances compared to the Weissman estimator and the BM-MLE estimators.

The second column of Figure 6, 7 and 8 present the results with Pareto innovations, which are quite similar to with Fréchet innovations. Thus the detailed discussion is skipped. Note that the BM-MLE estimators have higher biases with Pareto innovations.

Now we consider the results with student t innovations in the third column of Figure 6, 7 and 8. As mentioned in the MSE comparison section, the Weissman estimator outperforms the sliding BM-MLE estimator except in the ARMA (i) model. The sliding BM-MLE estimator has a lower variance than the Weissman estimator under strong serial dependence. However, its bias increases fast in k_n. By contrast, the Weissman estimator has a declining bias curve. The bias differences between the sliding BM-MLE estimator and the Weissman estimator further enlarge in the ARMA (ii) and ARMA (iii) model. Moreover, in the ARMA (iii) model with student t innovations, the variance reduction of the BM-MLE estimators is less than with Fréchet or Pareto innovations, such that the Weissman estimator becomes the most efficient quantile estimator. The variance curves of the BM-MLE estimators are U-shaped in the ARMA (iv) model. The biases rise rapidly due to the overestimation of γ, which is in line with the results in the i.i.d. student t sample. The POT-PWM estimator and the BM-PWM estimators have similar performances in the ARMA models with student t innovations as in the models with Fréchet and

Pareto innovations.

Finally, we observe that the performances and the γ estimations of six quantile estimators in the ARMA (iv) model are similar to the i.i.d. cases. When the serial dependence is locally strong but lack of persistence, each cluster of extreme values is relatively independent. Consequently, the serial dependence in the effective sample is weak. By contrast, if the serial dependence is persistent, either the block maxima from the BM method or the excesses from the POT method are strongly dependent. Under this circumstance, the quantile estimators are affected to a larger extent.

4.3 GARCH MODELS

Next we consider the GARCH models. Based on the results in the ARMA models, the POT-PWM and BM-PWM estimators may fail to capture the heavy-tailed feature. Therefore, the MSE should not solely decide the superiority of the quantile estimators based on the PWM method. We also take the γ estimation into consideration. In the following MSE comparison, a quantile estimator based on the PWM method is said to be optimal only if it provides the minimal MSE, and meanwhile a positive γ estimate. The true γ varies across the four GARCH models. Since we impose the GARCH models to be covariance stationary, the true γ of each GARCH model is in the range of 0 to $\frac{1}{2}$.

4.3.1 MSE COMPARISON

Figure 9 presents the simulation results in the GARCH models. The first three rows show the scaled MSE, scaled squared bias and scaled variance, respectively. The last row presents the γ estimates across the GARCH models. First of all, the POT-PWM estimator and the BM-PWM estimators fail to estimate the heavy tails in the GARCH (v) model where the nonlinear dependence is rather weak. So we would not consider the corresponding quantile estimators. Among the rest, the Weissman estimator outperforms the BM-MLE estimator, although the MSE increases fast as k_n increases. Secondly, the serial dependence is quite strong in the GARCH (vi) model. The Weissman estimator is again preferred. The POT-PWM estimator still fails to capture the heavy-tailed feature. The minimal MSEs of the Weissman estimator and the sliding BM-PWM estimator are comparable. The MSE curves of the BM-PWM estimators are more stable than the Weissman estimator. The BM-MLE estimators have increasing MSE curves. The MSEs have a large magnitude which cannot be fitted in the same graph with the MSEs of other estimators in the GARCH (vi) model when $n = 1000$ and $p_n = 0.001$. The results for other combinations of n and p_n show the MSEs of the BM-MLE estimators in the GARCH (vi) model.

Thirdly, the serial dependence is relatively weak in the GARCH (vii) model. The disjoint BM-PWM is the optimal method in this case. The BM-MLE estimators require a sufficiently large block size to perform well. On the contrary, the POT-PWM estimator provides the positive γ estimates only for high k_n. The MSE curves of the POT-PWM estimator and the BM-PWM estimators are flat. The Weissman estimator and the BM-MLE estimators have U-shaped MSE curves. Moreover, the minimal MSEs of the BM-MLE estimators are lower than that of the Weissman estimator. Finally, the disjoint BM-PWM estimator outperforms the other quantile estimators when the serial dependence becomes stronger from

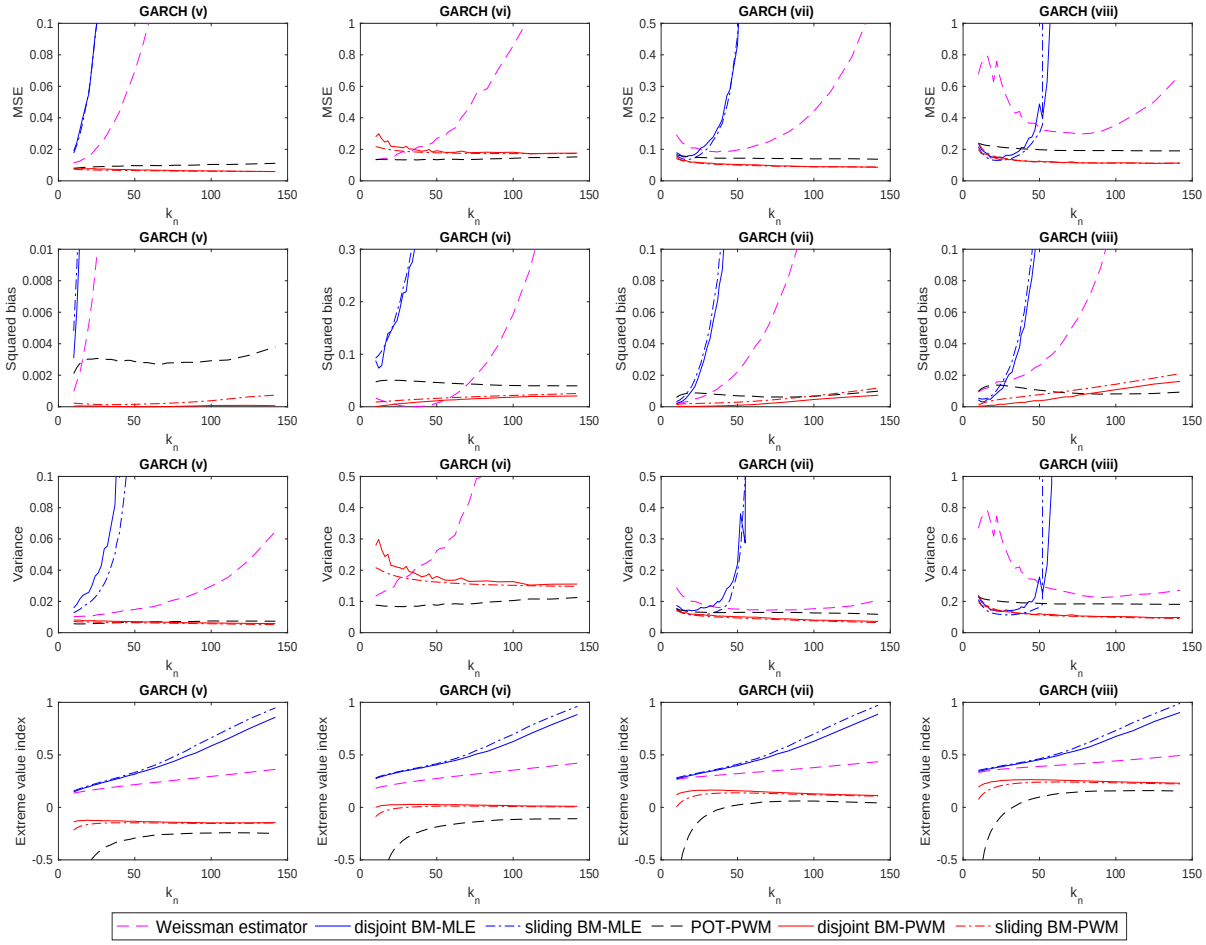

Figure 9: Performances (the scaled MSEs, scaled squared biases and scaled variances) and the extreme value index estimates in the GARCH models as a function of the effective sample size, respectively. $n = 1000$, $p_n = 0.001$

the GARCH (vii) model to GARCH (viii) model. Note that the POT-PWM estimator still fails to estimate the heavy tails for low k_n. The MSE curves of six quantile estimators have similar behaviours in the GARCH (vii) and the GARCH (viii) models. The magnitude of the MSEs is higher in the GARCH (viii) model due to the stronger serial dependence.

4.3.2 MSE DECOMPOSITION

The second and third rows in Figure 9 present the MSE decomposition into the squared bias and variance. First of all, the Weissman estimator has a lower bias and variance than the BM-MLE estimators in the GARCH (v) model where the degree of volatility clustering is low. The biases of the Weissman estimator and the BM-MLE estimators increase rapidly in k_n. The variance reduction is not observed as well. One possible explanation is as following. The true γ is positively related to the excess kurtosis. The excess kurtosis equals to 0.0049 which is close to zero in the GARCH (v) model. Thus, the true γ is small and overestimated by the Hill estimator and the BM-MLE estimators. The variance reduction is non-negligible only if γ is not overestimated, which is in line with the results in the i.i.d. student t sample and in the ARMA (iv) model with student t innovations. Moreover, the γ overestimation is more pronounced for the BM-MLE estimators than for the Hill estimator. The γ estimates based on the BM-MLE exceed

$\frac{1}{2}$ for high k_n. By contrast, the Hill estimator provides the γ estimates that are less than $\frac{1}{2}$. The POT-PWM estimator and the BM-PWM estimators in contrast to underestimate γ by providing the negative γ estimates. Consequently we would not take the corresponding quantile estimators into consideration. It suggests there is relatively high bias in the γ estimation of the POT-PWM and BM-PWMs. The small excess kurtosis could be the main contributor to the bias since the serial dependence is weak in this case.

The nonlinear serial dependence is persistent in the GARCH (vi) model. The excess kurtosis in this GARCH model is higher than in the GARCH (v) model, too. The biases and the variances of the BM-MLE estimators increase in k_n. Moreover, the BM-MLE estimators lose the advantage of low variance due to the overestimation of γ. Since the corresponding γ estimation is more accurate, the Weissman estimator outperforms the BM-PWM estimators due to a lower bias and variance. The POT-PWM estimator and the BM-PWM estimator fail to capture the heavy-tailed feature in the GARCH (vi) model. Thus, we do not consider the performances of their quantile estimations. By comparing the results in the GARCH (v) and (vi) models, it is the excess kurtosis that mainly affects the γ estimation rather than the degree of nonlinear serial dependence.

The excess kurtosis in the GARCH (vii) model is higher than in the GARCH (v) and (vi) models. The biases of the Weissman estimator and the BM-MLE estimators increase as k_n increases. The Weissman estimator has a lower bias than the BM-MLE estimators. The variance curve of the Weissman estimator is flat. By contrast, the variances of the BM-MLE estimators increase fast when k_n increases, as long as γ is overestimated. The BM-PWM estimators capture the heavy-tailed feature as the excess kurtosis is high. The disjoint BM-PWM estimator has a lower bias and a higher variance than its sliding counterpart. Nevertheless, the variance differences between the disjoint BM-PWM estimator and the sliding BM-PWM estimator are insignificant since the serial dependence in the GARCH (vii) model is relatively weak. The POT-PWM estimator is able to estimate the heavy tails when k_n is sufficiently high, though the γ estimates are close to zero.

Lastly, there are an infinite excess kurtosis and the relatively strong serial dependence in the GARCH (viii) model. The candidate quantile estimators have similar behaviours as in the GARCH (vii) model. Due to the true γ is closer to $\frac{1}{2}$, the BM-MLE estimators demand a higher block size such that γ is not overestimated. Since the overestimation of γ is less pronounced for the Weissman estimator in the GARCH (viii) model, the variance reduction is larger than in the first three GARCH models. The disjoint BM-PWM estimator outperforms the other quantile estimators due to a lower bias.

4.4 Summary

Figure 10 provides a brief summary on which extreme quantile estimator is preferable for different types of data. The sliding BM-MLE estimator outperforms in the i.i.d. sample due to its low variance. Its variance decreases as k_n increases, given γ is not overestimated. Thus, we suggest to apply the sliding BM-MLE estimator with high k_n in the i.i.d. sample. The simulation results show that serial dependence affects the estimators for high quantiles substantially. When the data are serial dependent, we consider the linear serial dependence and the nonlinear serial dependence separately. First of all, if the serial dependence is linear and weak, the sliding BM-MLE estimator is preferred. We can employ it with high

k_n, since the variance reduction is larger than the increase of bias. Given the minimal variances of the sliding BM-MLE estimator and of the Weissman estimator are comparable, the Weissman estimator is competitive in this case due to the low bias. When the the linear serial dependence is persistent, the sliding BM-MLE estimator performs best with a low variance. Similarly, we suggest to estimate the extreme quantile with high k_n.

Next, consider the nonlinear serial dependence. The excess kurtosis has a greater effect on the performance of the extreme quantile estimator than the degree of nonlinear dependence. If the excess kurtosis is low, the Weissman estimator outperforms the other estimators. Since the Hill estimator overestimates γ with high k_n, the Weissman estimator is preferable with low k_n. However, when the excess kurtosis is high or infinite, the disjoint BM-PWM estimator is the optimal extreme quantile estimator. It has a lower bias than other estimator and a lower variance than the Weissman estimator and the BM-MLE estimators. The disjoint BM-PWM estimator requires a sufficiently high k_n such that γ is not underestimated. A plot of the γ estimates of the POT-PWM can be an indicator of the magnitude of the excess kurtosis. If the γ estimates are always below zero for all values of k_n, then the excess kurtosis are likely to be low. Otherwise, if the γ estimates are positive for high k_n, the excess kurtosis can be considered as high.

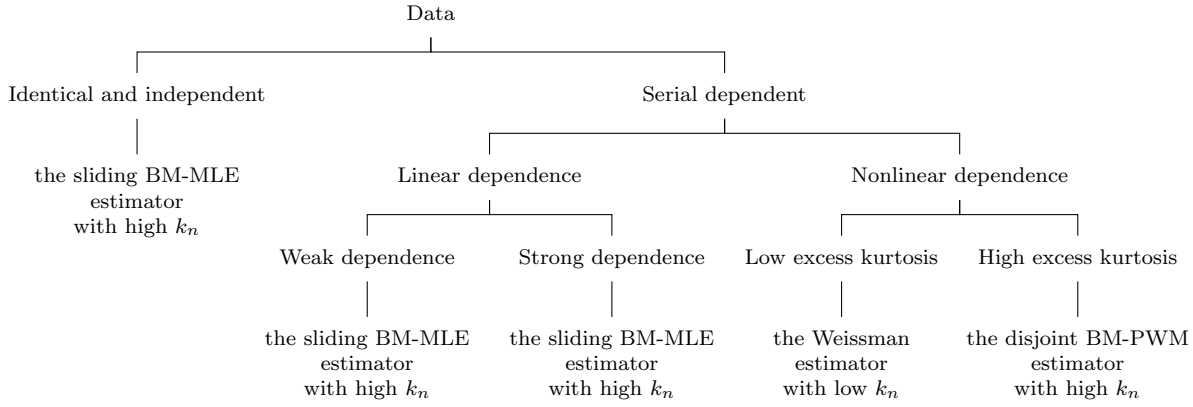

Figure 10: Preferable extreme quantile estimators in different scenarios

5 An efficiency improvement of the BM-MLE estimators in the GARCH models

In the quantile estimation procedure above, the block size m is identical across the estimation of θ, γ and x_{p_n}. We refer this situation to the *1-m procedure* in order to distinguish from the approach we propose below. Notice that in the simulation results for the GARCH model using the BM-MLE method, only a few large block sizes are valid for estimations. The variances of quantile estimators increase fast in k_n, once the BM-MLE estimators overestimate γ. In order to improve the efficiency of the BM-MLE estimators, we consider a *3-m procedure* that using different m for θ, γ and x_{p_n} estimations, namely m_θ, m_γ and m_x. We select the optimal m_θ and m_γ that provide the minimal variances of $\hat{\theta}$ and $\hat{\gamma}$, respectively. There are two reasons to employ the minimal variance criterion instead of the MSE criterion. Firstly, the calculation of the true γ of the GARCH model is not straightforward even if the GARCH model is

correctly specified. It requires to correctly fit the sample into the GARCH model, as well as specify the distribution of the innovations. Secondly, the results for the GARCH model show that the variance is the main contributor to the MSE. Hence, variance deduction is a second best choice. Last but not least, no true γ is known for application.

We present some simulation results to verify that the *3-m procedure* can indeed improve the finite-sample performances of the BM-MLE estimators in the GARCH models. The same GARCH model settings are adopted as in Section 3.8. We set the simulation runs to $S = 500$, the sample size to $n = 1000$, and the extreme quantiles x_{p_n} are estimated for $p_n = 0.001$.

Figure 11 compares the BM quantile estimators based on the *1-m procedure* and the *3-m procedure*. It plots the MSE ratio $\frac{MSE_{3\text{-}m}}{MSE_{1\text{-}m}}$, the squared bias ratio $\frac{bias^2_{3\text{-}m}}{bias^2_{1\text{-}m}}$ and the variance ratio $\frac{variance_{3\text{-}m}}{variance_{1\text{-}m}}$ against m_x in the first, second and third rows respectively. The results are similar across different GARCH models. The first row shows that the BM-MLE estimators have better performances based on the *3-m procedure*, and the MSE reduction is substantial when m_x is low. However, the *3-m procedure* fails to improve the performance of the disjoint BM-PWM estimator. The sliding BM-PWM estimator based on the *3-m procedure* outperforms the counterpart based on the *1-m procedure* only if m_x is low. The second row suggests that the minimal variance criterion does not always lead to a high bias. The bias ratios of the BM-MLE estimators are lower than one as long as m_x is at a low level. However, the *3-m procedure* distorts the BM-PWM estimators. More biases are introduced when m_x is higher. Next, consider the variance reduction by using the *3-m procedure*. As presented in the third row, the *3-m procedure* reduces the variances of the BM-MLE estimators and the BM-PWM estimators. The variance ratios of the BM-PWM estimators decrease as m_x increases, which means that the variances decrease more rapidly in the *3-m procedure*. However, the variance reduction is not sufficient to decrease the MSEs of the BM-PWM estimators, due to the MSEs are dominated by the biases.

Since the *3-m procedure* is not appropriate for the BM-PWM estimators, we compare the performances of the BM-MLE estimators based on the *3-m procedure* and the BM-PWM estimators based on the *1-m procedure* in Figure 12. The scaled MSE, scaled squared bias and scaled variance of each quantile estimator are depicted as functions of m_x. Although the BM-PWM estimators fail to capture the heavy-tailed feature in the first two GARCH models, we plot their performance curves as an indicator. The first row in Figure 12 shows the MSEs of corresponding quantile estimators. The BM-MLE estimators based on the *1-m procedure* have declining MSE curves in the GARCH (v) model. The level of MSEs is close to of the BM-PWM estimators as m_x increases. The MSE curves become U-shaped if the serial dependence is stronger in the GARCH (vi), (vii) and (viii) models. The disjoint BM-MLE estimator based on the *1-m procedure* is preferred when the serial dependence is persistent as in the GARCH (vi) model. The sliding BM-MLE estimator outperforms the disjoint counterpart when the serial dependence is less persistent in the GARCH (vii) and GARCH (viii) models.

The second and third rows show the decomposition of the MSE into the squared bias and variance. The BM-MLE estimators based on the *3-m procedure* have lower biases when m_x is higher in the GARCH (v) model. When the serial dependence is stronger and the excess kurtosis is higher in the GARCH (vi), (vii) and (viii) models, the U shape in the bias curve induces the similar shaped MSE curve. Since the

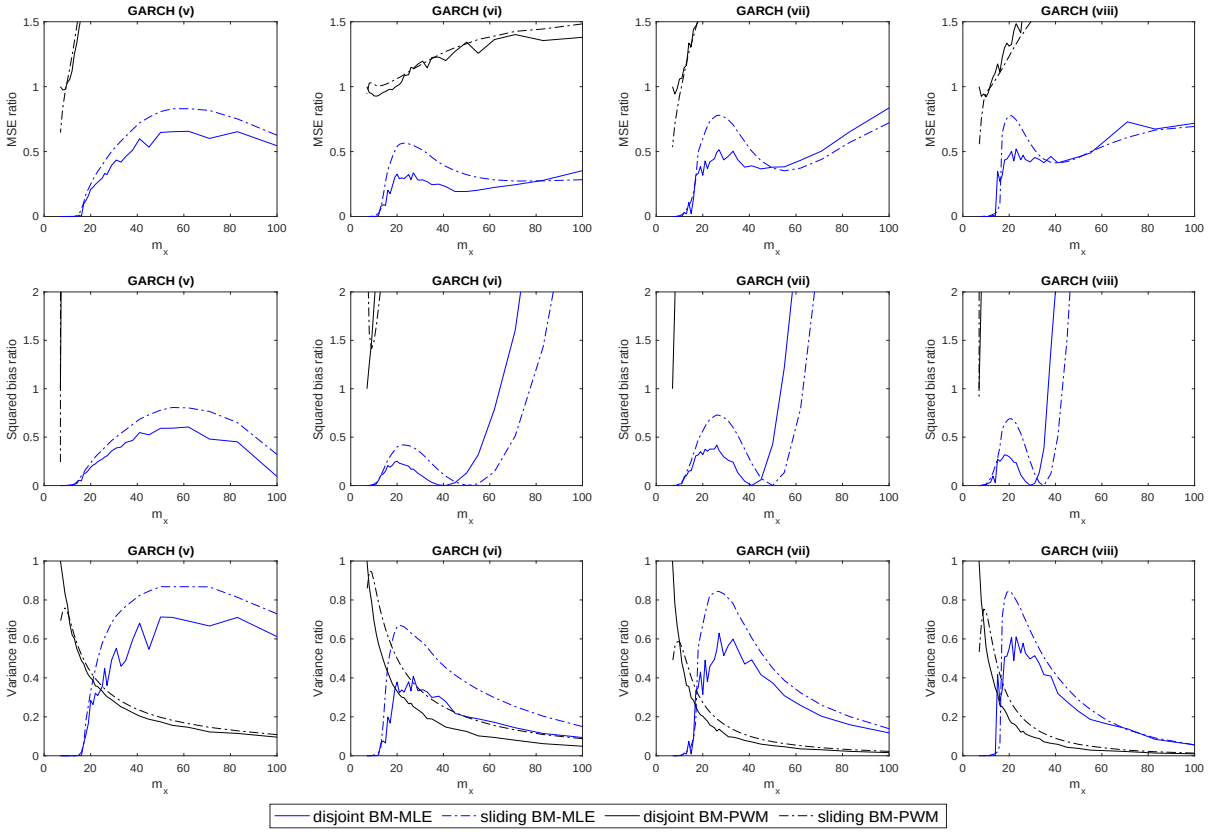

Figure 11: MSE ratio $\frac{MSE_{3\text{-}m}}{MSE_{1\text{-}m}}$, squared bias ratio $\frac{bias^2_{3\text{-}m}}{bias^2_{1\text{-}m}}$ and variance ratio $\frac{variance_{3\text{-}m}}{variance_{1\text{-}m}}$ in the GARCH models as a function of the block size of the quantile estimation, respectively. $n = 1000$, $p_n = 0.001$

BM-MLE estimators are declining functions of the m_x when $\hat{\theta}$ and $\hat{\gamma}$ are fixed, the U shape suggests that x_{p_n} is overestimated for low m_x but underestimated for high m_x. The x_{p_n} estimates decrease in m_x, consequently the bias reduces. As m_x further increases, the x_{p_n} estimates keep decreasing in m_x. Eventually the bias is high again after the x_{p_n} estimates is lower than the true x_{p_n}.

The third row presents that the *3-m procedure* improves the efficiency of the BM-MLE in the GARCH model by variance reduction, such that the BM-MLE becomes the most efficient method unless the serial dependence is weak as in the GARCH (v) model. Note that the disjoint BM-MLE estimator based on the *3-m procedure* has a lower variance than the sliding BM-MLE estimator in the GARCH (vi) model. We consider the reason as follows. Denote the optimal m_γ of the disjoint and sliding BM-MLE estimators as $m^*_{\gamma,dj}$ and $m^*_{\gamma,sl}$, respectively. The resulting effective sample sizes of disjoint blocks and sliding blocks are $k^*_{\gamma,dj} = \frac{n}{m^*_{\gamma,dj}}$ and $k^*_{\gamma,sl} = n - m^*_{\gamma,sl} + 1$, individually. The corresponding scale estimators are $\hat{a}^*_{m,dj}$ and $\hat{a}^*_{m,sl}$, and the modified scale estimators are $\hat{a}_{m,dj}$ and $\hat{a}_{m,sl}$. Regarding the γ and a_m estimations, the sliding BM-MLE is asymptotically more efficient than the disjoint BM-MLE. Meanwhile, we observe the advantage of efficiency of the sliding BM-MLE estimator in simulation results, when applying the same m_γ to the disjoint and sliding BM-MLE estimators. However, $\hat{a}_{m,sl}$ does not necessarily have a lower variance than $\hat{a}_{m,dj}$ in the *3-m procedure*. Define the asymptotic variance of an estimator as $\sigma^2(\cdot)$. From the asymptotic theory of the BM-MLE, $\sigma^2(\hat{\gamma})$ is a decreasing function of $\hat{\gamma}$. By contrast, $\sigma^2(\hat{a}^*_m)$ is an increasing function of $\hat{\gamma}$. There may exist some $m^*_{\gamma,dj}$ and $m^*_{\gamma,sl}$ yielding $\hat{\gamma}_{sl} > \hat{\gamma}_{dj}$, such that

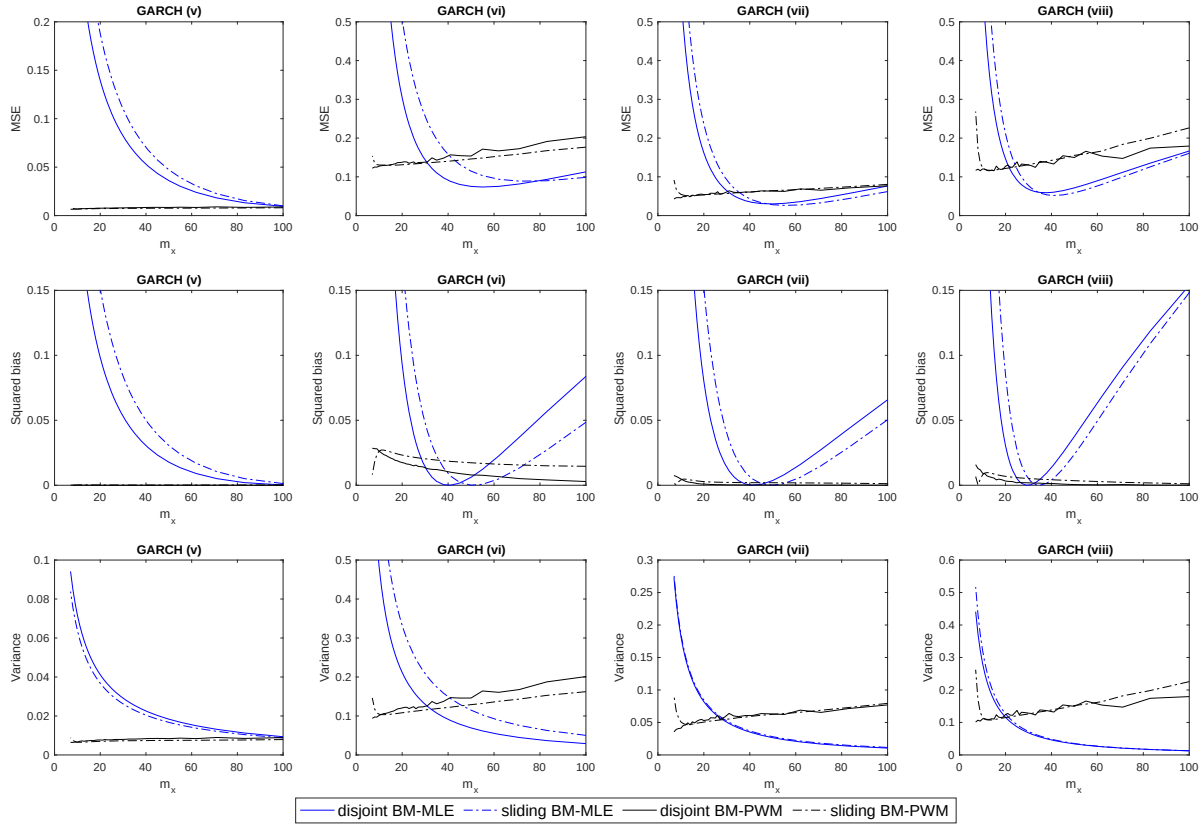

Figure 12: Performances (the scaled MSEs, scaled squared biases and scaled variances) of the BM-MLE estimators based on *3-m procedure* and of the BM-PWM estimators based on *1-m procedure* in the GARCH models as a function of the block size of the quantile estimation, respectively. $n = 1000$, $p_n = 0.001$.

$\sigma^2(\hat{\gamma}_{sl}) < \sigma^2(\hat{\gamma}_{dj})$ and $\sigma^2(\hat{a}^*_{m,sl}) > \sigma^2(\hat{a}^*_{m,dj})$ hold. Consequently, $\sigma^2(\hat{a}_{m,sl})$ could be larger than $\sigma^2(\hat{a}_{m,dj})$, although there is $\sigma^2(\hat{\theta}_{sl}) < \sigma^2(\hat{\theta}_{dj})$. The variances of $\hat{a}_{m,dj}$ and $\hat{a}_{m,sl}$ are approximately $\frac{\sigma^2(\hat{a}_{m,dj})}{k^*_{\gamma,dj}}$ and $\frac{\sigma^2(\hat{a}_{m,sl})}{k^*_{\gamma,sl}}$, respectively. Still, $\hat{a}_{m,dj}$ could have a lower variance than $\hat{a}_{m,sl}$ given that $k^*_{\gamma,dj} < k^*_{\gamma,sl}$. Under this circumstance, the disjoint BM-MLE can have a lower variance than the sliding BM-MLE in the extreme quantile estimation.

Heuristically, the asymptotic properties of $\hat{\theta}$ are established by splitting the ith block into a small block of length ℓ_n and a big block of length $m_\theta - \ell_n$. Then the Condition 2.1 in Berghaus and Bücher (2018) restricts the dependence between block maxima. It ensures that the mixing coefficients of both big block and small block decay at least with a hyperbolic rate. Furthermore, the size of ℓ_n ensures that the big blocks that are not adjacent are asymptotically independent and the contributions of the small blocks are negligible (Robert, 2009). Such restrictions are not necessary to establish the asymptotic normalities of γ estimators. Define m_γ and m_θ as n^{τ_γ} and n^{τ_θ} where $\tau \in (0, 1)$, respectively. In the future research, it could be possible to compare the lower-bounds of τ_γ and τ_θ which may give a concrete relationship between m_γ and m_θ.

6 Conclusion

We conduct Monte Carlo simulations in order to compare the finite-sample performances of six extreme quantile estimators under serial dependence. The POT method and the BM methods under the disjoint blocks and the sliding blocks are included in the comparison. We employ the probability-weighted moments and the maximum likelihood as estimation methods. Moreover, we consider different degrees of linear and nonlinear serial dependence in the data generating processes. Since we observe the large variances for the BM-MLE estimators under nonlinear serial dependence, we further investigate a *3-m procedure* allowing different block sizes for the extreme quantile estimation, the extreme value index estimation and the extremal index estimation for the BM-MLEs.

The sliding BM-MLE estimator outperforms other quantile estimators due to a low variance, in the i.i.d. sample or when the serial dependence is linear. Note that when the random variables or the innovations of the ARMA models are student t distributed, the disjoint and sliding BM-MLE estimators overestimate γ if the block size is sufficiently low. The POT-PWM estimator and the BM-PWM estimators fail to capture the heavy-tailed feature under strong linear serial dependence. When the serial dependence is not persistent, they require sufficiently high k_n to estimate heavy tails. Under nonlinear serial dependence, the biases and variances of the Weissman estimator and the BM-MLE estimators are high when they overestimate γ. The POT-PWM estimator and the BM-PWM estimators are able to estimate the heavy tails only if the excess kurtosis is large or infinite. Both the γ overestimation of the Weissman estimator and the BM-MLE estimators and the γ underestimation of the POT-PWM estimator and the BM-PWM estimators are less pronounced when the excess kurtosis is higher.

The *3-m procedure* is able to improve the performances of the BM-MLE estimators under nonlinear serial dependence. The sliding BM-MLE estimator based on the *3-m procedure* outperforms the disjoint BM-PWM estimator in the GARCH models. The bias reduction and variance reduction by using the *3-m procedure* are substantial when the block size for the quantile estimation is low. Once we observe the serial dependence in data, we can test the linearity of the serial dependence (Giannerini, Maasoumi, & Dagum, 2015). If the serial dependence is nonlinear, we suggest to apply the sliding BM-MLE estimator based on the *3-m procedure*. Afterwards, the selection of the block size of quantile estimation is essential. The extreme quantile is overestimated if the block size is low and underestimated if the block size is high. In order to find a *good* block size, we can plot the variance of the quantile estimator against the block size. Then we choose the block size that gives a large slope of the variance curve. The reason behind is as follows. The variance dominates the MSE when the block size is at a low level. Therefore, the MSE is lower with a higher block size, if the variance reduction by increasing the block size is large. When the block size increases further, the variance is low and the variance reduction is negligible. Consequently the bias becomes dominant, and the bias increases in the block size.

There are some limitations in this paper. First of all, the models in simulations can be enriched. For instance, the AR-MAX models, the asymmetric GARCH models and the GARCH models with different innovation distributions are of interests. Second, this paper does not provide an exact boundary between the weak and strong serial dependence, as well as the boundary between the small and large excess

kurtosis. It is left for future research. Lastly, the simulations are focused on the heavy-tailed series with a finite variance. Some more general cases where $\gamma \in \mathbb{R}$ and the variance is infinite are also worthy to investigate.

Appendix A: notation list

Here we list the important notations appeared in this paper and their explanations.

F: the continuous distribution function.

F_t: the conditional distribution function of excesses.

γ: the extreme value index.

θ: the extremal index.

$\hat{\sigma}^2$: the asymptotic variance estimator for θ.

G_γ: the GEV distribution function with the extreme value index γ.

H_γ: the GPD function with shape parameter $1/\gamma$.

a_n: the scale parameter of the GEV distribution.

b_n: the location parameter of the GEV distribution.

$\sigma(t)$: the scale of the GPD.

x_{p_n}: the true extreme quantile.

$x_{\frac{n}{k_n}}$: the true intermediate quantile.

$(X_n)_{n=1}^{\infty}$: an i.i.d. sequence or the associate i.i.d. sequence of a stationary sequence.

$(X_n)_{n\in\mathbb{Z}}$: an observed i.i.d. sequence.

$X_{n-i:n}$: the $i+1$th largest ordered statistic from $(X_n)_{n\in\mathbb{Z}}$.

M_n: the maximum of $(X_n)_{n=1}^{\infty}$.

$M_{i,m}^d$: the ith disjoint i.i.d. block maximum extracted from $(X_n)_{n\in\mathbb{Z}}$ with the block size m.

$M_{i:k_n}^d$: the $i+1$th largest disjoint i.i.d. block maximum.

$(\tilde{X}_n)_{n=1}^{\infty}$: an strictly stationary sequence.

$(\tilde{X}_n)_{n\in\mathbb{Z}}$: an observed strictly stationary sequence.

$\tilde{X}_{n-i:n}$: the $i+1$th largest ordered statistic from $(\tilde{X}_n)_{n\in\mathbb{Z}}$.

$\tilde{M}_n$: the maximum of $(\tilde{X}_n)_{n=1}^{\infty}$.

$\tilde{M}_{i,m}^d$: the ith disjoint block maximum extracted from $(\tilde{X}_n)_{n\in\mathbb{Z}}$ with the block size m.

$\tilde{M}_{i,m}^{sl}$: the ith sliding block maximum extracted from $(\tilde{X}_n)_{n\in\mathbb{Z}}$ with the block size m.

m_x: the block size that is used in the extreme quantile estimation based on the BM method.

m_γ: the block size that is used in γ estimations based on the BM method.

m_θ: the block size that is used in θ estimation based on the BM method.

Appendix B: additional simulation results

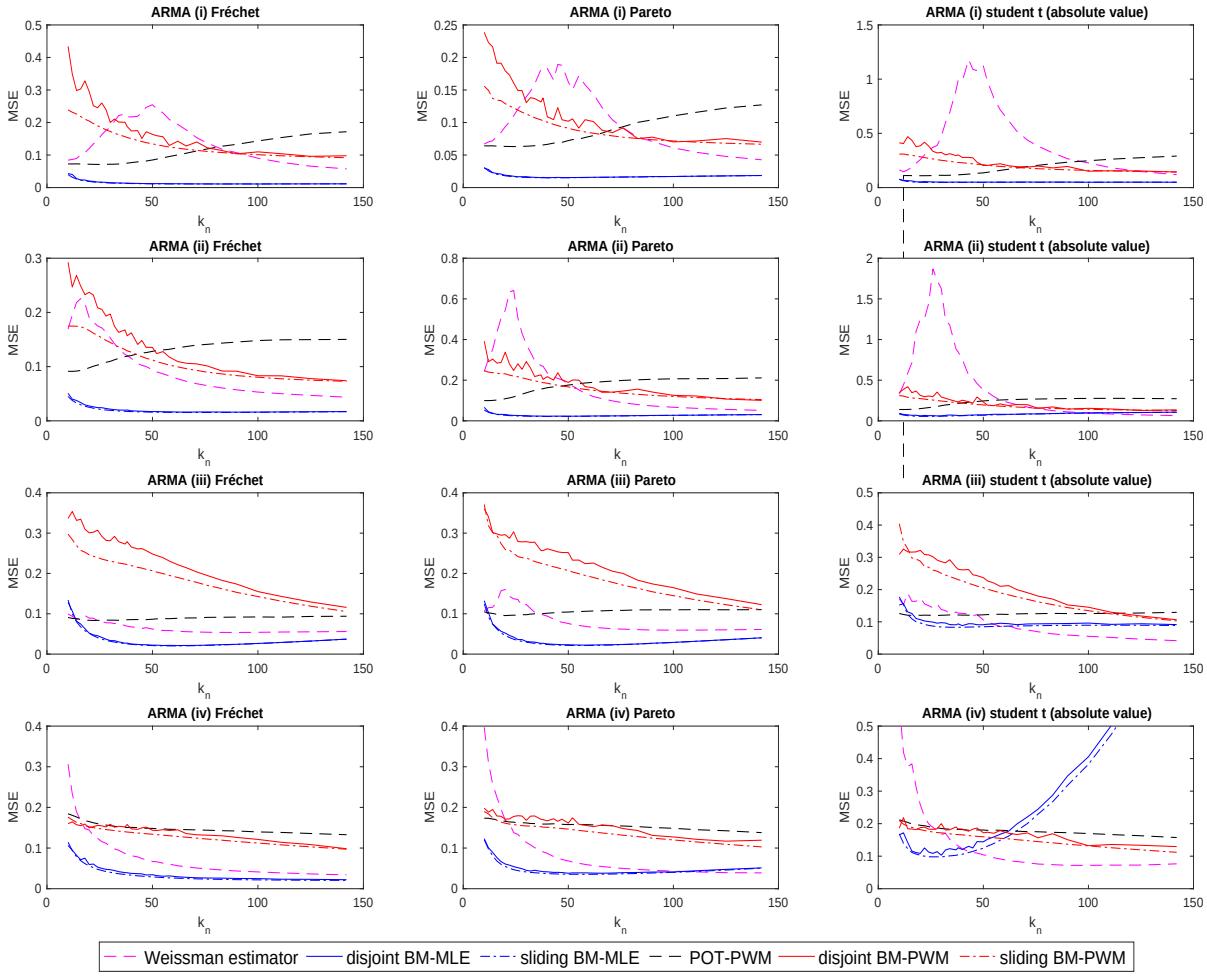

Figure 13: Scaled MSEs in the ARMA models as a function of the effective sample size. $n = 1000$, $p_n = 0.0005$

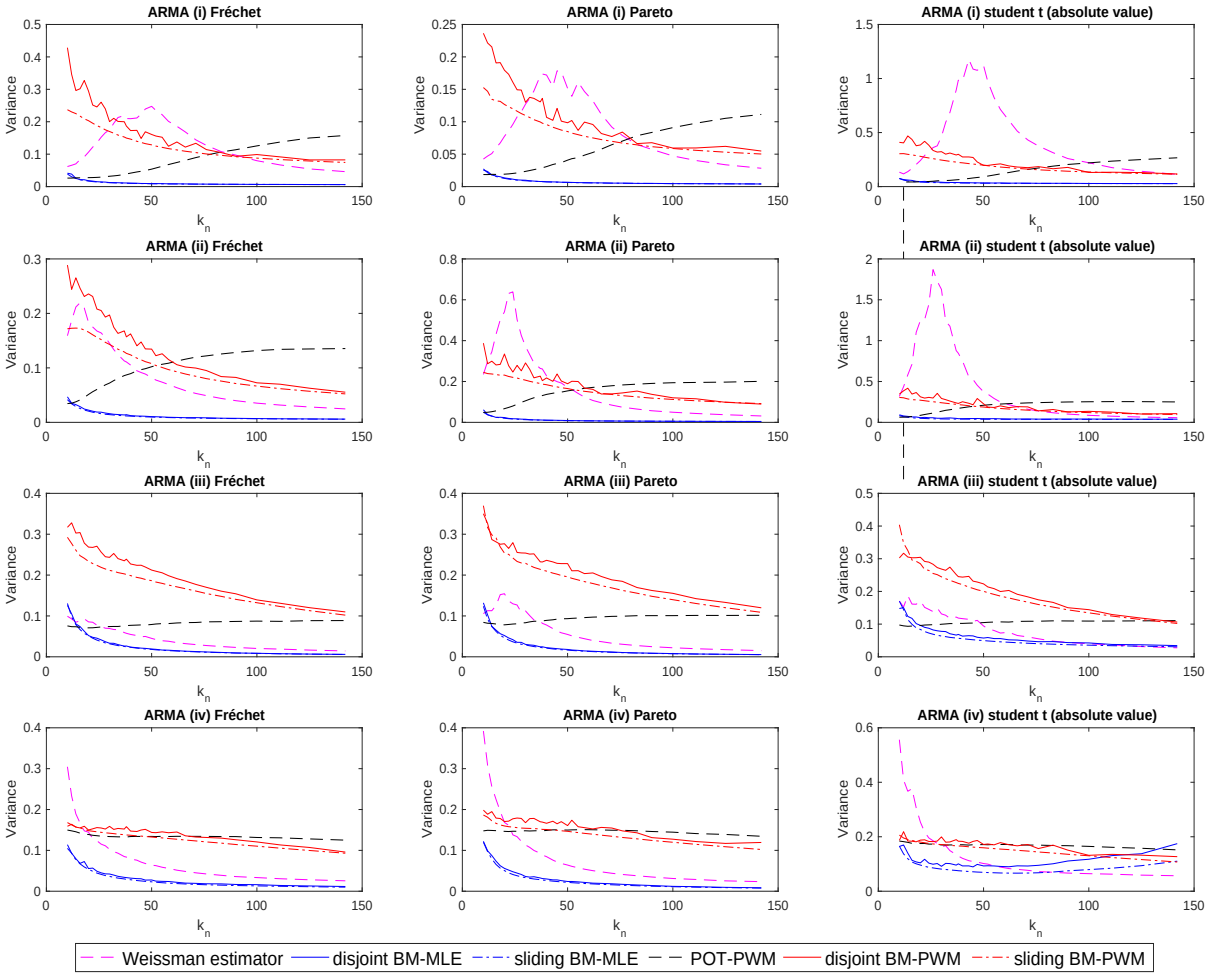

Figure 14: Scaled variances in the ARMA models as a function of the effective sample size. $n = 1000$, $p_n = 0.0005$

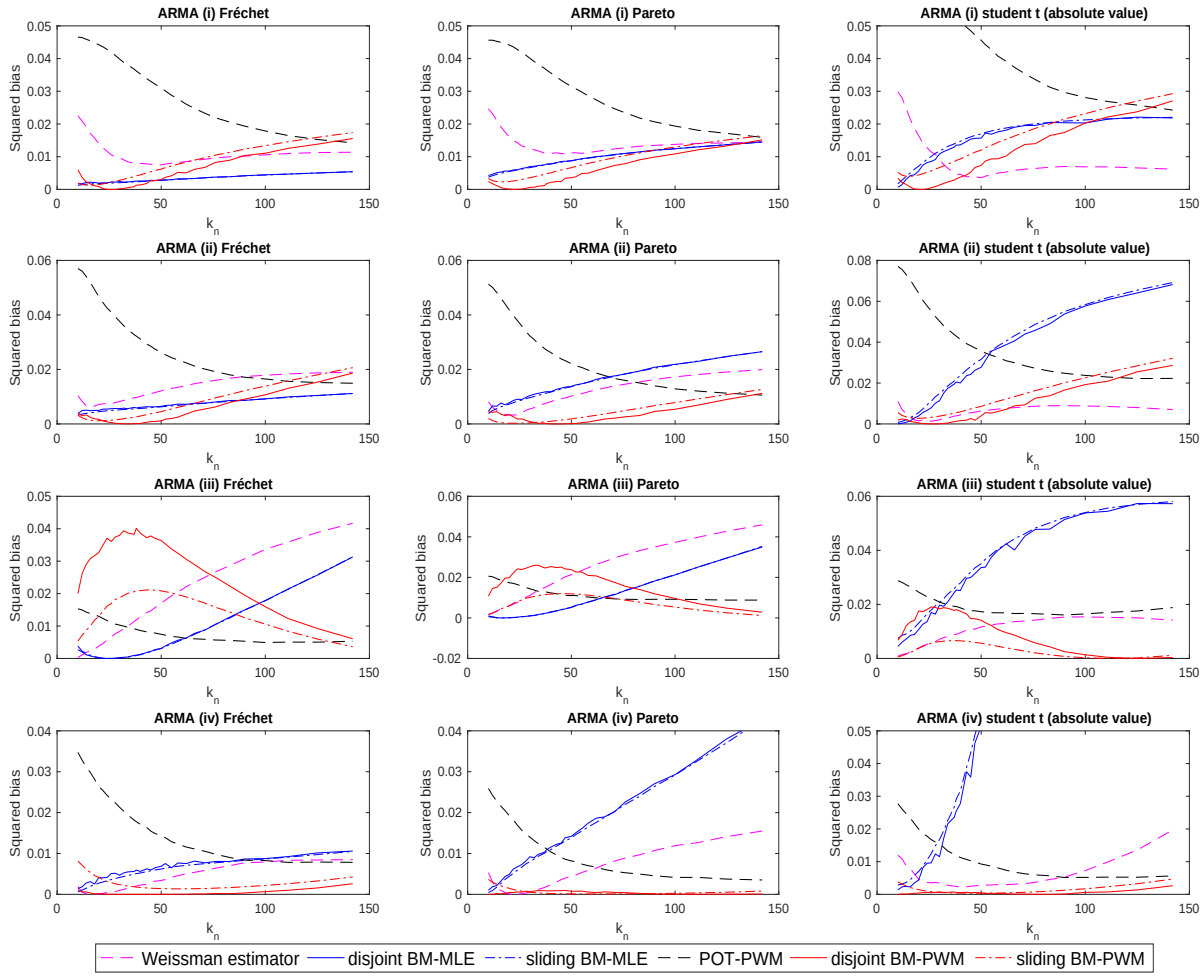

Figure 15: Scaled squared biases in the ARMA models as a function of the effective sample size. $n = 1000$, $p_n = 0.0005$

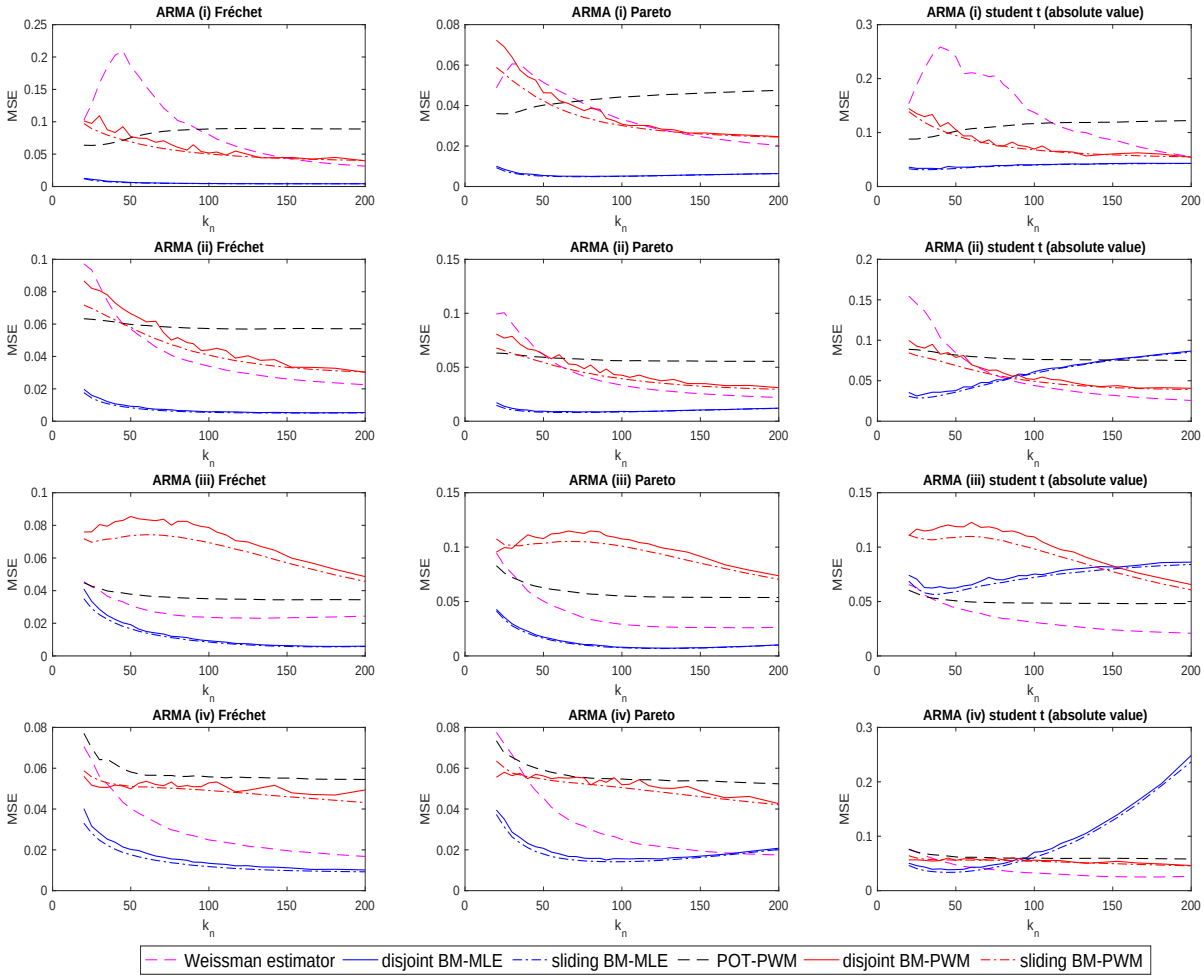

Figure 16: Scaled MSEs in the ARMA models as a function of the effective sample size. $n = 2000$, $p_n = 0.001$

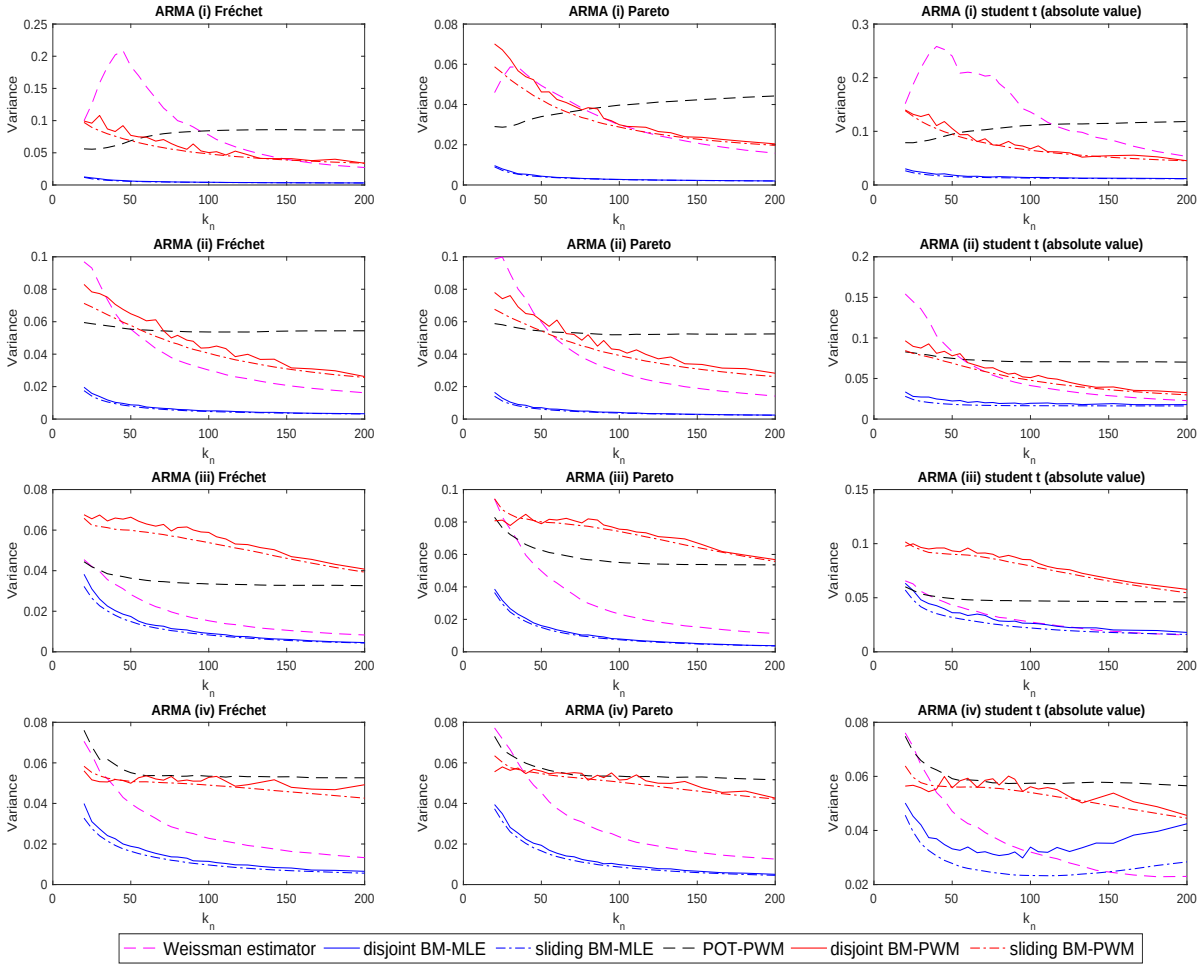

Figure 17: Scaled variances in the ARMA models as a function of the effective sample size. $n = 2000$, $p_n = 0.001$

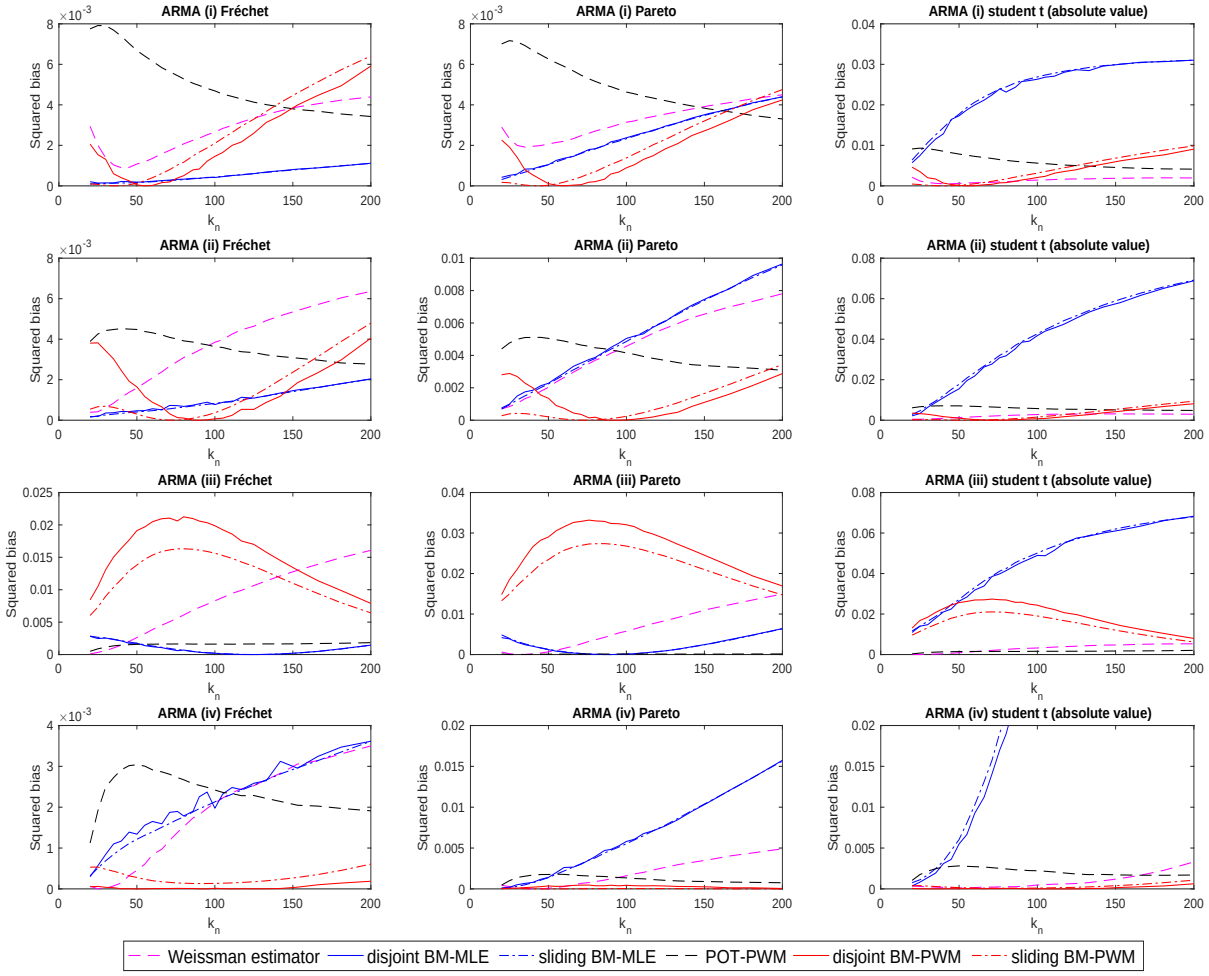

Figure 18: Scaled squared biases in the ARMA models as a function of the effective sample size. $n = 2000$, $p_n = 0.001$

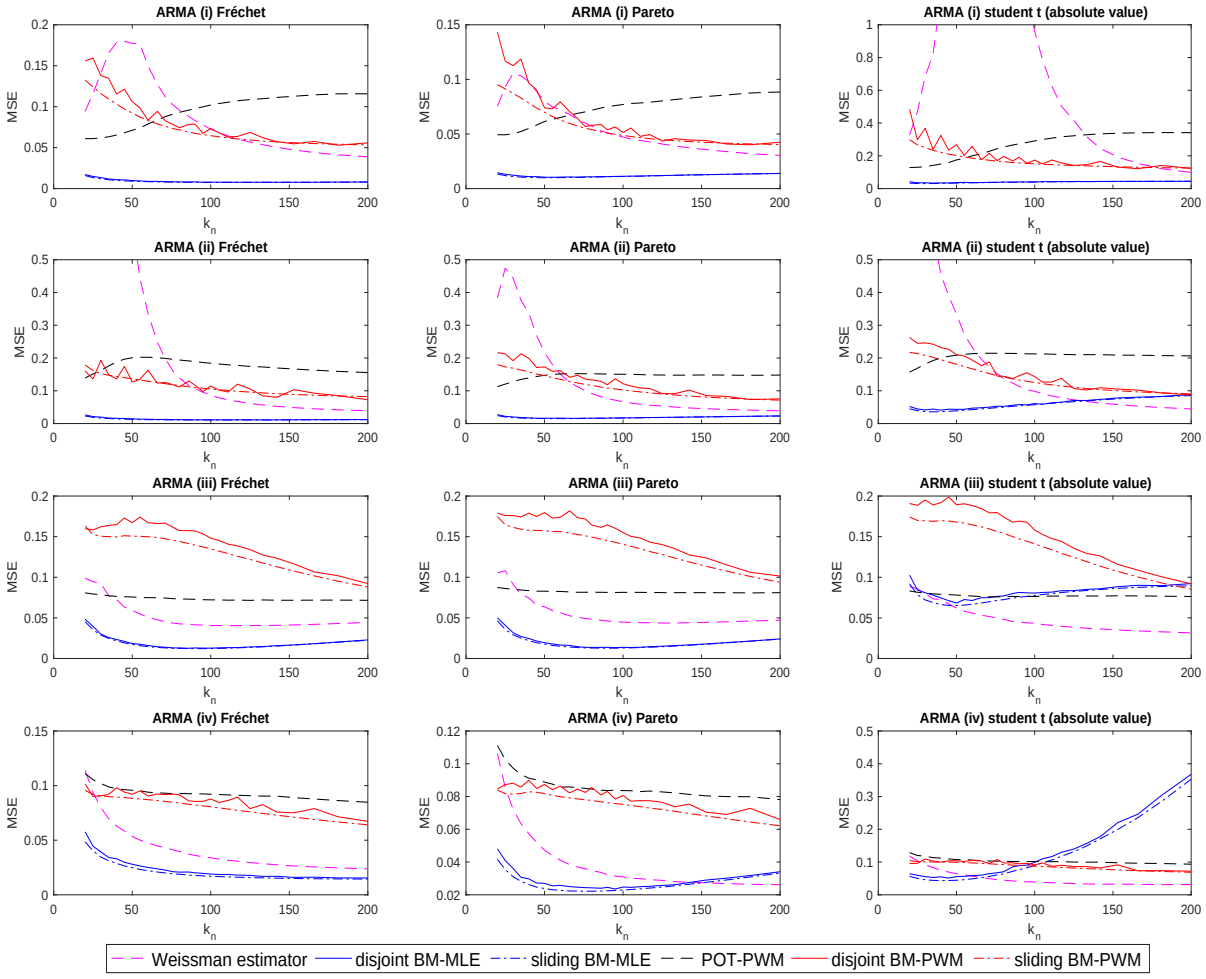

Figure 19: Scaled MSEs in the ARMA models as a function of the effective sample size. $n = 2000$, $p_n = 0.0005$

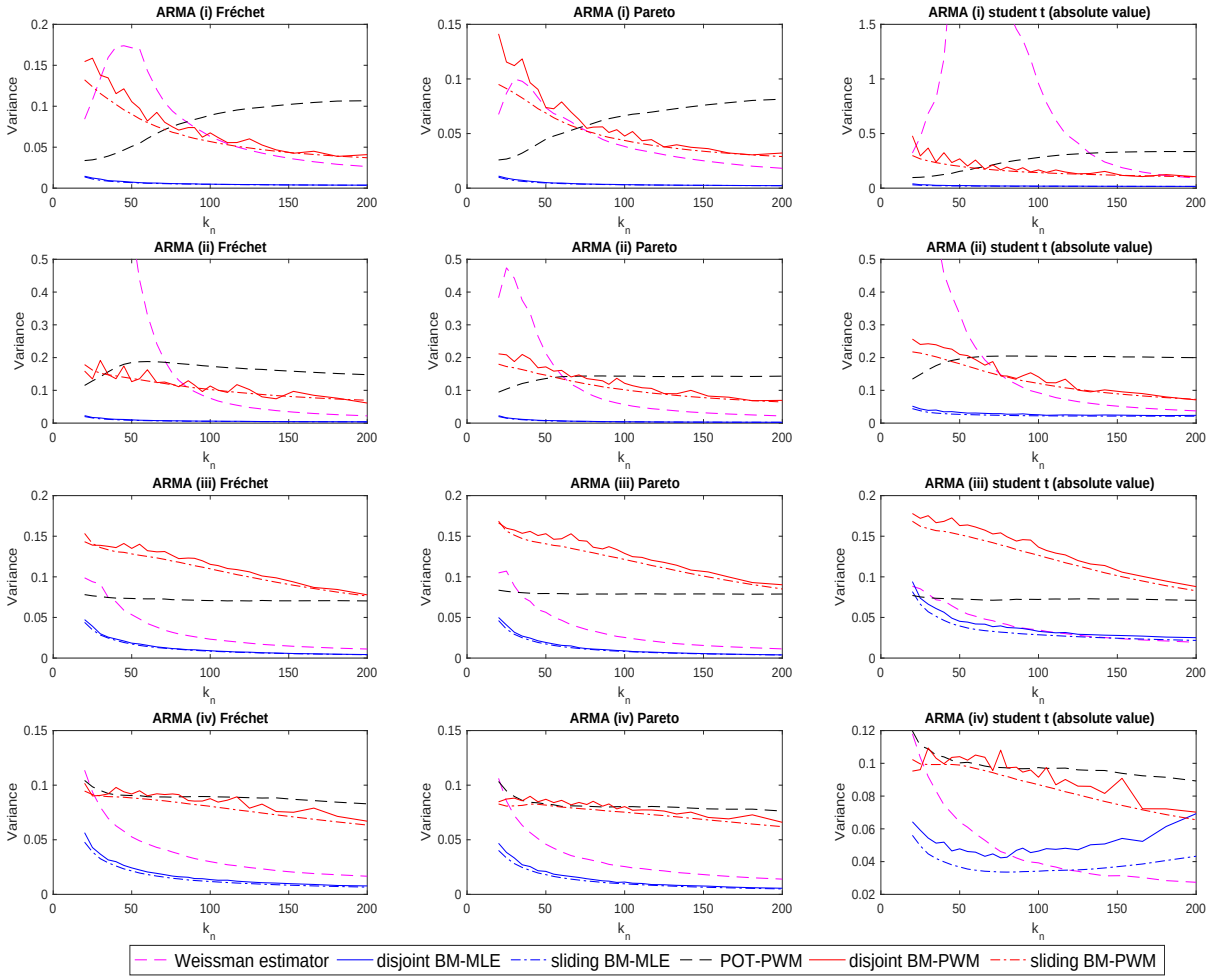

Figure 20: Scaled variances in the ARMA models as a function of the effective sample size. $n = 2000$, $p_n = 0.0005$

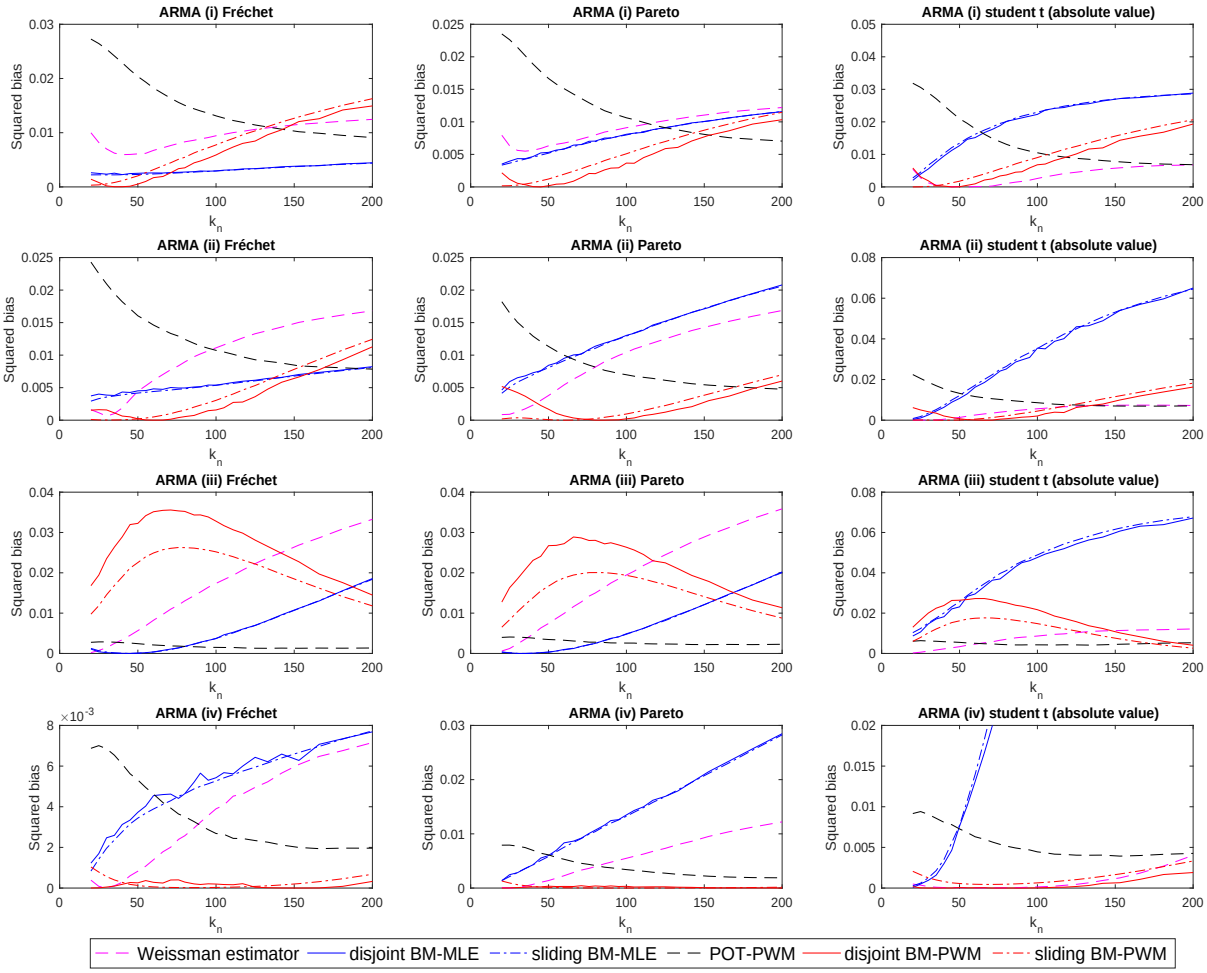

Figure 21: Scaled squared biases in the ARMA models as a function of the effective sample size. $n = 2000$, $p_n = 0.0005$

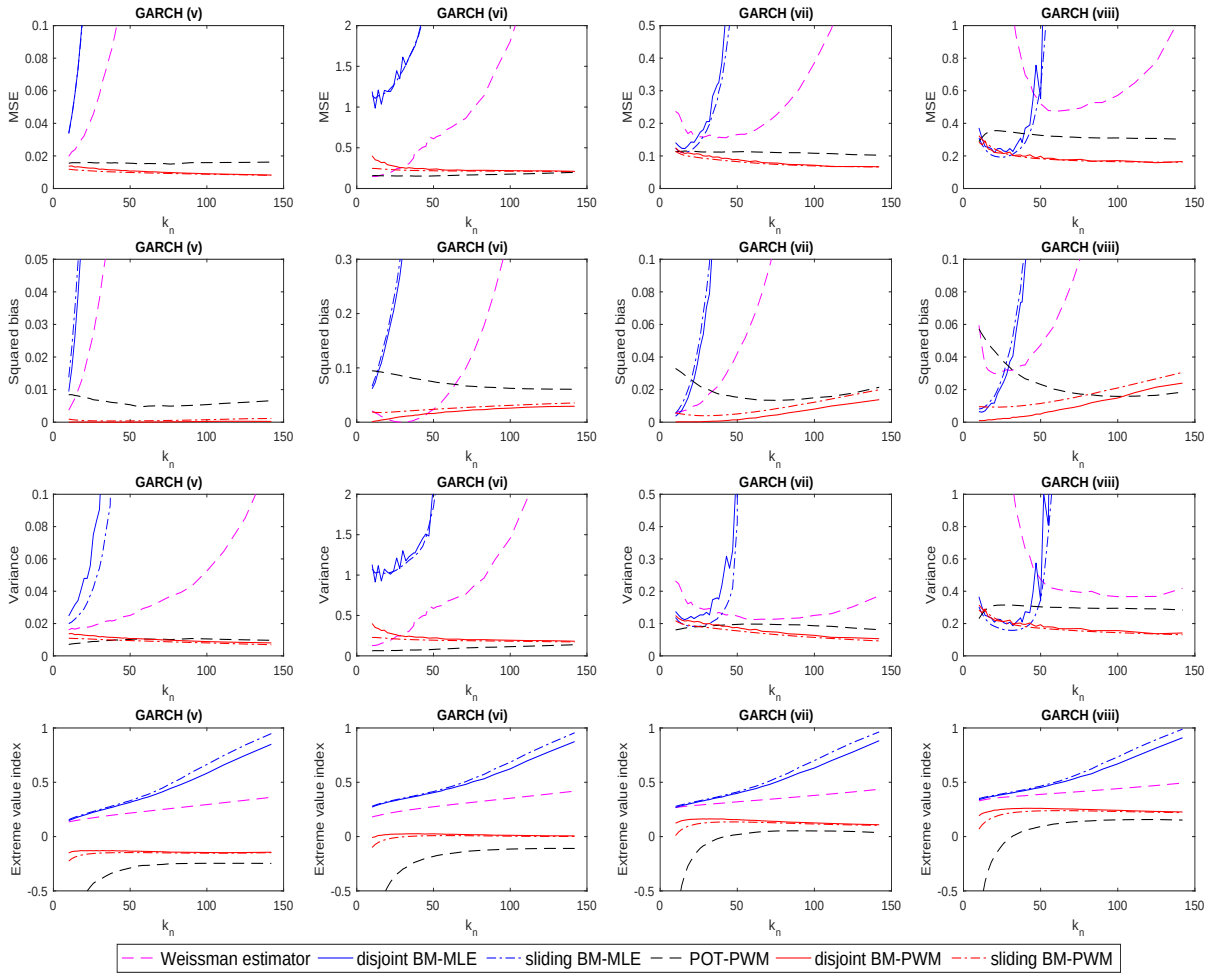

Figure 22: Performances (the scaled MSEs, scaled squared biases and scaled variances) and the extreme value index estimates in the GARCH models as a function of the effective sample size, respectively. $n = 1000$, $p_n = 0.0005$

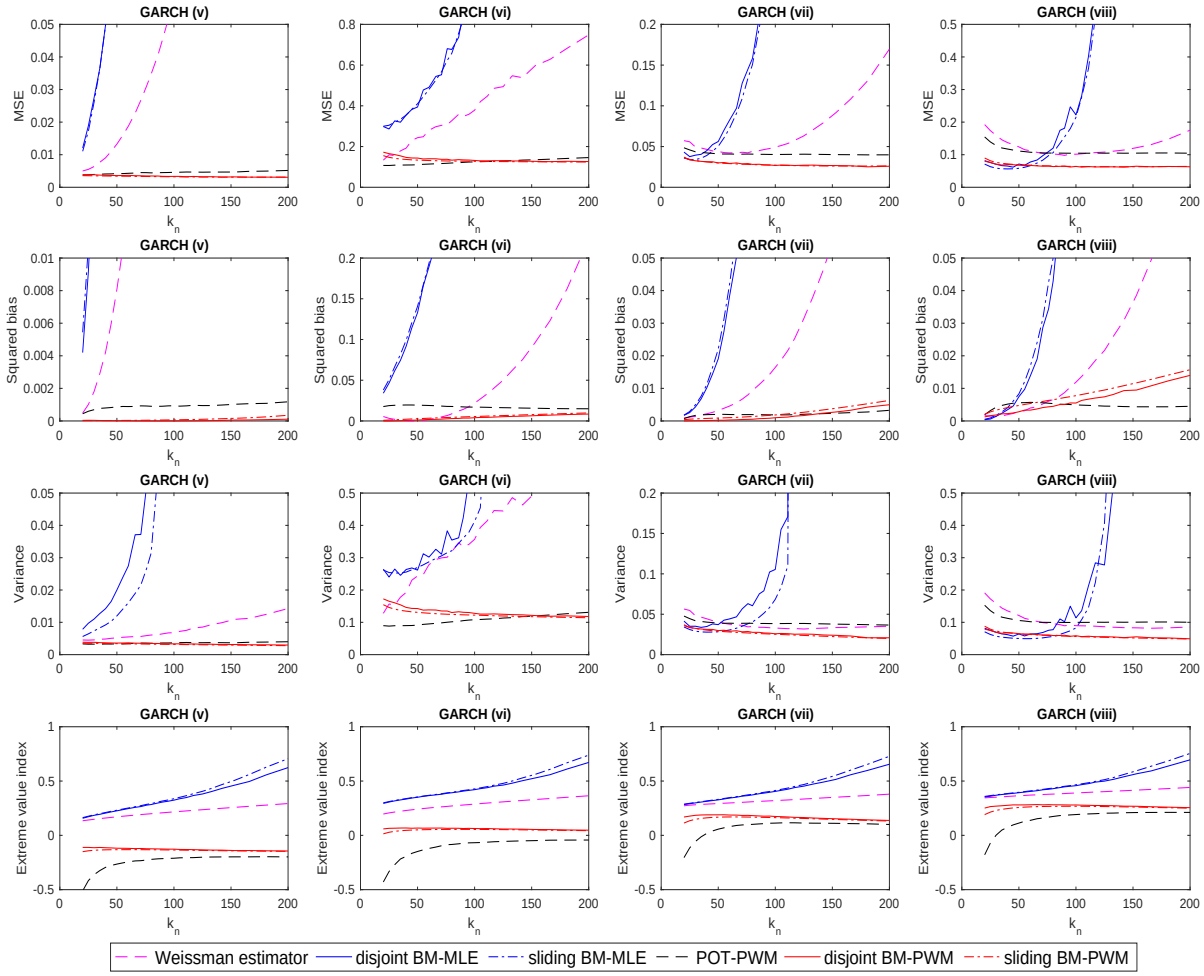

Figure 23: Performances (the scaled MSEs, scaled squared biases and scaled variances) and the extreme value index estimates in the GARCH models as a function of the effective sample size, respectively. $n = 2000$, $p_n = 0.001$

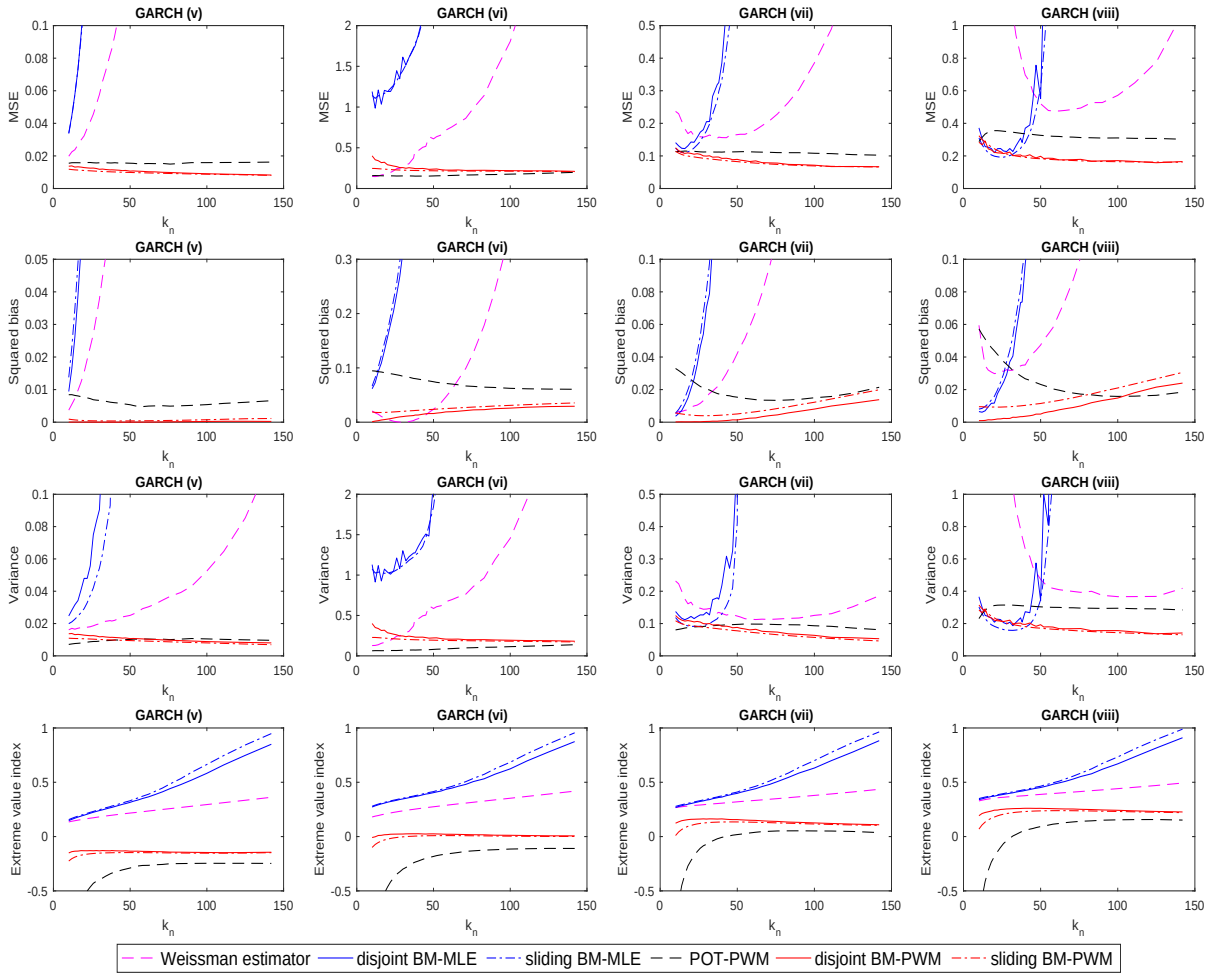

Figure 24: Performances (the scaled MSEs, scaled squared biases and scaled variances) and the extreme value index estimates in the GARCH models as a function of the effective sample size, respectively. $n = 1000$, $p_n = 0.0005$

References

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

Bücher, A., & Segers, J. (2018a). Inference for heavy tailed stationary time series based on sliding blocks. *Electronic Journal of Statistics*, *12*(1), 1098–1125.

Bücher, A., & Segers, J. (2018b). Maximum likelihood estimation for the échet distribution based on block maxima extracted from a time series. *Bernoulli*, *24*(2), 1427–1462.

Deheuvels, P., Haeusler, E., & Mason, D. M. (1988). Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society*, *104*(2), 371–381.

Dekkers, A., Einmahl, J., & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, *17*(4), 1833–1855.

Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, *21*(1), 420–436.

Dombry, C., & Ferreira, A. (2017). Maximum likelihood estimators based on the block maxima method. *arXiv preprint arXiv:1705.00465*.

Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli*, *9*(4), 617–657.

Drees, H., Ferreira, A., & de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *The Annals of Applied Probability*, *14*(3), 1179–1201.

Ferreira, A., & de Haan, L. (2015). On the block maxima method in extreme value theory: Pwm estimators. *The Annals of statistics*, *43*(1), 276–298.

Giannerini, S., Maasoumi, E., & Dagum, E. (2015). Entropy testing for nonlinear serial dependence in time series. *Biometrika*, *102*(3), 661–675.

Haeusler, E., & Teugels, J. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics*, *13*(2), 743–756.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*(5), 1163–1174.

Hosking, J. R., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, *29*(3), 339–349.

Hosking, J. R., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, *27*(3), 251–261.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Probability Theory and Related Fields*, *65*(2), 291–306.

Linsmeier, T. J., & Pearson, N. D. (2000). Value at Risk. *Financial Analysts Journal*, *56*(2), 47–67.

Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *The Annals of Probability*, *10*(3), 754–764.

McNeil, A. J. (1998). *Calculating quantile risk measures for financial return series using extreme value theory* (Tech. Rep.). ETH Zurich.

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

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131.

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

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, *73*(364), 812–815.

Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, *100*(4), 794–815.

Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, *101*(4), 971–983.